NCC-WCH

Version 1.2

Menopause

Appendix H: Evidence tables

Clinical guideline

Methods, evidence and recommendations

1 June 2015

Draft for Consultation

Commissioned by the National Institute for Health and Clinical Excellence

Disclaimer
Healthcare professionals are expected to take NICE clinical guidelines fully into account when exercising their clinical judgement. However, the guidance does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of each patient, in consultation with the patient and/or their guardian or carer.

Funding
The National Collaborating Centre for Women and Children’s Health was commissioned by the National Institute for Health and Care Excellence to undertake the work on this guideline

Contents

Appendices

Appendix H: Evidence tables

H.1 Diagnosis of perimenopause and menopause

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Full citation Blumel,J.E., Chedraui,P., Baron,G., Belzares,E., Bencosme,A., Calle,A., Danckers,L., Espinoza,M.T., Flores,D., Gomez,G., Hernandez-Bueno,J.A., Izaguirre,H., Leon-Leon,P., Lima,S., Mezones-Holguin,E., Monterrosa,A., Mostajo,D., Navarro,D., Ojeda,E., Onatra,W., Royer,M., Soto,E., Tserotas,K., Vallejo,M.S., Collaborative Group for Research of the Climacteric in Latin America (REDLINC), Menopausal symptoms appear before the menopause and persist 5 years beyond: a detailed analysis of a	Sample size N = 8394 total N = 8373 after exclusions n = 2655 premenopausal n = 1648 perimenopausal n = 4070 postmenopausal (subdivided into n = 2249 late postmenopause [1-4 years] and n = 1821 early postmenopause [≥5 years]) Characteristics Mean age (SD) = 49.1 (5.7) years · Premenopause 40-44 years category = 41.8 (1.4) years · Premenopause ≥45 years category = 47.9 (3.0) years · Perimenopause = 47.2 (4.1) years · Early postmenopause = 50.8 (4.4) years · Late postmenopause = 54.8 (3.9) years 14.7% users of hormone therapy · 3.0% premenopausal 40 - 44 years group · 4.9% premenopausal ≥ 45 years group · 10.4% perimenopausal group · 23.6% early postmenopausal group · 23.4% late postmenopausal group 17.4% current smokers BMI not reported Inclusion Criteria Mid aged women in 22 health centres located in 18 Latin American cities. Hispanic-Mestizo women aged 40 - 59 years who accompanied patients attending consultations at participating health centres.	Tests Women fulfilling the inclusion criteria were asked to complete the Menopause Rating Scale and a general data questionnaire (covering sociodemographic information, lifestyle and personal factors, current medical care and drug use). Definitions used Menopausal status defined according to STRAW criteria Premenopausal: women having regular menses Perimenopausal: women having menstrual irregularities >7 days from their usual cycle Postmenopausal: women no longer menstruating (subdivided into early postmenopause [1-4 years since final menstrual period] and late postmenopause [≥5 years since final menstrual period])	Methods Women completed the questionnaires, and the prevalence of different symptoms at specific stages of the menopause transition was calculated. The prevalence of severe or very severe symptoms in each category was also documented. Individual responses to MRS score for hot flushes/sweating was recorded. This was classified as any degree of symptoms (score 1,2,3 or 4 on the MRS) and as severe/very severe symptoms (score 3 or 4 on the MRS).	Results Symptoms of hot flushes/sweating to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 64 (63 to 66)[1] Specificity, % (95% CI) 41 (39 to 44)[1] Positive LR (95% CI) 1.08 (1.04 to 1.14)[1] Negative LR (95% CI) 0.87 (0.81 to 0.94)[1] Symptoms of severe hot flushes/sweating to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 12 (11 to 13)[1] Specificity, % (95% CI) 89 (88 to 91)[1] Positive LR (95% CI) 1.10 (0.93 to 1.29)[1] Negative LR (95% CI) 0.99 (0.97 to 1.01)[1] Symptoms of hot flushes/sweating to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 64 (63 to 66)[1]	Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1.A Could the selection of patients have introduced bias? LOW RISK OF BIAS 1.B Is there concern that the included patients do not match the review question? LOW CONCERN Index Test Were the index test results interpreted without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? N/A 2.A Could the conduct or

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
multinational study, Climacteric, 15, 542-551, 2012 Ref Id 266130 Country/ies where the study was carried out Ecuador (and 11 other Latin American countries) Study type Case-series Aim of the study To assess the prevalence and severity of menopausal symptoms and their impact over quality of life among mid-aged Latin American women. Study dates Not reported Source of funding None	Exclusion Criteria Women of other ethnic groups (non-Hispanic Mestizo) Mental or physical handicap impairing the capacity of understanding and/or providing answers during the interview Women unwilling to give written consent for participation. Incomplete data.			Specificity, % (95% CI) 63 (61 to 65)[1] Positive LR (95% CI) 1.73 (1.64 to 1.82)[1] Negative LR (95% CI) 0.57 (0.54 to 0.60)[1] Symptoms of severe hot flushes/sweating to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 12 (11 to 13)[1] Specificity, % (95% CI) 95 (94 to 95)[1] Positive LR (95% CI) 2.16 (1.81 to 2.58)[1] Negative LR (95% CI) 0.93 (0.92 to 0.95)[1] Symptoms of hot flushes/sweating to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 64 (63 to 66)[1] Specificity, % (95% CI) 55 (53 to 56)[1] Positive LR (95% CI) 1.41 (1.36 to 1.47)[1] Negative LR (95% CI) 0.66 (0.63 to 0.69)[1] Symptoms of severe hot flushes/sweating to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 12 (11 to 13)[1] Specificity, % (95% CI) 92 (92 to 93)[1] Positive LR (95% CI) 1.58 (1.38 to 1.80)[1] Negative LR (95% CI) 0.95 (0.94 to 0.97)[1] Symptoms of hot flushes/sweating to distinguish perimenopausal	interpretation of the index test have introduced bias? LOW RISK OF BIAS 2.B Is there concern that the index test, its conduct, or interpretation differ from the review question? LOW CONCERN Reference Standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3.A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3.B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW RISK Flow and Timing Was there an appropriate interval between index test(s) and reference standard? Yes Did all patients receive a reference standard? Yes

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				women from postmenopausal women Sensitivity, % (95% CI) 59 (57 to 61)[1] Specificity, % (95% CI) 36 (34 to 37)[1] Positive LR (95% CI) 0.92 (0.88 to 0.96)[1] Negative LR (95% CI) 1.15 (1.07 to 1.23)[1] Symptoms of severe hot flushes/sweating to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 11 (9 to 12)[1] Specificity, % (95% CI) 88 (87 to 89)[1] Positive LR (95% CI) 0.91 (0.77 to 1.07)[1] Negative LR (95% CI) 1.01 (0.99 to 1.03)[1] Symptoms of hot flushes/sweating to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 59 (57 to 61)[1] Specificity, % (95% CI) 63 (61 to 65)[1] Positive LR (95% CI) 1.59 (1.49 to 1.69)[1] Negative LR (95% CI) 0.65 (0.61 to 0.70)[1] Symptoms of severe hot flushes/sweating to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 11 (9 to 12)[1] Specificity, % (95% CI) 95 (94 to 95)[1] Positive LR (95% CI) 1.96	Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4.A Could the patient flow have introduced bias? LOW RISK Limitations Other information Women currently taking HRT were included in the study. This included 23% of all postmenopausal women. Women who had undergone surgical menopause were included in the study.

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(1.59 to 2.42)[1] Negative LR (95% CI) 0.94 (0.93 to 0.96)[1] Symptoms of hot flushes/sweating to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 59 (57 to 61)[1] Specificity, % (95% CI) 47 (45 to 48)[1] Positive LR (95% CI) 1.10 (1.05 to 1.15)[1] Negative LR (95% CI) 0.88 (0.83 to 0.94)[1] Symptoms of severe hot flushes/sweating to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 11 (9 to 12)[1] Specificity, % (95% CI) 91 (90 to 91)[1] Positive LR (95% CI) 1.15 (0.99 to 1.35)[1] Negative LR (95% CI) 0.98 (0.97 to 1.00)[1] LR = likelihood ratio [1] Calculated by the NCC WCH technical team from data reported in the article	
Full citation Brown,W.J., Mishra,G.D., Dobson,A., Changes in physical symptoms during the menopause transition, International Journal of Behavioral Medicine, 9, 53-67, 2002 Ref Id 266196	Sample size N = 8236 total. n = 4571 premenopausal n = 2092 perimenopausal n= 577 postmenopausal (remaining women were taking HRT preparations therefore not classifiable) Characteristics Mean age 47.7±1.5 years 15.6% smokers BMI 25.5±5.0	Tests Standardised questionnaire to ask about experiences of ten physical symptoms over the past 12 months: headaches/migraines, severe tiredness, stiff or painful joints, back pain, leaking urine, constipation, eyesight problems, difficulty sleeping, hot flashes and night sweats. Response options were never, rarely, sometimes or often. Survey was conducted once in	Methods Prevalence of different symptoms at each stage (premenopausal, perimenopausal and postmenopausal) was calculated using the response rates of "sometimes" and "often".	Results Hot flashes to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 55 (51 to 59)[1] Specificity, % (95% CI) 56 (54 to 58)[1] Positive LR (95% CI) 1.25 (1.15 to 1.36)[1] Negative LR (95% CI) 0.80 (0.73 to 0.89)[1] Night sweats to distinguish postmenopausal women	Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1. A Could the selection of patients

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Country/ies where the study was carried out Australia Study type Case-series Aim of the study To analyse different physical symptoms experienced in different stages of the menopause transition. The study aimed to test the hypothesis that there would be an association between the reporting of physical symptoms and menopausal status. Study dates National cohort study - the Australian Longitudain Study on Women's Health. Women completed two surveys - one in 1996 and the second in 1998. Source of funding Commonwealth Department of Health and Aged Care. Eli Lilly funded part of the analysis costs for this article.	Inclusion Criteria 45-50 years of age. Random selection of women from across Australia from national Medicare health insurance database. Exclusion Criteria For this analysis - excluded women taking HRT as menopausal status was not available. Excluded women with history of hysterectomy or oophorectomy.	1996 and again in 1998. Data from the first study were used for this analysis. Definitions used Premenopausal: menstrual bleeding in the last 3 months, and in the last 12 months, and with the same frequency as the year prior to that. Perimenopausal: menstrual bleeding in the last 12 months, but not in the last 3 months, or with different menstrual frequency compared with the previous year. Postmenopausal: no menstrual bleeding in the last 12 months.		from perimenopausal women Sensitivity, % (95% CI) 39 (35 to 43)[1] Specificity, % (95% CI) 67 (65 to 69)[1] Positive LR (95% CI) 1.18 (1.05 to 1.33)[1] Negative LR (95% CI) 0.91 (0.85 to 0.98)[1] Hot flashes to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 55 (51 to 59)[1] Specificity, % (95% CI) 84 (83 to 85)[1] Positive LR (95% CI) 3.44 (3.11 to 3.79)[1] Negative LR (95% CI) 0.54 (0.49 to 0.59)[1] Night sweats to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 39 (35 to 43)[1] Specificity, % (95% CI) 88 (87 to 89)[1] Positive LR (95% CI) 3.25 (2.86 to 3.69)[1] Negative LR (95% CI) 0.69 (0.65 to 0.74)[1] Hot flashes to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 55 (51 to 59)[1] Specificity, % (95% CI) 75 (74 to 76)[1] Positive LR (95% CI) 2.22 (2.04 to 2.41)[1] Negative LR (95% CI) 0.60 (0.55 to 0.66)[1] Night sweats to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 39	have introduced bias? LOW RISK OF BIAS 1. B Is there concern that the included patients do not match the review question? LOW CONCERN Index test Were the index test results interpreted without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? Unclear - threshold of response "sometimes" of "often" to report prevalence of symptoms. Not clear if this was pre-defined. 2. A Could the conduct or interpretation of the index test have introduced bias? LOW RISK OF BIAS 2. B Is there concern that the index test, its conduct or interpretation differ from the review question? LOW CONCERN Reference standard Is the reference standard likely to correctly classify the target condition? Yes

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(35 to 43)[1] Specificity, % (95% CI) 81 (80 to 82)[1] Positive LR (95% CI) 2.09 (1.87 to 2.34)[1] Negative LR (95% CI) 0.75 (0.70 to 0.80)[1] Hot flashes to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 44 (42 to 46)[1] Specificity, % (95% CI) 45 (41 to 49)[1] Positive LR (95% CI) 0.80 (0.73 to 0.87)[1] Negative LR (95% CI) 1.24 (1.13 to 1.37)[1] Night sweats to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 33 (31 to 35)[1] Specificity, % (95% CI) 61 (57 to 65)[1] Positive LR (95% CI) 0.85 (0.75 to 0.95)[1] Negative LR (95% CI) 1.10 (1.02 to 1.18)[1] Hot flashes to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 44 (42 to 46)[1] Specificity, % (95% CI) 84 (83 to 85)[1] Positive LR (95% CI) 2.75 (2.53 to 2.98)[1] Negative LR (95% CI) 0.67 (0.64 to 0.69)[1] Night sweats to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 33	Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3. A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3. B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW CONCERN Flow and timing Was there an appropriate interval between index test and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4. A Could the patient flow have introduced bias? LOW RISK OF BIAS Limitations Other information Women using HRT were excluded from this analysis as unable to determine

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(31 to 35)[1] Specificity, % (95% CI) 88 (87 to 89)[1] Positive LR (95% CI) 2.75 (2.49 to 3.03)[1] Negative LR (95% CI) 0.76 (0.74 to 0.79)[1] Hot flashes to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 44 (42 to 46)[1] Specificity, % (95% CI) 80 (79 to 81)[1] Positive LR (95% CI) 2.16 (2.01 to 2.32)[1] Negative LR (95% CI) 0.70 (0.68 to 0.73)[1] Night sweats to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 33 (31 to 35)[1] Specificity, % (95% CI) 85 (84 to 86)[1] Positive LR (95% CI) 2.20 (2.01 to 2.40)[1] Negative LR (95% CI) 0.79 (0.76 to 0.81)[1] LR = likelihood ratio [1] Calculated by the NCC WCH technical team from data reported in the article.	menopausal status. Women with surgical menopause were excluded from the study.
Full citation Burger,H.G., Cahir,N., Robertson,D.M., Groome,N.P., Dudley,E., Green,A., Dennerstein,L., Serum inhibins A and B fall differentially as FSH rises in perimenopausal	Sample size N = 110 n = 28 premenopausal n = 59 perimenopausal n = 23 postmenopausal Characteristics Age range 48 - 59 years Inclusion Criteria Women who were having regular or moderately irregular cycles or who had not bled for more than 3 months Exclusion Criteria	Tests Inhibin A Inhibin B Definitions used Premenopausal: not defined Perimenopausal: defined as self report of cycle change in the preceding 12 months, with a bleed in the preceding 12 months, or amenorrhoea for 3-11 months	Methods Samples were collected between cycle day 5 and 8 in women with regular or irregular cycles or at random in women with no cycles for over 3 months	Results Undetectable inhibin A to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 96 (78 to 100)[1] Specificity, % (95% CI) 39 (27 to 53)[1] Positive LR (95% CI) 1.57 (1.26 to 1.96)[1] Negative LR (95% CI) 0.11	Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Unclear - subgroup of women from larger study were enrolled, and recruitment to this sub-study was not reported.

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
women, Clinical Endocrinology, 48, 809-813, 1998 Ref Id 266215 Country/ies where the study was carried out Australia Study type Case-series Aim of the study To examine the behaviour of inhibin-A and inhibin-B in older peri-menopausal women in relation to changing levels of follicle-stimulating hormone, estradiol and immunoreactive inhibin. Study dates September - December 1994 Source of funding The Melbourne Women's Mid-Life Health Project is supported by the Victorian Health Promotion Foundation and the Public Health Research and Development Committee of the Australian National Health and Medical Research Council	Not reported	Postmenopausal: defined as ≥ 12 months amenorrhoea		(0.02 to 0.78)[1] Undetectable inhibin B to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 43 (23 to 66)[1] Specificity, % (95% CI) 54 (41 to 68)[1] Positive LR (95% CI) 0.95 (0.55 to 1.64)[1] Negative LR (95% CI) 1.04 (0.68 to 1.60)[1] Undetectable inhibin A to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 96 (78 to 100)[1] Specificity, % (95% CI) 54 (34 to 72)[1] Positive LR (95% CI) 2.06 (1.37 to 3.10)[1] Negative LR (95% CI) 0.08 (0.01 to 0.57)[1] Undetectable inhibin B to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 43 (23 to 66)[1] Specificity, % (95% CI) 78 (58 to 91)[1] LR+ (95% CI) 1.96 (0.84 to 4.56)[1] LR- (95% CI) 0.73 (0.48 to 1.10)[1] Undetectable inhibin A to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 96 (78 to 100)[1] Specificity, % (95% CI) 44 (33 to 55)[1] Positive LR (95% CI) 1.70 (1.38 to 2.08)[1]	Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1. A Could the selection of patients have introduced bias? LOW RISK 1. B Is there concern that the included patients do not match the review question? LOW CONCERN Index test Were the index test results interpreted without knowledge of the results of the reference standard? Unclear - blinding of investigators was not described, but unlikely to introduce bias as no subjective interpretation of results required. If a threshold was used, was it pre-specified? Yes 2. A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2. B Is there concern that the index test, its conduct or interpretation differ from the review question? LOW CONCERN

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Negative LR (95% CI) 0.10 (0.01 to 0.69)[1] Undetectable inhibin B to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 43 (23 to 66)[1] Specificity, % (95% CI) 62 (51 to 72)[1] Positive LR (95% CI) 1.14 (0.67 to 1.96)[1] Negative LR (95% CI) 0.91 (0.61 to 1.36)[1] Undetectable inhibin A to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 61 (47 to 73)[1] Specificity, % (95% CI) 4 (0 to 22)[1] Positive LR (95% CI) 0.64 (0.51 to 0.80)[1] Negative LR (95% CI) 8.97 (1.28 to 62.60)[1] Undetectable inhibin B to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 46 (32 to 59)[1] Specificity, % (95% CI) 57 (34 to 77)[1] Positive LR (95% CI) 1.05 (0.61 to 1.81)[1] Negative LR (95% CI) 0.96 (0.63 to 1.48)[1] Undetectable inhibin A to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 61 (47 to 73)[1] Specificity, % (95% CI) 54 (34 to 72)[1] Positive LR (95% CI) 1.31	Reference standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3. A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3. B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW CONCERN Flow and timing Was there an appropriate interval between index test and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4. A Could the patient flow have introduced bias? LOW RISK Limitations Women represented

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(0.84 to 2.06)[1] Negative LR (95% CI) 0.73 (0.45 to 1.16)[1] Undetectable inhibin B to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 46 (32 to 59)[1] Specificity, % (95% CI) 78 (58 to 91)[1] Positive LR (95% CI) 2.05 (0.96 to 4.39)[1] Negative LR (95% CI) 0.70 (0.51 to 0.96)[1] Undetectable inhibin A to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 61 (47 to 73)[1] Specificity, % (95% CI) 31 (19 to 46)[1] Positive LR (95% CI) 0.89 (0.67 to 1.17)[1] Negative LR (95% CI) 1.24 (0.74 to 2.08)[1] Undetectable inhibin B to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 46 (32 to 59)[1] Specificity, % (95% CI) 68 (54 to 80)[1] Positive LR (95% CI) 1.43 (0.87 to 2.34)[1] Negative LR (95% CI) 0.80 (0.59 to 1.08)[1] LR = likelihood ratio [1] Values calculated by the NCC WCH technical team from data reported in the paper	a subgroup of participants from a larger study (The Melbourne Women's Mid-Life Health Project). How this subgroup was identified and recruited is not described. Whether the index test was interpreted without knowledge of the reference standard is not made clear. However, this is unlikely to introduce bias as the index test result (inhibin B) was reported only as detectable or undetectable. Other information Not clear whether women with HRT and surgical menopause were included.
Full citation Chuni,N., Sreeramareddy,C.T.	Sample size N = 729 n = 267 premenopausal	Tests Frequency of menopausal symptoms reported according to	Methods Interviewer administered survey	Results Hot flushes/sweating to	Study quality - QUADAS 2 checklist Patient selection

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
, Frequency of symptoms, determinants of severe symptoms, validity of and cut-off score for Menopause Rating Scale (MRS) as a screening tool: a cross-sectional survey among midlife Nepalese women, BMC Women's Health, 11, 30-, 2011 Ref Id 228089 Country/ies where the study was carried out Nepal Study type Case-series Aim of the study To determine the validity of the Menopause Rating Scale as a screening tool for identification of women with severe menopausal symptoms and cut-off MRS score for referral to gynaecologist. Study dates February to August 2008. Source of funding Not reported	n = 215 perimenopausal n = 247 postmenopausal Characteristics Mean age (SD): 49.9 (5.6) years Mean age (SD) premenopausal women: 45.1 (2.78) years Mean age (SD) perimenopausal women: 49.14 (2.01) years Mean age (SD) postmenopausal women: 55.67 (5.6) years Inclusion Criteria All women aged between 40 and 65 years attending health screening camps in Bedabari Primary Health Centre and Batulechaur Health Post. Exclusion Criteria Pregnancy or lactation. History of cancer in remission or under treatment currently. History of drug or alcohol abuse. Mental disability or undergoing treatment for psychiatric disorders. Premature ovarian insufficiency or known genital malformations.	menopausal status. Identified using the Menopause Rating Scale (MRS). Definitions used Premenopausal: minor changes in cycle length, particularly decreasing cycle length Perimenopausal: increasing irregularity of menses without skipping periods (7 days difference from the beginnng of a given cycle to the next) (early perimenopausal) or menstruation in the past 2 - 12 months but not during the past 2 months (late perimenopausal) Postmenopausal: no menstrual bleeding in the past 12 months	to eligible women attending health screening camps in Western Development Region of Nepal. Questionnaire included socio-demographic characteristics, menopausal status, menstrual history, chronic diseases, HRT use, general health and well-being, and symptoms based on Menopause Rating Scale. Menopausal status was defined according to STRAW criteria, with early and late perimenopause categories combined.	distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 98 (96 to 100)[1] Specificity, % (95% CI) 5 (3 to 9)[1] Positive LR (95% CI) 1.04 (1.00 to 1.07)[1] Negative LR (95% CI) 0.32 (0.10 to 0.98)[1] Hot flushes/sweating to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 98 (96 to 100)[1] Specificity, % (95% CI) 77 (72 to 82)[1] Positive LR (95% CI) 4.31 (3.45 to 5.37)[1] Negative LR (95% CI) 0.02 (0.01 to 0.06)[1] Hot flushes/sweating to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 98 (96 to 100)[1] Specificity, % (95% CI) 45 (41 to 50)[1] Positive LR (95% CI) 1.79 (1.65 to 1.94)[1] Negative LR (95% CI) 0.04 (0.01 to 0.10)[1] Hot flushes/sweating to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 95 (91 to 97)[1] Specificity, % (95% CI) 2 (0 to 4)[1]	Was a consecutive or random sample of patients enrolled? Yes (consecutive) Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1. A Could the selection of patients have introduced bias? LOW RISK 1. B Is there concern that the included patients do not match the review question? LOW CONCERN Index test Were the index test results interpreted without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? Unclear - threshold for symptoms not reported in paper, but assumed to be score of ≥ 1 on MRS 2. A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2. B Is there concern that the index test, its conduct or interpretation differ from the review

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Positive LR (95% CI) 0.96 (0.93 to 1.00)[1] Negative LR (95% CI) 3.16 (1.02 to 9.78)[1] Hot flushes/sweating to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 95 (91 to 97)[1] Specificity, % (95% CI) 77 (72 to 82)[1] Positive LR (95% CI) 4.15 (3.32 to 5.19)[1] Negative LR (95% CI) 0.07 (0.04 to 0.12)[1] Hot flushes/sweating to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 95 (91 to 97)[1] Specificity, % (95% CI) 41 (37 to 45)[1] Positive LR (95% CI) 1.60 (1.48 to 1.73)[1] Negative LR (95% CI) 0.13 (0.07 to 0.22)[1] LR = likelihood ratio [1] Calculated by the NCC WCH technical team from data reported in the article.	question? LOW CONCERN Reference standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3. A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3. B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW CONCERN Flow and timing Was there an appropriate interval between index test and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4. A Could the patient flow have introduced

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					bias? LOW RISK Limitations Other information Article does not report whether a threshold score on the MRS was used to identify prevalence of symptoms. It is assumed that a score of ≥ 1 would be taken as symptomatic. No description of whether women using HRT or those with surgical menopause were included.
Full citation Cooper,G.S., Baird,D.D., The use of questionnaire data to classify peri- and premenopausal status, Epidemiology, 6, 625-628, 1995 Ref Id 266473 Country/ies where the study was carried out USA Study type Case-series Aim of the study To assess how well questionnaire data could classify peri- versus premenopausal status in women aged 38-49 years. Study dates Not reported Source of funding American Institute	Sample size N = 280 after exclusions (see below) n = 39 perimenopausal women n = 241 premenopausal women Characteristics Mean age (SD) = 44.2 (3.0) 11% African American 20/280 women (7%) current users of HRT Inclusion Criteria Women between the ages of 38 and 49. Exclusion Criteria Previous hysterectomy or oophorectomy. Post menopausal women (12 or more months since last menstrual period)	Tests Serum FSH was measured on the morning of day 2, 3 or 4 of a menstrual cycle for women who had a period within the preceding 2 months. Other women were scheduled at their convenience. Each participant completed a self administered questionnaire that included sections on reproductive and menstrual history. Definitions used Premenopausal: FSH < 15 IU/L Perimenopausal: FSH ≥ 15 IU/L	Methods Participants completed a self administered questionnaire that included sections on reproductive and menstrual history. Prevalence of specific symptoms was then calculated for women who were classified as pre and perimenopausal.	Results Diagnostic accuracy of either a single symptom, or a combination of symptoms was assessed. Age ≥ 42 years to distinguish perimenopausal from premenopausal women Sensitivity, % (95% CI) 90 (76 to 97)[1] Specificity, % (95% CI) 29 (23 to 35)[1] Positive LR (95% CI) 1.26 (1.10 to 1.45)[1] Negative LR (95% CI) 0.36 (0.14 to 0.93)[1] Age ≥ 46 years to distinguish perimenopausal from premenopausal women Sensitivity, % (95% CI) 54 (37 to 70)[1] Specificity, % (95% CI) 73 (67 to 79)[1] Positive LR (95% CI) 2.00 (1.40 to 2.85)[1] Negative LR (95% CI) 0.63 (0.45 to 0.89)[1] Hot flashes/night sweats during the past 6 months ≥1	Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Unclear - women responded to advertisements for participants. Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes (N.B. study excluded menopausal women as aim was to classify only perimenopausal and premenopausal status) 1. A Could the selection of patients have introduced bias? LOW RISK 1. B Is there concern that the included patients do not match the review question?

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
for Cancer Research Reproductive Hazards in the Workplace, Home, Community and Environment Research National Cancer Institute Research Service Award Division of Research Resources, NIH.				per day Sensitivity, % (95% CI) 29 (15 to 43) Specificity, % (95% CI) 97 (95 to 99) Positive LR (95% CI) 9.43 (3.90 to 22.80)[1] Negative LR (95% CI) 0.73 (0.60 to 0.90)[1] Longer menstrual cycle during past 5 years Sensitivity, % (95% CI) 28 (13 to 42) Specificity, % (95% CI) 91 (87 to 95) Positive LR (95% CI) 3.11 (NC)[2] Negative LR (95% CI) 0.79 (NC)[2] More variable menstrual cycle during past 5 years Sensitivity, % (95% CI) 58 (42 to 74) Specificity, % (95% CI) 84 (79 to 89) Positive LR (95% CI) 3.63 (NC)[2] Negative LR (95% CI) 0.50 (NC)[2] Length of last menstrual cycle ≥60 days Sensitivity, % (95% CI) 33 (16 to 50) Specificity, % (95% CI) 99 (98-100) Positive LR (95% CI) 38.00 (8.74 to 165.22)[1] Negative LR (95% CI) 0.67 (0.52 to 0.87)[1] At least one of the following symptoms: hormone replacement therapy begun when periods irregular, hot flashes/night sweats ≥1 per day or last menstrual cycle more than	LOW CONCERN Index test Were the index test results interpreted without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? No - a variety of thresholds were presented within the article. 2. A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2. B Is there concern that the index test, its conduct or interpretation differ from the review question? LOW CONCERN Reference standard Is the reference standard likely to correctly classify the target condition? No - serum FSH used as the gold standard for perimenopause. Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3. A Could the reference standard,

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				60 days. Sensitivity, % (95% CI) 56 (41 to 72) Specificity, % (95% CI) 95 (93 to 98) Positive LR (95% CI) 12.36 (6.52 to 23.44)[1] Negative LR (95% CI) 0.46 (0.32 to 0.65)[1] At least one of the following symptoms: hormone replacement therapy begun when periods irregular, hot flashes/night sweats ≥1 per day, last menstrual cycle more than 60 days or menstrual cycles longer or more variable during the past 5 years. Sensitivity, % (95% CI) 69 (55 to 84) Specificity, % (95% CI) 75 (70 to 81) Positive LR (95% CI) 2.78 (2.05 to 3.77)[1] Negative LR (95% CI) 0.41 (0.25 to 0.66)[1] LR = likelihood ratio NC = not calculable [1] Likelihood ratios and confidence intervals calculated by the NCC WCH technical team from data presented in the article [2] Confidence intervals unable to be calculated around the point estimate due to the limited data available for this measure	its conduct, or its interpretation have introduced bias? LOW RISK 3. B Is there concern that the target condition as defined by the reference standard does not match the review question? HIGH RISK Flow and timing Was there an appropriate interval between index test and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4. A Could the patient flow have introduced bias? LOW RISK Limitations FSH was used as the gold standard for perimenopausal status. Other information 7% of participants were current users of HRT.
Full citation El,Shafie K., Al,Farsi Y., Al,Zadjali N., Al,Adawi S., Al,Busaidi Z., Al,Shafaee M.,	Sample size N = 479 total N = 472 after 7 exclusions for data error or inconsistency · n = 190 premenopausal · n = 73 perimenopausal	Tests The Menopause Rating Scale was used to identify frequency and severity of current symptoms. Definitions used	Methods Data were collected through face to face interviews by health educators trained to read the	Results Hot flashes to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 55 (48 to 61)[1]	Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled?

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Menopausal symptoms among healthy, middle-aged Omani women as assessed with the Menopause Rating Scale, Menopause, 18, 1113-1119, 2011 Ref Id 266687 Country/ies where the study was carried out Oman Study type Case-series Aim of the study To assess the frequency and severity of menopausal symptoms among a cohort of healthy, middle-aged Omani women using the Menopause Rating Scale. Study dates March and April 2010 Source of funding None reported	· n = 209 postmenopausal Characteristics Age range: 40 - 60 years Smoking status: Not reported BMI: Not reported Inclusion Criteria Healthy women between the age of 40 and 60 who were not pregnant or lactating, had an intact uterus and had no history of chronic disease Exclusion Criteria Women aged over 60, or who had a chronic illness or declined to participate	Premenopausal: having regular menses and ≥12 menses in previous 12 months Perimenopausal: irregular menses and at least 1 but less than 12 menses in previous 12 months Postmenopausal: no menses in previous 12 months	questionnaire and to document the responses.	Specificity, % (95% CI) 51 (39 to 63)[1] Positive LR (95% CI) 1.11 (0.85 to 1.44)[1] Negative LR (95% CI) 0.90 (0.68 to 1.18)[1] Hot flashes to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 55 (48 to 61)[1] Specificity, % (95% CI) 74 (67 to 80)[1] Positive LR (95% CI) 2.07 (1.59 to 2.71)[1] Negative LR (95% CI) 0.62 (0.52 to 0.73)[1] Hot flashes to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 55 (48 to 61)[1] Specificity, % (95% CI) 67 (61 to 73)[1] Positive LR (95% CI) 1.67 (1.35 to 2.06)[1] Negative LR (95% CI) 0.68 (0.57 to 0.80)[1] Hot flashes to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 49 (37 to 61)[1] Specificity, % (95% CI) 45 (39 to 52)[1] Positive LR (95% CI) 0.90 (0.69 to 1.18)[1] Negative LR (95% CI) 1.12 (0.85 to 1.46)[1] Hot flashes to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 49 (37 to 61)[1] Specificity, % (95% CI) 74	Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1. A Could the selection of patients have introduced bias? LOW RISK 1. B Is there concern that the included patients do not match the review question? LOW CONCERN Index test Were the index test results interpreted without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? Unclear - threshold for symptoms was not described in article, but assumed to be MRS score of >0. 2. A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2. B Is there concern that the index test, its conduct or interpretation differ from the review question? LOW CONCERN

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(67 to 80)[1] Positive LR (95% CI) 1.87 (1.34 to 2.61)[1] Negative LR (95% CI) 0.69 (0.54 to 0.88)[1] Hot flashes to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 49 (37 to 61)[1] Specificity, % (95% CI) 59 (54 to 64)[1] Positive LR (95% CI) 1.20 (0.92 to 1.56)[1] Negative LR (95% CI) 0.86 (0.68 to 1.09)[1] LR = likelihood ratio [1] Calculated by the NCC WCH technical team from data reported in the article	Reference standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3. A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3. B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW CONCERN Flow and timing Was there an appropriate interval between index test and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4. A Could the patient flow have introduced bias? LOW RISK Limitations Other information

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					MRS grading system from 0 (not present) to 4 (1, mild; 2, moderate; 3, severe; 4, very severe) MRS score used to identify prevalence of symptoms is not reported, but assumed that a score of ≥ 1 equates to symptom prevalence. Women with hysterectomy excluded. No comment on women with bilateral salpingoophorectomy, or on current use of HRT.
Full citation Giacobbe,M., Mendes Pinto-Neto,A., Simoes Costa-Paiva,L.H., Martinez,E.Z., The usefulness of ovarian volume, antral follicle count and age as predictors of menopausal status, Climacteric, 7, 255-260, 2004 Ref Id 266886 Country/ies where the study was carried out Brazil Study type Case-series Aim of the study To compare the accuracy of ovarian volume, antral	Sample size N = 204 N = 192 after exclusions (see below) n = 121 premenopausal n = 71 postmenopausal Characteristics Mean age (all women) 46.8 years Mean age premenopausal women 44.3 years Mean age postmenopausal women 50.9 years Ethinicity: 74% white, 36% non-white Smoking status: 27% smokers, 73% non-smokers Hormonal contraception use: 36% non-users, 64% past users Hormone replacement therapy use: 80% non-users, 20% past or current users Inclusion Criteria Premenopausal and postmenopausal women aged between 40 and 55 years from the gynaecology division of Leonor Mendes do Barros Maternity Hospital, Sao Paolo, Brazil. Exclusion Criteria Unilateral oophorectomy, presence of cysts or	Tests Women were interviewed about demographic, social and medical conditions. They then underwent an ovarian scan with a 5-7MHz transvaginal multifrequency probe, by a single observer. Definitions used Premenopausal: the period of time in a women over 40 years of age when she had regular or irregular menstruation accompanied or not by climacteric symptoms Postmenopausal: absence of vaginal bleeding for one year	Methods Ovarian scans were conducted during the early follicular phase of the cycle (day 4 to 7) for premenopausal women. Antral follicle count obtained after scanning the ovaries for small echo-free areas of approximately 3-8mm diameter. Average follicle count was taken if both ovaries were visible, or the count was obtained from the only visible ovary.	Results Age ≥ 48 to distinguish menopausal women from all other women Sensitivity, % (95% CI) 79 (68 to 88)[1] Specificity, % (95% CI) 76 (67 to 83)[1] Positive LR (95% CI) 3.29 (2.34 to 4.62)[2] Negative LR (95% CI) 0.28 (0.18 to 0.44)[2] Age ≥ 50 to distinguish menopausal women from all other women Sensitivity, % (95% CI) 68 (55 to 78)[2] Specificity, % (95% CI) 94 (88 to 98)[2] Positive LR (95% CI) 11.69 (5.59 to 24.42)[2] Negative LR (95% CI) 0.34 (0.25 to 0.48)[2] Ovarian volume <4cm³ to distinguish menopausal women from all other women	Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Unclear - patient recruitment not described in detail. Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1. A Could the selection of patients have introduced bias? LOW RISK 1. B Is there concern that the included patients do not match the review question? LOW CONCERN Index test Were the index test

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
follicle count and age in predicting menopausal status in healthy women. Study dates July - November 2002 Source of funding Not reported	ovarian masses larger than 20mm diameter, pregnancy, polycystic ovary syndrome, inflammatory pelvic disease, gonadal dysgenesis, premature menopause and undetermined menopausal status.			Sensitivity, % (95% CI) 73 (61 to 83)[1] Specificity, % (95% CI) 81 (73 to 88)[1] Positive LR (95% CI) 3.85 (2.60 to 5.71)[2] Negative LR (95% CI) 0.33 (0.22 to 0.49)[2] Antral follicle count cut-point ≤ 2 follicles to distinguish menopausal women from all other women Sensitivity, % (95% CI) 89 (79 to 95)[1] Specificity, % (95% CI) 42 (33 to 51)[1] Positive LR (95% CI) 1.53 (1.29 to 1.82)[2] Negative LR (95% CI) 0.27 (0.13 to 0.53)[2] [1] Point estimate only provided in article. 95% CI calculated by the NCC WCH technical team from data reported. [2] Calculated by the NCC WCH technical team from data reported in the article.	results interpreted without knowledge of the results of the reference standard? Unclear - two measures utilised ovarian ultrasonography which involves some subjectivity in reporting images. If the sonographer was not blinded this may have the potential to introduce bias. If a threshold was used, was it pre-specified? No - a variety of cut-points were assessed in the article to identify the optimum threshold. 2. A Could the conduct or interpretation of the index test have introduced bias? UNCLEAR 2. B Is there concern that the index test, its conduct or interpretation differ from the review question? LOW CONCERN Reference standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					results of the index test? Yes 3. A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3. B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW CONCERN Flow and timing Was there an appropriate interval between index test and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4. A Could the patient flow have introduced bias? LOW RISK Limitations Recruitment of participants was not described in detail. The authors do not described whether the individual performing the ultrasonography was blinded to menopausal status.

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					As sonography involves subjective interpretation of images, a lack of blinding may introduce bias. A variety of possible cut-points for antral follicle count are presented in the paper, rather than using a pre-specified threshold. Other information 20% of women past or current HRT users. No comment on inclusion/exclusion of women with surgical menopause (hysterectomy).
Full citation Gold,E.B., Sternfeld,B., Kelsey,J.L., Brown,C., Mouton,C., Reame,N., Salamone,L., Stellato,R., Relation of demographic and lifestyle factors to symptoms in a multi-racial/ethnic population of women 40-55 years of age, American Journal of Epidemiology, 152, 463-473, 2000 Ref Id 266916 Country/ies where the study was carried out United States Study type	Sample size N = 12396 total For the purposes of this analysis women with surgical menopause were excluded, n = 1988. Therefore N = 10408 after exclusions. n = 4497 premenopausal n = 4158 perimenopausal n = 1753 postmenopausal Characteristics Age range: 40 - 55 Smoking status: · 23.3% past history of smoking · 23.4% current smokers Ethnicity: African American: 29.5% Caucasian: 46.5% Japanese: 5.7% Chinese: 4.4% Hispanic: 13.8% Inclusion Criteria Women aged between 40 and 55 years. Exclusion Criteria Women whose menstrual periods had stopped because of	Tests Self-reported symptoms reported included Hot flushes/night sweats Urine leakage Vaginal dryness Difficult sleep Stiff/sore Heart pounding Forgetfulness Definitions used Postmenopausal: menses had stopped for at least 12 months without surgery Perimenopausal: menses had occurred in the past 3 months but had become less predictable (early perimenopause) or menses had occurred in the past 12 months but not in the last 3 months (late perimenopause) Premenopausal: menses had	Methods Baseline data on the number of women who had experienced each of the menopause-related symptoms in the previous two weeks was collected by computer-assisted telephone interviews or in-person interviews	Results Hot flashes/night sweats in previous 2 weeks to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 49 (46 to 51)[1] Specificity, % (95% CI) 60 (59 to 62)[1] Positive LR (95% CI) 1.22 (1.15 to 1.30)[1] Negative LR (95% CI) 0.85 (0.81 to 0.90)[1] Heart pounding in previous 2 weeks to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 20 (18 to 21)[1] Specificity, % (95% CI) 80 (79 to 81)[1] Positive LR (95% CI) 0.97 (0.86 to 1.08)[1] Negative LR (95% CI) 1.01	Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1. A Could the selection of patients have introduced bias? LOW RISK 1. B Is there concern that the included patients do not match the review question? LOW CONCERN Index test Were the index test results interpreted

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Case-series Aim of the study To investigate the relation of sociodemographic and lifestyle factors to a number of specific symptoms or conditions in a large, multiethnic, community-based sample of women from across the USA. Study dates Original cross sectional study was carried out from 1995 to 1997 Source of funding The orginal study was funded by the National Institute on Aging, the National Institute of Nursing research, and the Office on Women's Health of the National Institutes of Health	medication, radiotherapy, pregnancy or lactation, or extreme weight change who reported use of exogenous female hormones in the past three months who reported their race/ethnicity as mixed/other	occurred in the past 3 months with no decrease in predictability		(0.98 to 1.04)[1] Hot flashes/night sweats in previous 2 weeks to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 49 (46 to 51)[1] Specificity, % (95% CI) 81 (79 to 82)[1] Positive LR (95% CI) 2.52 (2.33 to 2.72)[1] Negative LR (95% CI) 0.64 (0.61 to 0.67)[1] Heart pounding in previous 2 weeks to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 20 (18 to 21)[1] Specificity, % (95% CI) 85 (84 to 86)[1] Positive LR (95% CI) 1.33 (1.18 to 1.49)[1] Negative LR (95% CI) 0.94 (0.92 to 0.97)[1] Hot flashes/night sweats in previous 2 weeks to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 49 (46 to 51)[1] Specificity, % (95% CI) 71 (70 to 72)[1] Positive LR (95% CI) 1.67 (1.58 to 1.77)[1] Negative LR (95% CI) 0.72 (0.69 to 0.76)[1] Heart pounding in previous 2 weeks to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 20 (18 to 21)[1] Specificity, % (95% CI) 83 (82 to 83)[1]	without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? n/a 2. A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2. B Is there concern that the index test, its conduct or interpretation differ from the review question? LOW CONCERN Reference standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3. A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3. B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW CONCERN

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Positive LR (95% CI) 1.13 (1.01 to 1.25)[1] Negative LR (95% CI) 0.97 (0.95 to 1.00)[1] Hot flashes/night sweats in previous 2 weeks to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 40 (38 to 41)[1] Specificity, % (95% CI) 51 (49 to 54)[1] Positive LR (95% CI) 0.82 (0.77 to 0.87)[1] Negative LR (95% CI) 1.17 (1.12 to 1.24)[1] Heart pounding in previous 2 weeks to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 20 (19 to 21)[1] Specificity, % (95% CI) 80 (79 to 82)[1] Positive LR (95% CI) 1.03 (0.92 to 1.16)[1] Negative LR (95% CI) 0.99 (0.96 to 1.02)[1] Hot flashes/night sweats in previous 2 weeks to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 40 (38 to 41)[1] Specificity, % (95% CI) 81 (79 to 82)[1] Positive LR (95% CI) 2.05 (1.91 to 2.20)[1] Negative LR (95% CI) 0.75 (0.73 to 0.77)[1] Heart pounding in previous 2 weeks to distinguish perimenopausal women from premenopausal women	Flow and timing Was there an appropriate interval between index test and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4. A Could the patient flow have introduced bias? LOW RISK Limitations Other information For the purposes of this review data reported for early perimenopausal and late perimenopausal women was combined into one category of perimenopausal. Women with surgical menopause (periods ceased due to hysterecomy and/or oophorectomy) were omitted from the analysis for the purposes of this review. HRT users were excluded from the study.

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Sensitivity, % (95% CI) 20 (19 to 21)[1] Specificity, % (95% CI) 85 (84 to 86)[1] Positive LR (95% CI) 1.37 (1.25 to 1.51)[1] Negative LR (95% CI) 0.94 (0.92 to 0.95)[1] Hot flashes/night sweats in previous 2 weeks to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 40 (38 to 41)[1] Specificity, % (95% CI) 72 (71 to 73)[1] Positive LR (95% CI) 1.44 (1.36 to 1.52)[1] Negative LR (95% CI) 0.83 (0.81 to 0.86)[1] Heart pounding in previous 2 weeks to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 20 (19 to 21)[1] Specificity, % (95% CI) 84 (83 to 85)[1] Positive LR (95% CI) 1.26 (1.16 to 1.37)[1] Negative LR (95% CI) 0.95 (0.93 to 0.97)[1] [1] Calculated by the NCC WCH technical team from data reported in the article	
Full citation Henrich,J.B., Hughes,J.P., Kaufman,S.C., Brody,D.J., Curtin,L.R., Limitations of follicle-stimulating hormone in assessing menopause status:	Sample size N = 576 after exclusions (see below) n = 304 premenopausal n = 93 perimenopausal n = 179 postmenopausal Characteristics Population based sample of women aged 35 to 60 years. Mean age, total (SE) = 45.8 (0.4), range 35-60	Tests Serum FSH level measured by microparticle enzyme immunoassay Definitions used Premenopausal: menses occurred regularly, or were "usually irregular" but had occured within the last 12 months	Methods Participants completed a reproductive health questionnaire administered as a face to face interview. Serum FSH and LH were also collected.	Results FSH level to distinguish perimenopause from reproductive stage: cut-point 13mIU/mL Sensitivity, % (95% CI) 67 (50 to 81) Specificity, % (95% CI) 88 (81 to 92) Positive LR (95% CI) 5.72	Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
findings from the National Health and Nutrition Examination Survey (NHANES 1999-2000)*, Menopause, 13, 171-177, 2006 Ref Id 267109 Country/ies where the study was carried out USA Study type Case-series Aim of the study To assess the efficacy of FSH levels in distinguishing among women in the reproductive, menopause transition and postmenopausal stages. Study dates 1999-2000 Source of funding National Institute of Child Health and Human Development, NIH Centers for Disease Control and Prevention, National Center for Health Statistics	Mean age, premenopausal (SE) 41.4 (0.3), range 35-52 Mean age, perimenopausal (SE) 49.1 (0.7), range 38-60 Mean age, postmenopausal (SE) 53.4 (0.4) 40-60 Ethnicity: 67.2% non-hispanic white, 11.8% non-hispanic black, 6.4% Mexican American 21.8% current smokers Mean BMI (SE) 28.8 (0.5) Inclusion Criteria Women aged 35-60 years. Exclusion Criteria Pregnancy, breast feeding, current users of Depo-Provera or oral contraceptive pill, surgical or medical amenorrhoea, or could not provide useful information about menstrual history.	Perimenopausal: menses had been irregular in the past 12 months, with such irregularity reportedly due to "going/gone through the menopause" Postmenopausal: last menstrual period took place ≥12 months earlier, was attributed to the menopause and was not surgically induced		(4.08 to 8.01)[1] Negative LR (95% CI) 0.37 (0.28 to 0.49)[1] FSH level to distinguish postmenopause from perimenopause: cut-point 45mIU/mL Sensitivity, % (95% CI) 74 (60 to 84) Specificity, % (95% CI) 71 (52 to 84) Positive LR (95% CI) 2.54 (1.83 to 3.53)[1] Negative LR (95% CI) 0.37 (0.28 to 0.49)[1] LR = likelihood ratio [1] Calculated by the NCC WCH technical team from data reported in the article	inappropriate exclusions? Yes 1. A Could the selection of patients have introduced bias? LOW RISK 1. B Is there concern that the included patients do not match the review question? LOW CONCERN Index test Were the index test results interpreted without knowledge of the results of the reference standard? Unclear - blinding of investigators was not described, but level of FSH should not depend on subjective interpretation. If a threshold was used, was it pre-specified? No - appropriate threshold was deteremined during the course of the study. 2. A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2. B Is there concern that the index test, its conduct or interpretation differ from the review question? LOW CONCERN

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Reference standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3. A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3. B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW CONCERN Flow and timing Was there an appropriate interval between index test and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4. A Could the patient flow have introduced bias? LOW RISK Limitations Whether the index

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					test (FSH) was interpreted without knowledge of menopausal status is not clear. However, the index test in this study involved a laboratory measurment of FSH level, and therefore there is a low risk of bias being introduced due to a lack of blinding. No pre-specified threshold for FSH level was given. Instead, the authors determined the optimum cut-point as part of the study. Other information 12.5% of participants were current users of HRT. Women with surgical menopause were excluded.
Full citation Johnson,B.D., Merz,C.N., Braunstein,G.D., Berga,S.L., Bittner,V., Hodgson,T.K., Gierach,G.L., Reis,S.E., Vido,D.A., Sharaf,B.L., Smith,K.M., Sopko,G., Kelsey,S.F., Determination of menopausal status in women: the NHLBI-sponsored	Sample size N = 515 n = 507 after exclusions (see below) n = 186 after excluding women automatically classed as pos menopausal (≥55 years and amenorrhoea for a year or more) - these women were not included in the populations for analysis of diagnostic accuracy. n = 122 premenopausal n = 33 perimenopausal n = 31 postmenopausal Characteristics Age range 21 to 55 Ethnicity: 72% white 50% obese 30% current smokers	Tests Blood levels of estradiol and FSH taken at any phase of the menstrual cycle. Reproductive status questionnaire completed by participants. Definitions used Classification of women as pre, peri and postmenopausal was performed by expert consensus opinion by the WISE hormone committee, comprising two reproductive endocrinologists, two clinical cardiologists, a statistician and a nurse, as follows: "Each member of the hormone	Methods Menopausal status (pre, peri or menopausal) was allocated by expert consensus (as described above) after review of individual patient data by a committee of 6 experts. This was then taken as the reference standard, against which the diagnostic algorithms were compared. Two established	Results Diagnostic accuracy measures are presented separately for women with and without a hysterectomy. Menstrual algorithm to distinguish postmenopausal women from all other women (women with hysterectomy excluded) Sensitivity, % (95% CI) 90 (70 to 99)[1] Specificity, % (95% CI) 98 (93 to 99)[1] Positive LR (95% CI) 36.19 (11.74 to 111.58)[1]	Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Unclear - recruitment not described in detail, but all individuals were under investigation for possible myocardial ischaemia. Was a case-control design avoided? Yes Did the study avoid

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Women's Ischemia Syndrome Evaluation (WISE) Study, Journal of Women's Health, 13, 872-887, 2004 Ref Id 229576 Country/ies where the study was carried out USA Study type Case-series Aim of the study To develop a new algorithm for the diagnosis of perimenopause and menopause, using hormonal measurements in addition to menstrual cycle regularity and age. Study dates Not reported Source of funding National Heart Lung and Blood Institute	27% known coronary artery disease 69% had at least two cardiac risk factors 24% had previous hysterectomy with ovarian preservation. Inclusion Criteria Women undergoing clinically ordered angiogram for suspected myocardial infarction. No current use of oral contraceptive pill or hormone replacement therapy. Exclusion Criteria Missing data on at least one relevant reproductive variable (current HRT use, BSO, hysterectomy, menstrual history)	committee examined the complete data available for each patient, including the patient's age, BMI, smoking, whether she had a hysterectomy with or without bilateral or unilateral oophorectomy, whether the cycles (if present) were regular or irregular, months or days since last menstrual period, and levels of serum FSH, LH, estradiol, estrone and progesterone. Each member then classified the patient into premenopausal (follicular, luteal or midcycle, if possible), postmenopausal, perimenopausal, or unclear, including a group of women were eventually classified as having hypothalamic hypoestrogenemia or hypothalamic amenorrhoea or both. Following these preliminary classifications, the committee as a group reviewed and adjudicated menopausal status for each of 186 individual women who could not definitely be classified as postmenopausal"	algorithms were used (menstrual and historical), and a new algorithm was developed (hormonal). 1. Menstrual algorithm: postmenopausal defined as 12 months amenorrhoea perimenopausal defined as amenorrhoea for 3-12 months all other women defined as premenopausal 2. Historical algorithm: post menopausal defined as amenorrhoea for ≥ 12 months plus a) known bilateral salpingoophorectomy ; b) age ≥ 55 years; c) age <55 years but uterus intact. All other women (menstruation within last 12 months, or no menstruation within 12 months but previous hysterectomy with ovarian conservation and age <55 years) defined as premenopausal. This algorithm was unable to classify women as perimenopausal. 3. Hormonal algorithm: two arms,	Negative LR (95% CI) 0.09 (0.03 to 0.37)[1] Historical algorithm to distinguish postmenopausal women from all other women (women with hysterectomy excluded) Sensitivity, % (95% CI) 90 (70 to 99)[1] Specificity, % (95% CI) 98 (93 to 99)[1] Positive LR (95% CI) 36.19 (11.74 to 111.58)[1] Negative LR (95% CI) 0.09 (0.03 to 0.37)[1] Hormonal algorithm to distinguish postmenopausal women from all other women (women with hysterectomy excluded) Sensitivity, % (95% CI) 90 (70 to 99)[1] Specificity, % (95% CI) 100 (97 - 100)[1] Positive LR (95% CI) ∞ (NC)[2] Negative LR (95% CI) 0.10 (0.03 to 0.36)[1] Menstrual algorithm to distinguish perimenopausal women from all other women (women with hysterectomy excluded) Sensitivity, % (95% CI) 96 (78 to 100)[1] Specificity, % (95% CI) 98 (94 to 100)[1] Positive LR (95% CI) 56.43 (14.24 to 223.63)[1] Negative LR (95% CI) 0.04 (0.01 to 0.30)[1] Hormonal algorithm to distinguish perimenopausal women from all other women (women with hysterectomy excluded) Sensitivity, % (95% CI) 91	inappropriate exclusions? Yes 1. A Could the selection of patients have introduced bias? LOW RISK 1. B Is there concern that the included patients do not match the review question? HIGH RISK Index test Were the index test results interpreted without knowledge of the results of the reference standard? Unclear - however, measurement of hormone levels should not be influenced by subjectivity, therefore unlikely to introduce bias. If a threshold was used, was it pre-specified? No - an appropriate hormonal algorithm was devised during the course of the study with thresholds for allocation determined as part of the research. 2. A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2. B Is there concern that the index test, its conduct or

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
			for women with last menstrual period (LMP) within 12 months, and LMP more than 12 months ago. LMP within 12 months: premenopausal if a) regular periods and LMP < 3 months, with FSH < 20 or; b) irregular periods or LMP ≤ 6 months with FSH < 10 and estradiol < 200. postmenopausal if LMP > 6 months, age > 50 and FSH >30. perimenopausal for all other women - including a) regular periods and LMP <3 months with FSH ≥20 or; b) irregular periods or LMP ≥ 3 months with FSH <10 and either LMP > 6 months or estradiol ≥ 200 or; c) irregular periods or LMP ≥ 3 months with FSH ≥10, but not yet reaching criteria for menopause (FSH > 30, plus age > 50, plus LMP >6 months). LMP more than 12 months ago: premenopausal if previous hysterectomy and a) FSH < 10 or; b) FSH	(72 to 99)[1] Specificity, % (95% CI) 98 (94 to 100)[1] Positive LR (95% CI) 53.87 (13.55 to 214.11)[1] Negative LR (95% CI) 0.09 (0.02 to 0.33)[1] Menstrual algorithm to distinguish postmenopausal women from all other women (including women with hysterectomy) Sensitivity, % (95% CI) 94 (79 to 99)[3] Specificity, % (95% CI) 76 (69 to 83)[3] LR+ (95% CI) 3.92 (2.92 to 5.27)[1] LR- (95% CI) 0.08 (0.02 to 0.32)[1] Historical algorithm to distinguish postmenopausal women from all other women (including women with hysterectomy) Sensitivity, % (95% CI) 59 (39 to 75)[3] Specificity, % (95% CI) 97 (93 to 99)[3] LR+ (95% CI) 18.00 (7.23 to 44.84)[1] LR- (95% CI) 0.43 (0.29 to 0.66)[1] Hormonal algorithm to distinguish postmenopausal women from all other women (including women with hysterectomy) Sensitivity, % (95% CI) 85 (66 to 95)[3] Specificity, % (95% CI) 99 (95 to 100)[3] LR+ (95% CI) 65.00 (16.26 to 259.82)[1] LR- (95% CI) 0.16 (0.07 to 0.36)[1]	interpretation differ from the review question? LOW RISK Reference standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3. A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3. B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW CONCERN Flow and timing Was there an appropriate interval between index test and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
			= 10-20 with estradiol ≥50. postmenopausal if a) previous BSO or age ≥55 years or; b) estradiol <50 and FSH ≥20 or; c) previous hysterectomy and FSH >30 and estradiol <50. perimenopausal if previous hysterectomy and a) estradiol ≥200 and age >45 or; b) FSH = 10-20 and estradiol <50 or; c) FSH = 20-30 or; d) FSH >30 and estradiol ≥50. This algorithm also contained a branch for "hand classification" where the individual patient data and circumstances would need to be scrutinised to allow correct classification - women were assigned to this category if they had an LMP more than 12 months ago, no hysterectomy but estradiol ≥50 or FSH <20.	Menstrual algorithm to distinguish perimenopausal women from all other women (including women with hysterectomy) Sensitivity, % (95% CI) 6 (1 to 20)[3] Specificity, % (95% CI) 99 (95 to 100)[3] Positive LR (95% CI) 4.64 (0.68 to 31.74)[1] Negative LR (95% CI) 0.95 (0.87 to 1.04)[1] Hormonal algorithm to distinguish perimenopausal women from all other women (including women with hysterectomy) Sensitivity, % (95% CI) 88 (72 to 97)[3] Specificity, % (95% CI) 97 (93 to 99)[3] Positive LR (95% CI) 26.89 (11.25 to 64.27)[1] Negative LR (95% CI) 0.13 (0.05 to 0.31)[1] LR = likelihood ratio NC = not calculable [1] Calculated by the NCC WCH technical team from data reported in the article [2] Specificity 100%, therefore positive LR = infinity and 95% CI not calculable. [3] Point estimate reported in the paper. 95% CI calculated by the NCC WCH technical team	4. A Could the patient flow have introduced bias? LOW RISK Limitations Recruitment not described in detail - only that all women were undergoing investigation for possible myocardial ischaemia. This population may therefore differ from the general population of women, and there is significant concern that the included patients do not match the review question. Knowledge of the reference standard during the conduct of the index test is not described. However, the algorithm presents fixed options to determine menopausal status and therefore it is unlikely that women would be misclassified because of a lack of blinding. A pre-determined "threshold" was not described. The authors used the data to produce a hormonal algorithm to classify women. Other information All women in study population were under investigation

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					for possible myocardial ischaemia. Separate analysis was conducted for classification of women without a hysterectomy, and classification of all women. This was reported as due to the "inherently low agreement for women with hysterectomy". Users of HRT were excluded from the study.
Full citation Kapur,P., Sinha,B., Pereira,B.M., Measuring climacteric symptoms and age at natural menopause in an Indian population using the Greene Climacteric Scale, Menopause, 16, 378-384, 2009 Ref Id 267312 Country/ies where the study was carried out India Study type Case-series Aim of the study To establish the age at onset of natural menopause and the prevalence of symptoms and identify any socio-demographic,	Sample size N=129 Premenopause, n= 70; Early post-menopause: n=33 (1-5 yr after last menstrual cycle) Late post-menopause: n=26 (> 5 yr after last menstrual cycle) Characteristics Age (range): 30-65 years Menopausal group, n (%): Premnopause: 70 (54.26) Early postmenopause (1-5 yr): 33 (25.58) Late postmenopause (>5yr): 26 (20.15) BMI, n (%) Underweight: 6 (4.65) Normal: 87 (67.44) Overweight: 30 (23.25) Obese: 6 (4.65) Socioeconomic status, n (%): Poor: 29 (22.48) Middle: 100 (77.5) Inclusion Criteria Not reported Exclusion Criteria	Tests -The Greene Climacteric Scale was used to assess the nature and severity of occurrence of climacteric symptoms among the selected participants; Definitions used Menopausal status of the participants was defined using World Health Organization (WHO) criteria. Premenopause: women who had regular menstruation cycles during the last 3 months Postmenopause: women who had no cycle in the previous 12 months Early and late menopause status was defined using the STRAW staging system;	Methods -Women self-related their menopausal symptoms using the Greene Climacteric Scale; prevalence of symptoms was documented in groups.	Results Symptoms of hot flushes to distinguish early Postmenopausal (1-5yr) from pre-menopausal women: Sensitivity: n/N, % (95%CI): 19/33, 58 (40 to 74) Specificity: n/N, %, (95%CI): 58/70, 83 (74 to 92) Positive LR (95% CI): 3.36 (1.86 to 6.07) Negative LR (95%CI): 0.51 (0.34 to 0.77) Symptoms of hot flushes to distinguish late Postmenopausal (>5 yr) women from pre-menopausal women: Sensitivity: n/N, % (95%CI): 12/26, 46 (27 to 64) Specificity: n/N, %, (95%CI): 58/70, 83 (71 to 92) Positive LR (95% CI): 2.69 (1.39 to 5.22) Negative LR (95%CI): 0.65 (0.44 to 0.94)	Study quality - QUADAS 2 checklist Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1.A Could the selection of patients have introduced bias? LOW RISK OF BIAS 1.B Is there concern that the included patients do not match the review question? LOW CONCERN Index Test Were the index test results interpreted without knowledge of the results of the

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
physical, or other factors that may influence the onset of menopause among women in the Haridwar district of Uttarakhand, a state located in northern India. Study dates Not reported Source of funding The University Grants Commission, Government of India	Women who -1) had surgical menopause; 2) had serious illness like hyptertension, fibroids, migranies, diabetes, spondylitis; 3) were users of any type of medication for menopause; 4) were unable to understand the questionnaire; and 5) returned forms with missing information.			Symptoms of night sweating to distinguish early Postmenopausal (1-5 yr) women from premenopausal women: Sensitivity: n/N, % (95%CI): 12/26, 46 (27 to 64) Specificity: n/N, %, (95%CI): 64/70, 91.4 (85 to 98) Positive LR (95% CI): 5.38 (2.25 to 12.85) Negative LR (95%CI): 0.59 (0.41 to 0.85) Symptoms of night sweating to distinguish late Postmenopausal women from Premenopausal women (>5 yr): Sensitivity: n/N, % (95%CI): 8/26, 31 (13 to 49) Specificity: n/N, %, (95%CI): 64/70, 91.4 (85 to 98) Positive LR (95% CI): 3.59 (1.38 to 9.36) Negative LR (95%CI): 0.76 (0.58 to 0.99) (LR = likelihood ratio Calculated by the NCC WCH technical team from data reported in the article)	reference standard? Yes If a threshold was used, was it pre-specified? N/A 2.A Could the conduct or interpretation of the index test have introduced bias? LOW RISK OF BIAS 2.B Is there concern that the index test, its conduct, or interpretation differ from the review question? LOW CONCERN Reference Standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3.A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3.B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW RISK Flow and Timing Was there an

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					appropriate interval between index test(s) and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4.A Could the patient flow have introduced bias? LOW RISK
Full citation Shin,S.Y., Lee,J.R., Noh,G.W., Kim,H.J., Kang,W.J., Kim,S.H., Chung,J.K., Analysis of serum levels of anti-Mullerian hormone, inhibin B, insulin-like growth factor-I, insulin-like growth factor binding protein-3, and follicle-stimulating hormone with respect to age and menopausal status, Journal of Korean Medical Science, 23, 104-110, 2008 Ref Id 268528 Country/ies where the study was carried out Korea Study type Case-control study Aim of the study To determine which	Sample size N = 144 total n = 33 postmenopausal (physiologic menopause for at least one year) n = 111 pre-menopausal (regular menstrual cycles of 24-35 days) Characteristics Mean age (range) of premenopausal women = 31 (20-49) years Mean age (range) of postmenopausal women = 56 (50-59) years Inclusion Criteria All required to have BMI of 19-26kg/m², both ovaries present, no use of hormonal medication, no evidence of polycystic ovarian syndrome, normal prolactin and thyroid stimulating hormone levels and no medical or reproductive disorders (including any history of subfertility). Exclusion Criteria None described	Tests Serum levels of FSH measured by immunoradiometric assay and estrogen with radioimmunoassay. AMH and inhibin B measured with ELISA. Definitions used Premenopausal: regular menstrual cycles of 24-35 days Postmenopausal: physiologic menopause for at least one year	Methods Blood collected by venepuncture on cycle day 3 for menstruating women, or randomly for postmenopausal women.	Results FSH cut-point >22.3mIU/mL to distinguish menopausal from premenopausal women: Sensitivity, % (95% CI) 99 (89 to 100)[1] Specificity, % (95% CI) 97 (92 to 99)[1] Positive LR (95% CI) 33.04 (11.47 to 95.21)[2] Negative LR (95% CI) 0.01 (0.00 to 0.33)[2] AMH cut-point <0.5ng/mL to distinguish menopausal from premenopausal women Sensitivity, % (95% CI) 92 (80 to 98)[1] Specificity, % (95% CI) 97 (92 to 99)[1] Positive LR (95% CI) 30.88 (10.62 to 89.83)[2] Negative LR (95% CI) 0.08 (0.03 to 0.26)[2] Estradiol cut-point <34.5pg/mL to distinguish menopausal from premenopausal women: Sensitivity, % (95% CI) 84 (68 to 93)[1]	Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Unclear - recruitment not described clearly. Was a case-control design avoided? No Did the study avoid inappropriate exclusions? Yes 1. A Could the selection of patients have introduced bias? HIGH RISK 1. B Is there concern that the included patients do not match the review question? HIGH CONCERN Index test Were the index test results interpreted without knowledge of the results of the reference standard? Unclear - but

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
of several serum markers best reflects the reproductive ageing process in women, including AMH, inhibin B, estradiol and FSH. Study dates Not reported Source of funding Korean Science and Engineering Foundation, Seoul National University College of Medicine				Specificity, % (95% CI) 97 (92 to 99)[1] Positive LR (95% CI) 28.23 (9.65 to 82.58)[2] Negative LR (95% CI) 0.17 (0.08 to 0.36)[2] Inhibin B cut-point <0.4pg/mL to distinguish menopausal from premenopausal women: Sensitivity, % (95% CI) 91 (80 to 98)[1] Specificity[1], % (95% CI) 100 (97 to 100)[1] Positive LR (95% CI) ∞ (NC)[23] Negative LR (95% CI) 0.09 (0.03 to 0.27)[2] LR = likelihood ratio NC = not calculable [1] Point estimate presented in paper, confidence intervals calculated by the NCC WCH technical team from data reported in the article [2] Calculated by the NCC WCH technical team from data reported in the article [3] Specificity = 100%, therefore positive LR = infinity, and 95% CI not calculable [3] specificity 100%, therefore positive likelihood ratio = infinity, and 95% CI not calculable	objective testing of serum markers therefore unlikely to be subject to interpretation bias. If a threshold was used, was it pre-specified? No - the appropriate threshold was determined in the study. 2. A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2. B Is there concern that the index test, its conduct or interpretation differ from the review question? LOW CONCERN Reference standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3. A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3. B Is there concern that the target condition as defined by the reference

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					standard does not match the review question? LOW CONCERN Flow and timing Was there an appropriate interval between index test and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4. A Could the patient flow have introduced bias? LOW RISK Limitations No description of recruitment in the article. The majority of premenopausal women in this study were aged under 40 (81 of 111 premenopausal women). Therefore this population is likely to be less applicable to the population in whom a test for menopause or perimenopause would be used in clinical practice. Unclear if index test was interpreted without knowledge of

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					the reference standard, but laboratory values are reported for the index tests, which should not be at risk of misinterpretation and bias. No predetermined threshold was reported; instead the optimum cut-point for the tests was determined in the study. Other information Only women with regular cycles included in premenopausal group. Mean age was significantly different between the two groups. HRT users were excluded from the study. Whether women with surgical menopause were included is unclear.
Full citation Sierra,B., Hidalgo,L.A., Chedraui,P.A., Measuring climacteric symptoms in an Ecuadorian population with the Greene Climacteric Scale, Maturitas, 51, 236-245, 2005 Ref Id 227336 Country/ies where the study was	Sample size N=385 Characteristics Age, mean (SD): 47.6 (5.5) Menopausal status in percentages: Pre-menopausal: 38.9% Peri-menopausal: 28.8% Postmenopausal: 32.3% Education: Schooling < 12 years: 67.3% Inclusion Criteria Not reported; Exclusion Criteria -Hysterectomized women -those who couldn't fill out the Greene Climacteric	Tests Definitions used Premenopause: women having regular menses and >= 12 menses during the last 12 months Perimenopause: irregular menses, less than 12 menses during the last 12 months; Postmenopause: no more menses in the last 12 months	Methods Survey by questionnaire using the Greene Climacteric Scale	Results Symptoms of heart beating to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI): 64 (2 to 10) Specificity, % (95% CI): 95 (91 to 99) Positive LR (95% CI): 1.44 (0.48 to 1.28) Negative LR (95% CI): 0.97 (0.92 to 1.04) Symptoms of heart beating to distinguish	Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1. A Could the selection of patients have introduced bias? LOW RISK

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
carried out Ecaudor Study type Case-series Aim of the study To measure climacteric symptoms in a low socio-economic Ecuadorian population with the Greene Climacteric Scale and determine risk factors involved with higher scorings. Study dates November 2001 to April 2002 Source of funding the Foundation for Health and Well Being, Ecuador	Scale due to illiteracy			postmenopausal women from premenopausal women Sensitivity, % (95% CI): 64 (2 to 10) Specificity, % (95% CI): 99 (98 to 100) Positive LR (95% CI): 9.6 (1.21 to 75.8) Negative LR (95% CI): 0.94 (0.89 to 0.98) Symptoms of heart beating to distinguish postmenopausal women from all other women Sensitivity, % (95% CI): 64 (2 to 10) Specificity, % (95% CI): 97 (95 to 99) Positive LR (95% CI): 2.8 (0.99 to 7.9) Negative LR (95% CI): 0.95 (0.91 to 1.00) Symptoms of heart beating to distinguish peri from postmenopausal women: Sensitivity, % (95% CI): 4 (0 to 8) Specificity, % (95% CI): 93 (89 to 97) Positive LR (95% CI): 0.69 (0.23 to 2.05) Negative LR (95% CI): 1.02 (0.96 to 1.08) Symptoms of heart beating to distinguish peri from premenopausal women Sensitivity, % (95% CI): 4 (0 to 8) Specificity, % (95% CI): 99 (98 to 100) Positive LR (95% CI): 6.6 (0.78 to 56.1)	1. B Is there concern that the included patients do not match the review question? LOW CONCERN Index test Were the index test results interpreted without knowledge of the results of the reference standard? N/A If a threshold was used, was it pre-specified? No - 2. A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2. B Is there concern that the index test, its conduct or interpretation differ from the review question? LOW CONCERN Reference standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? N/A 3. A Could the reference standard, its conduct, or its interpretation have introduced bias?

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Negative LR (95% CI): 0.96 (0.92 to 1.00) Symptoms of heart beating to distinguish peri from all other women Sensitivity, % (95% CI): 4 (0 to 8) Specificity, % (95% CI): 0.96 (94 to 98) Positive LR (95% CI): 1.35 (0.46 to 3.95) Negative LR (95% CI): 0.98 (0.94 to 1.03) Symptoms of hot flashes to distinguish post from perimenopausal women: Sensitivity, % (95% CI): 45 (36 to 53) Specificity, % (95% CI): 45 (36 to 54) Positive LR (95% CI): 0.82 (0.64 to 1.07) Negative LR (95% CI): 1.20 (0.93 to 1.55) Symptoms of hot flashes to distinguish post from premenopausal women: Sensitivity, % (95% CI): 45 (36 to 53) Specificity, % (95% CI): 50 (42 to 58) Positive LR (95% CI): 0.90 (0.70 to 1.17) Negative LR (95% CI): 1.08 (0.86 to 1.35) Symptoms of hot flashes to distinguish postmenopausal from all other women: Sensitivity, % (95% CI): 45 (36 to 53) Specificity, % (95% CI): 48 (42 to 54)	LOW RISK 3. B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW CONCERN Flow and timing Was there an appropriate interval between index test and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4. A Could the patient flow have introduced bias? LOW RISK

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Positive LR (95% CI): 0.87 (0.69 to 1.09) Negative LR (95% CI): 1.13 (0.9 to 1.39) Symptoms of hot flashes to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI): 54 (45 to 63) Specificity, % (95% CI): 54 (46 to 63) Positive LR (95% CI): 1.20 (0.93 to 1.56) Negative LR (95% CI): 0.83 (0.64 to 1.07) Symptoms of hot flashes to distinguish perimenopausal from premenopausal women Sensitivity, % (95% CI): 54 (45 to 63) Specificity, % (95% CI): 50 (42 to 58) Positive LR (95% CI): 1.09 (0.86 to 1.38) Negative LR (95% CI): 0.90 (0.96 to 1.17) Symptoms of hot flashes to distinguish perimenopausal from all other women Sensitivity, % (95% CI): 54 (45 to 63) Specificity, % (95% CI): 52 (46 to 58) Positive LR (95% CI): 1.14 (0.92 to 1.41) Negative LR (95% CI): 0.86 (0.68 to 1.09) Symptoms of night sweat to distinguish postmenopausal women from perimenopausal women	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Sensitivity, % (95% CI): 23 (15 to 30) Specificity, % (95% CI): 66 (57 to 74) Positive LR (95% CI): 0.68 (0.45 to 1.03) Negative LR (95% CI): 1.15 (0.98 to 1.36) Symptoms of night sweat to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI): 23 (15 to 30) Specificity, % (95% CI): 80 (74 to 86) Positive LR (95% CI): 1.20 (0.76 to 1.89) Negative LR (95% CI): 0.95 (0.83 to 1.07) Symptoms of night sweat to distinguish postmenopausal women from all other women Sensitivity, % (95% CI): 23 (15 to 30) Specificity, % (95% CI): 74 (69 to 79) Positive LR (95% CI): 0.91 (0.62 to 1.33) Negative LR (95% CI): 1.03 (0.91 to 1.16) Symptoms of night sweat to distinguish perimenopausal from postmenopausal women Sensitivity, % (95% CI): 33 (25 to 42) Specificity, % (95% CI): 76 (69 to 84) Positive LR (95% CI): 1.45 (0.92 to 2.18) Negative LR (95% CI): 0.86	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(0.73 to 1.01) Symptoms of night sweat to distinguish perimenopausal from premenopausal women Sensitivity, % (95% CI): 33 (25 to 42) Specificity, % (95% CI): 80 (74 to 86) Positive LR (95% CI): 1.74 (1.14 to 2.64) Negative LR (95% CI): 0.82 (0.70 to 0.95) Symptoms of night sweat to distinguish perimenopausal from all other women Sensitivity, % (95% CI): 33 (25 to 42) Specificity, % (95% CI): 78 (73 to 83) Positive LR (95% CI): 1.59 (1.13 to 2.25) Negative LR (95% CI): 0.83 (0.72 to 0.97)	
Full citation Williams,R.E., Kalilani,L., DiBenedetti,D.B., Zhou,X., Granger,A.L., Fehnel,S.E., Levine,K.B., Jordan,J., Clark,R.V., Frequency and severity of vasomotor symptoms among peri- and postmenopausal women in the United States, Climacteric, 11, 32-43, 2008	Sample size N = 4402 after exclusions (see below) n = 1267 premenopausal n = 432 perimenopausal n = 2703 postmenopausal Characteristics Age range: 40 to 65 years Smoking status: 34.5% Ethnicity: • 77.8% White, non-Hispanic • 11.3% Black/African-American, non-Hispanic • 7.5% Hispanic • 3.4% other non-Hispanic Inclusion Criteria Women aged between 40 and 65 years Exclusion Criteria	Tests The confidential self-administered survey consisted of 2 parts. Part 1 included baseline characteristics such as participant characteristics, menstrual history, severity of premenstrual symptoms, pregnancy history, Menopause Quality of Life Instrument (MENQOL) and other symptoms. Part 2 (completed by perimenopausal and postmenopausal women) included detailed assessment of menopausal symptoms, healthcare seeking and medication use.	Methods Number of women with the symptom in each stage (premenopausal, perimenopausal and postmenopausal)	Results Age ≥ 45 to distinguish menopausal women from perimenopausal women Sensitivity, % (95% CI) 95 (94 to 96)[1] Specificity, % (95% CI) 9 (7 to 12)[1] Positive LR (95% CI) 1.04 (1.01 to 1.08)[1] Negative LR (95% CI) 0.55 (0.39 to 0.77)[1] Age ≥ 50 to distinguish menopausal women from perimenopausal women Sensitivity, % (95% CI) 84 (83 to 85)[1] Specificity, % (95% CI) 47 (43 to 52)[1]	Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1. A Could the selection of patients have introduced bias? LOW RISK 1. B Is there concern that the included patients do not match

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Ref Id 269042 Country/ies where the study was carried out United States Study type Case-series Aim of the study The focus of this paper (part of a wider study) was to describe frequency and severity of vasomotor symptoms in detail for peri- and postmenopausal women age 40 - 65 years. Study dates April 1st to April 20th 2005 Source of funding GlaxoSmithKline funded the study	Women were excluded due to unknown menopausal status, missed periods for reasons other than menopause or hysterectomy (such as pregnancy in the last year, intrauterine device, chemotherapy, strenuous exercise, anorexia, or other medical condition that resulted in a lack of a menstrual period).	Information on vasomotor symptoms in the past 4 weeks was obtained from several questions as follows Hot flushes or flashes in the past month (yes/no) Night sweats in the past month (yes/no) In the past 4 weeks, how often did you have hot flashes (never, 1-3 days in the past month, 1-2 days a week, 3-4 days a week, 5-6 days a week, every day) In the past 4 weeks, how often did you have night sweats (never, 1-3 days in the past month, 1-2 days a week, 3-4 days a week, 5-6 days a week, every day) Definitions used Premenopausal: had a period every month for the past 12 months Perimenopausal: did not have a period every month but at least 1 period in the past 12 months Postmenopausal: did not have a period in the past 12 months		Positive LR (95% CI) 1.60 (1.46 to 1.75)[1] Negative LR (95% CI) 0.34 (0.30 to 0.38)[1] Age ≥ 55 to distinguish menopausal women from perimenopausal women Sensitivity, % (95% CI) 62 (60 to 64)[1] Specificity, % (95% CI) 89 (85 to 91)[1] Positive LR (95% CI) 5.44 (4.17 to 7.09)[1] Negative LR (95% CI) 0.43 (0.41 to 0.46)[1] Age ≥ 60 to distinguish menopausal women from perimenopausal women Sensitivity, % (95% CI) 33 (31 to 35)[1] Specificity, % (95% CI) 98 (96 to 99)[1] Positive LR (95% CI) 15.84 (8.28 to 30.30)[1] Negative LR (95% CI) 0.68 (0.66 to 0.71)[1] Occurrence of hot flashes or night sweats in the past four weeks to distinguish menopausal women from perimenopausal women Sensitivity, % (95% CI) 60 (58 to 62)[1] Specificity, % (95% CI) 25 (21 to 29)[1] Positive LR (95% CI) 0.80 (0.75 to 0.85)[1] Negative LR (95% CI) 1.60 (1.35 to 1.90)[1] Occurrence of night sweats in the past four weeks to distinguish menopausal women from perimenopausal women Sensitivity, % (95% CI) 44 (42 to 46)[1]	the review question? LOW CONCERN Index test Were the index test results interpreted without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? Yes 2. A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2. B Is there concern that the index test, its conduct or interpretation differ from the review question? LOW CONCERN Reference standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3. A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3. B Is there concern that the target

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Specificity, % (95% CI) 44 (39 to 49)[1] Positive LR (95% CI) 0.79 (0.72 to 0.86)[1] Negative LR (95% CI) 1.27 (1.14 to 1.42)[1] Age ≥ 45 to distinguish menopausal women from premenopausal women Sensitivity, % (95% CI) 95 (94 to 96)[1] Specificity, % (95% CI) 53 (50 to 56)[1] Positive LR (95% CI) 2.03 (1.92 to 2.16)[1] Negative LR (95% CI) 0.09 (0.08 to 0.11)[1] Age ≥ 50 to distinguish menopausal women from premenopausal women Sensitivity, % (95% CI) 84 (83 to 85)[1] Specificity, % (95% CI) 88 (86 to 90)[1] Positive LR (95% CI) 6.92 (5.96 to 8.03)[1] Negative LR (95% CI) 0.18 (0.17 to 0.20)[1] Age ≥ 55 to distinguish menopausal women from premenopausal women Sensitivity, % (95% CI) 62 (60 to 64)[1] Specificity, % (95% CI) 99 (98 to 99)[1] Positive LR (95% CI) 45.99 (28.66 to 73.81)[1] Negative LR (95% CI) 0.39 (0.37 to 0.41)[1] Age ≥ 60 to distinguish menopausal women from premenopausal women Sensitivity, % (95% CI) 33 (31 to 35)[1] Specificity, % (95% CI) 100 (99 to 100)[1]	condition as defined by the reference standard does not match the review question? LOW CONCERN Flow and timing Was there an appropriate interval between index test and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4. A Could the patient flow have introduced bias? LOW RISK Limitations Other information Women with hysterectomy were included in this study. It is unclear if current users of HRT were also included.

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Positive LR (95% CI) 69.69 (31.31 to 155.10)[1] Negative LR (95% CI) 0.67 (0.65 to 0.69)[1] Occurrence of hot flashes or night sweats in the past four weeks to distinguish menopausal women from premenopausal women Sensitivity, % (95% CI) 60 (58 to 62)[1] Specificity, % (95% CI) 60 (57 to 63)[1] Positive LR (95% CI) 1.50 (1.39 to 1.61)[1] Negative LR (95% CI) 0.67 (0.63 to 0.71)[1] Occurrence of night sweats in the past four weeks to distinguish menopausal women from premenopausal women Sensitivity, % (95% CI) 44 (42 to 46)[1] Specificity, % (95% CI) 70 (67 to 76)[1] Positive LR (95% CI) 1.47 (1.33 to 1.61)[1] Negative LR (95% CI) 0.80 (0.76 to 0.84)[1] Age ≥ 45 to distinguish menopausal women from all other women Sensitivity, % (95% CI) 95 (94 to 96)[1] Specificity, % (95% CI) 42 (40 to 44)[1] Positive LR (95% CI) 1.64 (1.57 to 1.71)[1] Negative LR (95% CI) 0.12 (0.10 to 0.14)[1] Age ≥ 50 to distinguish menopausal women from all other women Sensitivity, % (95% CI) 84 (83 to 85)[1]	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Specificity, % (95% CI) 78 (76 to 80)[1] Positive LR (95% CI) 3.75 (3.43 to 4.10)[1] Negative LR (95% CI) 0.21 (0.19 to 0.22)[1] Age ≥ 55 to distinguish menopausal women from all other women Sensitivity, % (95% CI) 62 (60 to 64)[1] Specificity, % (95% CI) 96 (95 to 97)[1] Positive LR (95% CI) 15.89 (12.52 to 20.16)[1] Negative LR (95% CI) 0.40 (0.38 to 0.42)[1] Age ≥ 60 to distinguish menopausal women from all other women Sensitivity, % (95% CI) 33 (31 to 35)[1] Specificity, % (95% CI) 99 (99 to 100)[1] Positive LR (95% CI) 37.38 (22.52 to 62.04)[1] Negative LR (95% CI) 0.68 (0.66 to 0.69)[1] Occurrence of hot flashes or night sweats in the past four weeks to distinguish menopausal women from all other women Sensitivity, % (95% CI) 60 (58 to 62)[1] Specificity, % (95% CI) 51 (47 to 53)[1] Positive LR (95% CI) 1.23 (1.16 to 1.30)[1] Negative LR (95% CI) 0.78 (0.73 to 0.84)[1] Occurrence of night sweats in the past four weeks to distinguish menopausal women from all other women Sensitivity, % (95% CI) 44	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(42 to 46)[1] Specificity, % (95% CI) 63 (61 to 66)[1] Positive LR (95% CI) 1.20 (1.11 to 1.30)[1] Negative LR (95% CI) 0.88 (0.84 to 0.93)[1] Age < 45 to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 9 (7 to 12)[1] Specificity, % (95% CI) 95 (94 to 96)[1] Positive LR (95% CI) 1.82 (1.29 to 2.56)[1] Negative LR (95% CI) 0.96 (0.93 to 0.99)[1] Age < 50 to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 47 (43 to 52)[1] Specificity, % (95% CI) 84 (83 to 85)[1] Positive LR (95% CI) 2.98 (2.61 to 3.40)[1] Negative LR (95% CI) 0.62 (0.57 to 0.68)[1] Age < 55 to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 89 (85 to 91)[1] Specificity, % (95% CI) 62 (60 to 64)[1] Positive LR (95% CI) 2.32 (2.18 to 2.46)[1] Negative LR (95% CI) 0.18 (0.14 to 0.24)[1] Age < 60 to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 98 (96 to 99)[1] Specificity, % (95% CI) 33	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(31 to 35)[1] Positive LR (95% CI) 1.46 (1.42 to 1.51)[1] Negative LR (95% CI) 0.06 (0.03 to 0.12)[1] Occurrence of hot flashes or night sweats in the past four weeks to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 75 (71 to 79)[1] Specificity, % (95% CI) 40 (38 to 42)[1] Positive LR (95% CI) 1.25 (1.17 to 1.33)[1] Negative LR (95% CI) 0.63 (0.53 to 0.74)[1] Occurrence of night sweats in the past four weeks to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 56 (51 to 61)[1] Specificity, % (95% CI) 56 (54 to 58)[1] Positive LR (95% CI) 1.27 (1.16 to 1.40)[1] Negative LR (95% CI) 0.79 (0.70 to 0.88)[1] Age ≥ 45 to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 91 (88 to 94)[1] Specificity, % (95% CI) 53 (50 to 56)[1] Positive LR (95% CI) 1.95 (1.82 to 2.08)[1] Negative LR (95% CI) 0.17 (0.13 to 0.23)[1] Age ≥ 50 to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 53	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(48 to 57)[1] Specificity, % (95% CI) 88 (86 to 90)[1] Positive LR (95% CI) 4.32 (3.64 to 5.14)[1] Negative LR (95% CI) 0.54 (0.49 to 0.60)[1] Age ≥ 55 to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 11 (9 to 15)[1] Specificity, % (95% CI) 99 (98 to 99)[1] Positive LR (95% CI) 8.45 (4.92 to 14.52)[1] Negative LR (95% CI) 0.90 (0.87 to 0.93)[1] Age ≥ 60 to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 2 (1 to 4)[1] Specificity, % (95% CI) 100 (99 to 100)[1] Positive LR (95% CI) 4.40 (1.58 to 12.29)[1] Negative LR (95% CI) 0.98 (0.97 to 1.00)[1] Occurrence of hot flashes or night sweats in the past four weeks to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 75 (71 to 79)[1] Specificity, % (95% CI) 60 (57 to 63)[1] Positive LR (95% CI) 1.87 (1.72 to 2.04)[1] Negative LR (95% CI) 0.42 (0.35 to 0.49)[1] Occurrence of night sweats in the past four weeks to distinguish perimenopausal women from premenopausal	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				women Sensitivity, % (95% CI) 56 (52 to 61)[1] Specificity, % (95% CI) 70 (67 to 73)[1] Positive LR (95% CI) 1.87 (1.66 to 2.10)[1] Negative LR (95% CI) 0.63 (0.56 to 0.70)[1] Occurrence of hot flashes or night sweats in the past four weeks to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 75 (71 to 79)[1] Specificity, % (95% CI) 46 (45 to 48)[1] Positive LR (95% CI) 1.40 (1.31 to 1.49)[1] Negative LR (95% CI) 0.54 (0.46 to 0.64)[1] Occurrence of night sweats in the past four weeks to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 56 (52 to 61)[1] Specificity, % (95% CI) 60 (59 to 62)[1] Positive LR (95% CI) 1.42 (1.29 to 1.55)[1] Negative LR (95% CI) 0.72 (0.65 to 0.81)[1] LR = likelihood ratio [1] Calculated by the NCC WCH technical team from data reported in the article	
Full citation Maartens,L.W., Leusink,G.L., Knottnerus,J.A., Smeets,C.G., Pop,V.J., Climacteric complaints in the	Sample size Initial sample population, N = 5896 N = 2450 total after exclusions (see below) n = 526 premenopausal n = 1250 perimenopausal n = 674 postmenopausal	Tests Standard questionnaire sent to all participants. Validated questionnaire covering 24 different possible complaints (pins and needles, dizziness, night-time sweating, day time	Methods Frequency of complaints recorded for different menopausal states.	Results Hot flushes to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 66 (62 to 70)[1] Specificity, % (95% CI) 51	Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Yes

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
community, Family Practice, 18, 189-194, 2001 Ref Id 282180 Country/ies where the study was carried out The Netherlands Study type Case-series Aim of the study To investigate the relationship between climacteric complaints and the menstrual pattern during the transition. Study dates September 1994 to September 1995 Source of funding Dutch Preventiefonds	Characteristics 76.4 % married Inclusion Criteria Women born between 1941 and 1947, living in the city of Eindhoven. Exclusion Criteria Previous hysterectomy (n = 1117), previous bilateral oophorectomy (n = 11), users of oestrogens/progestagens (n = 1433). Non-compliance with one or more items in the questionnaire (n = 1622). Non-Dutch Causcasian women excluded due to possible language problems (n = 734).	sweating, muscle pain, palpitations, vaginal itching, vaginal discharge, burning on micturition, loss of urine, tiredness, shortness of breath, flushing, agitation, headache, tiredness on waking, irritability, forgetfulness, insomnia, depressed mood, migraine, lack of energy, restless legs, lack of self confidence) and added vaginal dryness, pain during intercourse and waking at night. Definitions used Premenopausal: regular menstrual pattern Perimenopausal: irregular menstrual cycle (at least one period in the last year) Postmenopausal: amenorrhoea for one year prior to screening		(49 to 54)[1] Positive LR (95% CI) 1.36 (1.26 to 1.47)[1] Negative LR (95% CI) 0.66 (0.59 to 0.74)[1] Night sweats to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 58 (54 to 61)[1] Specificity, % (95% CI) 50 (47 to 52)[1] Positive LR (95% CI) 1.14 (1.05 to 1.24)[1] Negative LR (95% CI) 0.86 (0.77 to 0.95)[1] Palpitations to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 38 (35 to 42)[1] Specificity, % (95% CI) 66 (64 to 69)[1] Positive LR (95% CI) 1.14 (1.01 to 1.29)[1] Negative LR (95% CI) 0.93 (0.87 to 1.00)[1] Hot flushes to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 66 (62 to 70)[1] Specificity, % (95% CI) 88 (85 to 91)[1] Positive LR (95% CI): 5.51 (4.35 to 6.99)[1] Negative LR (95% CI): 0.39 (0.35 to 0.43)[1] Night sweats to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 58 (54 to 61)[1] Specificity, % (95% CI) 74 (70 to 78)[1] Positive LR (95% CI) 2.23	Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1. A Could the selection of patients have introduced bias? LOW RISK 1. B Is there concern that the included patients do not match the review question? LOW CONCERN Index test Were the index test results interpreted without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? N/A 2. A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2. B Is there concern that the index test, its conduct or interpretation differ from the review question? LOW CONCERN Reference standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(1.90 to 2.61)[1] Negative LR (95% CI) 0.57 (0.52 to 0.63)[1] Palpitations to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 38 (35 to 42)[1] Specificity, % (95% CI) 75 (71 to 79)[1] Positive LR (95% CI) 1.53 (1.28 to 1.83)[1] Negative LR (95% CI) 0.82 (0.76 to 0.89)[1] Hot flushes to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 66 (62 to 70)[1] Specificity, % (95% CI) 62 (60 to 65)[1] Positive LR (95% CI) 1.75 (1.61 to 1.90)[1] Negative LR (95% CI) 0.55 (0.49 to 0.61)[1] Night sweats to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 58 (54 to 61)[1] Specificity, % (95% CI) 57 (54 to 59)[1] Positive LR (95% CI) 1.33 (1.23 to 1.45)[1] Negative LR (95% CI) 0.75 (0.68 to 0.82)[1] Palpitations to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 38 (35 to 42)[1] Specificity, % (95% CI) 69 (67 to 71)[1] Positive LR (95% CI) 1.23 (1.09 to 1.39)[1] Negative LR (95% CI) 0.89	interpreted without knowledge of the results of the index test? Yes 3. A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3. B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW CONCERN Flow and timing Was there an appropriate interval between index test and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4. A Could the patient flow have introduced bias? LOW RISK Limitations Other information Women with hysterectomy were excluded, as were those using HRT.

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(0.84 to 0.96)[1] Hot flushes to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 49 (46 to 51)[1] Specificity, % (95% CI) 34 (30 to 38)[1] Positive LR (95% CI) 0.74 (0.68 to 0.80)[1] Negative LR (95% CI) 1.51 (1.35 to 1.70)[1] Night sweats to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 50 (48 to 53)[1] Specificity, % (95% CI) 42 (39 to 46)[1] Positive LR (95% CI) 0.88 (0.81 to 0.95)[1] Negative LR (95% CI) 1.17 (1.05 to 1.30)[1] Palpitations to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 34 (31 to 36)[1] Specificity, % (95% CI) 62 (58 to 65)[1] Positive LR (95% CI) 0.88 (0.78 to 0.99)[1] Negative LR (95% CI) 1.08 (1.00 to 1.16)[1] Hot flushes to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 49 (46 to 51)[1] Specificity, % (95% CI) 88 (85 to 91)[1] Positive LR (95% CI) 4.05 (3.19 to 5.15)[1] Negative LR (95% CI) 0.58 (0.55 to 0.62)[1] Night sweats to distinguish	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				perimenopausal women from premenopausal women Sensitivity, % (95% CI) 50 (48 to 53)[1] Specificity, % (95% CI) 74 (70 to 78)[1] Positive LR (95% CI) 1.96 (1.67 to 2.28)[1] Negative LR (95% CI) 0.67 (0.62 to 0.72)[1] Palpitations to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 33 (31 to 36)[1] Specificity, % (95% CI) 75 (71 to 79)[1] Positive LR (95% CI) 1.35 (1.14 to 1.59)[1] Negative LR (95% CI) 0.88 (0.83 to 0.94)[1] Hot flushes to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 49 (46 to 51)[1] Specificity, % (95% CI) 58 (55 to 60)[1] Positive LR (95% CI) 1.15 (1.05 to 1.25)[1] Negative LR (95% CI) 0.89 (0.83 to 0.96)[1] Night sweats to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 50 (48 to 53)[1] Specificity, % (95% CI) 56 (53 to 59)[1] Positive LR (95% CI) 1.16 (1.06 to 1.26)[1] Negative LR (95% CI) 0.88 (0.82 to 0.95)[1] Palpitations to distinguish perimenopausal women from all other women	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Sensitivity, % (95% CI) 34 (31 to 36)[1] Specificity, % (95% CI) 67 (65 to 70)[1] Positive LR (95% CI) 1.04 (0.93 to 1.16)[1] Negative LR (95% CI) 0.98 (0.93 to 1.04)[1] LR = likelihood ratio [1] Calculated by the NCC WCH technical team from data reported in the article	
Full citation Stellato,R., Crawford,S.L., McKinlay,S.M., Long-cope,C., Can follicle-stimulating hormone be used to define menopausal status?, Endocrine Practice, 4, 137-141, 1998 Ref Id 289730 Country/ies where the study was carried out Study type Case-series Aim of the study To assess the ability of FSH levels to distinguish between premenopausal, perimenopausal and postmenopausal women. Longitudinal study following premenopausal and perimenopausal women over the course of 6 years. Study dates 1986 to 1987.	Sample size N = 345 after exclusions n = 99 premenopausal n = 179 perimenopausal n = 67 postmenopausal Characteristics Mean age = 52 years. Inclusion Criteria Living within one hour's drive of Boston. Intact uterus with at least one ovary. No more than 11 consecutive months of amenorrhoea at baseline. 50 - 60 years old. Exclusion Criteria Baseline menopausal status could not be determined. Blood samples collected more than one month after the interview at which menopausal status was assessed. Estrogen users.	Tests Serum FSH was measured at baseline. Definitions used Premenopausal: recent bleeding (0 to 3 months before the baseline interview) and no report of cycle irregularity. Perimenopausal: less than 3 months of amenorrhoea but increasing irregularity, or 3 - 11 months amenorrhoea. Postmenopausal: 12 or more months of amenorrhoea.	Methods Data from the baseline interview was used to assess the ability of serum FSH levels to diagnose the perimenopause and menopause.	Results Serum FSH cut-point ≥ 38 IU/L to distinguish postmenopausal from perimenopausal women Sensitivity, % (95% CI) 63 (50 to 74)[1] Specificity, % (95% CI) 64 (57 to 71)[1] Positive LR (95% CI) 1.75 (1.34 to 2.30)[2] Negative LR (95% CI) 0.58 (0.42 to 0.81)[2] Serum FSH cut-point ≥ 24 IU/L to distinguish perimenopausal from premenopausal women Sensitivity, % (95% CI) 65 (57 to 72)[1] Specificity, % (95% CI) 69 (59 to 78)[1] Positive LR (95% CI) 2.07 (1.52 to 2.82)[2] Negative LR (95% CI) 0.51 (0.41 to 0.65)[2] LR = likelihood ratio [1] Point estimate reported in the article. 95% CI calculated by the NCC WCH technical team. [2] Calculated by the NCC WCH technical team from	Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1. A Could the selection of patients have introduced bias? LOW RISK 1. B Is there concern that the included patients do not match the review question? LOW CONCERN Index test Were the index test results interpreted without knowledge of the results of the reference standard? Unclear, but level of FSH is unlikely to be subject to bias as objectively recorded as absolute value.

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Source of funding The National Institute of Aging of the NIH.				data reported in the article.	If a threshold was used, was it pre-specified? No - thresholds were determined as part of the study. 2. A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2. B Is there concern that the index test, its conduct or interpretation differ from the review question? LOW CONCERN Reference standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3. A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3. B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW CONCERN

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
					Flow and timing Was there an appropriate interval between index test and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4. A Could the patient flow have introduced bias? LOW RISK Limitations Other information Women with surgical menopause or HRT use were excluded from the study.
Full citation Chompootweep,S., Tankeyoon,M., Yamarat,K., Poomsuwan,P., Dusitsin,N., The menopausal age and climacteric complaints in Thai women in Bangkok, Maturitas, 17, 63-71, 1993 Ref Id 226320 Country/ies where the study was carried out Thailand Study type Case-series	Sample size N = 2354 n = 735 premenopausal n = 292 perimenopausal n = 1327 postmenopausal Characteristics Mean age (SD) = 51.4 (4.7) years 12.4% smokers Inclusion Criteria Women aged 45 to 59 years who live in Bangkok. Exclusion Criteria Not reported.	Tests Prevalence of menopausal symptoms (hot flushes, night sweats and palpitations). Definitions used Premenopausal: regular menstruation Perimenopausal: irregular menstruation Postmenopausal: ≥ 12 months amenorrhoea	Methods A standardised questionnaire was administered through interview with a trained nurse, either at a health centre or on a home visit to enquire about climacteric symptoms. The timing of the symptoms was not described (i.e. whether the symptom had to have occurred within a specific time period, or at any point).	Results Hot flushes to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 6 (5 to 7)[1] Specificity, % (95% CI) 78 (73 to 82)[1] Positive LR (95% CI) 0.26 (0.19 to 0.35)[1] Negative LR (95% CI) 1.21 (1.14 to 1.29)[1] Night sweats to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 5 (4 to 7)[1] Specificity, % (95% CI) 83 (78 to 87)[1] Positive LR (95% CI) 0.30	Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1. A Could the selection of patients have introduced bias? LOW RISK 1. B Is there concern that the included patients do not match the review question?

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Aim of the study To determine the prevalence of climacteric symptoms of Thai women in Bangkok. Study dates October 1987 - January 1988 Source of funding The Institute of Health Research, Chulalongkorn University.				(0.21 to 0.42)[1] Negative LR (95% CI) 1.15 (1.09 to 1.21)[1] Palpitations to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 15 (13 to 17)[1] Specificity, % (95% CI) 66 (60 to 71)[1] Positive LR (95% CI) 0.44 (0.36 to 0.54)[1] Negative LR (95% CI) 1.29 (1.19 to 1.41)[1] Hot flushes to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 6 (5 to 7)[1] Specificity, % (95% CI) 90 (87 to 92)[1] Positive LR (95% CI) 0.55 (0.41 to 0.75)[1] Negative LR (95% CI) 1.05 (1.02 to 1.08)[1] Night sweats to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 5 (4 to 7)[1] Specificity, % (95% CI) 93 (91 to 95)[1] Positive LR (95% CI) 0.80 (0.56 to 1.14)[1] Negative LR (95% CI) 1.01 (0.99 to 1.04)[1] Palpitations to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 15 (13 to 17)[1] Specificity, % (95% CI) 77 (74 to 80)[1] Positive LR (95% CI) 0.65 (0.54 to 0.78)[1] Negative LR (95% CI) 1.11	LOW CONCERN Index test Were the index test results interpreted without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? N/A 2. A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2. B Is there concern that the index test, its conduct or interpretation differ from the review question? LOW CONCERN Reference standard Is the reference standard likely to correctly classify the target condition? Unclear - perimenopause defined only as irregular menstruation. Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3. A Could the reference standard, its conduct, or its interpretation have

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(1.06 to 1.16)[1] Hot flushes to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 6 (4 to 7)[1] Specificity, % (95% CI) 86 (84 to 88)[1] Positive LR (95% CI) 0.42 (0.32 to 0.54)[1] Negative LR (95% CI) 1.09 (1.06 to 1.12)[1] Night sweats to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 5 (4 to 7)[1] Specificity, % (95% CI) 90 (88 to 92)[1] Positive LR (95% CI) 0.54 (0.40 to 0.73)[1] Negative LR (95% CI) 1.05 (1.02 to 1.07)[1] Palpitations to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 15 (13 to 17)[1] Specificity, % (95% CI) 74 (71 to 76)[1] Positive LR (95% CI) 0.57 (0.48 to 0.67)[1] Negative LR (95% CI) 1.15 (1.10 to 1.20)[1] Hot flushes to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 22 (18 to 27)[1] Specificity, % (95% CI) 94 (93 to 95)[1] Positive LR (95% CI) 3.89 (2.86 to 5.28)[1] Negative LR (95% CI) 0.82 (0.77 to 0.88)[1] Night sweats to distinguish	introduced bias? LOW RISK 3. B Is there concern that the target condition as defined by the reference standard does not match the review question? UNCLEAR Flow and timing Was there an appropriate interval between index test and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4. A Could the patient flow have introduced bias? LOW RISK Limitations Definition of perimenopause includes all women with irregular cycles, which may include some women with long standing cycle irregularity (not necessarily due to perimenopause). Other information Unclear whether women with surgical menopause or users of HRT were included.

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 17 (13 to 22)[1] Specificity, % (95% CI) 95 (93 to 96)[1] Positive LR (95% CI) 3.36 (2.39 to 4.71)[1] Negative LR (95% CI) 0.87 (0.82 to 0.92)[1] Palpitations to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 34 (29 to 40)[1] Specificity, % (95% CI) 85 (83 to 87)[1] Positive LR (95% CI) 2.28 (1.86 to 2.80)[1] Negative LR (95% CI) 0.77 (0.71 to 0.84)[1] Hot flushes to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 22 (18 to 27)[1] Specificity, % (95% CI) 90 (87 to 92)[1] Positive LR (95% CI) 2.15 (1.59 to 3.87)[1] Negative LR (95% CI) 0.87 (0.81 to 0.93)[1] Night sweats to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 17 (13 to 22)[1] Specificity, % (95% CI) 93 (91 to 95)[1] Positive LR (95% CI) 2.67 (1.85 to 3.87)[1] Negative LR (95% CI) 0.88 (0.83 to 0.93)[1] Palpitations to distinguish perimenopausal women from premenopausal women	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Sensitivity, % (95% CI) 34 (29 to 40)[1] Specificity, % (95% CI) 77 (74 to 80)[1] Positive LR (95% CI) 1.48 (1.20 to 1.82)[1] Negative LR (95% CI) 0.86 (0.78 to 0.94)[1] Hot flushes to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 22 (18 to 27)[1] Specificity, % (95% CI) 93 (91 to 94)[1] Positive LR (95% CI) 3.04 (2.34 to 3.96)[1] Negative LR (95% CI) 0.84 (0.79 to 0.89)[1] Night sweats to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 17 (13 to 22)[1] Specificity, % (95% CI) 94 (93 to 95)[1] Positive LR (95% CI) 3.08 (2.27 to 4.18)[1] Negative LR (95% CI) 0.88 (0.83 to 0.92)[1] Palpitations to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 34 (29 to 40)[1] Specificity, % (95% CI) 82 (80 to 84)[1] Positive LR (95% CI) 1.91 (1.59 to 2.30)[1] Negative LR (95% CI) 0.80 (0.74 to 0.87)[1] LR = likelihood ratio [1] Calculated by the NCC WCH technical team from data reported in the article.	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Full citation Punyahotra,S., Dennerstein,L., Lehert,P., Menopausal experiences of Thai women. Part 1: Symptoms and their correlates, Maturitas, 26, 1-7, 1997 Ref Id 289733 Country/ies where the study was carried out Thailand Study type Case-series Aim of the study To examine the relationship between menopausal symptoms and menopausal status Study dates January to February 1994 Source of funding Not reported.	Sample size N = 268 N = 248 after exclusions (see below) n = 127 premenopausal n = 22 perimenopausal n = 99 postmenopausal Characteristics Mean age (SD) = 49.35 (6.11) years Inclusion Criteria Women who accompanied patients to the Royal Irrigation Hospital. Exclusion Criteria Previous hysterectomy and/or bilateral oophorectomy. Current users of HRT or OCP.	Tests Prevalence of specific symptoms at different stages of the menopause. Definitions used Premenopausal: menses occurred with usual regularity during the year preceding the survey. Perimenopausal: menstrual cycles have changed in frequency during the previous year. Postmenopausal: no menses in the previous 12 months.	Methods A semi-structured questionnaire was conducted by interview with a Thai gynaecologist. Participants were asked whether they suffered from a variety of symptoms during the previous 2 weeks.	Results Hot flushes to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 33 (24 to 44)[1] Specificity, % (95% CI) 45 (24 to 68)[1] Positive LR (95% CI) 0.61 (0.38 to 0.98)[1] Negative LR (95% CI) 1.47 (0.91 to 2.37)[1] Night sweats to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 32 (23 to 42)[1] Specificity, % (95% CI) 73 (50 to 89)[1] Positive LR (95% CI) 1.19 (0.57 to 2.48)[1] Negative LR (95% CI) 0.93 (0.70 to 1.24)[1] Rapid heart beat to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 41 (32 to 52)[1] Specificity, % (95% CI) 64 (41 to 83)[1] Positive LR (95% CI) 1.14 (0.62 to 2.08)[1] Negative LR (95% CI) 0.92 (0.64 to 1.23)[1] Hot flushes to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 33 (24 to 44)[1] Specificity, % (95% CI) 83 (75 to 89)[1] Positive LR (95% CI) 1.92 (1.20 to 3.08)[1] Negative LR (95% CI) 0.81	Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? No - a "convenience sample" of patients were enrolled. Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1. A Could the selection of patients have introduced bias? HIGH RISK 1. B Is there concern that the included patients do not match the review question? LOW CONCERN Index test Were the index test results interpreted without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? N/A 2. A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2. B Is there concern that the index test, its conduct or interpretation differ from the review

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(0.69 to 0.95)[1] Night sweats to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 32 (23 to 42)[1] Specificity, % (95% CI) 83 (75 to 89)[1] Positive LR (95% CI) 1.87 (1.16 to 3.00)[1] Negative LR (95% CI) 0.82 (0.70 to 0.96)[1] Rapid heart beat to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 41 (32 to 52)[1] Specificity, % (95% CI) 74 (65 to 81)[1] Positive LR (95% CI) 1.59 (1.09 to 2.32)[1] Negative LR (95% CI) 0.79 (0.65 to 0.96)[1] Hot flushes to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 33 (24 to 44)[1] Specificity, % (95% CI) 77 (70 to 84)[1] Positive LR (95% CI) 1.46 (0.97 to 2.19)[1] Negative LR (95% CI) 0.86 (0.73 to 1.02)[1] Night sweats to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 32 (23 to 42)[1] Specificity, % (95% CI) 81 (74 to 87)[1] Positive LR (95% CI) 1.72 (1.11 to 2.67)[1] Negative LR (95% CI) 0.83 (0.71 to 0.97)[1]	question? LOW CONCERN Reference standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3. A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3. B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW CONCERN Flow and timing Was there an appropriate interval between index test and reesference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4. A Could the patient flow have introduced

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Rapid heart beat to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 41 (32 to 52)[1] Specificity, % (95% CI) 72 (65 to 79)[1] Positive LR (95% CI) 1.51 (1.06 to 2.14)[1] Negative LR (95% CI) 0.81 (0.67 to 0.98)[1] Hot flushes to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 55 (32 to 76)[1] Specificity, % (95% CI) 67 (56 to 76)[1] Positive LR (95% CI) 1.64 (1.02 to 2.62)[1] Negative LR (95% CI) 0.68 (0.42 to 1.10)[1] Night sweats to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 27 (11 to 50)[1] Specificity, % (95% CI) 68 (58 to 77)[1] Positive LR (95% CI) 0.84 (0.40 to 1.77)[1] Negative LR (95% CI) 1.07 (0.80 to 1.44)[1] Rapid heart beat to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 36 (17 to 59)[1] Specificity, % (95% CI) 59 (48 to 68)[1] Positive LR (95% CI) 0.88 (0.48 to 1.60)[1]	bias? LOW RISK Limitations Non-random recruitment of participants through convenience sampling approach may introduce bias. Other information Women with surgical menopause or HRT use were excluded.

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Negative LR (95% CI) 1.09 (0.76 to 1.55)[1] Hot flushes to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 55 (32 to 76)[1] Specificity, % (95% CI) 83 (75 to 89)[1] Positive LR (95% CI) 3.15 (1.84 to 5.39)[1] Negative LR (95% CI) 0.55 (0.35 to 0.87)[1] Night sweats to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 27 (11 to 50)[1] Specificity, % (95% CI) 83 (75 to 89)[1] Positive LR (95% CI) 1.57 (0.72 to 3.44)[1] Negative LR (95% CI) 0.88 (0.67 to 1.15)[1] Rapid heart beat to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 36 (17 to 59)[1] Specificity, % (95% CI) 74 (65 to 81)[1] Positive LR (95% CI) 1.40 (0.75 to 2.62)[1] Negative LR (95% CI) 0.86 (0.62 to 1.20)[1] Hot flushes to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 55 (32 to 76)[1] Specificity, % (95% CI) 76 (70 to 82)[1] Positive LR (95% CI) 2.28 (1.46 to 3.57)[1] Negative LR (95% CI) 0.60	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(0.38 to 0.95)[1] Night sweats to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 27 (11 to 50)[1] Specificity, % (95% CI) 77 (70 to 82)[1] Positive LR (95% CI) 1.16 (0.57 to 2.39)[1] Negative LR (95% CI) 0.95 (0.73 to 1.24)[1] Rapid heart beat to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 36 (17 to 59)[1] Specificity, % (95% CI) 67 (61 to 73)[1] Positive LR (95% CI) 1.11 (0.62 to 1.99)[1] Negative LR (95% CI) 0.95 (0.68 to 1.31)[1] LR = likelihood ratio [1]Calculated by the NCC WCH technical team from data reported in the article.	
Full citation Ho,S.C., Chan,S.G., Yip,Y.B., Cheng,A., Yi,Q., Chan,C., Menopausal symptoms and symptom clustering in Chinese women, Maturitas, 33, 219-227, 1999 Ref Id 289734 Country/ies where the study was carried out Hong Kong Study type Case-series	Sample size N = 2125 N = 1900 after exclusions (see below) n = 1258 premenopausal n = 92 perimenopausal n = 540 postmenopausal Characteristics Mean age (SD) premenopausal women 47.27 (3.22) years Mean age (SD) perimenopausal women 49.26 (6.02) years Mean age (SD) postmenopausal women 51 59 (5.30) years Inclusion Criteria Age 44 to 55 years. Hong Kong Chinese residents. Exclusion Criteria Women who had stopped menstruating as a result	Tests Prevalence of a variety of symptoms during different stages of the menopause transition. Definitions used Premenopausal: still having menses (regular or irregular). Perimenopausal: cessation of menstrual periods for at least three months within the previous 12 months, but not due to hysterectomy, oophorectomy or pregnancy. Postmenopausal: cessation of menstruation for at least 12 months.	Methods A standardised questionnaire was conducted over the telephone, to enquire about specific symptoms. Presence of symptoms was recorded as "yes" or "no" to experience of the symptom during the past two weeks.	Results Hot flushes to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 12 (9 to 15)[1] Specificity, % (95% CI) 78 (68 to 86)[1] Positive LR (95% CI) 0.54 (0.34 to 0.84)[1] Negative LR (95% CI) 1.13 (1.01 to 1.26)[1] Cold sweats to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 6 (4 to 8)[1] Specificity, % (95% CI) 96	Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1. A Could the selection of patients have introduced bias? LOW RISK 1. B Is there concern that the included

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Aim of the study To report the prevalence of symptoms in Hong Kong Chinese perimenopausal women, and to clarify whether symptom groups are associated with menopausal status. Study dates 1996 Source of funding Health Services Research Committee.	of hysterectomy or radio/chemotherapy. Menstrual status could not be determined due to missing data.			(89 to 99)[1] Positive LR (95% CI) 1.36 (0.49 to 3.76)[1] Negative LR (95% CI) 0.98 (0.94 to 1.03)[1] Rapid heart beat to distinguish postmenopausal women from perimenopausal women Sensitivity, % (95% CI) 12 (9 to 15)[1] Specificity, % (95% CI) 84 (75 to 91)[1] Positive LR (95% CI) 0.73 (0.43 to 1.22)[1] Negative LR (95% CI) 1.05 (0.96 to 1.16)[1] Hot flushes to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 12 (9 to 15)[1] Specificity, % (95% CI) 91 (90 to 93)[1] Positive LR (95% CI) 1.33 (1.00 to 1.79)[1] Negative LR (95% CI) 0.97 (0.93 to 1.00)[1] Cold sweats to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 6 (4 to 8)[1] Specificity, % (95% CI) 96 (94 to 97)[1] Positive LR (95% CI) 1.33 (0.87 to 2.03)[1] Negative LR (95% CI) 0.98 (0.96 to 1.01)[1] Rapid heart beat to distinguish postmenopausal women from premenopausal women Sensitivity, % (95% CI) 12 (9 to 15)[1] Specificity, % (95% CI) 86	patients do not match the review question? LOW CONCERN Index test Were the index test results interpreted without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? N/A 2. A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2. B Is there concern that the index test, its conduct or interpretation differ from the review question? LOW CONCERN Reference standard Is the reference standard likely to correctly classify the target condition? Unclear - premenopausal women included those with irregular menstruation, who may be perimenopausal by other definitions. Were the reference standard results interpreted without knowledge of the results of the index

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(84 to 88)[1] Positive LR (95% CI) 0.84 (0.64 to 1.10)[1] Negative LR (95% CI) 1.03 (0.99 to 1.07)[1] Hot flushes to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 12 (9 to 15)[1] Specificity, % (95% CI) 90 (89 to 92)[1] Positive LR (95% CI) 1.21 (0.91 to 1.61)[1] Negative LR (95% CI) 0.98 (0.94 to 1.01)[1] Cold sweats to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 6 (4 to 8)[1] Specificity, % (95% CI) 96 (94 to 97)[1] Positive LR (95% CI) 1.33 (0.88 to 2.02)[1] Negative LR (95% CI) 0.98 (0.96 to 1.01)[1] Rapid heart beat to distinguish postmenopausal women from all other women Sensitivity, % (95% CI) 12 (9 to 15)[1] Specificity, % (95% CI) 86 (84 to 88)[1] Positive LR (95% CI) 0.83 (0.64 to 1.09)[1] Negative LR (95% CI) 1.03 (0.99 to 1.07)[1] Hot flushes to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 22 (14 to 32)[1] Specificity, % (95% CI) 88 (85 to 91)[1] Positive LR (95% CI) 1.86	test? Yes 3. A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3. B Is there concern that the target condition as defined by the reference standard does not match the review question? UNCLEAR Flow and timing Was there an appropriate interval between index test and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the analysis? Yes 4. A Could the patient flow have introduced bias? LOW RISK Limitations Premenopausal women included those with regular and irregular menstruation, whilst perimenopausal women were those with at least 3 months amenorrhoea. Therefore there may be overclassification

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(1.19 to 2.93)[1] Negative LR (95% CI) 0.89 (0.79 to 0.99)[1] Cold sweats to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 4 (1 to 11)[1] Specificity, % (95% CI) 94 (92 to 96)[1] Positive LR (95% CI) 0.73 (0.27 to 1.03)[1] Negative LR (95% CI) 1.02 (0.97 to 1.07)[1] Rapid heart beat to distinguish perimenopausal women from postmenopausal women Sensitivity, % (95% CI) 16 (9 to 25)[1] Specificity, % (95% CI) 88 (85 to 91)[1] Positive LR (95% CI) 1.38 (0.82 to 2.31)[1] Negative LR (95% CI) 0.95 (0.86 to 1.04)[1] Hot flushes to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 22 (14 to 32)[1] Specificity, % (95% CI) 91 (90 to 93)[1] Positive LR (95% CI) 2.49 (1.62 to 3.81)[1] Negative LR (95% CI) 0.86 (0.77 to 0.96)[1] Cold sweats to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 4 (1 to 11)[1] Specificity, % (95% CI) 96 (94 to 97)[1] Positive LR (95% CI) 0.98 (0.36 to 2.63)[1]	of some perimenopausal women as premenopausal. Other information Women with hysterectomy were excluded. It is unclear whether users of HRT were included in this study.

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Negative LR (95% CI) 1.00 (0.96 to 1.05)[1] Rapid heart beat to distinguish perimenopausal women from premenopausal women Sensitivity, % (95% CI) 16 (9 to 25)[1] Specificity, % (95% CI) 86 (84 to 88)[1] Positive LR (95% CI) 1.16 (0.72 to 1.88)[1] Negative LR (95% CI) 0.97 (0.89 to 1.07)[1] Hot flushes to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 22 (14 to 32)[1] Specificity, % (95% CI) 90 (89 to 92)[1] Positive LR (95% CI) 2.26 (1.50 to 3.41)[1] Negative LR (95% CI) 0.87 (0.78 to 0.97)[1] Cold sweats to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 4 (1 to 11)[1] Specificity, % (95% CI) 95 (94 to 98)[1] Positive LR (95% CI) 0.89 (0.33 to 2.37)[1] Negative LR (95% CI) 1.01 (0.96 to 1.05)[1] Rapid heart beat to distinguish perimenopausal women from all other women Sensitivity, % (95% CI) 16 (9 to 25)[1] Specificity, % (95% CI) 87 (85 to 88)[1] Positive LR (95% CI) 1.22 (0.75 to 1.96)[1] Negative LR (95% CI) 0.97	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(0.88 to 1.06)[1] LR = likelihood ratio [1]Calculated by the NCC WCH technical team from data reported in the article	
Full citation Dennerstein,L., Smith,A.M., Morse,C., Burger,H., Green,A., Hopper,J., Ryan,M., Menopausal symptoms in Australian women, Medical Journal of Australia, 159, 232-236, 1993 Ref Id 255899 Country/ies where the study was carried out Australia Study type Case-series Aim of the study To describe Australian-born women's experience of symptoms during the natural menopause transition. Study dates Not reported Source of funding Victorian Health Promotion Foundation.	Sample size N = 1220 n = 316 premenopausal n = 549 perimenopausal n = 355 postmenopausal Characteristics Inclusion Criteria Age 45 to 55 years. Australian born women from the Melbourne metropolitan region. Exclusion Criteria Use of oral contraceptive pill. Using hormone replacement therapy. Surgical menopause (hysterectomy and/or bilateral oophorectomy).	Tests Each subject was asked whether she had been bothered in the previous 2 weeks with a variety of symptoms. Definitions used Premenopausal: no changes in menstrual frequency of flow in the prior 12 months. Perimenopausal: changes in menstrual frequency or flow in the prior 12 months. Menopausal: no menses in the prior 12 months.	Methods A 20 - 25 minute telephone interview was conducted by trained interviewers to enquire about symptoms.	Results Hot flushes to distinguish between postmenopausal and perimenopausal women Sensitivity , % (95% CI) 39 (34 to 45)[1] Specificity, % (95% CI) 68 (64 to 72)[1] Positive LR (95% CI) 1.25 (1.05 to 1.50)[1] Negative LR (95 % CI) 0.88 (0.80 to 0.98)[1] Cold sweats to distinguish between postmenopausal and perimenopausal women Sensitivity , % (95% CI) 1 (0 to 3)[1] Specificity, % (95% CI) 90 (88 to 93)[1] Positive LR (95% CI) 0.15 (0.06 to 0.36)[1] Negative LR (95 % CI) 1.09 (1.06 to 1.12)[1] Rapid heart beat to distinguish between postmenopausal and perimenopausal women Sensitivity , % (95% CI) 10 (7 to 13)[1] Specificity, % (95% CI) 88 (85 to 90)[1] Positive LR (95% CI) 0.80 (0.54 to 1.17)[1] Negative LR (95 % CI) 1.03 (0.98 to 1.08)[1] Hot flushes to distinguish between postmenopausal and premenopausal women Sensitivity , % (95% CI) 39 (34 to 45)[1]	Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Yes Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes 1. A Could the selection of patients have introduced bias? LOW RISK 1. B Is there concern that the included patients do not match the review question? LOW CONCERN Index test Were the index test results interpreted without knowledge of the results of the reference standard? Yes If a threshold was used, was it pre-specified? N/A 2. A Could the conduct or interpretation of the index test have introduced bias? LOW RISK 2. B Is there concern that the index test, its

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Specificity, % (95% CI) 90 (86 to 93)[1] Positive LR (95% CI) 4.02 (2.81 to 5.75)[1] Negative LR (95 % CI) 0.67 (0.61 to 0.74)[1] Cold sweats to distinguish between postmenopausal and premenopausal women Sensitivity , % (95% CI) 1 (0 to 3)[1] Specificity, % (95% CI) 98 (95 to 99)[1] Positive LR (95% CI) 0.64 (0.20 to 1.98)[1] Negative LR (95 % CI) 1.01 (0.99 to 1.03)[1] Rapid heart beat to distinguish between postmenopausal and premenopausal women Sensitivity , % (95% CI) 10 (7 to 13)[1] Specificity, % (95% CI) 93 (89 to 95)[1] Positive LR (95% CI) 1.35 (0.82 to 2.24)[1] Negative LR (95 % CI) 0.97 (0.93 to 1.02)[1] Hot flushes to distinguish between postmenopausal and all other women Sensitivity , % (95% CI) 39 (34 to 45)[1] Specificity, % (95% CI) 76 (73 to 79)[1] Positive LR (95% CI) 1.67 (1.40 to 1.99)[1] Negative LR (95 % CI) 0.79 (0.72 to 0.87)[1] Cold sweats to distinguish between postmenopausal and all other women Sensitivity , % (95% CI) 1 (0. to 3)[1] Specificity, % (95% CI) 93	conduct or interpretation differ from the review question? LOW CONCERN Reference standard Is the reference standard likely to correctly classify the target condition? Yes Were the reference standard results interpreted without knowledge of the results of the index test? Yes 3. A Could the reference standard, its conduct, or its interpretation have introduced bias? LOW RISK 3. B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW CONCERN Flow and timing Was there an appropriate interval between index test and reference standard? Yes Did all patients receive a reference standard? Yes Did patients receive the same reference standard? Yes Were all patients included in the

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(91 to 95)[1] Positive LR (95% CI) 0.20 (0.08 to 0.50)[1] Negative LR (95 % CI) 1.06 (1.04 to 1.08)[1] Rapid heart beat to distinguish between postmenopausal and all other women Sensitivity , % (95% CI) 10 (7 to 13)[1] Specificity, % (95% CI) 89 (87 to 91)[1] Positive LR (95% CI) 0.94 (0.65 to 1.36)[1] Negative LR (95 % CI) 1.01 (0.97 to 1.05)[1] Hot flushes to distinguish between perimenopausal and postmenopausal women Sensitivity , % (95% CI) 32 (28 to 36)[1] Specificity, % (95% CI) 61 (55 to 66)[1] Positive LR (95% CI) 0.80 (0.67 to 0.96)[1] Negative LR (95 % CI) 1.13 (1.02 to 1.25)[1] Cold sweats to distinguish between perimenopausal and postmenopausal women Sensitivity , % (95% CI) 10 (7 to 12)[1] Specificity, % (95% CI) 99 (97 to 100)[1] Positive LR (95% CI) 6.85 (2.77 to 16.98)[1] Negative LR (95 % CI) 0.93 (0.89 to 0.94)[1] Rapid heart beat to distinguish between perimenopausal and postmenopausal women Sensitivity , % (95% CI) 12 (10 to 15)[1] Specificity, % (95% CI) 90	analysis? Yes 4. A Could the patient flow have introduced bias? LOW RISK Limitations Other information Women with surgical menopause or using HRT were excluded from this study.

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				(87 to 93)[1] Positive LR (95% CI) 1.26 (0.85 to 1.85)[1] Negative LR (95 % CI) 0.97 (0.93 to 1.02)[1] Hot flushes to distinguish between perimenopausal and premenopausal women Sensitivity , % (95% CI) 32 (28 to 36)[1] Specificity, % (95% CI) 90 (86 to 93)[1] Positive LR (95% CI) 3.21 (2.25 to 4.59)[1] Negative LR (95 % CI) 0.76 (0.71 to 0.81)[1] Cold sweats to distinguish between perimenopausal and premenopausal women Sensitivity , % (95% CI) 10 (7 to 12)[1] Specificity, % (95% CI) 98 (95 to 99)[1] Positive LR (95% CI) 4.36 (2.01 to 9.47)[1] Negative LR (95 % CI) 0.92 (0.89 to 0.95)[1] Rapid heart beat to distinguish between perimenopausal and premenopausal women Sensitivity , % (95% CI) 12 (10 to 15)[1] Specificity, % (95% CI) 93 (89 to 95)[1] Positive LR (95% CI) 1.70 (1.08 to 2.67)[1] Negative LR (95 % CI) 0.95 (0.90 to 0.99)[1] Hot flushes to distinguish between perimenopausal and all other women Sensitivity , % (95% CI) 32 (28 to 36)[1] Specificity, % (95% CI) 75 (71 to 78)[1]	

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Positive LR (95% CI) 1.24 (1.03 to 1.48)[1] Negative LR (95 % CI) 0.92 (0.86 to 0.99)[1] Cold sweats to distinguish between perimenopausal and all other women Sensitivity , % (95% CI) 10 (7 to 12)[1] Specificity, % (95% CI) 98 (97 to 99)[1] Positive LR (95% CI) 5.40 (2.91 to 10.00)[1] Negative LR (95 % CI) 0.92 (0.89 to 0.95)[1] Rapid heart beat to distinguish between perimenopausal and all other women Sensitivity , % (95% CI) 12 (10 to 15)[1] Specificity, % (95% CI) 91 (89 to 93)[1] Positive LR (95% CI) 1.43 (1.03 to 2.00)[1] Negative LR (95 % CI) 0.96 (0.92 to 1.00)[1] LR = likelihood ratio [1] Calculated by the NCC WCH technical team from data reported in the article.	
Full citation Bener, A., Falah, A., A measurement-specific quality-of-life satisfaction during premenopause, perimenopause and postmenopause in Arabian Qatari women, Journal of Mid-life Health, 5, 126-34, 2014 Ref Id 337335	Sample size N=1158 n=334 perimenopausal n=629 menopausal n=195 postmenopausal Characteristics Age (years, mean, SD): Perimenopausal: 50.6 (6.1) Menopausal: 42.5 (1.9) Postmenopausal: 51.9 (2.5) Level of education (n) (perimenopausal/menopausal/postmenopausal): Elementary:66/120/44 Secondary:77/165/46	Tests -Menopause-specific quality of life questionnaire (MENQOL) -Symptoms or problems experienced were recorded on the Likert scale (physical, emotional (vasomotor), psycho-social and sexual areas, and additional socio-demographic sections) Definitions used Peri-menopause: around the	Methods -Cross-sectional primary health care centre based study -MENQOL questionnaire: the data was collected through the validated questionnaire by qualified nurses between July 2012 and November 2013. -Sample size of 1500 participants was	Results Symptoms of hot flushes to distinguish post menopause from all hot flushes Sensitivity (%): 43 (36-50) Specificity (%): 68 (65-71) LR+: 1.39 (1.15-1.67) LR- : 0.82 (0.72-0.93) Symptoms of hot flushes to distinguish post menopause from peri menopause Sensitivity (%): 43 (36-50) Specificity (%): 68 (64-72) LR+: 1.38 (1.13-1.68)	Study quality - QUADAS 2 checklist Study quality - QUADAS 2 checklist Patient selection Was a consecutive or random sample of patients enrolled? Unclear Was a case-control design avoided? Yes Did the study avoid inappropriate exclusions? Yes

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
Country/ies where the study was carried out Qatar Study type Nested case-control study Aim of the study To use the menopause -specific quality of life satisfaction in the state of Qatar for the premenopausal, menopause and postmenopausal period. Study dates July 2012-November 2103 Source of funding Qatar national research fund	University:77/103/14 Occupation (n) (perimenopausal/menopausal/postmenopausal): Housewife: 167/337/123 Sedentary and professional: 63/75/17 Clerk: 71/119/34 Business/private: 17/49/11 Inclusion Criteria Women aged 40-60 years who had not had a hysterectomy , and who had not used hormone replacement therapy during the preceding 6 months. Exclusion Criteria Women with contraindications to oestrogen use and, women who had a current unstable medical or social problem.	menopause (menopause transition years, a span of time both before and after the date of the final episode of flow). Post-menopause: women who have not experienced any menstrual flow for a minimum of 12 months, assuming they still have a uterus, and are not pregnant or lactating. In women without a uterus, menopause or post-menopause can be identified by a blood test for follicle stimulating hormone levels.	determined a priori on the assumption that the prevalence rate of postpartum depression would be similar to prevalence rates in other eastern Mediterranean countries (20%, 95%CI 2.5%). -Data was analysed using student t test to ascertain significance of differences between mean values of two continuous variables and confirmed by non-parametric Mann-Whitney test. Chi squared test and Fisher exact test (two-tailed) were performed to test for differences in the proportion of categorical variables between two or more groups. Kruskal Wallis ANOVA was employed for comparison of several group means. Spearman's correlation coefficient was used to evaluate strength of concordance between variables. For all statistical tests, a P value <0.05 was considered statistically significant.	LR-: 0.82 (0.71-0.94) Symptoms of hot flushes to distinguish post menopause from pre menopause Sensitivity (%): 43 (36-50) Specificity (%): 69 (64-74) LR+: 1.41 (1.12-1.77) LR-: 0.81 (0.70-0.94) Symptoms of hot flushes to distinguish perimenopause from all hot flushes Sensitivity (%): 31 (27-35) Specificity (%): 64 (60-68) LR+: 0.88 (0.75-1.04) LR- : 1.06 (0.97-1.15) Symptoms of hot flushes to distinguish peri menopause from post menopause Sensitivity (%): 31 (27-35 Specificity (%): 56 (49-63) LR+: 0.72 (0.59-0.87) LR-: 1.21 (1.06-1.38) Symptoms of hot flushes to distinguish perimenopause from pre menopause Sensitivity (%): 31 (27-35) Specificity (%): 69 (64-74) LR+: 1.02 (0.83-1.24) LR- : 0.99 (0.90-1.08) Symptoms of sweating to distinguish post menopause from all sweating Sensitivity (%): 72 (66-79) Specificity (%): 34 (31-37) LR+: 1.10 (1.00-1.21) LR- : 0.79 (0.62-1.02) Symptoms of sweating to distinguish post menopause from perimenopause Sensitivity (%):89 (86-92) Specificity (%): 32 (28-35) LR+: 1.31 (1.23-1.39) LR- :0.33 (0.25-0.44) Symptoms of sweating to distinguish post menopause from premenopause	1.A Could the selection of patients have introduced bias? LOW RISK OF BIAS 1.B Is there concern that the included patients do not match the review question? LOW CONCERN Index Test Were the index test results interpreted without knowledge of the results of the reference standard? N/A If a threshold was used, was it pre-specified? N/A 2.A Could the conduct or interpretation of the index test have introduced bias? UNCLEAR RISK OF BIAS 2.B Is there concern that the index test, its conduct, or interpretation differ from the review question? LOW CONCERN Reference Standard Is the reference standard likely to correctly classify the target condition? N/A Were the reference standard results interpreted without knowledge of the results of the index test? N/A

Bibliographic details	Participants	Tests	Methods	Outcomes and results	Comments
				Sensitivity (%): 72 (66-79) Specificity (%): 37 (32-42) LR+: 1.16 (1.03-1.31) LR- : 0.72 (0.55-0.94) Symptoms of sweating to distinguish peri menopause from all sweating Sensitivity: (%): 67 (64-71) Specificity (%): 33 (29-37) LR+: 1.02 (0.94-1.10) LR-: 0.94 (0.80-1.11) Symptoms of sweating to distinguish perimenopause from post menopause Sensitivity (%): 62 (57-67) Specificity (%): 27 (20-33) LR+: 0.85 (0.25-0.96) LR- :1.38 (1.06-1.81) Symptoms of sweating to distinguish perimenopause from premenopause Sensitivity (%): 67 (64-71) Specificity (%): 37 (32-42) LR+: 1.09 (0.98-1.20) LR- : 0.85 (0.71-1.01)	3.A Could the reference standard, its conduct, or its interpretation have introduced bias? UNCLEAR RISK 3.B Is there concern that the target condition as defined by the reference standard does not match the review question? LOW RISK Flow and Timing Was there an appropriate interval between index test(s) and reference standard? N/A Did all patients receive a reference standard? N/A Did patients receive the same reference standard? N/A Were all patients included in the analysis? Yes 4.A Could the patient flow have introduced bias? UNCLEAR RISK

H.2 Classification systems for the diagnosis of menopause

H.3 Information and advice

H.3.1 What information about the menopause do women find helpful?

Study details	Summary of study	Results	Other
Full citation Alfred,A., Esterman,A., Farmer,E., Pilotto,L., Weston,K., Women's decision making at menopause - a focus group study, Australian Family Physician, 35, 270-272, 2006 Ref Id 302967 Country/ies where the study was carried out Australia Study type Qualitative (content)	Aim of the study To explore women's views about menopause support needs Characteristics Aged 40 - 64 Inclusion criteria Women with diverse demographic backgrounds. Exclusion criteria Women seeking medical support for menopause issues. Intervention None Data collection 4 focus groups of 31 women explored their experience about menopause, its management and decision support needs. Data analysis A phenomological, grounded theory approach produced bullet-pointed themes with example-quotations.	Results relevant to protocol Women found the following things from their doctors useful: Comprehensive information on self-management practices; alternative options; acknowledgement of therapy risks and referral to reliable information sources. Acknowledgement of evidence uncertainty. Adequate time for discussion. Female practitioners for menopause issues. Information on 'natural' treatments. Information that was personalised to their own 'individual chemistry'. Information about incontinence as it was embarrassing to bring it up. Aviodance of the 'myth of certainty around what is inherently uncertain.' GPs perceived as 'so busy' that women did not want to 'wear them out' with all the information they required	Comments Limitations Themes were subjectively titled and not enough examples quoted. The paper was too short to adequately represent women's voices. Quality checklist NICE Appendix H: Methodology checklist for qualitative studies Is a qualitative approach appropriate? Yes How well was the data collection carried out? Under-reported Were the methods reliable? Yes Are the data 'rich'? No Is the analysis reliable? Yes Is the role of the researcher clearly described? No
Full citation Andrist,L.C., The impact of media attention, family history, politics and maturation on women's decisions regarding hormone replacement therapy, Health Care for Women International, 19, 243-260, 1998 Ref Id 302992 Country/ies where the study was carried out USA Study type Qualitative (content)	Aim of the study An exploration of how women make decisions about HRT for natural menopause. Characteristics 21 Well-educated European Americans. Characteristic: n In favour of HRT: 6 Undecided: 10 Opposed to HRT: 5 Had college degrees: 17 Were healthcare professionals: 11 Had administrative, legal or consulting roles: 10 Pre-menopausal: 1 Peri-menopausal (cycle changes and VSM): 11 Menopausal (menses cessation during study): 4 Post-menopausal (Amenorhea >12 months): 5	Results relevant to protocol An admin assistant said she needed 'more education' to take fully informed decisions regarding HRT. Another woman said she would like her HCP to lay out options and help her make a decision. One woman said that "Risk reduction was a compelling piece of information." Women favoured balancing their own family histories with research findings. A professor of nursing said that even academic HCPs feel confused because "I notice that some people have very strong opinions on it when I've asked professional people." One woman said she felt 'intimidated' by reading because "What you read you can turn it around in to something else." Access to information is not enough on its own as	Comments Limitations Possible bias in favour of not using HRT. Quality checklist NICE Appendix H: Methodology checklist for qualitative studies Is a qualitative approach appropriate? Yes How well was the data collection carried out? The role of focus group facilitator was under-reported. Were the methods reliable? Yes Are the data 'rich'? No - they do not adequately fit the aim of the study

Study details	Summary of study	Results	Other
	Inclusion criteria · Women with intact uterus and ovaries · Aged 40-55 Exclusion criteria Intervention None Data collection A purposeful study consisting of semi-structured and open-ended 1 hour interviews (one per woman). Data analysis Interview tapes were transcribed and Content-analysed (Field and Morse 1985). Validity was maintained by sharing data and 'checking in' with women and researchers over time. Fieldnotes and data-trails were kept with the expectation of further interviews (not reported here).	it is so confusing. Some women did not want information that was related to money-making (e.g. doctors with interests or drug-manufacturers). "Women are consumers now, and women need to be more educated to see through it (vested interests in keeping women on hormones). The researchers' conclusions state that women need help to understand aspects of ageing, chronic disease and life-transitions in relation to menopause.	
Full citation Armitage,G.D., Suter,E., Verhoef,M.J., Bockmuehl,C., Bobey,M., Women's needs for CAM information to manage menopausal symptoms, Climacteric, 10, 215-224, 2007 Ref Id 303007 Country/ies where the study was carried out Canada Study type Quantitative. Content/method	Aim of the study To identify information needs of women regarding complementary and alternative medicine (CAM) Characteristics Not reported Inclusion criteria Women using Calgary women's health centre. Immigrant and 'at-risk' women were particularly encouraged to take part. Exclusion criteria None reported Intervention None Data collection A self-administered mail-out survey questionnaire. Questions were informormed by qualitative results of an earlier phase of the study. Questionnaires were mailed out to 413 women who were predominantly white and well educated (despite efforts to recruit a diverse range). Women were asked to choose a score of 1 to 5 (1 = strongly disagree; 5 = strongly agree) regarding statements about trustworthiness of information and what 'ideal' infomormation about CAM would consist of. Data analysis	Results relevant to protocol Strongly disagree - strongly agree Lickert scale answers (what good information consists of): Good information is based on government/not-for-profit information: 1=11 (2.7); 2 = 16 (4.0); 3=50 (12.3); 4=93 (23.0); 5=235 (58) Good information includes views of doctors: 1=17 (4.2); 2=31 (7.7); 3=104 (25.7); 4=144 (35.6); 5=109 (26.9) Good information includes personal accounts women who have taken treatment: 1=9 (2.2); 2=33 (8.0); 3=74 (18.0); 4=114 (27.8); 5=180 (43.9) Good information includes views of CAM practitioners: 1=9 (2.2); 2=30 (7.3); 3=84 (20.5); 4=148 (36.1); 5=139 (33.9) Not important - very important Lickert scale (relevance of information topics): Which treatments relate to which symptoms: 1=0 (0); 2=0 (0); 3=7 (1.7); 4=40 (9.9); 5=358	Comments Limitations There was no hierarchy of how important information information-topics in relation to each other. No women's characteristics list despite researchers targeting vulnerable women to achieve diversity. Quality checklist NICE appendix C methodology checklist for RCTs: A. Selection bias (systematic differences between the comparison groups): Unclear B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation): Unclear C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants): Unclear D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified): The assessment was self-administered and subjective.

Study details	Summary of study	Results	Other
	Descriptive analysis was performed (frequencies and means). Multivariate modeling was used to determine if there were any significant differences (p<0.05) among the preferred information sources. Percentages were recorded alongside frequency scores for each point on the Lickert scale.	(88.4) How a therapy works: 1=3 (0.7); 2=5 (1.2); 3=32 (7.8); 4=99 (24.2); 5=270 (66.0) How long it takes to work: 1=2 (0.5); 2=6 (1.5); 3=41 (10.1); 4=122 (30.0); 5=235 (68.0) How long should I take the treatment after seeing results: 1=2 (0.5); 2=4 (1.0); 3=34 (8.3); 4=91 (22.2); 5=279 (68.0) Side-effects: 1=0 (0); 2=0 (0); 3=4 (1.0); 4=16 (3.9); 5=388 (95.1) Which treatments can be combined (e.g. complementary and conventional): 1=2 (0.5); 2=1 (0.2); 3=11 (2.7); 4=49 (12.0); 5=344 (84.5) A list of places I can get further information: 1=4 (1.0); 2=8 (2.0); 3=35 (8.6); 4=101 (24.9); 5=258 (63.5) How to evaluate the quality of a therapy: 1=4 (1.0); 2=5 (1.2); 3=30 (7.4); 4=102 (25.2); 5=264 (65.2)	
Full citation Becker,H., Stuifbergen,A.K., Dormire,S.L., The effects of hormone therapy decision support for women with mobility impairments, Health Care for Women International, 30, 845-854, 2009 Ref Id 303070 Country/ies where the study was carried out Texas Study type Quantitative RCT (methods)	Aim of the study To evaluate tailored HT decision support to women with mobility impairments. Characteristics Ethnicity African American 6% White 87% Other 7% Mean age 53 At least a college degree 58% HRT use at baseline %	Results relevant to protocol Time 1; time 2; time 3 Mean±SD DCS total score Tailored DS group (n=86): 2.68±0.78; 2.14±0.65; 2.13±0.70 NAMS booklet group (n=90): 2.49±0.83; 1.99±0.58; 1.94±0.73 Knowledge score Tailored DS group (n=86): 9.44±4.62; 14.77±3.62; 12.42±4.13 NAMS booklet group (n=90): 10.17±3.98; 15.03±3.20; 13.28±3.47	Comments Limitations Mean±SD baseline characteristics not reported for each group. Sample size calclation not reported. Quality checklist NICE appendix C methodology checklist for RCTs: A. Selection bias (systematic differences between the comparison groups): None B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation): Unclear C. Attrition bias (systematic differences between the comparison groups with

Study details	Summary of study	Results	Other
	Never 47 Previous 30 Current 23 Inclusion criteria · Aged 40 to 65 · Have at least two of four mobility limitations identified in the National Health Interview Survey or indicate that they used adaptive equipment because of mobility limitations (Not required to indicate they presently were making a HT decision to participate) Exclusion criteria Only inclusion criteria reported Intervention Once baseline questionnaires were returned, participants were randomly assigned to one of the two interventions. Tailored support decision booklet Outlined risk factors associated with heart disease, osteoporosis, and cancer prevention and early detection strategies. The booklet includes current guidelines (American College of Obstetricians and Gynaecologists, US Federal Drug Administration and North American Menopause Society) as well as specialised information for this population. Provide information about the National Centre on Physical Activity and Disability to help women with disabilities to become more physically active. Case studies describing women with physical impairments are also provided. North American Menopause Society (NAMS) Menopause guidebook Contains a general explanation of menopause, latest clinical guidelines for menopause treatment, and strategies for achieving optimal long-term health. Does not provide information specific to women with mobility impairments. Data collection Participants were mailed materials for their group and a questionnaire packet that included the DCS and knowledge test. Follow-up telephone calls were made if		respect to loss of participants): Unclear D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified): None

Study details	Summary of study	Results	Other
	questionnaires were not returned. 6 months after participants indicated they had completed their second questionnaire packet, the last questionnaire packet was mailed to them. Data analysis The DCS (O'Connor et al., 1998) is a 16-item scale assessing uncertainty about the choice to use HRT, values clarity, perceived support, information and decision-making effectiveness. Higher scores reflect greater decision conflict. If a scale had missing data for less than 15% of the items, the mean score for the individual on the scale was imputed; otherwise, the entire scale was treated as missing for the individual.		
Full citation Bravata,D.M., Rastegar,A., Horwitz,R.I., How do women make decisions about hormone replacement therapy?, American Journal of Medicine, 113, 22-29, 2002 Ref Id 303163 Country/ies where the study was carried out USA Study type Qualitative (method)	Aim of the study An investigation into how patients make decisions and the role clinicians can play in the process - in the context of deciding about HRT. Characteristics Women contacted: N = 35 (10 excluded for not meeting inclusion criteria; 2 refused informed consent) Women interviewed: N = 23 White: 96% Professional/managerial: 74% Age range: 35 - 72 Inclusion criteria · Currently making medically complex decisions regarding HRT. · Menopausal (including surgical menopause). · English speakers. Exclusion criteria Past experience of HRT. Intervention None Data collection 23 women who were deciding on hormone therapy, but not begun treatment, took part in semi-structured interviews (in groups of 2 - 5). They were either identified by their primary healthcare providers or responded to posters in community clinics. Questions included: "What role would you want your physician to play	Results relevant to protocol Helpful information from gynaecologist: "I would have confidence in him, leading me in the direction of what he thought was best from a physician's point of view, but still leaving me to make up my own mind." "I would like the doctor to be strong one way or the other. Not to waver too much. So I think scientific data is important, but also the doctor should take a position." Women would have liked their doctors to be mindful that they pay for prescriptions.	Comments Limitations The coding was done by computerised keyword-identification which is not as accurate as manual coding which recognises nuances and synonyms. Quality checklist NICE Appendix H: Methodology checklist for qualitative studies Is a qualitative approach appropriate? Yes How well was the data collection carried out? Unclear Were the methods reliable? They were well reported, but no citations given which indicates the methods were not standardised. Are the data 'rich'? No Is the analysis reliable? Unclear - it appears to have been over-processed by the analysts. Is the role of the researcher clearly described? No

Study details	Summary of study	Results	Other
	in helping you to make the decision?" "What kind of information would you like your doctor to give you to help you make the decision?". Data analysis Transcripts of interviews were converted into a database using 'Folio VIEWS', and coded with descriptive labels using women's language. Labels were derived from key words, and checked for completeness and accuracy by a second researcher. Patterns and common themes were developed by identifying recurring categories and combinations of themes. Themes were organised into a model of patient decision making.		
Full citation Clinkingbeard,C., Minton,B.A., Davis,J., McDermott,K., Women's knowledge about menopause, hormone replacement therapy (HRT), and interactions with healthcare providers: an exploratory study, Journal of Womens Health and Gender-Based Medicine, 8, 1097-1102, 1999 Ref Id 303318 Country/ies where the study was carried out USA Study type Quali/quanti (content)	Aim of the study To elicit women's preferences for presentation and framing of complex risk information. Characteristics All 665 women lived in Boise, Idaho. Inclusion criteria Peri and post-menopausal women recruited through hospital advertising. Exclusion criteria Intervention Data collection The survey consisted of 22 items: checklist, open-ended and multiple choice. Open-ended responses were analysed using standard content analysis (Kerlinger 1973). Outcomes were Sources of information about menopause; Knowledge of health risks associated with menopause; Knowledge about HRT. Data analysis	Results relevant to protocol % of women who endorsed menopausal information from the following sources: Magazines: 76%; Healthcare providers (HCP): 68%; Friends: 52%; TV: 44%; Mother: 44%; Public lectures: 10%; Library: 7%. Menopausal topics women wanted to discuss with HCP: HRT: 37%; General symptoms: 33%; "Other things": 12%. Women who felt their questions were not answered by HCP: 36% Women who wished they had received better information about alternative treatments for symptoms: 10% Women who preferred other sources of information to HCP: 13% Many women left doctor's appointments without the information they needed due to short consultations and verbal-only communication. Others received denigrating comments such as "It's not such a big deal", and "You're like an old chicken that's not laying eggs anymore." Questions women wanted their HCP to answer: When will periods end with HRT? Why do I feel so lousy when I'm taking hormones? What does one believe with all the conflicting reports one hears? Will all my questions be answered?	Comments 99% of women were Caucasian. Limitations Quality checklist Is a qualitative approach appropriate? Yes How well was the data collection carried out? The number of unreturned questionnaires was not reported. Were the methods reliable? Yes Are the data 'rich'? Not enough direct quotations from women. Is the analysis reliable? Yes Is the role of the researcher clearly described? There is no report of how the questions were phrased.

Study details	Summary of study	Results	Other
		Reassurance was needed that: Male doctors are well versed in women's issues.	
Full citation Connelly,M.T., Ferrari,N., Hagen,N., Inui,T.S., Patient-identified needs for hormone replacement therapy counseling: a qualitative study, Annals of Internal Medicine, 131, 265-268, 1999 Ref Id 303338 Country/ies where the study was carried out USA Study type Quantitative. Content/method	Aim of the study To understand women's concerns and better align the content of counselling with women themselves. Characteristics Eligible: N = 114 Declined: n = 34 Interviewed: N = 26 Median age (range) 53 (42-70) White 85% Median household income 46,313$ Hysterectomised 31% Inititiated HRT discussion with provider 54% Inclusion criteria Member of Harvard Pilgrim healthcare maintenance organisation in Boston. Exclusion criteria Women excluded after saturation of N = 26. Intervention None Data collection At interview, women were asked to describe their decision-making process and identify the factors regarding HRT that were of greatest concern to them. Data analysis The interviewer transcribed the interviews which were checked for accuracy by two further researchers. The panel then identified content domains by a process of consensus.	Results relevant to protocol Topics which women felt should be included in guidelines for menopause counselling (ranked by popularity) %: Risk of breast cancer: 77 Medication: 73 Osteoporosis: 69 Prevention of heart disease: 58 Insomnia: 54 Living with medical uncertainty: 54 Genitourinary symptoms: 50 96% thought provider opinion was an important part of information, 81% valued media reports, 77% found experiences and opinions of friends useful (family: 60%). A secondary outcome was which of these topics (or 'domains') women would recommend to the medical practices and medication-'counsellors'.	Comments Limitations No copy of interview schedule is included in the paper. Quality checklist NICE Appendix H: Methodology checklist for qualitative studies How well was the data collection carried out? Well Were the methods reliable? Yes Are the data 'rich'? No Is the analysis reliable? Yes Is the role of the researcher clearly described? Yes
Full citation Deschamps,M.A., Taylor,J.G., Neubauer,S.L., Whiting,S., Green,K.,	Aim of the study To compare the effects of pharmacist consultation versus a decision aid (DA) on women's decision	Results relevant to protocol DCS score including the "informed" subscale items Baseline; survey 2	Comments Sample size: 64 women in each group required to detect a 0.5 effect size in

Study details	Summary of study	Results	Other
Impact of pharmacist consultation versus a decision aid on decision making regarding hormone replacement therapy, International Journal of Pharmacy Practice, 12, 21-28, 2004 Ref Id 282884 Country/ies where the study was carried out Canada Study type Quantitative RCT (method)	conflict regarding the use of HRT and subsequent satisfaction with the decision-making process. Characteristics n(%) White 104(99.0) Greater than high school education 85(35.2) Employment Technical: 37(35.2) Professional: 37(35.2) Pharmacist group (n=49); DA group (n=56) HRT use Current: 11(22.4); 9(16.1) Previous: 4(8.2); 7(12.5) Never: 34(69.4); 40(71.4) Menopausal status Peri: 32(65.3); 40(71.4) Post: 12(24.5); 11(19.7) Hysterectomy with at least one ovary: 4(8.2); 5(8.9) Inclusion criteria · Aged 48 to 52 · Recruited from a family medicine clinic · English speaking peri- and post-menopausal women regardless of current or previous HRT use Exclusion criteria • Already consulted the study pharmacist • Premenopausal HRT contraindicated Intervention Pharmacist consultation The pharmacist held a postgraduate Phar.D. with several years' experience in women's health; they had access to the patient's medical chart. The 40-minute private consultation reviewied the risks and benefits of HRT and was based on the prescribing guidelines produced by the Society of Obstretricians and Gynaecologists of Canada. Charts and graphs were used to visually represent population data and to provide consistency between patient encounters.	"I am aware of the choices to reduce my risk of heart disease and osteoporisis" Pharmacist group: 2.7; 1.7 DA group: 2.7; 1.7 "I feel I know the benefits of HT" Pharmacist group: 3.0; 1.8 DA group: 3.0; 1.7 "I feel I know the risks of HT" Pharmacist group: 3.2; 1.8 DA group: 3.2; 1.8 Averge "informed" score Pharmacist group: 3.0; 1.8 DA group: 3.0; 1.7 DSC score Pharmacist group: 3.0; 2.0; p<0.05 DA group: 3.0; 1.9; p<0.05	decision conflict with 80% power and alpha=0.05. Financial support by an unrestricted grant from Eli Lilly. Limitations 77 women randomised to the pharmacist group and 61 to the DA group. 20 women failed to make or keep appointments to receive their intervention, 3 baseline surveys were incomplete, 13 did not make or attend appointments, 1 moved away, 3 saw their doctor too late to be included and 1 withdrew their consent. DA not described in any detail. DCS items not described. Unclear when the second survey was completed. Quality checklist NICE appendix C methodology checklist for RCTs: A. Selection bias (systematic differences between the comparison groups): Randomisation not decribed B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation): Unclear C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants): 91 out of 138 women completed the study D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified): None

Study details	Summary of study	Results	Other
	At the end of the consultation, the pharmacist and patient agreed on a provisional plan regarding HRT. DA Titled "Making Choices: hormones after menopause" Ottawa Health Decision Centre. Communicate the risks and benefits of therapies to assist the patient in clarifying values and expectations. After each intervention, patients were instructed to see their doctor within two to four weeks. Data collection The DCS contains 16 items measured on a scale of 1 (strongly agree) to 5 (strongly disagree) capable of discriminating between women making or delaying decisions and between different educational interventions. The three question "informed" subscale of the DCS assessed the perception of being informed. Data analysis Differences between the intervention groups were analysed with t-tests of indepdendent means while dependent means t-tests were used to detect changes within groups.		
Full citation Doubova,S.V., Infante-Castaneda,C., Martinez-Vega,I., Perez-Cuevas,R., Toward healthy aging through empowering self-care during the climacteric stage, Climacteric, 15, 563-572, 2012 Ref Id 266636 Country/ies where the study was carried out Mexico Study type Qualitative (content)	Aim of the study To identify the changes in women's discourse regarding their concerns and needs about the climacteric stage and self-care after they had participated in an integrative women-centred healthcare model with empowerment for self-care. Characteristics N = 121 Mean age ±SD 49.3 ± 3.0 %: Up to secondary school level: 39.6 Beyond secondary school level: 60.3 Professionals: 4.1 Low-skilled or craft workers: 30.5 Housewives: 60.3 Retired: 5.1	Results relevant to protocol Peer discussion as a way of learning how to approach the menopause: Information which women found empowering: "I learnt that we do not have to leave everything up to the doctor" "For me (the menopause) is one more stage, another stage of my life." On groupwork: "We get to know ourselves through others." "It is very important to start working with ourselves: taking care, exercising. (If) we are not aware of this we will always continue living for others." Learning to live for themselves, not just others. "I am responsible for (my health)." The importance of getting information from reliable sources. Motivation to transmit acquired knowledge of menopause to others. At the end of the sessions women were less	Comments Limitations No citation for a standardised analytical method. Quality checklist NICE Appendix H: Methodology checklist for qualitative studies Is a qualitative approach appropriate? Yes How well was the data collection carried out? Well Were the methods reliable? Methodology non-standardised and un-cited Are the data 'rich'? Yes Is the analysis reliable? Yes Is the role of the researcher clearly described? Yes

Study details	Summary of study	Results	Other
	Inclusion criteria Women who had attended a consultation at family medical practice. Exclusion criteria Intervention Data collection A research-based bio-psycho-social care model for information provision by a doctor, a nurse and a psychologist centred on women's information needs, doubts and personal experiences orientated towards the empowerment for self-care and applicable in family clinics. (Described in full in Doubrova 2011). Women's narratives were analysed during the sessions. Data analysis 4 mixed disciplinary researchers carried out coding with continual iteration between complete dataset and codified extracts.	concerned with the social and sexual stigma of menopause. They found it a less taboo subject which meant they were able to share ideas and learn from each other. The importance of limiting food. "If I control my food, I control other's food. If I am well emotionally we are all well." (speaking of the advantages of self-care when one is the "nucleus" of the family). "By myself, I would not know what to do. Hearing others, I have another perspective to do other things."	
Full citation Forouhari,S., Khajehei,M., Moattari,M., Mohit,M., Rad,M.S., Ghaem,H., The Effect of Education and Awareness on the Quality-of-Life in Postmenopausal Women, Indian Journal of Community Medicine, 35, 109-114, 2010 Ref Id 266790 Country/ies where the study was carried out Iran Study type Quantitative RCT (method)	Aim of the study To evaluate the effect of an information-giving course about menopause on women's quality of life. Characteristics Age, mean±SD 50.63±2.7 Study group; control group n(%) Menopause status Premenopause: 5(13.6); 5(13.6) Perimenopause: 6(21.9); 7(25.1) Postmenopause: 20(64.5); 19(61.3) Occupation Housewife: 25 (80.64); 24 (77.41) Employed: 6 (19.36); 7 (22.59) High school education 5 (15.8); 3 (13.1) Inclusion criteria · Healthy pre/peri/post-menopausal women were selected by simple random sampling · Aged 44 to 55 · Symptoms of moderate to severe hot flushes at	Results relevant to protocol Mean quality of life score Before intervention; 3 months after intervention Study group 81.7; 75.3 SD (within group change) = 6.4 P= 0.001 Control group 74.8; 75.8 SD (within group change) = 1.4 P= 0.001	Comments The study took place in Shiraz which is a wealthy area of Iran. Limitations It is not reported whether the questionnaire was translated from English. Unable to calculate 95% CIs from the SDs reported. Quality checklist NICE appendix C methodology checklist for RCTs: A. Selection bias (systematic differences between the comparison groups): Unclear exclusion criteria B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation): None C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants): Unclear D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified): Unclear - knowledge score is not described in detail

Study details	Summary of study	Results	Other
	least once a day · Not using any kinds of medication and/or HRT 6 months prior to the study · Not completing ay physical exercise (<20 minutes/week) · Married · Lack of illnesses creating hot flash like symptoms or impairing quality of life Exclusion criteria See inclusion criteria Intervention Randomised by assigning each participant a number and then using a random table pointed a finger in order to choose an arbitary and random starting point, they were the first participant in the study group. Then moved across the row of numbers to select the first participant in the control group. Continued to assign every number to each of the groups until there were two groups with 31 participants in each. An educational intervention 45 to 60 minute weekly sessions for 6 weeks in the form of 8-person discussion groups. Information about female organs, what menopause is, symptoms and complications, approaches to complications, exercise, relaxation and their effect on symptoms. The control group received no education and they had no contact with the study personnel (or other participants) beyond recruitment and data collection. Data collection All women's scores for Quality of Life were obtained using a 26-question questionnaire (Hilditch 1996) before and 3 months after the education course. The quality of life questionnaire contained 4 domains including: vasomotor, psychosocial, physical and sexual aspects. Women made their responses via a Lickert Scale from 1 (no problems) to 6 (problems causing severe distress).		

Study details	Summary of study	Results	Other
	Minimum score = 26 and highest = 156. The higher the point score the more severe the symptoms. Data analysis Powering (using pilot study): 31 women were needed for each group (with at least 25 completing the study) for 95% power to detect at least a 5% difference in quality of life.		
Full citation Fortin,J.M., Hirota,L.K., Bond,B.E., O'Connor,A.M., Col,N.F., Identifying patient preferences for communicating risk estimates: a descriptive pilot study, BMC Medical Informatics and Decision Making, 1, 2-, 2001 Ref Id 229300 Country/ies where the study was carried out USA Study type Qualitative and quantitative	Aim of the study To elicit women's preferences for the presentation and framing of complex risk information Characteristics Age Mean (range): 51 (38-67) <45: 6 45-55: 24 >55: 10 Race Non-white: 20 White: 20 Income $ <25,000: 11 25,000 - 49,000: 13 >49,000: 16 Education Low (<grade 13/vocational): 9 High (2-4 years of college/post-grad): 10 Inclusion criteria Peri and post-menopausal women. Exclusion criteria Not reported Intervention None Data collection 40 women were recruited via hospital advertising in March - May 1999. 8 focus groups and 15 interviews were conducted to assess women's preferences for menopausal risk communication. Women were shown different graphical formats, metrics and time-horizons illustrating a fictional patient's risk of coronoary heart disease, hip fracture and breast cancer with and without HRT. Women's preferences were assessed using	Results relevant to protocol Bar graphs were preferred by 83% of women over line graphs, thermometer graphs, 100 faces and survival curves. Lifetime risk estimates were preferred over 10 or 20 year horizons. Absolute risks were preferred over relative risks and numbers needed to treat. Preference of n±SD Bar graph: 4±1; Linegraph: 3.1±0.9; Thermometer chart: 2.6±1.1; "100 faces" (visual Lickert): 2.4±1.5; Survival curves: 2.5±1.1 Preferences for Risk Information Presentations (column boundaries marked by dashes): a. Time Horizon: 1st Choice (n = 4O) / 2nd Choice (n = 33) 10-year 23% / 12% 20-year 20% / 58% Lifetime 55% / 27% No response 3% / 3% b. Multiple diseases and multiple time Preference: Horizons (n = 40) Set A: I disease over 3 time horizons 53% Set B: 3 diseases over I time horizon 43% No response 5% c. Relative v absolute risk: Graph Preference (n = 25) / (n 20) Relative risk: 28% / 30% Absolute risk: 72% / 65% No response: 0% / 5% d. NNT Preference (n-40) / Standard explanation (1 in x) 28% Alternative explanation (x out of I 00) 45% Neither 25% No response 3%	Comments This paper is very graphically presented, and is best understood by seeing it as it presents the graphical reporting styles being assessed. Limitations A pilot study. Quality checklist How well was the data collection carried out? Well Were the methods reliable? Yes Is the role of the researcher clearly described? This is under-reported, especially the analysis which apprears to be a mixture of qualitative and quantitative. No inclusion of the "worksheet" format in paper.

Study details	Summary of study	Results	Other
	Lickert scales, ranking and abstractions of discussions. They indicated preferences via individual 'worksheets' prior to focus groups. Data analysis Descriptive statistics were performed on sub-groups stratified according to race, income and education. Means for differences in preference were assessed using a Wilcoxon signed-rank test.	Preferences for Risk Information Presentations a. Time Horizon: 1st Choice (n = 4O) / 2nd Choice (n = 33) 10-year 23% / 12% 20-year 20% / 58% Lifetime 55% / 27% No response 3% / 3% b. Multiple diseases and multiple time: Preference Horizons (n = 40) Set A: I disease over 3 time horizons: 53% Set B: 3 diseases over I time horizon: 43% No response: 5% c. Relative v absolute risk: Graph preference (n=25) / Text preference (n=20) Relative risk: 28% / 30% Absolute risk: 72% / 65% No response: 0% / 5% d. NNT Preference (n=40) Standard explanation (1 in x): 28% Alternative explanation (x out of 100): 45% Neither: 25% No response 3%	
Full citation Fox-Young,S., Sheehan,M., O'Connor,V., Cragg,C., Del,Mar C., Women's perceptions and experience of menopause: a focus group study, Journal of Psychosomatic Obstetrics and Gynecology, 16, 215-221, 1995 Ref Id 303556 Country/ies where the study was carried out Australia Study type Qualitative	Aim of the study To investigate women's perception and experience of HRT, osteoporosis and doctor-patient relationships. Characteristics Volunteers: N = 260 Selected: N = 148 Dropouts were explained as failure to keep appointments or inability to be contacted. Focus groups: N = 40: Aged 45 - 55 (mean: 48.4) Highest secondary school education: 56.3% Pre-menopausal: 22.5% Perimenopausal: 20% Post-menopausal: 17.5% Hysterectomy: 40% Have used HRT: 42.5% Ceased HRT: 47.1% Inclusion criteria	Results relevant to protocol Women needed information that was clear and uncontradictory: "You hear such divergent opinions." Women felt that the menopause is a taboo subject and not generally discussed, so therefore led to fear. This led to a need for reassurance and reassurance of not being alone. Women's need for information of menopause was inseparable from their loneliness and empathy with their mothers' suffering with no HRT option. Women wanted doctors to treat them as partners in decision-making*. They wanted to be told more about the pros and cons of treatments. Women who had been hysterectomised felt their doctors had not prepared them for menopause beforehand: "I was very angry about the lack of preparation for the (menopausal) changes I experienced after my operation."	Comments *This links to generic treatment guidelines. Limitations Very poor reporting of method. It was not clear how many researchers were involved in the data collection or analysis. No standardised analytical method was reported. In spite of the above limitation, thorough descriptions of women's views are reported. Quality checklist

Study details	Summary of study	Results	Other
	Sample randomly selected from electoral role. Focus group participants were selected to proportionately represent different HRT statuses (used successfully, used unsuccessfully, never used, had changed doctors in serch of HRT). Exclusion criteria Intervention None Data collection Allocation to 7 focus groups was based on knowledge and experience of HRT to maximise homogeity of groups. The relevant semi-structured FG topic was 'Current access to information and recommended improvements." The FGs were facilitated two researchers:one moderator and one scribe. Data analysis A summary of statements made during focus groups were compiled by the scribe and checked for completeness by the the moderator and other members of the research team. This data was then analysed for themes.		
Full citation Hallowell,N., A qualitative study of the information needs of high-risk women undergoing prophylactic oophorectomy, Psycho-Oncology, 9, 486-495, 2000 Ref Id 303722 Country/ies where the study was carried out UK Study type Qualitative (content)	Aim of the study To determine the information needs of women who had undergone surgical menopause (bilateral oophorectomy). Characteristics Mean (range) or n(%) Age 44.4 (32 to 62) Age at surgery 38.8 (31 to 45) Time since surgery 5.5 (0.5 to 25) School leaving age 15-16: 17 (74%) 17-18: 3 (13%) Occupational diplomas/further education 2 (9%) Degree 1 (4%) Inclusion criteria	Results relevant to protocol 6 women could not recall being told they would need HRT before surgery. For instance, a doctor gave a woman 'a patch' to 'change on Sunday', but did not tell her what it was. Women needed to have known that their oestrogen would fluctuate and they might have menopausal symptoms following surgery as none were told this. They also needed to have known how long to take HRT for (some HCPs did not know this). They would also like to have been informed of the likely cost of prescriptions for HRT as money was an issue and they had assumed it would be free. Although most women were informed that they would have to take HRT following surgery, many said this was the only information they received: "My information from the hospital was about the operation ...it just tells you what it does. That was it. It didn't say - it said a bit about, you will be given HRT, and that was it."	Comments Recommendations include gynaecology nurses to be available for information-provision both pre and post surgery. Limitations The authors note a potential for sample bias in that women with issues about information provision might have been more likely to take up the offer of a interview, (but this is similar in other interview studies). Quality checklist NICE Appendix H: Methodology checklist for qualitative studies Is a qualitative approach appropriate? Yes How well was the data collection carried out? Well reported Were the methods reliable? Yes, standardised with citations. Are the data 'rich'? Reasonably Is the analysis reliable? Yes Is the role of the researcher clearly described? Yes

Study details	Summary of study	Results	Other
	· Prophylactic bilateral oophorectomy before age 46 · Pre-menopausal prior to surgery · No previous history of cancer · 2 or more relations with ovarian cancer Exclusion criteria Not reported Intervention None Data collection Recruitment was conducted from the UK Co-ordinating Committee for Cancer Research's Familial Ovarian Cancer Register. Invited to respond: N = 33 Recruited: N = 23 Recruitment ceased once saturation was reached in the data analysis. Women were asked, by interview, a series of questions on their understanding of ovarian function and menopause. They were also asked for their understanding and recall of information they received pre and post surgery, the sources of this information and what further information they wanted or needed. Data analysis Following transcription of interview tapes, thematic analysis was undertaken. The data were indexed on a case by case basis, which allowed patterns and relationships between codes to emerge within the dataset. Coding was refined by comparing interviews and identifying deviant cases (Silverman 1993). The resulting set of categories were then collapsed into higher order themes (including Knowledge of the menopause and Information needs). The analysis was then validated by the respondents. Some frequency data were reorded, not to indicate a hierarchy of import, but to summarise the data.	Only 1 woman recalled being given a choice about the different forms of HRT. 3 women were not given a choice about HRT, with 1 having a hormonal patch inserted under anaesthetic. Women wanted the information to make the decision for themselves. Women with implanted patches had to delay decision-making by 6 months. There was a conflict between information given by gynaecologists and information given by GPs. The researchers compared a drop in HRT compliance (after 18 months) with an American study with a 100% compliance. They infered this as being a result of poor information provision regarding risks of surgically induced menopause i.e. cardio-vascular incidents and osteoporosis (Schrag et al., 1997).	
Full citation Hunter,M., O'Dea,I., An evaluation of a health education intervention for mid-aged women: five year follow-up	Aim of the study An evaluation of the long term impact of a healthcare intervention in primary care for pre-menopausal women.	Results relevant to protocol Knowledge of menopause (mean ± SD): Intervention: 5.16±2.23; Control: 3.74±2.11 The intervention group had significantly greater	Comments Limitations No measurement of pre-intervention knowledge reported (this may be because

Study details	Summary of study	Results	Other
of effects upon knowledge, impact of menopause and health, Patient Education and Counseling, 38, 249-255, 1999 Ref Id 303830 Country/ies where the study was carried out UK Study type Quanti (RCT). Method	Characteristics Post-intervention: n = 45 Post-control: n = 41 Peri-menopausal: 55% Post-menopausal: 12% Taking HRT: 29% There were no significant group differences in terms of socio-demographic/menopausal status. All women had been pre-menopausal during the intervention-phase of the study (as it was a preventative intervention). Inclusion criteria Women aged 50. All women had been in the study for 5 years, and had been exposed to either the intervention or control in 1991. Exclusion criteria Pre-menopausal Intervention Two 90 minute workshops which included: Health education (information about the menopause, self-help and medical treatments) Discussion of expectations and beliefs about menopause General health (reducing stress, exercise, smoking and diet). Data collection Questionnaires sent: N = 86 Returned questionnaires: N = 78 (91% response rate) Sample: N = 68 (10 excluded for being pre-menopausal). 4 questionnaires were self-administered: Socio-demographic questions; knowledge about menopause (Hunter and Liaho 1994); Menopause Representation Questionnaire (O'Dea and Hunter 19?), and Women's Health Questionnaire (Hunter 1992), and an evaluation of study-participation. Data analysis Mean questionnaire scores (with SDs) were calculated for each group. The significance of differences in outcome between groups was measured with t-tests and chi-square tests.	knowledge than the control group (t=2.57; df=65; p<0.01) Influene of study on experience of the menopause: Intervention: 4.15±0.83; Control: 3.38±1.36 The intervention group said study-participation had influenced their experience of the menopause to a significantly greater extent than the control group (t=2.46; df=66; p<0.01) % of intervention group who rated the course as follows: Helpful: 88; Informative: 92; Optimistic: 86.5; Supportive: 96; Helped deal emotionally with menopause: 75; Helped deal with practical aspects of menopause: 87	women were pre-menopausal then). No overall quality-of-life score. Ambiguous outcome = 'influence' of menopause (no % given for the extent to which this was positive. Quality checklist NICE appendix C methodology checklist for RCTs: A. Selection bias (systematic differences between the comparison groups): None. Good response rate from the original women. B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation): Unclear C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants): None (though a 4:1 ratio of women were peri-menopausal (compared with post-menopausal) D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified): Seriously biased because it is not known what other events had taken place over the 5 years since the study started. The researchers analysing the data were not reported as blinded. The researchers had a strong interest in both the intervention and the questionnaires. Outcomes were often ambiguous (see Limitations).
Full citation Kiatpongsan,S., Carlson,K., Feibelmann,S., Sepucha,K., Decision aid reduces misperceptions about	Aim of the study To evaluate the role of an up-to-date decision aid (DA) a 44-minute DVD and booklet in improving women's knowledge of menopausal symptom	Results relevant to protocol Knowledge scores Mean difference (95% CI) between the two arms	Comments Sample size: 100 participants required in each of the four arms to detect a difference in total knowledge of 6% assuming a

Study details	Summary of study	Results	Other
hormone therapy: a randomized controlled trial, Menopause, 21, 33-38, 2014 Ref Id 303976 Country/ies where the study was carried out USA Study type Quantitative RCT (method)	management, benefits of HT and risks of HT. Characteristics Control arm (n=213); DA arm (n=188) Mean±SD or n(%) Age 51±5.1; 51±5.5 Race White: 131(61.5); 120(64.5) Black: 58(27.2); 47(25.3) Other: 15(8.1); 21(9.9) Unkown: 4(2.2); 4(1.4) Education Higher than college graduate: 34(16.0); 28(14.9) College graduate: 44(20.7); 40(21.3) Some college: 74(34.7); 84(44.7) High school or less: 49(23.0); 28(14.9) Income US$ ≤30,000: 89(41.8); 71(37.8) >60,000: 54(25.4); 59(31.4) Inclusion criteria · Aged 40 to 60 · Menopausal symptoms · Discussed symptom management with their healthcare providers within the past 12 months or had taken any medicine or supplements to manage their menopausal symptoms Exclusion criteria Prior diagnosis of breast cancer Surgically or medically induced menopause (ovaries removed) Intervention Used a 2x2 factorial design. Participants were assigned to one of four arms (with DA or without DA; telephone survey administered either by an interviewer or by an automated voice recognition system). All participants were suryed by telephone 2 weeks after enrolling or receiving the DA. Assigned to one of four arms in blocks of four, in sequential order with the blocks, until all eligible participants had been assigned to an arm. DA 44-minute DVD and booklet "Managing	Total knowledge score 5.8 (2.3 to 9.3) P=0.001 DA arm: Mean 63.3% (SD 18.4%) Control arm: Mean 57.5% (SD 16.4%) P=0.001 Risks of HT subscore 2.1 (-3.0 to 7.2) P=0.422 Benefits of HT subscore 4.2 (0.03 to 8.5) P=0.048 General menopausal symptom managment subscore 11.0 (5.3 to 16.6) P<0.001 The DA arm had greater knowledge of menopausal symptom management than the control arm. Scores on knowledge about HT risks were not different between arms.	common SD of 20% with 80% power. Assignment: · Control & interviewer n=128 · Control & voice recognition n=127 · DA & interviewer n=130 · DA & voice recognition n=130 Analysed: · Control & interviewer n=115 · Control & voice recognition n=98 · DA & interviewer n=102 · DA & voice recognition n=86 Participants received a small incentive payment for participation (US$10 to US$20). Limitations The study staff were not blinded to assignment arms. Reasons for comparing a survey administered by an interview or automated voice recognition system appear irrelevant to the aim of the study. Quality checklist NICE appendix C methodology checklist for RCTs: A. Selection bias (systematic differences between the comparison groups): None B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation): None C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants): Yes: 42 participants lost to follow-up in the control arm and 72 participants lost to follow-up in the DA arm. D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified): None

Study details	Summary of study	Results	Other
	Menopause: Choosing Treatments for Menopause Symptoms" (2008 Health Dialog, Informed Medical Decisions Foundation). Provides evidence based information about symptoms of menopause, treatment options including HT, nonhormone prescription medications, herbal remedies and lifestyle changes, the benefits and risks of each treatment option, and vignettes about how women with menopause symptoms made decision about treatment options. This DA scored 23 out of 25 points in the IPDAS quality criteria. Data collection The knowledge test included 13 questions covering general menopausal symptoms and the benefits and risks associated with HT. Data analysis Calculated the total knowledge score by summing up the number of correct responses, dividing by the total number of items. Missing items were considered incorrect. Any respondent who had more than half of the knowledge items missing was not given a score. Student t-test was used to compare mean scores in the control and DA arms. For missing items from responders, calculated knowledge scores using nonskipped items only and reran the analysis. For nonresponders, used a conservative estimate of mean knowledge score for the control arm and reran the analysis.		
Full citation Legare,F., Stacey,D., Dodin,S., O'Connor,A., Richer,M., Griffiths,F., LeBlanc,A., Rousseau,J.L., Tapp,S., Women's decision making about the use of natural health products at menopause: a needs assessment and patient decision aid, Journal of Alternative and Complementary Medicine, 13, 741-749, 2007 Ref Id 227793 Country/ies where the study was carried out	Aim of the study To identify the decision-making needs of women about the use of natural health products (NHP) Characteristics N = 40 Median age (range) 56 (44-67) Education, % Secondary education or less: 12.5 Post-secondary education: 87.5	Results relevant to protocol Women were ambivalent regarding doctors as sources of information: sometimes women were given all the information they needed from their physician, but they did not understand it. Women wanted information from doctors to be free from the doctor's own strong opinions. They wanted information to be objective, reliable and credible. Internet not considered a useful source of information because women needed help to distinguish what information is science from information that is marketing (especially re	Comments Limitations Quality checklist NICE Appendix H: Methodology checklist for qualitative studies Is a qualitative approach appropriate? Yes How well was the data collection carried out? Unclear how 'informants' were involved in the process. Were the methods reliable? Yes Are the data 'rich'? No Is the analysis reliable? Yes

Study details	Summary of study	Results	Other
Canada Study type Qualitative (method)	Decision making, n Preferred role in decision: Prefer to make decision alone: 12.5 Make decision with advice from doctor: 55 Share decision with doctor: 25 Prefer doctor to make decision alone: 0 Inclusion criteria · Aged 45 to 64 · Peri or postmenopausal women from 2 cities in Ottawa · Considering the use of NHP for menopausal reasons A purposeful sampling stratergy sought to recruit 15 key informants representing groups of individuals who may advise and/or guide women on use of NHPs (e.g. physicians, nurses, pharmacists etc). To recruit these a snowball approach was used by asking "well suited people" in each group to identify potential individuals. Exclusion criteria Not reported. Intervention N/A Data collection Women were recruited by local media (radio, newspapers, notice boards) and word of mouth. 6 focus groups and individual interviews with semi-structured questions. The questions were from a standardised schedule: OSDF (Cranny 2002). Data analysis Content analysis was carried out on the transcripts of interviews and focus groups. Women were sent their transcripts with a summary of the themes in order to verify the accuracy. Resulting categories were tabulated alongside illustrative quotations.	internet). 3/6 focus groups agreed they wanted education sessions (with a telephone information line). 2/5 focus groups agreed they wanted a trustworthy website as a way of providing information. Difficult decisions about the use of NHPs at menopause identified by focus groups: What to take and which product? Whether or not to take NHPs Take nothing at all? HRT or NHP? NHP in combination with HRT? Who to consult Changing from HRT to NHP Information sources focus groups said they needed: · Education sessions · Telephone line · More time with doctor · Trustworthy website.	Is the role of the researcher clearly described? Yes
Full citation Legare,F., Dodin,S., Stacey,D., Leblanc,A., Tapp,S., Patient decision aid on natural health products for menopausal symptoms: randomized controlled trial, Menopause	Aim of the study To evaluate the impact of a patient decision aid (PDA) regarding the use of natural health products (NHPs) at menopause on decision conflict, knowledge of NHPa, congruence between values and choice, persistence with an	Results relevant to protocol Pre intervention; post intervention; p value Mean±SD Control group n=41 PDA group n=43	Comments Sample size: 35 women in each group required to detect a 0.4 improvement in the DCS with a power of 80% and alpha=0.05. Taking into account possible dropouts (30%) aimed at recruiting 100 women.

Study details	Summary of study	Results	Other
International, 14, 105-110, 2008 Ref Id 304075 Country/ies where the study was carried out France Study type Quantitative RCT (method)	option, intention to disclose the use of NHPs to a physician or a pharmacist and intention to use decision support interventions in the future. Characteristics Control group (n=41); DA group (n=44) Mean±SD or n(%) Age 53.4±3.9; 54.3±4.7 Education No high school diploma: 2(5); 9(20) High school diploma: 21(51); 19(44) College/university diploma: 18(44); 16(36) Personal or household income, CAN$ <30,000: 4(10); 5(11) ≥60,000: 23(56); 20(45) Curent use HT: 13(32); 11(25) NHPs: 20(49); 25(57) Menopausal 30(73); 32(73) Inclusion criteria · Aged 45 to 64 years · Suffering from symptoms of the menopause · Considering NHPs for their menopausal symptoms · Able to read, understand and write French at grade 8 level · Capable of giving free, informed consent for their participation (Did not exclude women who reported using NHPs because they can reconsider their choice) Exclusion criteria • Women who reported symptoms for which there was no precise diagnosis • Owners and/or managers of natural health food stores • Pharmaceutical companies or pharmacies • Women with a close relationship with a study investigator Intervention Randomisation A biostatistician used computer generated	DCE score Total score Control group: 2.60±0.84; 2.08±0.61; p<0.0001 PDA group: 2.47±0.69; 1.92±0.57; p<0.0001 Uncertainty subscore Control group: 2.93±1.10; 2.33±1.01; p<0.0001 PDA group: 2.68±1.04; 2.06±0.92; p<0.0001 Inadequate knowledge subscore Control group: 2.98±1.16; 2.37±1.04; p=0.0022 PDA group: 2.71±1.00; 2.19±0.91; p=0.0060 Improvement in knowledge test Control group: 0.86±1.77 p=0.002 PDA group: 0.51±1.47 p=0.031 Difference between groups: p=0.162	Limitations The six stage process described in the DA intervention describes how the DA works but does not describe the content. 43 participants had a personal or household income ≥60,000 CAN$. 45 participants were already using NHPs. Quality checklist NICE appendix C methodology checklist for RCTs: A. Selection bias (systematic differences between the comparison groups): None B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation): Unclear C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants): 45 particpants in each group were enrolled, 41 completed the study in the control group and 43 completed the study in the DA group D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified): None

Study details	Summary of study	Results	Other
	unequal blocks. Sealed envelopes containing one of the two interventions were prepared by another individual external to the study. The investigators and research assistants involved in data collection and analysis were blinded to the participants' assignment. Paper-based PDA Developed by their research team using International PDA standards and the Ottawa Decision Support Framework. It consisted of a six stage process: be clear about the decision made, get the facts based on the best evidence avaliable, identify the avaliable questions, clarify what is important, select the role in making the decision and the next steps. Control group Paper-based general information brochure distributed by a community-based women's group. Focued on the physcological aspects on a diverse range of ways to manage these. It did not focus on making a decision regarding the use of NHPs for menopausal symptoms, but mentioned a few aspects regarding a smaller number of NHPs than the PDA. It did not assess risks and benefits regarding NHPs that had been identified. It did not address the lack of presence of evidence regarding the NHPs. Women were given two weeks to use their intervention, as a reminder women were given a call after the first week. Data collection The DCS comprised of 16 items divided into subscales: uncertainty, inadequate knowledge, unclear values, lack of support and ineffective choice. Each item is measured on a Likert scale from 1 (strongly agree) to 5 (strongly disagree). The total DCS score was obtained by summing up the 16 items and dividing by 16, resulting in a score which ranged from 1 (low decision conflict) to 5 (high decision conflict).		

Study details	Summary of study	Results	Other
	Knowledge of NHPs was assessed with a 10 item test on a response scale of yes (correct answer), no and unsure (wrong answer). The knowledge score was obtained by summing up the 10 items: 0= no correct answers to 10= all correct answers. The last data collection was preformed at the end of the second week, during a telephone interview conducted by a research assistant who was blinded to the intervention group. Data analysis A paired t-test was used to compare the results within each group. intention-to-treat analysis was performed. Analysis of covariance (ANCOVA) was used to compare results between each group while controlling for baseline scores.		
Full citation Liao,K.L., Hunter,M.S., Preparation for menopause: prospective evaluation of a health education intervention for mid-aged women, Maturitas, 29, 215-224, 1998 Ref Id 304101 Country/ies where the study was carried out UK Study type Quantitative RCT (method)	Aim of the study To assess the effects of a health education intervention on knowledge of menopause 3 months and 15 months later, and to assess whether the intervention would modify overly negative beliefs and menopause and health related behaviours. Characteristics Education group (n=45); control group (n=41); second control group (n=44) White British, % 76; 78; 79 Employed, % 89; 88; - Inclusion criteria 45 year old women (born 1946) registered at 5 general practices in south London Exclusion criteria • Taking HRT • Post-menopausal Intervention 50 women were randomly allocated to a second control group to be contacted at a later phase of the study to control for the effects of completing questionnaires by the original control group. Intervention The preparation intervention consisted of two	Results relevant to protocol Knowledge score Mean±SD Baseline; 3 months; 15 months Education group: 2.58±1.80; 5.56±2.60 ab; 5.19±2.06 ab Control group: 2.71±2.05; 3.05±2.08; 3.03±1.91 b Second control group: -; -; 3.52±2.04 a Significant within-group difference p<0.000 b Significant between-group difference p<0.001	Comments 106 out of 178 returned questionnaires giving a response rate of 60%. 11 of the 106 were excluded based on the criteria. Sample size at: baseline; 3 months; 15 months Education group: 45; 44; 43 Control group: 41; 3; 35 Second control group: -; -; 44 Limitations Knowledge score not described in detail. Control intervention and randomisation not described. Few baseline demographics are reported. Unclear if pre and peri menopausal women are included. Quality checklist NICE appendix C methodology checklist for RCTs: A. Selection bias (systematic differences between the comparison groups): Unclear B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation): Unclear C. Attrition bias (systematic differences between the comparison groups with

Study details	Summary of study	Results	Other
	educational sessions. Every 15 minute talk was followed by a 10 to 15 minute question and discussion session by the group. Group sizes varied between 4 and 8. The two sessions each lasted 1.5 hours. Workshop 1 · Warm-up exercise where each woman talked briefly about her concerns · "Menopause: facts and myths" talk on the menstrual cycle, hormonal and menstrual changes, hot flushes and vaginal changes, birth control and health issues in the post menopause (e.g. osteoporosis) · "Preparing for menopause" talk with particular attention to diet, exercise, smoking, alcohol, managing tension and stress · Homework: read handout, note questions and consider a health behaviour target Workshop 2 · Feedback and queries on the last session and handout · "Self-help and treatment at menopause" talk on self-help for hot flushes, relaxation, vaginal remedies, peer support, alternative therapies, the facts and myths of HRT · "Changing lifestyle" talk on goal-planning, sustaining effort and what to do if we lose interest · 20 minute practice session on goal-planning with example targets from participants Handout · Information on topics discussed in greater detail · Audio-cassette on stress and relaxation · Worksheets to aid goal-planning · List of useful addresses and telephone numbers Data collection Knowledge was assessed using 10 mulitple choice items chosen from Hunter et al., 1994 & Liao et al., 1995. A score of 1 was given to each correct response and 0 for each incorrect response resulting in a total score from 0 to 10. Data analysis For related samples t-tests were used to examine		respect to loss of participants): 6 participants in the control group were lost at the 15-month follow-up D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified): None

Study details	Summary of study	Results	Other
	within-group differences in the knowledge score. Independent t-tests (post-hoc sheffe) and analysis of variance (ANOVA) examined between-group differences.		
Full citation Mahon,S.M., Williams,M., Information needs regarding menopause. Results from a survey of women receiving cancer prevention and detection services, Cancer Nursing, 23, 176-185, 2000 Ref Id 295079 Country/ies where the study was carried out USA Study type Quanti. Method & Content	Aim of the study To describe women's information needs at menopause, and evaluate an education brochure. Characteristics N = 161 Age range: 26 -69 (mean 48) Self-identified menopause (or might have menopause): n = 86 (55%) Pre-menopausal: n = 69 (45%). Inclusion criteria Women attending a cancer screening and wellness centre who were given a copy of the brochure to read (questionn. Exclusion criteria Intervention The brochure, Understanding menopause and beyond was developed as an adjunct to patient-education regarding menopause (rather than a sole source). The manual was developed by 4 doctors (different specialties), a psychologist and a nurse. The brochure contained information on menopause-definition, symptoms & risk factors, HRT (benefits and side-effects), community-resources, suggested reading, and information to share with 'my' doctor. Data collection The brochure was evaluated by self-administered questionnaire. The women were a convenience sample of women seeking wellness services and education from a nurse-managed cancer screening centre in an urban mid-western city. Women were asked to spend 5 minutes completing 10 multiple-choice questions which had been slotted into brochures given out at the centre. Questionnaires distributed: N = 200 Returned questionnaires: N = 161 Data analysis Percentages of the women who found each topic	Results relevant to protocol Proportions of women who found the the brochure-information valuable in the following ways N (%) Risk factors for osteoporosis: 70 (45) Risks of HRT: 45 (71) Benefits of HRT: 54 (35) Expected tests at menopause: 29 (19) Risk factors for breast cancer: 24 (15) Physical and emotional changes at menopause: 19 (12) Self-management techniques: 28 (18) Risk factors for uterine cancer: 15 (24) Risk factors for heart disease: 10 (6) Definition of menopause: 11 (7) Information about VSM was not seen as important by the women, which the authors noted as a departure from previous interviews. Pre-menopausal women were more likely to prefer information on 'natural' remedies to HRT. Post-menopausal women were more likely to prefer HRT information. Pre-menopausal women were more likely to discuss the risks and benefits of HRT, osteoporosis, BMD and heart disease. In contrast, post-menopausal women seemed more focused on discussing these and non-hormonal treatments. Women felt the information in the brochure would motivate a discussion with a healthcare provider. Nearly 1/3 of post-menopausal women still had questions and concerns related to the risks of HRT.	Comments The brochure was intended to promote the seeking of further information from clinicians rather than be a standalone intervention. The population was women receiving a cancer detection service. Limitations No objective assessment of women's knowledge pre and post intervention. Women's level of knowledge pre-intervention was self-judged subjectively and retrospectively. Informal methodology, e.g. no powering, no comparator, minimal characteristics-list. Strong risk of bias. Quality checklist

Study details	Summary of study	Results	Other
	important were calculated and tabulated.		
Full citation Mingo,C., Herman,C.J., Jasperse,M., Women's stories: Ethnic variations in women's attitudes and experiences of menopause, hysterectomy, and hormone replacement therapy, Journal of Women's Health and Gender-Based Medicine, 9, S27-S38, 2000 Ref Id 304293 Country/ies where the study was carried out USA Study type Qualitative	Aim of the study To increase understanding of women's midlife changes Characteristics N = 165 (49 white, 75 non-white) Mean age Non-Hispanic white (n=29): 49 Hispanic (n=70): 50 Navajo (n=57): 59 Menopause status Pre/peri: 139 Natural: 89 Surgical: 182 Pending surgical: 11 Inclusion criteria Women who self-identified as peri, post or currently menopausal recruited between Jan 1996 and March 1997. Exclusion criteria Intervention None Data collection Bilingual (Spanish, English and Navajo) researchers ran 23 focus single-ethnicity focus groups using open-ended ethnographic techniques. The diversity of cultures meant that structured questions would have been culturally biased. They were asked: "Tell me about your menopause/hysterectomy experience". This was because 'story-telling' was considered the natural way in which women communicate. Data analysis QSR NUD*IST (non-numerical unstructured data indexing searching and theorizing) was used to code, identify and explore relationships and patterns, and compare/contrast	Results relevant to protocol The women felt health professionals (HPs) 'ligitimised' a very limited number of their perimenopausal concerns. Symptoms which women felt were menopausal were disregarded as ageing. Women felt they needed information on more than the 'core' symptoms of menpause (change in menstrual pattern, hot flushes, vaginal dryness, urinary incontinence). They would like HPs to give them information on memory loss, changes in skin, 'feeling blue', tender breasts, metalic taste, hot feet, burning head, mental lapses, formication ('bugs crawling'), chills, shape-changing, weight-gain, moodiness ('hating your husband'), change in libido and muscle pain (including waist). "I want to get the names of all these people who would actually give (HRT) out." Women in some ethic populations (e.g. Mexican) benefited from learning about the menopause in peer groups: "The idea was to develop leaders, so the group is led by women of the area. When we spoke about sexuality, everyone was very quiet, everyone looked around to see who would speak first. What's worked for us is that we tell our story to the rest. Then everyone opens up and builds trust and confidence. Then they realise that (friends) have the same problem, but they never talked about it. The thing is (non white) women are more submissive...we have many taboos. We haven't woken up." Women found it helpful to have a gynaecologist who gave information about coming off HRT. Some did not give information on discontinuing and some did.	Comments Limitations No citation for women-as-story-tellers evidence. Quality checklist NICE Appendix H: Methodology checklist for qualitative studies Is a qualitative approach appropriate? Yes How well was the data collection carried out? Well, though no evidence for elicitation method. Were the methods reliable? Yes Are the data 'rich'? Yes Is the analysis reliable? Yes, though translating from different languages may have affected accuracy. Is the role of the researcher clearly described?
Full citation Murray,E., Davis,H., Tai,S.S., Coulter,A., Gray,A., Haines,A., Randomised controlled trial of an interactive multimedia decision aid on hormone replacement therapy in primary care, BMJ, 323, 490-493,	Aim of the study To determine whether a decision aid on hormone replacement therapy influences decision-making and health outcomes. Outcome measures included decisional conflict scores, menopausal symptoms and perception of who made decisions.	Results relevant to protocol Acceptability of decision aid to women n = 101 (%) Effect on difficulty of decision making: Easier to decide 56 (54) Neither easier nor harder to decide 37 (36) Harder to decide 8 (8)	Comments Funded jointly by BUPA and King's Fund. Limitations Researchers not blinded and randomisation unclear. Quality checklist A. Selection bias (systematic differences

Study details	Summary of study	Results	Other
2001 Ref Id 256774 Country/ies where the study was carried out UK Study type Quantitative RCT (method)	Characteristics Referred by GPs: N = 259 Randomised: N = 205 (n = 102 in each arm) Intervention group; control group Mean age (years) 50.75; 50.11 Ethnicity, white 95 (92); 93 (93) Educated to secondary level 40 (39); 24 (24)4340 Educated beyond secondary level 63 (61); 78 (77) Mean (SD) decisional conflict score: Uncertainty: 3.61 (0.73); 3.69 (0.87) Factors contributing to uncertainty: 2.70 (0.45); 2.65 (0.46) Inclusion criteria Women on lists of GPs in two urban (Oxford and London) areas and one suburban (Harrow) and one semi-rural (Thame and the Chilterns). Peri-/menopausal and needing to make a decision to start, stop or continue using HRT. Good knowledge of English. Exclusion criteria Women with contraindication to hormone replacement therapy or if they had breast or pelvic cancer, severe visual or hearing impairment, or severe learning difficulties or mental illness. Intervention An interactive multimedia programme, with booklet and printed summary. 16 information comprised quantified probabilities of the risks and benefits of hormone replacement therapy taken from systematic reviews and other published data available in 1996 and updated in 1998. Topics discussed were menopausal symptoms, mood changes, skin changes, changes in energy, vaginal dryness, changes in libido, heart disease, osteoporosis, breast cancer, and endometrial cancer.	Effect on understanding of issues around hormone replacement therapy: Understand more 88 (87) Understand same 13 (13) Understand less 0 Decisional conflict scores at three months Mean(SD) and mean difference Uncertainty Intervention group 3.1 (1.0) Control group 3.4 (1.1) MD (95% CI) -0.3 (-0.7 to -0.04) Factors contributing to uncertainty Intervention group 2.4 (0.5) Control group 2.8 (0.6) MD (95% CI) -0.4 (-0.5 to -0.2) Perceived effective decision making Intervention group 2.2 (0.6) Control group 2.5 (0.7) MD (95% CI) -0.3 (-0.5 to -0.2) Total decisional conflict score Intervention group 2.5 (0.5) Control group 2.8 (0.6) MD (95% CI) -0.3 (-0.5 to -0.2)	between the comparison groups): None B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) Uncertain C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) None D. Other bias: Uncertain - Possible bias from part-private funding. Subjective data collection. Non-blinded study.

Study details	Summary of study	Results	Other
	After viewing the programme the patients were given a summary of the information; a copy was also sent to their general practitioners. Data collection Data collected from women at baseline and at 3 months after randomisation, by self-administered questionnaire. Data analysis A retrospective calculation showed that the power to determine the observed difference in decisional conflict score between the two groups at the final assessment was 95% at the 5% significance level. Comparison were made of the change in scores from baseline to final assessment for the MenQol and Spielberger scales between study groups, and comparison of decisional conflict score was made between the two groups at three and nine months. Data was based on intention to treat. Sample powering reported.		
Full citation Roberts,P.J., The menopause and hormone replacement therapy: views of women in general practice receiving hormone replacement therapy, British Journal of General Practice, 41, 421-424, 1991 Ref Id 304622 Country/ies where the study was carried out UK Study type Quali and quanti. (method)	Aim of the study To explore women's expectations of the menopause and their attitudes towards it, and women's sources of information about HRT, their accuracy of knowledge, and their expectations of HRT. Characteristics Questionnaires returned: N = 64 Mean age (range) 50 (34-65) Hysterectomies, n(%) 26 (41) Class (based on the 1981 census) A smaller proportion of women in this study were found to be in social classes 1 and 2 as compared with the north west region (16% versus 24%). 61% of women were in social class 3N and 3M compared with 41% identified in the census in the north west region. Inclusion criteria · Aged 40 - 65	Results relevant to protocol 37% of women wanting information would like to have known the long term effects of HRT, and 26% would have liked information about the optimal duration of therapy. When asked what worries about HRT they had (in an information-receiving context), 2% said Weight gain. No other specific worries were mentioned. The largest proportion of women (61%) sourced information from the Media (TV, magazines, newspapers etc). The authors concluded that women often find this innacurate, and that doctors should be aware of what women are reading. Surgically menopausal women had not received information from their gynaecologists during surgery-contact. This was in spite of 81% of women saying they would like to have received information before the onset of menopause.	Comments Questionnaires were given to 95 women and 64 replies were received giving a response rate of 67%. This authors had a keen consciousness of the influence of class on their population sample and survey-responses. However, this was compromised by their use of a non-standardised social demographic nomenclature with no citations. Limitations This study had good data on different sources of knowledge, but did not stratify the women's knowledge-gained data accordingly, this meant the amount of knowledge gained could not be linked to its source. No analysis of variance. Quality checklist NICE Appendix H: Methodology checklist for qualitative studies Is a qualitative approach appropriate? Yes How well was the data collection carried

Study details	Summary of study	Results	Other
	· Using HRT · Registered with one named GP practice in Wigan Exclusion criteria Not reported. Intervention None Data collection Data was collected over six months in 1990. Demographic and 'views' data were collected by self-administered questionnaires which consisted of open and closed questions. The first set of questions asked for background information. The second set asked about the women's expectations of the menopause, whether she would have liked more information about the menopause, and whether she had received any other advice or treatments before commencing HRT. The third set concentrated on HRT asking the perceived reason for commencing it, expectations, her sources of information and accuracy of knowledge. Data analysis Means, ranges and percentages for characteristics and survey data were calculated and tabulated.		out? Appropriate Were the methods reliable? Yes Are the data 'rich'? No Is the analysis reliable? Unclear Is the role of the researcher clearly described? Unclear
Full citation Rostom,A., O'Connor,A., Tugwell,P., Wells,G., A randomized trial of a computerized versus an audio-booklet decision aid for women considering post-menopausal hormone replacement therapy, Patient Education and Counseling, 46, 67-74, 2002 Ref Id 304651 Country/ies where the study was carried out Canada Study type Quantitative RCT (method)	Aim of the study To compare the efficacy of an interative computerised decision aid (DA) for women considering long-term hormone replacement therapy, to that of a validated audio-booklet version of the same intervention Characteristics Computer DA group (n=25); audio-booklet DA (n=26) Mean±SD or n(%), (95% CI) Age 50.6±7.67, (47.6 to 53.6); 53.8±8.13, (50.0 to 56.9) High school degree 6(24.0), (7.3 to 40.7); 7(26.9), 9.5 to 43.9) University of college degree	Results relevant to protocol Knowledge score Computer DA group (n=25); audio-booklet DA (n=26) Mean±SD (95% CI) Pre-intervention 76.4±14.9 (70.2 to 82.5); 78.7±16.7 (72.0 to 85.4) Post-intervention 93.8±9 (90.1 to 97.5); 87.1±11.8 (82.3 to 91.8) Difference 17.5±13.4 (11.9 to 23.0); 8.4±13.3 (3.0 to 13.8) Opinions on computerised DA Formats participants felt would be best suited to inform women about menopause and HRT: · Booklet with or without audio 43.1% (29.5 to 57.6) · Videotape 25% (14.4 to 39.4)	Comments Sample size estimate based on the realistic expectations score (not extracted for this protocol): 50 patients required to achieve 80% power to detect a difference of 20% in the expectations score between the two groups Limitations Questions asked in the knowledge score are not described. Interventions may be repeated by participants since no restrictions on the number of times they can be completed is described. Follow-up time for post data collection not described. Quality checklist NICE appendix C methodology checklist for RCTs:

Study details	Summary of study	Results	Other
	19(76.0), (56.8 to 91.2); 19(73.1), (56.1 to 90.1) Currently not using HRT 19(76.0), (59.3 to 92.7); 13(50.0), (30.8 to 69.2) Menses 16(64.0), (45.2 to 82.8); 7(26.9), (9.9 to 43.9) Inclusion criteria · Aged 40 to 70 · Peri- and post-menopausal period · Fully fluent in spoken and written English · No evidence of cognitive impairment or overt psychiatric illness Exclusion criteria Only inclusion criteria reported Intervention Randomisation was performed using a table of random numbers and allocation concealment was maintained through the use of consecutively numbered sealed envelopes. Audio DA The HRT audio-booklet DA is a self-administered self-paced, 40 minute audio-tape that guides a women through a 32-page ilustrated booklet. Provides detailed information (including their risk factors and functional impact) about coronary heart disease, osteoporosis, endometrial cancer and breast cancer. The risks and benefits of HRT are presented along with the probabilities of disease both with and without HRT, tailored to the individual's ris of disease and hysterectomy status. Computerised DA Designed to present the validated HRT DA in a format that is intuitive and appealing to patients, while maintaining the exact factual content and visual "feel" of the audio-booklet. Presents a self-test and feedback module after each section for participants to complete. Data collection Participants were recruited from various medical clinics of the Ottawa Hospital. Knowledge was assessed by an 11-item multiple choice questionnaire designed to determine the	· Computer/Internet 23.5% (13.2 to 37.8) · Formats are equally effective 7.8% (2.5 to 19.7)	A. Selection bias (systematic differences between the comparison groups): None B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation): Unclear C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants): None D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified): Unclear - knowledge score is not described in detail

Study details	Summary of study	Results	Other
	patient's understanding of the symptoms and risks of menopause and the risks and benefits of HRT. All post-study questionnaire data were collected within a single contact. Data analysis The pre- and post-changes in the knowledge score between the two intervention groups were analysed with an independent sample t-test with two-sided alpha=0.05. Statistically significant group differences were maintained after re-analysing the data using a non-parametric test, and after adjusting for baseline characteristics.		
Full citation Rothert,M.L., Holmes-Rovner,M., Rovner,D., Kroll,J., Breer,L., Talarczyk,G., Schmitt,N., Padonu,G., Wills,C., An educational intervention as decision support for menopausal women, Research in Nursing and Health, 20, 377-387, 1997 Ref Id 232971 Country/ies where the study was carried out USA Study type Quantitative RCT (method)	Aim of the study To develop and test a decision support intervention to assist women to make and act on informed decisions that are consistent with their values in the area of menopause and HRT Characteristics Age 40 to 45: 37% 46 to 50: 46% White 94% College educated 49% Income $ 15,000 to 49,000: 40% 50,000 to 99,000: 46% Inclusion criteria Not reported. Exclusion criteria Not reported. Intervention Group A - brochure Three-part brochure addressing the physiology of menopause and self-care, the pros and cons of HRT and communication with health care professionals. Group B - lecture Three one and a half hour sessions using a lecture/discussion combined with a question and	Results relevant to protocol Group: A; B; C Mean±SD Decision conflict Time 1: not reported Time 2: (n=89) 3.0±1.00; (n=80) 2.7±0.90; (n=83) 2.6±0.98 Time 3: (n=75) 2.6±0.91; (n=65) 2.6±0.89; (n=63) 2.7±0.97 Time 4: (n=74) 2.5±1.00; (n=65) 2.6±0.78; (n=62) 2.5±0.83 Satisfaction with provider Time 1: (n=89) 3.5±0.68; (n=78) 3.4±0.86; (n=83) 3.4±0.77 Time 2: not reported Time 3: (n=75) 3.6±0.76; (n=65) 3.7±0.80; (n=63) 3.5±0.68 Time 4: (n=74) 3.6±0.76; (n=65) 3.7±0.70; (n=62) 3.6±0.75	Comments A raffle for cash prizes ($25, $50 and $75) was offered to participants. Limitations Demographics not reported for each group. Randomisation not described. Non standardised tests used for measuring outcomes. Decision support 3-item subscale not described in detail. Quality checklist NICE appendix C methodology checklist for RCTs: A. Selection bias (systematic differences between the comparison groups): Unclear B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation): Unclear C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants): 208 out of 238 participants completed the study until time 4 D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified): None

Study details	Summary of study	Results	Other
	answer. Programme content was parallel to the brochure. Group C - additional activities Personalised decision intervention which provided information and experience in an active involvement format. Parallel in programme B to time and parallel to A and B in content. They were assisted to assess their risks and values using a Personal Risk Assessment form and a Problem Significance Assessment form. Asked to aggregate and combine risks and values as a basis of their decision making using a Relevance Chart. Given practical information and strategies for a health care visit. Programme instructors were members of the Decision Making in Menopause Study research team. Two instructors team-taught each intervention session for programmes B and C and attended the data collection sessions for programme A. The clinicians were a physician and three nurses and non clinicians were two psychologists and a health services researcher. Data collection Information/knowledge of menopause was measured using a 24-item multiple choice and true/false scale developed for the study. Content was taken from the interventions and included physiological process of menopause, changes in risk factors postmenopause, common symptoms and their treatments, and pros and cons of HRT. The instrument was reviewed by a panel of experts (nurses and physicians) for content validity and a group of lay women for face validity. Decision conflict was measured using a 3-item subscale of O'Connor's 1995 DCS. Time 1 = preintervention Time 2 = end of intervention / week 3 Time 3 = 6 months Time 4 = 12 months Data analysis Missing data were handled by taking the mean of the nonmissing values if greater than 50% of the items were present.		

Study details	Summary of study	Results	Other
	(The longitudinal data were analysed using multiple regression for repeated measures, to test differences among the three intervention groups. Nominal variables were dummy coded).		
Full citation Theroux,R., Women's decision making during the menopausal transition, Journal of the American Academy of Nurse Practitioners, 22, 612-621, 2010 Ref Id 304938 Country/ies where the study was carried out USA Study type Qualitative	Aim of the study To develop a rich understanding of decision making during or after menopause as constructed by women. Characteristics Seven European women aged 48 to 58. All participants had health insurance and were well educated. Inclusion criteria · Recruited participants via brochures placed in 10 NPs offices · Spoke English · Experiencing changes of menopause · Postmenopausal · Recently made a decision about menopause management and had discussed the decision with an NP Exclusion criteria Not reported Intervention Qualitative interview Data collection The initial interviews were tape recorded and lasted approximately 1 hour using a semi-structured guide with several open ended questions. Data analysis Audio tapes were transcribed verbatim, the transcripts were then compared with the auditotape for accuracy. After each interview, the data was coded line by line using quantitative content analysis (Downe-Wambolt 1992) and constant comparison (Glaser & Strauss 1967). Similar groups were coded into categories. After each interview new codes were compared with previous codes across all categories to explore new and emerging issues with subsequent participants. The initial 25 categories that emerged from the data were subsumed into four major categories: experiencing changes, searching for answers,	Results relevant to protocol Sources of information · Women sourced information from written materials (newspapers, magazines and books) by popular physicians, celebrities and herbalists. · Women who decided for or against HRT received relevant information from the following sources: WHI findings, Current clinical guidelines, and Interactions with a healthcare practitioner. · Women could not make the decision about what information was useful and what was not because they were unable to judge its quality. This was particularly the case with online information where search engines retrieved "millions of hits on menopause". "You need to narrow down your search, but it's difficult when you don't know what you're looking for." For this reason the internet was not a primary resource. · All participants had heard about the findings of the Women's Health Initiative (WHI) through media reports, which highlighted their concerns about HRT safety: "I can remember when the WHI first came out, hearing how women were running from HT. I had the feeling that it was unsafe to go on HT, so I needed to know more about that...I think that fear is a huge thing for women around this whole issue." · All participants reported that the NP's focus on helping them figure out the best option for their situation was "empowering". They valued being treated by the NP as partners in the healthcare process: "It's a matter of having someone listen to you and put all the pieces together. Women need a comfortable place to share experiences."	Comments In this study menopause and HRT information was only part of the issues involved in decision-making, emotions and family played a significant part as well. This study seems to show that American lay-women are familiar with the WHI and use it as a useful resource for HRT information. Limitations Results may not be generalizable from this single NP practice. Quality checklist NICE Appendix H: Methodology checklist for qualitative studies Is a qualitative approach appropriate? Yes How well was the data collection carried out? Appropriate for study Were the methods reliable? Yes Are the data 'rich'? Yes Is the analysis reliable? Yes Is the role of the researcher clearly described? Yes

Study details	Summary of study	Results	Other
	making the decision and womens' needs.	Useful content Women thought the information on the following were important: Lifestyle changes to manage symptoms; Safety of menopausal treatments (especially HRT); Explanation/translation of recent research results about HRT and help with decision-making.	
Full citation Thewes,B., Meiser,B., Rickard,J., Friedlander,M., The fertility- and menopause-related information needs of younger women with a diagnosis of breast cancer: a qualitative study, Psycho-Oncology, 12, 500-511, 2003 Ref Id 304939 Country/ies where the study was carried out Australia Study type Qualitative. Content & sources	Aim of the study Identify degree of satisfaction among younger breast cancer patients with menopause information. Identify what information they seek and their preferred communication strategies. Characteristics N = 36 (invited) N = 24 (66% participation rate) Reasons for not taking part were busyness, lack of interest or pain at addressing fertility issues. Number of women with no children: 14 Inclusion criteria 18-45 years old with fluent English. Early stage breast cancer in past 5 years and pre-menopausal at time of diagnosis. Exclusion criteria Intervention Commenced or completed chemo/radio/hormone therapy for cancer causing early menopause, menopausal symptoms or potential menopause. Data collection Focus groups, or telephone interviews if too ill to attend FG. Data analysis Transcripts were thematically analysed using 'transcendental realism' (Miles and Huberman 1994). This method was considered comprehensive, explicit and protective against threats to validity.	Results relevant to protocol Women without children wanted information on the impact of treatment on fertility. Fertility became a bigger issue for women as over time (a year was mentioned). This was because the cancer took priority until it was abated. Women wanted more menopause information than they were currently getting. The biggest concerns were not having had this information at the right time and receiving conflicting information: "The information didn't come until I was about to start my chemo, or it was scattered." "Nobody handed you anything; you had to go and look for it." Women wanted clarity about their fertility and menopause status following treatment: "There was no clear answer on anything." They wanted to know if tests could be performed to establish these parameters: "Even if there are no answers to my questions, well then I want to read information which says at this stage we don't know x,y, z." Women wanted doctors to take seriously their need for fertility and menopause information. They had experienced 'discord' with doctors over this issue. "Aggressive" and "blase" were adjectives used: "They (doctors) have their priorities in curing you buth they just thought it (fertility/menopause) wasn't that important." Women wanted menopause information prior to treatment. Most women had been given information orally which left them feeling 'bombarded' and 'overwhelmed' when it was immediately after diagnosis. They felt 'something in writing' would have made it easier to digest. Questions which women thought were important on reflection after treatment Will my periods stop? How will that affect my life? How do I know if I'm menopausal or not? What tests diagnose menopause?	Comments Limitations Quality checklist Is a qualitative approach appropriate? Yes How well was the data collection carried out? Quite well Were the methods reliable? Yes Are the data 'rich'? Yes Is the analysis reliable? Yes Is the role of the researcher clearly described? Fairly well

Study details	Summary of study	Results	Other
		How do I manage symptoms? What does 'menopause' mean? How will treatment affect my bone density? What does a hot flush feel like? Can I have children during menopause? What effect does menopause have on my body? Who do I talk to about sexuality issues? Preferred method of information (in order of rank): 1 most preferred, 9 least preferred Information video: 3.61 (2.35) Decision aid: 4.09 (2.27) Talks and information sessions by experts: 4.70 (2.46) Support groups: 5.61 (2.19) Internet: 6.09 (2.09) Question prompt sheet: 6.30 (1.84) Leaflet: 6.35 (2.53) CD-Rom: 6.48 (2.25)	
Full citation Walter,F.M., Britten,N., Patients' understanding of risk: a qualitative study of decision-making about the menopause and hormone replacement therapy in general practice, Family Practice, 19, 579-586, 2002 Ref Id 305047 Country/ies where the study was carried out UK Study type Qualitative	Aim of the study Uses risk discussions about the menopause and HRT to explore women's understanding of risk issues. The aim is to inform our comprehension of the meaning of specific risks to the primary care patient, and thereby to enhance risk communication in the consultation. Characteristics N = 40 Education, n Some secondary education: 10 Completed O levels: 6 Completed A levels: 9 University graduate: 15 Inclusion criteria · Recruited from two Cambridge practices · Aged 50 to 55 · The practice computers randomly selected 30 patients from each HRT usage group (current, never or previous) who were invited to participate in a focus group Exclusion criteria GP excluded all patients with psychological, psychiatric or chronic medical conditions Intervention N/A Data collection Using 6 focus groups including 5 to 8 participants	Results relevant to protocol Regarding risk-education, women... viewed their family history as 'unique and individual'. found it useful to ignore "statistics on other people and just go from my own experience." found it confusing when experts changed their minds about what is good for you. understood information presented in words and numbers (some preferred words, some preferred numbers). saw numbers as being abstract and scientific. Some felt numbers to be 'truthful', and some saw statistics as always changeable. liked words and numbers to be ranked in their order of magnitude. needed context to give meaning and comprehension. interpreted presentation of risk as binary: "We turn it into acceptable or not acceptable really." wanted truth and knowledge rather than opinions (but added that is probably not possible). (some) felt the opinions of others could take their own risk-judgement away*. "In order to get a correct perception, you've got to have both numbers and your verbal interpretation of what those numbers mean." "I think by saying that it's one in a million, you're	Comments Limitations Quality checklist NICE Appendix H: Methodology checklist for qualitative studies Is a qualitative approach appropriate? Yes How well was the data collection carried out? Well - focus group process was well reported. Not all data recorded in the same way though (some women interviewed). Were the methods reliable? Yes Are the data 'rich'? Yes Is the analysis reliable? Yes Is the role of the researcher clearly described? It was not reported how many field-workers facilitated focus groups. If just one, field notes could be biased.

Study details	Summary of study	Results	Other
	(n=36) or semi-structured interviews (n=4) participants could complete at home. A risk game derived from Kitzinger aimed to develop a friendly atmosphere and familiarise participants with some of the key concepts. The game lasted 15 minutes and involved 16 laminated cards, each of which bore a single legend of a phrase or figure for the group to dicsuss. The ensuing discussion lasted up to one hour, the facilitator asked three questions to initiate the discussion, sometimes using probes to elucidate participants' idea, redirect the discussion or summarise: 1) "How do you view your personal risks of general risk factors such as smoking, alcohol, diet, exercise or family history of breast cancer?" 2) "How do you view your personal risks of the disorders that the menopause might bring, or HRT might prevent, such as osteoporosis, cardiovascular disease, Alzheimer's disease, breast cancer or uterine cancer?" 3) "How do you view the risks and benefits of different menopausal options?" Data analysis All patient contacts were audio-taped, professionally transcribed in full, and usbjected to "Framework" analysis (Ritchie 1994). The transcripts were read repeatedly, and an iterative process followed, involving the stages of familiarisation with the data, identification of a thematic framework, and coding using ATLAS Ti software.	able to make up your own mind rather than someone having made it up for you by saying, 'this is a minimal risk.'"..."In other words you feel as if you're trying to be talked into something." "I associate numbers with personal experiences. When I heard '1 in 100' I immediately thought of my twins (1 in 100 chance)." "I think it's increased knowledge and increased awareness that makes you more averse to risk." Women's perceptions of risk was largely informed by experiences of their own families. Personal experience was often given more weight than expert opinion*. Life events (such as bereavement and unemployment) were seen as risk factors.	
Full citation Walter,F.M., Emery,J.D., Rogers,M., Britten,N., Women's views of optimal risk communication and decision making in general practice consultations about the menopause and hormone replacement therapy, Patient Education and Counseling, 53, 121-128, 2004 Ref Id 305048	Aim of the study To gain insight into the range of women's views on risk and decision-making in GP consultations about menopause/HRT. Characteristics 30 women (with a diversity of HRT status) were selected from GP lists. First language (English:non-English): 34:6 Pre O level education: 10 Completed O levels: 6 Completed A-levels: 9	Results relevant to protocol Women found it useful to have an expert to summarise information for them as otherwise it was just a list of 'opinions'. This was useful in making the decision to use HRT or not. They needed something to take away from the surgery as otherwise they would forget the information straight away. Women wanted assurance that information given to them was the "full truth" i.e. "applicable to themselves, unbiased and trustworthy."	Comments This study has common results to other papers re peer-information-sharing and the menopausal years as being socially vulnerable. Limitations No number of study-decliners was reported. Quality checklist NICE Appendix H: Methodology checklist for qualitative studies

Study details	Summary of study	Results	Other
Country/ies where the study was carried out UK Study type Qualitative (content)	Graduate: 15 Inclusion criteria 30 women (with a diversity of HRT status) were selected from 2 Cambridge general practices, and were aged 50 - 55. The practices were in contrasting areas of Cambridge, one of which was under-privileged (Jarman Area Index J1). Exclusion criteria Intervention None Data collection Women were divided into 7 focus groups with a variety of HRT statuses in each group to promote optimal discussion. Individual views were then explored in-depth through interviews. Data analysis Interviews and FGs were transcribed, then codes were used to categorise key issues, concepts and themes. This was an iterative process using Framework Analysis (Ritchie and Spencer 1994).	It was appreciated when GPs presented both sides of 'the story' regarding HRT. Women wanted their risk information to be individualised and personalised as they perceived every woman's body and menopause was unique. Other approaches were seen as 'blunderbuss'. Women who received information about their own bone density or blood tests felt that the information they were given contained more 'truth'. Women felt they did not have enough 'dedicated time' to discuss information with their GPs. As the women were 'not urgent and not ill' they felt their GPs were too busy with ill people to prioritise explaining HRT to them. Women felt the most helpful information came from Menopause Clinics as they gave 'more up-to-date' information. They were seen as more informed with higher expertise than GPs. It was felt this led to more individualised risk information. "A special clinic...whereby you're not mixed in with the general things." Women felt that listening was a big part of information-giving, and wanted information-giving to be twinned with reassurance. Young male doctors were seen as more ignorant and less sympathetic information-givers than female doctors: "'Oh your hormones! It's all in the head." Women wanted a peer-group for women to meet and exchange information on HRT. This was partly due to feeling unsupported and isolated during their menopausal years: "I think a group would be quite a nice way of doing it. Having it set up so people could talk to each other, to get you into the idea of seeing other people's experiences, before you say 'Yes, it's what I'll do."	Is a qualitative approach appropriate? How well was the data collection carried out? Well Were the methods reliable? Yes Are the data 'rich'? Yes Is the analysis reliable? Yes Is the role of the researcher clearly described? Fairly well described.
Full citation Wathen,C.N., Health information seeking in context: how women make decisions regarding hormone replacement therapy, Journal of Health Communication, 11, 477-493, 2006 Ref Id 305060 Country/ies where the study was carried out	Aim of the study To examine women's information behaviour and decision making regarding HRT, and in particular decision to start and stop HRT and use complementary and alternative approaches. Characteristics Characteristics for the interview sample (n=20) Mean age 55.4	Results relevant to protocol The vast majority of women (n=17) (including those "put on" HRT by their physician without specific consultation) felt that their doctor was the most influential source of information when they decided to start HRT. The remaining (n=3) had been convinced of the need to take HRT prior to consulting their physicians sourcing information from formal sources (books, seminars), media and informal sources.	Comments Women received a $40 honorarium for participating. Another sample of participants received a questionnaire, this has not been extracted because it is not relevant to this protocol. Limitations Quality checklist NICE Appendix H: Methodology checklist for qualitative studies Is a qualitative approach appropriate? Yes

Study details	Summary of study	Results	Other
Canada Study type Qualitative (methods)	Education Completed high school: 95% Some college or university: 30% Completed college or university: 20% Caucasian, n 19 Inclusion criteria · Aged 45 to 65 · Self-identified as being peri or postmenopausal, current or former HRT users Exclusion criteria Not reported. Intervention N/A Data collection Interviews averaged 60 minutes in length, and were tape recorded. The qualitative interview guide addressed a number of areas to determine women's experiences with menopause, HRT, and use of CAM therapies to manage menopausal symptoms. Data analysis The data sources for the interview were verbatim transcripts of interview tapes and a synthesis of written notations made during the interview with expanded summary notes made immediately following each interview. A blended inductive/deductive coding scheme was used, consistent with the pre-identified key questions derived from the existing literature and pilot interviews conducted prior to the main study, and with the categories and themes emerging from the data during an initial process of open coding.	Medical sources were the most influential in terms of decision making, women did consult a number of other sources including books, libraries, or local information sessions (n=9), media stores or the Internet (n=8). Informal sources and often the media, were not particularly helpful compared with medical sources and books etc.: "I read things and I get frustrated when I hear things on the YV and then see it in the paper and it's twisted around or you don't get all, you never get all the facts" The internet was seen as untrustworthy, inaccurate and contradictory: "I did a few times go into the Internet but not knowing how reliable the sites were that I was looking at... and there's so much contradiction." Some women found the medical perspective from a doctor troubling because of the many related diseases to consider: "Well, maybe we shouldn't be doing this... the breast cancer problems are minor compared to the other things that might develop if you didnt take it" Women were affected by the WH1 news: "If I stop taking estrogen, because of the possibility after what I saw in the news report on the television last night" but they were also annoyed by the news: "People will quote half of it you know, and the same with television, they only have so much time and you do not have all those factors that have gone into these studies" Women felt they needed to be self-reliant regarding information-sourcing. Women did not view doctors as appropriate sources for information on complementary/alternative therapies, even though such therapies were seen as slightly more useful than HRT. Women were suspicious that information they received was about people who did not have the 'same factors' as themselves.	How well was the data collection carried out? Self-administered questionnaire Were the methods reliable? Yes Are the data 'rich'? Yes Is the analysis reliable? Yes Is the role of the researcher clearly described? No

Study details	Summary of study	Results	Other
		Usefulness % Where women went for information about CAM alternatives to HRT Doctor Very: 38 Somewhat: 43 Not: 17 Other health professional Very: 46 Somewhat: 43 Not: 11 Internet Very: 47.5 Somewhat: 47.5 Not: 5 Magazines and news media Very: 27 Somewhat: 69 Not: 4	

H.3.2 Information needs of women with menopause

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Full citation Kernohan,A.F., Sattar,N., Hilditch,T., Cleland,S.J., Small,M., Lumsden,M.A., Connell,J.M., Petrie,J.R., Effects of low-dose continuous combined hormone replacement therapy on glucose homeostasis and markers of cardiovascular risk in women with type 2 diabetes, Clinical Endocrinology, 66, 27-34, 2007 Ref Id 202962 Country/ies where the study was carried out UK Study type Randomised, double-blind placebo controlled trial	Sample size N=30 randomised (n=15 in HRT group, n=15 in placebo group) N=28 analysed (n=14 in HRT group, n=14 in placebo group Characteristics HRT/placebo Mean age, year (SD) 62.2 (5.8)/62.1 (3.8) Years since menopause, mean year (SD) 13.0 (1.4)/14.0 (4.7) Weight, mean kg (SD) 82.0 (16.4)/80.5 (20.3) BMI, mean kg/m2 (SD) 34.0 (6.3)/33.0 (8.9) Hypertension, % 78.6/78.6 Mean number of antihypertensive drugs 1.6/1.9	Interventions Oral 17β oestradiol (1mg) and norethisterone (0.5mg) Matching placebo tablet	Details Setting Diabetes centres of North Glasgow University Hospitals NHS trust Randomisation method Participants were randomly assigned to HRT or placebo in blocks of six, stratified for presence or absence of hypertension, method not clearly reported Statistical methods Baseline and after treatment data were reported as means and SDs, or median and interquartile range for parameters not exhibiting normal distribution Results after treatment	Results HbA1c Reported as mean percentage (SD) HRT/placebo Baseline: 7.4 (1.1)/ 7.6 (0.9) 3 months treatment (final): 7.4 (1.3)/ 8.1 (1.1) P= 0.11 Fasting glucose Reported as mean mmol (SD) HRT/placebo Baseline: 8.1 (1.9)/8.5 (2.1) 3 months treatment (final): 7.2 (1.9)/ 8.9 (1.6)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes, reported, but method of randomisation not reported A2 - Was there adequate concealment - Unclear, methods of concealment not reported A3 - Were groups comparable at baseline - Yes Level of bias: Moderate B Performance bias B1 - Did groups get same level of care - Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Aim of the study To assess the effects on glucose homeostasis and cardiovascular risk factors of continuous oral 17b oestradiol (1mg) and norethisterone (0.5mg) in postmenopausal women with type 2 diabetes Study dates Not reported Source of funding British Heart Foundation	Inclusion criteria Postmenopausal women, >1 year from last menstrual period Age <70 years and had type 2 diabetes according to national guidelines Women on stable oral anti-diabetic therapy and/or diet for at least 3 months prior to entry and regular medication was not changed during the study Exclusion criteria Poor glycaemic control, (glycated haemoglobin (HbA1c) >10%), severe hypertriglyceridaemia (>70 mmol/l), serum creatinine >120µmol/l, blood pressure >160/110 mmHg, HRT use within 2 years, insulin therapy, or other standard contraindication to HRT		expressed as mean (or median) and as percentage change from baseline. Between group differences assessed by two-sample t test or Mann-Whitney U test P value of <0.05 was considered significant Pearson's correlation coefficients (r) were calculated using Minitab A priori power calculation based on previous studies in subjects with type 2 diabetes estimated that a sample size of n=15 in each group would give 80% power to detect a 10-15% change in EGP, fasting plasma glucose, HbA1c and total cholesterol (α=0.05, two-sided)	P=0.02	B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Unclear, not reported Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Yes D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Unclear, not reported D5 - Were investigators blinded to confounding factors - Unclear, not reported Level of bias: Moderate Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information
Full citation Darko,D.A., Dornhorst,A., Kennedy,G., Mandeno,R.C., Seed,M., Glycaemic control	Sample size N=41 recruited, N=33 completed study Characteristics	Interventions Three cycles were taken continuously for 12 weeks Oral preparation: 28 day	Details Randomisation method At visit one, participants were randomised and	Results HbA1c Reported as mean percentage (SD)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
and plasma lipoproteins in menopausal women with Type 2 diabetes treated with oral and transdermal combined hormone replacement therapy, Diabetes Research and Clinical Practice, 54, 157-164, 2001 Ref Id 203073 Country/ies where the study was carried out UK Study type Randomised open parallel study Aim of the study To compare the effect of a fixed combination of an oestrogen (17b-oestradiol) with cyclical progestogen (norethisterone) on glycaemic control, plasma lipoproteins and haemostatic factors in women with type 2 diabetes Study dates Not reported Source of funding Coronary Thrombosis Trust at Charing Cross Hospital	HRT (oral)/HRT (transdermal)/control BMI, mean kg/m2 (SD) 28.2 (6.8)/33.5 (8.0)/33.5 (9.1) Fasting plasma glucose, mean mmol (SD) 8.2 (1.6)/11.2 (5.5)/8.7 (3.9) HbA1c, mean percentage (SD) 7.4 (1.4)/7.8 (1.7)/7.4 (1.2) Inclusion criteria Postmenopausal women (cessation of menses for >1 year in the presence of climacteric symptoms, or biochemically, follicular stimulating hormone >25IU with serum oestradiol <100pmol-1) with type 2 diabetes (diagnosed after age of 40 years and treated with either diet alone or diet and oral hypoglycaemic agents) recruited from outpatient clinics from hospital or from local GPs Exclusion criteria Women taking insulin or lipid lowering therapy within the last 6 months or HRT within the last 3 months Women consuming >20 units of alcohol a week or had significant medical co-morbidity	cycle of 17β oestradiol 2mg for 16 days followed by norethisterone 1 mg for 12 days Transdermal preparation: patch releasing 17β oestradiol 50µg per 24 hours transdermally for 14 days followed by a second patch releasing both 17β oestradiol 50µg and norethisterone 170µg per 24 hours for 14 days Control group: no treatment	allocated to one of the three study groups, and biochemical, demographic and clinical data was recorded At visit two (at 12 weeks), all measurements were repeated Samples were obtained at start of HRT use and also at the second visit for future analysis Statistical methods All values were expressed as mean (SD) ANOVA was used to analyse paired data and P value of <0.05 as significant	Oral HRT/transdermal HRT/control At 12 weeks: 6.8 (1.2)/ 7.8 (1.8)/ 7.4 (1.6) Control P value at baseline and 12 weeks: not significant Oral HRT P value at baseline and 12 weeks: <0.005 Transdermal HRT P value at baseline and 12 weeks: not significant Fasting plasma glucose Reported as mean mmol/l (SD) Oral HRT/transdermal HRT/control 8.4 (2.4)/ 10.7 (3.0)/ 9.2 (4.2) P value for all treatment groups at baseline and 12 weeks: not significant	trials A Selection bias A1 - Was there appropriate randomisation - Yes, randomisation by drawing of lots into one of three treatment groups A2 - Was there adequate concealment - No. The study was an open parallel study A3 - Were groups comparable at baseline - Unclear, not reported Level of bias: High B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- No. The study was an open trial B3 - Were individuals administering care blinded to treatment allocation- No, the study was an open trial Level of bias: High C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Unclear, not reported Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Yes D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					D4 - Were investigators blinded to intervention - Unclear, not reported D5 - Were investigators blinded to confounding factors - Unclear, not reported Level of bias: High Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information
Full citation Ferrara,A., Karter,A.J., Ackerson,L.M., Liu,J.Y., Selby,J.V., Northern California Kaiser Permanente Diabetes Registry., Hormone replacement therapy is associated with better glycemic control in women with type 2 diabetes: The Northern California Kaiser Permanente Diabetes Registry, Diabetes Care, 24, 1144-1150, 2001 Ref Id 323433 Country/ies where the study was carried out USA Study type Cross sectional study of cohort from the Kaiser Permanente Diabetes Registry Aim of the study To examine whether HbA1c levels varied by current HRT among women with type 2 diabetes Study dates Diabetes registry was started in	Sample size N=15,435 women with T2DM Characteristics Characteristics during 2 year study period HRT/no HRT Mean age, years (SD) 61.2 (7.6)/65.9 (8.8) BMI, mean kg/m2 (SD) 30.7 (6.5)/30.4 (6.8) HbA1c, mean %, SD 8.1 (1.7)/8.4 (2.0) Ethinicity, % Non-Hispanic: 60.9/53.2 African-American: 9.4/15.0 Hispanic: 12.9/12.3 Asian/Pacific Islanders: 9.4/11.5 Other/unknown: 7.4/8.0 Therapy, % Diet: 13.9/12.2 OHA: 51.5/53.4 Insulin: 34.6/34.4 Diabetes duration, % <5 years: 38.0/36.2 5-9 years: 23.9/21.6 ≥10 years: 38.1/42.2 SMBG practice, % Never: 19.9/26.4 <1/week: 18.2/17.1	Interventions Current HRT (oestrogen and/or progestin) No current HRT	Details Setting Kaiser Permanente Medical Care Programme of Northern California, group practice pre-paid health plan Statistical methods Two sample t test was used to compare current HRT and no current HRT use for continuous variables and X2 for categorical variables HbA1c and BMI means were age-adjusted (ANOVA) Generalised estimating equation model was constructed to assess association between HRT and HbA1c level (after taking into account clustering of patients characteristics treated by the same physician and adjusting for age, ethnicity, education, BMI, hypoglycaemic therapy, diabetes duration, SMBG,	Results Age adjusted mean (SE) HbA1c (%) during 2 year study HRT/no HRT 7.9 (0.03)/8.5 (0.02) P=0.0001 Regression coefficient for HRT in predicting HbA1c: HRT use/HbA1c: β coefficient= -0.475 (SE 0.04), P=0.0001	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies 1 Objectives 1.1 Are the objectives of the study clearly stated? Yes 2 Design 2.1 Is the research design clearly specified and appropriate for the research aims? Yes 2.2 Were the subjects recruited in an acceptable way? Yes 2.3 Was the sample representative of a defined population? Yes Risk of bias: Low 3 Measurement and observation 3.1 Is it clear what was measured, how it was measured and what the outcomes were? Yes 3.2 Are the measurements valid? Partly. Duration of HRT use prior to study was not reported. 3.3 Was the setting for data

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
1993, patients included in study from 1995 to 1997 Source of funding American Heart Association and SmithKline Beecham Pharmaceuticals	≥1/week: 61.8/56.5 Smoking,% Current: 9.7/8.9 Former: 36.0/31.6 Never: 54.3/59.5 Exercise, % 52.4/46.9 Inclusion criteria Women aged ≥50 years age who were members of the diabetes registry, Women who filled an HRT prescription, women who were continuously enrolled in the health plan (without gaps), confirmed type 2 diabetes, HbA1c measured at least once Exclusion criteria Women not continuously enrolled in the health plan, women who stated that they did not have diabetes on the survey, women with type 1 diabetes or unclassified for type of diabetes		and exercise Confounders were included in the GEE models if their inclusion resulted in appreciable changes in the HRT coefficient or if the variable was shown by previous scientific publications to be associated with both outcome and exposure All P values were for two-tailed tests with statistical significance defined as P≤0.05		collection justified? Yes 3.4 Were all important outcomes/results considered? Partly. Only HbA1c was considered, not blood glucose levels. Risk of bias: Low 4 Analysis 4.1 Are tables/graphs adequately labelled and understandable? Yes 4.2 Are the authors' choice and use of statistical methods appropriate, if employed? Yes, they want to see the correlation of HbA1c in women currently taking HRT 4.3 Is there an in-depth description of the analysis process? Yes 4.4 Are sufficient data presented to support the findings? Partly. This is a cross-sectional study, but the HbA1c results are reported at an unknown time point during the 2 year study Risk of bias: Low 5 Discussion 5.1 Are the results discussed in relation to existing knowledge on the subject and study objectives? Yes, other studies are also discussed 5.2 Can the results be generalised? Yes Risk of bias: Low Indirectness Does the study match the review protocol in terms of; Population:Yes Outcome: Yes Indirectness: None Other information

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Full citation McKenzie,J., Jaap,A.J., Gallacher,S., Kelly,A., Crawford,L., Greer,I.A., Rumley,A., Petrie,J.R., Lowe,G.D., Paterson,K., Sattar,N., Metabolic, inflammatory and haemostatic effects of a low-dose continuous combined HRT in women with type 2 diabetes: potentially safer with respect to vascular risk?, Clinical Endocrinology, 59, 682-689, 2003 Ref Id 203263 Country/ies where the study was carried out Scotland, UK Study type Double-blind, randomized placebo-controlled trial. Aim of the study To assess the metabolic effects of a continuous combined HRT containing 1 mg oestradiol and 0·5 mg norethisterone or matching placebo Study dates Study only stated women with type 2 diabetes aged under 70 years of age were recruited between December 1998 to September 2000 Source of funding Not reported	Sample size n=50 Active n=25 randomized/22 completed trial/19 demonstrated compliance Placebo n=25 randomized/23 completed trial Characteristics Active/placebo Mean age, year (SD): 60.7 (5.5)/61.3 (4.8) BMI (kg/m2) (SD): 30.5 (6.5)/29.8(5.61) Waist circumference,cm (SD): 93.9 (11.3)/93.7 (13.6) Years postmenopausal (SD): 14.6 (8.5)/14.2(6.3) Inclusion criteria -women with type 2 diabetes aged under 70 years of age -clinically and biochemically postmenopausal, i.e. at least 1 year since last menses and a FSH concentration of greater than 20 IU/l. Menopause could be either natural or surgically induced Exclusion criteria -poor glycaemic control -severe hypertriglyceridaemia (> 10 mmol/ l) -moderate to severe hypertension (systolic > 160 mmHg, diastolic > 110 mmHg) -renal impairement (serum creatinine greater than twice the upper limit of normal range) -liver disease (serum transaminases and bilirubin greater than twice the upper limit of normal range) -established cardiovascular, cerebrovascular, or peripheral vascular disease -subjects with either a personal history of – or first-degree relative with – breast cancer	Interventions Active medication (1 mg oestradiol plus 0·5 mg norethisterone) or identical placebo daily for 6 months	Details Setting General diabetic clinics in Glasgow Hospitals Randomisation method In blocks of four using computer-generated number Statistical methods Mean differences in changes from baseline between the two treatment groups were compared using the unpaired t-test; 95% confidence interval for change in active group data relative to change in control group data are presented. Adjustment for baseline concentrations was made by linear regression. Baseline data are presented as mean and SD or median and interquartile range (IQR) for parameters exhibiting skewed distribution.	Results Glycaemic control -HbA1c (%) Reported as mean (SD) Active/Placebo Baseline: 10.2 (1.8) / 10.2 (1.3) Mean change: -0.37/0.22 Mean difference for change active relative to change placebo (95%CI) / p: -0.59 (-1.45 to 0.27)/ 0.17 -Blood glucose Reported as Glycaemia glucose (mmol/l), mean (SD) Active/Placebo Baseline: 12.4 (4.2) / 11.3 (3.2) Mean change: -1.74/0.42 Mean difference for change active relative to change placebo (95%CI) / p: -2.16 (-4.06 to -0.28)/ 0.026 Health related quality of life Not reported Mortality Not reported Adverse events (complications resulting from diabetes) Not reported	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Unclear, methods of concealment not reported A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Unclear, methods of blinding not reported B3 - Were individuals administering care blinded to treatment allocation- Unclear, methods of blinding not reported Level of bias: High C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Unclear, not reported Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Yes D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					method used to assess outcome - Unclear, not reported D4 - Were investigators blinded to intervention - Unclear, not reported D5 - Were investigators blinded to confounding factors - Unclear, not reported Level of bias: High Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information Study does not report the sample size analysed for each treatment outcome.
Full citation Perera,M., Sattar,N., Petrie,J.R., Hillier,C., Small,M., Connell,J.M.C., Lowe,G.D.O., Lumsden,M.A., The effects of transdermal estradiol in combination with oral norethisterone on lipoproteins, coagulation, and endothelial markers in postmenopausal women with type 2 diabetes: A randomized, placebo-controlled study, Journal of Clinical Endocrinology and Metabolism, 86, 1140-1143, 2001 Ref Id 311478 Country/ies where the study was carried out Scotland, UK Study type Randomised placebo-controlled trial Aim of the study	Sample size Continuous combined HRT [transdermal oestradiol (80-µg patches) in combination with oral norethisterone (1 mg daily; n = 22] or identical placebos (n = 21) Characteristics HRT/Placebo Mean age, year (SD): 61.2 (3.7)/62.8(4.9) Duration of diabetes, median year (ranges): 2 (1-20)/4 (1-14) Mean BMI (kg/m2), (SD): 31 (7.8)/31.6(4.3) Inclusion criteria Not reported Exclusion criteria Not reported	Interventions Continuous transdermal oestradiol (80-µg patches) in combination with oral norethisterone (1 mg daily) or identical placebos for 6 months	Details Setting Diabetes Centers in Glasgow Randomisation method Not reported Statistical methods The adequacy of the randomization process was checked by comparing the baseline values in the two groups (unpaired t test or Mann-Whitney U test as appropriate). Differences in changes from baseline between the two treatment groups were compared using t tests if the changes were normally distributed. Baseline values in parameters of interest and in age, smoking status, and	Results Glycaemic control -HbA1c (%): Reported as mean (SD) HRT/placebo Baseline: 6.6(1.3)/6.4(1.3) 6 months (final): 6.6(1.2)/6.8(1.6) p value change (differences in changes from baseline between groups): 0.35 -Blood glucose: Reported as mean fasting blood glucose (mmol/L) (SD) HRT/placebo Baseline: 8.1 (1.7)/8.5(2.7)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Unclear, not reported A2 - Was there adequate concealment - Unclear, not reported A3 - Were groups comparable at baseline - Yes Level of bias: High B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Unclear, not reported B3 - Were individuals administering care blinded to

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
To assess the effect of transdermal oestradiol (80-µg patches) in combination with continuous oral norethisterone (1 mg daily) on conventional anthropometric parameters, lipoprotein concentrations, coagulation (fibrinogen, factor VII, and fibrin D dimers), and endothelial factors [tissue plasminogen activator (t-PA), and von Willebrand factor (vWF)] in postmenopausal women with type 2 diabetes. Study dates Not reported Source of funding Not reported			diabetes duration were adjusted for using linear regression. Correlation analysis was performed using the Spearman rank correlation. Data are presented as the mean and SD for normally distributed data and as the median and range for data with a nonparametric distribution.	6 months (final): 8.6(2.5)/8.6(2.6) p value change (differences in changes from baseline between groups): 0.57 Health related quality of life Not reported Mortality Not reported Adverse effects (complications resulting from diabetes) Not reported	treatment allocation- Unclear, not reported Level of bias: High C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear, not reported C3 - Were groups comparable for missing data - Unclear, not reported Level of bias: High D Detection bias D1 - Was follow-up appropriate length - Yes D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Unclear, not reported D4 - Were investigators blinded to intervention - Unclear, not reported D5 - Were investigators blinded to confounding factors - Unclear, not reported Level of bias: High Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information
Full citation Sutherland, W. H., Manning, P. J., de Jong, S. A., Allum, A. R., Jones, S. D., Williams, S. M., Hormone-replacement therapy increases serum paraoxonase	Sample size N=47 HRT group=28 Placebo group=19 Characteristics Age (years, mean, SD):	Interventions HRT: conjugated equine oestrogen (Premarin 0.625mg) and medroxyprogesterone acetate (Provera 2.5 mg)	Details Treatment: Written informed consent obtained from participants HRT was titrated upward over a 4-week period to	Results Glycaemic control -HbA1c (%) Reported as mean (SD) HRT/Placebo	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
arylesterase activity in diabetic postmenopausal women, Metabolism: Clinical & ExperimentalMetabolism, 50, 319-24 Ref Id 325988 Country/ies where the study was carried out New Zealand Study type Randomised placebo-controlled, cross-over study Aim of the study To test the effect of HRT on plasma concentrations of lipids, lipoproteins, and apolipoproteins in postmenopausal diabetic women Study dates Recruitment of participants ended in 1996 Source of funding Health Research Council of New Zealand	64±8 BMI (kg/mg2, mean, SD): 32.3±5.7 HbA1c (%, mean, SD): 7.5±1.9 Fasting glucose (mmol, mean, SD): 10.2±3.9 Inclusion criteria Postmenopausal women with type 2 diabetes (postmenopausal defined as absence of menstrual periods for more than 2 years Cardiovascular disease was present in 14% of the diabetic women Exclusion criteria Poorly controlled diabetes (glycosylated [HbA1c] >10%) Concomitant significant medical disorder Contraindications to HRT (history of breast or endometrial cancer) Undiagnosed vaginal bleeding Uncontrolled hypertension Severe liver dysfunction or they met the current national criteria for lipid-lowering therapy with statins	combined in a single capsule Placebo (single capsule identical to HRT)	minimise acute side effects. At end of 4 weeks women were taking either HRT or placebo treatment (1 capsule/daily)Patients were seen at 3 month intervals to check for adverse effects (reaction to medication, suffered serious concurrent illness contraindicating HRT or receiving lipid-lowering therapy), compliance (capsule counting: defined as tablet count >80%), record body weight, measure blood lipids Laboratory methods: Plasma gluocose was measured enzymatically by automated methods using a commercial kit HbA1c was measured using a commercial kit Statistics: Values expressed as means±SD Multivariate linear regression analysis with final (6 month) and baseline values to test for differences between HRT and placebo treatment Paired t test was used to estimate treatment effect if significant difference was observed between HRT and placebo treatments Two-tailed tests of significance were used, and a P value of <0.05 was considered statistically significant	Baseline: 7.3 (1.6) / 7.8 (2.3) 6 months: 7.9 (1.6) / 8.5 (2.1) -Blood glucose Reported as glucose (mmol/l), mean (SD) HRT/Placebo Baseline: 9.97 (3.30) / 10.66 (4.69) 6 months: 8.37 (2.1) / 10.38 (4.1)	A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Unclear, methods of blinding not reported B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Moderate C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - No. 13 participants (40%) in the placebo group dropped out compared with 1 in the HRT group C3 - Were groups comparable for missing data - Unclear, not reported Level of bias: High D Detection bias D1 - Was follow-up appropriate length - Yes D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Unclear, not reported D5 - Were investigators blinded

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					to confounding factors - Unclear, not reported Level of bias: High Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no indirectness Other information

H.4 Management short-term symptoms

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Full citation Al-Akoum,M., Maunsell,E., Verreault,R., Provencher,L., Otis,H., Dodin,S., Effects of Hypericum perforatum (St. John's wort) on hot flashes and quality of life in perimenopausal women: a randomized pilot trial, Menopause, 16, 307-314, 2009 Ref Id 226059 Country/ies where the study was carried out	Sample size St John's wort n=22 randomised, 20 completed the study Placebo n=25 randomised, 20 completed the study Characteristics St John's wort / Placebo Mean age, year (SD): 53.4 (4.8) / 54.0 (5.8) Breast cancer survivor, n (%): 11 (55) / 15 (68.2) -With tamoxifen, n (%): 6 (30) / 9 (40.9) Prior hysterectomy, n (%): 5 (25) / 8	Interventions 900 mg of St. John's wort (300mg TID) or placebo (T1D) for 3 months	Power calculation Not reported Intention to treat Yes Details Setting Centre Menopause Quebec in Canada Randomisation method Computer generated by the Clinical Unit of the Hopital St. Francois d'Assise Research Centre Statistical	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported -Depression Not reported -Cognitive function Not reported -Sleep disturbance Reported as mean (SD) Sleep Problems Scale St John's wort/Placebo Baseline: 1.7 (0.8)/1.7 (0.6) Month 3: 1.2 (0.8)/1.6 (0.6)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: Low	Main outcome classification -Sleep disturbance-Sleep Problems Scale -Quality of life-psychological-Menopause-Specific Quality of Life Psychosocial domain -Quality of life-musculoskeletal-Menopause-Specific Quality of Life Physical domain Main interventions classification Herbal preparations - St. John's wort Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Canada Study type Double-blind, randomized clinical trial Aim of the study To obtain preliminary evidence of the effect of Hypericum perforatum extract (St. John's wort extract) compared with placebo on symptoms and quality of life of symptomatic perimenopausal women Study dates Between October 2003 to September 2005 Source of funding Quebec Breast Cancer Foundation	(36.4) Inclusion criteria -3 or more hot flashes a day -FSH concentrations of 40 mIU/mL or more -At least 6 months of amenorrhea in the year preceding study entry -Normal mammogram in preceding 2 years Exclusion criteria -Used St John's wort or antidepressants within the preceding 6 months -Ingested phytoestrogens from soybean or soy product food supplements on a regular basis -Had received HT in the preceding 3 months -Had a history of recurrent or metastatic cancer -Had uncontrolled hyperthyroidism or hypothryoidism or a severe psychiatric disorder -Used or planned to use other agents for treating hot flashes or used other oral herbal therapies or medications that could cause potential interactions with St. John's wort		methods Difference between the placebo and St. John's wort groups at 3 months was calculated using Student's t test. Intragroup and intergroup differences were computed as d, the standardised mean difference, or effect size (ES).	Difference: 0.5 (0.8)/0.07 (0.58) Between-group effect size:-0.67 p-value for within groups, baseline vs month 3: 0.009/ 0.589 p-value for between groups, St John's wort vs placebo:0.05 -Quality of life Reported as mean (SD) Menopause-Specific Quality of Life Psychosocial domain St John's wort/Placebo Baseline: 2.9 (1.4)/ 3.2 (1.4) Month 3: 2.2 (1.1) / 3.1 (1.2) Difference: -0.8 (1.4)/-0.1 (1.0) Between-group effect size:-0.75 p-value for within groups, baseline vs month 3: 0.02/ 0.69 p-value for between groups, St John's wort vs placebo: 0.01 Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not reported -Quality of life Reported as mean (SD) Menopause-Specific Quality of Life Physical domain St John's wort/Placebo Baseline: 3.5 (1.5) / 3.7 (1.3) Month 3: 2.8 (1.1) / 3.6 (1.4) Difference: -0.7 (0.9)/ -0.1 (1.0) Between-group effect size:-0.57 p-value for within groups, baseline vs month 3: 0.003/0.56 p-value for between groups, St John's wort vs placebo: 0.06 Safety outcomes -Discontinuation Not reported	B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
				-Major adverse events Not reported -Minor adverse events Not reported	blinded to intervention - Unclear D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information	
Full citation Brunner,R.L., Aragaki,A., Barnabei,V., Cochrane,B.B., Gass,M., Hendrix,S., Lane,D., Ockene,J., Woods,N.F., Yasmeen,S., Stefanick,M., Menopausal symptom experience before and after stopping estrogen therapy in the Women's Health Initiative randomized, placebo-controlled trial, Menopause, 17, 946-954, 2010 Ref Id 226240 Country/ies where the study was carried out	Sample size 10,739 women randomised. 5310 received conjugated equine oestrogens. 5429 assigned to placebo. Characteristics Baseline characteristics not reported in this study as they have been described in previous studies. The study reported: -Women aged between 50 to 79 years -Participants were an average of nearly 20 years post hysterectomy at baseline -One-third of trial participants reported the presence of one	Interventions 0.625 mg/day conjugated equine oestrogens (CEE-Premarin) or a matching placebo.	Power calculation Not reported Intention to treat Yes Details Setting Women's Health Initiative CEE trial at 40 clinical centers in the United States Randomisation method Not reported Statistical methods Intention-to-treat analyses of 10,739 postmenopausal women focused on incidence of symptoms at year 1. Comparisons of	Results Frequency of hot flushes (including night sweats) Not reported Frequency of sexual intercourse Not reported Psychological symptoms Not reported Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Reported as risk ratio (95% CI) of incident symptoms at year 1 of taking CEE compared with placebo by prevalence of symptoms at baseline Joint pain not present at baseline: 0.91 (0.81-1.01) Joint pain present at baseline: 0.98 (0.93-1.03) P-value for test of main effect=0.04 -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not reported	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Unclear A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Unclear Level of bias: Unclear B Performance bias B1 - Did groups	Main outcome classification Musculoskeletal: Symptom relief Main interventions classification Oestrogen (oral)-CEE Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
United States Study type Randomised, placebo-controlled Women's Health Initiative (WHI) oestrogen plus progestin trial Aim of the study To assess vasomotor and other menopausal symptoms before, one year later, again at trial closure and after stopping estrogens or placebo. The role of baseline symptoms and age was examined as was the frequency and determinants of hormone use and symptom management strategies after discontinuing conjugated equine estrogens or placebo. Study dates Exact study dates not reported. Randomisation conducted between 1993 and 1998. Analyses were conducted before and 1 year after randomisation. Source of funding National Heart, Lung, and Blood Institute, National Institutes of Health, Department of	or more moderate-to-severe menopause-associated symptoms at baseline Inclusion criteria Post-menopausal women, aged 50 to 79 years at initial screening, were eligible if they had a prior hysterectomy and met specific health criteria (not reported in the study). Exclusion criteria Not reported		active to placebo, stratified by presence or absence of baseline symptoms, are presented as relative risks (RRs) and 95% confidence intervals (CIs) along with p-values for the main effect of CEE on symptom incidence and p-values for the interaction between CEE and the presence or absence of baseline symptoms (p-int). Estimated RR (95%CIs) and p-values were obtained from generalized linear models. Further analyses were conducted of these relative risks as modified by age. Follow-up Outcomes were recorded before and 1 year after randomisation to CEE or placebo	-Quality of life Not reported Safety outcomes -Discontinuation Not reported -Major adverse events Not reported -Minor adverse events Not reported	get same level of care - Unclear B2 - Were participants blinded to treatment allocation- Unclear B3 - Were individuals administering care blinded to treatment allocation- Unclear Level of bias: Unclear C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Unclear D4 - Were investigators blinded to	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Health and Human Services					intervention - Unclear D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: some Other information Rated down for indirectness as one-third of participants reported at least one moderate-to-severe symptom at baseline.	
Full citation Carranza-Lira,S., Cortes-Fuentes,E., Modification of vasomotor symptoms after various treatment modalities in the postmenopause, International Journal of Gynaecology and Obstetrics, 73, 169-171, 2001 Ref Id 226284 Country/ies where the study was carried out	Sample size Conjugated equine oestrogens (CEE) n=15 Clonidine n=15 Placebo n=15 Characteristics Not reported other than they were postmenopausal for greater than or equal to 1-5 years with vasomotor symptoms and insomnia Inclusion criteria -Postmenopausal women (greater	Interventions Interventions relevant to protocol are reported here: 0.625 mg/day CEE for hysterctomised patients. Those with contraindication for CEE were randomly distributed to: 0.10mg/day clonidine A placebo/day	Power calculation Not reported Intention to treat Not reported Details Setting Mexico Randomisation method Not reported for CEE, the study only reported random distribution of subjects to other treatment groups.	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported -Depression Not reported -Cognitive function Not reported -Sleep disturbance Reported as insomnia presence (% yes)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - No A2 - Was there adequate concealment - No A3 - Were groups comparable at	Main outcome classification Sleep disturbance-insomnia (presence) Main interventions classification Oestrogen (oral) Clonidine Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Mexico Study type Study does not state the study type, however it seems like a semi-RCT (randomisation for all treatment groups except oestrogen group) Aim of the study To evaluate the efficiency of various treatments in postmenopausal women with vasomotor symptoms Study dates Not reported Source of funding Not reported	than or equal to 1-5 years) -FSH and oestradiol levels were in the postmenopausal range Exclusion criteria Not reported		Statistical methods Mann-Whiteney U-test and Wilcoxon test were used	Oestrogen/ clonidine/ placebo Baseline: 80/87/73.3 3rd month: 8*/22**/46.7 * p <0.01, ** p <0.05 -Quality of life Not reported Musculoskeletal symptoms Not reported Safety outcomes -Discontinuation Not reported -Major adverse events Not reported -Minor adverse events Not reported	baseline - Unclear Level of bias: High B Performance bias B1 - Did groups get same level of care - Unclear B2 - Were participants blinded to treatment allocation- No B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Unclear C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome -	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					Unclear D4 - Were investigators blinded to intervention - No D5 - Were investigators blinded to confounding factors - Unclear Level of bias: High Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information This is a low quality study that does not state randomisation methods	
Full citation Demetrio,F.N., Renno,J.,Jr., Gianfaldoni,A., Goncalves,M., Halbe,H.W., Filho,A.H., Gorenstein,C., Effect of estrogen replacement therapy on symptoms of depression and anxiety in non-depressive menopausal women: a randomized double-blind, controlled study, Archives of Women's	Sample size N = 76 Characteristics Age (mean ± SD) CEE (N = 30): 49.9 ± 3.25 Placebo (N = 36): 50.83 ± 2.71 Type of menopause Natural (non-bilateral oophorectomy): CEE: N = 24 (80%) Placebo: N = 26 (72.2%) Surgical (bilateral oophorectomy)	Interventions - CEE (0.625 mg/da) - Placebo Both orally, for 6 sycles of 28 days each.	Power calculation 30 participants per group for 80% power, significance = 5% Intention to treat Not reported. Details Setting Participants attending the Division of Endocrine Gynaecology of the Department of Gynaecology, Clinical Hospital, School of Medicine, San	Results State-Trait Anxiety Inventory Significant differences seen in active group (CEE) compared to baseline. CEE Baseline mean score: 37.5 Endpoint: 32.2, p = 0.01 Placebo Baseline: 39.1 Endpoint: 34.2, p = 0.001 *No differnces were seen between groups.	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Unclear - not reported A2 - Was there adequate concealment - Unclear	Main outcome classification Psychological Main interventions classification HRT

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Mental Health, 14, 479-486, 2011 Ref Id 226407 Country/ies where the study was carried out Brazil Study type Double-blind, randomised, placebo-controlled study Aim of the study To investigate the efficacy of ERT for improving mood and anxiety of non-depressive postmenopausal women. Study dates Not reported. Source of funding Not reported.	CEE: N = 6 (20%) Placebo: N = 10 (27%) Inclusion criteria - Hysterectomy for non-malignant causes, with or without unilateral or bilateral oophorectomy - In menopause for at least 2 years but no more than 10. - Only mild to moderate hot flashes and < 5 severe hot flashes over a 2 week period. - Aged 45 - 56 Exclusion criteria - Major or minor depression (according to SADS-L) - Severe hot flashes on more than 5 days over a 2 week period - Procoagulant disorders - History of CVd and other comorbidities - Smoking		Paulo Randomisation method Not reported. Statistical methods For comparing proportions between groups: the chi squared test and Fisher's exact test (small expected number of events) . For variables with normal distribution: ANOVA.		A3 - Were groups comparable at baseline - Yes Level of bias: high B Performance bias B1 - Did groups get same level of care - Unclear B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Yes Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Unclear D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Medium Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Derman,R.J., Dawood,M.Y., Stone,S., Quality of life during sequential hormone replacement therapy -- a placebo-controlled study, International Journal of Fertility and Menopausal Studies, 40, 73-78, 1995 Ref Id 226410 Country/ies where the study was carried out Not reported. Study type Placebo-controlled, parallel group, double-blind RCT. Aim of the study To confirm the efficacy of	Sample size N = 82 Sequential estrogen / progestin (Trisequens) = 40 Placebo = 42 Characteristics Average age = 50 yrs Average weight = 68 kg Inclusion criteria - Women aged 40 - 60 yrs who complained of menopausal symptoms Exclusion criteria - Women who had estrogen therapy within last 3 months, steroid therapy within last 3 months, history of major diseases	Interventions Sequential 17 beta - estradiol and norethindrone acetate (Trisequens)	Power calculation Not reported. Intention to treat Yes Details Setting 3 centers Randomisation method Computer generated randomisation schedule. Statistical method Qualitative variables - Mantel-Haenszel test in contingency table Continuous variables - ANOVA	Results Greene Psychological Index Pretreatment / baseline Mean (SD) Trisequens (N = 39) = 14.2 (9.52) Placebo (N = 39) = 17.6 (11.87) Posttreatment mean (SD) Trisequens (N = 39) = 8.0 (9.04) Placebo (N = 39) = 16.7 (9.43) Beck Depression Inventory Pretreatment / baseline Mean (SD) Trisequens (N = 39) = 5.1 (4.66) Placebo (N = 39) = 6.5 (6.54) Posttreatment mean (SD) Trisequens (N = 39) = 3.1 (3.79) Placebo (N = 39) = 6.4 (5.90) Greene Somatic Index Pretreatment mean (SD) Trisequens (N = 39) = 4.1 (3.50) Placebo (N = 39) = 5.9 (3.85) Posttreatment mean (SD) Trisequens (N = 39) = 3.3 (3.47) Placebo (N = 39) = 5.4 (3.60)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: medium B Performance bias B1 - Did groups	Main outcome classification Psychological Muscoloskeletal Main interventions classification HRT

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Trisequens in comparison with placebo in the relief of vasomotor symptoms, to assess alterations in quality of life by patient questionnaires, to evaluate cycle control, and to compare dropout rates between groups. Study dates Not reported. Source of funding Novo Pharmaceuticals Inc., Princeton, NJ					get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Yes Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Unclear D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: Unclear Intervention: yes Outcomes: yes Indirectness: unclear	
Full citation Elfituri,A., Sherif,F., Elmahaishi,M., Chrystyn,H., Two hormone replacement therapy (HRT) regimens for middle-eastern postmenopausal women, Maturitas, 52, 52-59, 2005 Ref Id 226445 Country/ies where the study was carried out Libya Study type 12-month randomised prospective study Aim of the study To evaluate the 12-month effects of two different HRT regimens on postmenopausal symptoms of Middle-Eastern women. Study dates Not reported Source of funding Not reported	Sample size Tibolone n=50 17 beta-Oestradiol/dydrogesterone n=50 Characteristics Tibolone /17 beta-Oestradiol/dydrogesterone Mean age (years), SD: 43.8±7.6 / 44.8±8.7 Inclusion criteria -Healthy non-hysterectomised Libyan women naturally or surgically menopausal, with menopausal symptoms - In naturally menopausal women, it was at least 12 months since the last menstrual period (LMP) and at least 3 months after the bilateral oophorectomy in	Interventions 2.5 mg Livial® (2.5 mg tibolone) oral tablets 2/10 mg Femoston® (2 mg 17-beta oestradiol sequentially combined with 10 mg dydrogesterone) oral tablets	Power calculation Not reported Intention to treat Not reported Details Setting Faculty of Medicine, University of Alfateh, Tripoli, Libya Randomisation method Not reported Statistical methods The statistical significant differences between the groups were performed using one-way unrelated analysis of variance (ANOVA), with Bonferroni correction to highlight the differences between the	Results Frequency of hot flushes (including night sweats) Not reported Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported -Depression Reported as mean scores (SD) of depression using scores similar to those of 'The Green Climacteric Scale'. Severity of the symptoms was classified as none, mild, moderate and severe, and scored as 0, 1, 2, 3, respectively. Tibolone group / oestradiol/dydrogesterone group Month 0: 0.46 (.76) / 0.36 (0.56) Month 12: 0 (0)* / 0 (0)* *P < 0.001: reference is made to month 0. -Cognitive function Reported as mean scores (SD) of loss of memory using scores similar to those of 'The Green Climacteric Scale'. Severity of the symptoms was classified as none, mild, moderate and severe, and scored as 0, 1, 2, 3, respectively. Tibolone group / oestradiol/dydrogesterone group Month 0: 0.24 (.48) / 0.34 (0.68) Month 12: 0 (0)* / 0 (0)* *P < 0.001: reference is made to month 0.	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Unclear A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: High B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- No B3 - Were	Main outcome classification -Depression -Cognitive function -Sleep disturbance -Symptom relief (joint pain and muscular pain [with and without] stiffness) *reported using scales similar to Greene -Discontinuation -Minor adverse event bleeding Main interventions classification Tibolone Combined oestrogen with progesterone (17-beta oestradiol sequentially combined with 10 mg dydrogesterone)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
	surgically menopausal women Exclusion criteria -Pregnancy -Significant past or present medical illness with the exception of mild controlled diabetes, stabilised hypothyrodism, mild controlled hypertension and mild stabilised obstructive pulmonary disease -Concomitant administration of a medication that is likely to interfere with the treatment use; the contraindications to oestrogen or progestogen therapy; the known hypersensitivity, intolerance or severe side effects to prior therapy -Presence of abnormal vaginal bleeding of unknown aetiology during the last 6 months		individual pairs. Contingency tables were presented and $\chi 2$ test was used for the comparisons of those with and without symptoms withinthe groups between each visit.	-Sleep disturbance Reported as mean scores (SD) of insomnia using scores similar to those of 'The Green Climacteric Scale'. Severity of the symptoms was classified as none, mild, moderate and severe, and scored as 0, 1, 2, 3, respectively. Tibolone group / oestradiol/dydrogesterone group Month 0: 0.82 (.52) / 0.92 (0.66) Month 12: 0 (0)* / 0 (0)* *P < 0.001: reference is made to month 0. -Quality of life Not reported Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Reported as mean scores (SD) of joint pain using scores similar to those of 'The Green Climacteric Scale'. Severity of the symptoms was classified as none, mild, moderate and severe, and scored as 0, 1, 2, 3, respectively. Tibolone group / oestradiol/dydrogesterone group Month 0: 1.04 (1.03) / 0.70 (0.79) Month 12: 0 (0)* / 0 (0)* *P < 0.001: reference is made to month 0. -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not reported -Quality of life Not reported Safety outcomes -Discontinuation Withdrew due to adverse events by third month Tibolone group n=1 Oestradiol/dydrogesterone group n=1 -Major adverse events	individuals administering care blinded to treatment allocation- No Level of bias: High C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: High D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Unclear D4 - Were investigators blinded to intervention - No D5 - Were investigators blinded to confounding factors - Unclear Level of bias: High Indirectness Does the study match the review	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
				Not reported -Minor adverse events Bleeding Tibolone n=3 Oestradiol/dydrogesterone group n=4	protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: some, study used Middle Eastern women only	
Full citation Evans,M., Elliott,J.G., Sharma,P., Berman,R., Guthrie,N., The effect of synthetic genistein on menopause symptom management in healthy postmenopausal women: a multi-center, randomized, placebo-controlled study, Maturitas, 68, 189-196, 2011 Ref Id 226467 Country/ies where the study was carried out Canada Study type Randomized double-blind, placebo-controlled study Aim of the study To evaluate the efficacy of synthetic genistein for reducing the frequency and severity of hot flushes Study dates Not reported	Sample size Genistein n=42 assigned, n=40 intention-to-treat Placebo n=42 assigned and intention-to-treat Characteristics Genistein/placebo Age mean ± SD: 53.39 ± 5.05 / 53.50 ± 4.44 Natural menopause (%): 63.4/69.1 Surgical menopause (%): 36.6/31 Inclusion criteria Subjects had to have a minimum of 40 hot flushes per week, be between the ages of 40 and 65 and be in a physiological state of natural or surgical menopause Exclusion criteria -Clinical or laboratory abnormalities -Had used conventional hormone therapy or selective estrogen receptor modulators within 4 weeks of study start -Had known allergy	Interventions Placebo or a single 30 mg dose of synthetic genistein daily for 12 weeks	Power calculation Assuming a standard deviation of 50% and allowing for a 20% rate of withdrawal, 42 subjects per group were required to detect a clinically important difference of 35% at the 5% level of significance (two-sided) with 80% power. Intention to treat Yes Details Setting 5 study sites in southwestern Ontario, Canada Randomisation method Subjects were randomly assigned to one of two treatment groups in blocks of six and a treatment code was randomly allocated in the order in which a subject was enrolled. Each	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Reported as mean Greene Climacteric Scale-anxiety (SD) Genistein/Placebo/p-value Week 0 (baseline): 4.79 (3.13) / 5.76 (3.84) Week 4: 3.64 (3.38) / 4.56 (3.34) / 0.581 Week 8: 3.43 (2.63) / 4.54 (3.03) / 0.250 Week 12: 3.00 (2.25) / 4.32 (3.34) / 0.142 -Depression Reported as mean Greene Climacteric Scale-depression (SD) Genistein/Placebo/p-value Week 0 (baseline): 4.36 (3.19) / 4.83 (3.74) Week 4: 2.95 (3.35) / 4.19 (3.56) / 0.070 Week 8: 2.94 (2.13) / 3.62 (3.25) / 0.543 Week 12: 2.48 (2.06) / 3.35 (3.55) / 0.389 -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Mean Greene Climacteric Scale-psychological subscale (SD) reported but study did not report it as psychological quality of life Genistein/Placebo/p-value Week 0 (baseline): 9.08 (5.90) / 10.45 (7.46) Week 4: 6.59 (6.50) / 8.61 (6.63) / 0.248 Week 8: 6.38 (4.20) / 8.15 (6.06) / 0.484	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes	Main outcome classification Anxiety Depression Psychological quality of life Physical activity All measured by Greene Climacteric Scale Discontinuation Minor adverse events Main interventions classification Phytoestrogens-genistein Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Source of funding DSM Nutritional Products, Inc., the manufacturer of the genistein tested, fully funded this study but played no role in its execution and analysis of findings.	or hypersensitivity to soy, peanuts, purified isoflavones, genistein, lactose and/or cow's milk -Had consumed soy products within 4 weeks prior to the screening visit -Reported unpredictable vaginal bleeding (i.e., leiomyoma or endometrial polyps), uterine fibroids or endometriosis that required treatment; untreated polycystic ovary syndrome (PCOS) -History of abnormal pap smear -Use of gonadotropin agonists within 24 weeks -Glucocorticoids or chronic high dose (>7.5 mg/day) prednisone or equivalent for the past 12 weeks		treatment code was associated with either the genistein or placebo. Statistical methods The statistical analysis was a modified intent-to-treat analysis in which all subjects receiving the test product for a period of four weeks were included in the efficacy analysis, and all subjects taking at least one dose of the test product were included in an analysis of safety. A per protocol analysis of the results was also conducted for both efficacy and safety endpoints and included all subjects completing 12 weeks of treatment. Where subjects terminated early, data from the withdrawal date were used as study completion data. The distribution of baseline characteristics in the two groups	Week 12: 5.48 (3.91) / 7.65 (6.68) / 0.182 Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Reported as mean Greene Climacteric Scale-somatic (SD) Genistein/Placebo/p-value Week 0 (baseline): 3.36 (2.69) / 4.17 (3.19) Week 4: 2.28 (1.97) / 3.26 (3.16) / 0.254 Week 8: 2.51 (2.23) / 2.71 (2.74) / 0.617 Week 12: 2.30 (1.95) / 2.73 (3.00) / 0.608 -Quality of life Not reported Safety outcomes -Discontinuation Genistein: n=2 due to adverse events Placebo: n=1 due to adverse event -Major adverse events Not reported -Minor adverse events Bleeding: genistein n=4 / placebo n=1 Headache: genistein n=1 / placebo n=1 Increasingly emotional: placebo n=1	Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
			was compared descriptively. Treatment group comparisons for primary and secondary outcomes, the percentage change in the number of hot flushes, the change in the duration and severity of hot flushes, the change in Greene Climacteric Scale scores, endometrial thickness, serum FSH and 17β-estradiol concentrations were analysed using analysis of covariance (ANCOVA). Descriptive statistics present the mean values and associated standard deviations for all available data by treatment groups. Calculations of within group changes were made using data for subjects having both baseline and applicable endpoint values. A t-test was used to determine probability values		Outcomes: yes Indirectness: no	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
			for within group differences.			
Full citation Geller,S.E., Shulman,L.P., van Breemen,R.B., Banuvar,S., Zhou,Y., Epstein,G., Hedayat,S., Nikolic,D., Krause,E.C., Piersen,C.E., Bolton,J.L., Pauli,G.F., Farnsworth,N.R., Safety and efficacy of black cohosh and red clover for the management of vasomotor symptoms: a randomized controlled trial, Menopause, 16, 1156-1166, 2009 Ref Id 226551 Country/ies where the study was carried out USA Study type Randomised control trial Aim of the study To evaluate the safety and efficacy of black cohosh and red clover compared with placebo for the relief of menopausal vasomotor symptoms. Study dates February 2003 to December 2007 Source of funding	Sample size Placebo arm: n = 22 randomised Placebo arm: n = 21 included in analysis Ostrogens + progestin arm (CEE/MPA): n = 23 randomised and included in analysis Black cohosh arm (BC): n = 22 randomised BC: n = 21 included in analysis Red clover arm (RC): n = 22 randomised and included in analysis Characteristics Placebo / CEE,MPA / Black cohosh / Red clover / P-value Mean age, year (SD): 52 (4.2) / 53.3 (4.0) / 54.4 (3.9) / 52.4 (4.6) / 0.24 Mean BMI (SD): 30.1 (4.9) / 26 (3.9) / 28.3 (4.5) / 30.5 (4.3) / 28.7 (4.7) / 0.004 Race n (%) p-value = 0.005, statistically significant difference between groups African American: 16 (72.7) / 7 (30.4) / 8 (38.1) / 13 (59.1) White: 5 (22.7) / 16 (69.6) / 13 (61.9) / 5 (22.7) Hispanic: 1 (4.6)/ 0 / 0 / 3 (13.6)	Interventions Capsules were taken twice daily for 12 months -0.625 mg conjugated equine oestrogens plus 2.5 mg medroxyprogesterone acetate (CEE/MPA) -Black cohosh -Red clover -Placebo	Power calculation The sample size calculation for the primary outcome (reduction in vasomotor symptoms) was based on prior research and powered with the following assumptions. Botanical treatments would reduce vasomotor symptoms by approximately 60%, for example, from 35 hot flashes to 13 hot flashes per week, with a probability of at least 0.80, SD of 10, and an anticipated placebo effect of 35%. The null hypothesis to be tested was the equality of reduction in the number of hot flashes between placebo and the botanical groups. This was a two-sided test with an alpha error rate of 5% and a 5% dropout rate anticipated during the 12-month intervention period. The	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Reported as Greene Anxiety Score difference in mean reduction (SD) Placebo vs black cohosh/ p-value: 3 month: -0.20 (0.74) / 0.78 12 month: -0.47 (0.81) / 0.56 Placebo vs red clover/ p-value: 3 month: 1.14 (0.73) / 0.12 12 month: 1.64 (0.80) / 0.04 (statistically significant difference) Placebo vs CEE/MPA/ p-value: 3 month: 1.01 (0.72) / 0.16 12 month: 0.83 (0.79) / 0.29 -Depression Not reported -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Not reported Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not reported -Quality of life	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups	Main outcome classification Anxiety-Greene anxiety scale Discontinuation Minor adverse events-headache Main interventions classification -Oestrogen combined with progesterone (CEE/MPA) -Herbal preparation (Black cohosh) -Phytoestrogens (Red clover) -Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Not stated	Pacific islander: 0 / 0 / 0 / 1 (4.6) Last period in years (SD): 2.8 (2.9) / 3.6 (2.9) / 3.4 (2.6) / 4.1 (2.8) / 0.52 Inclusion criteria -Perimenopausal or postmenopausal -Intact uterus ->34 vasomotor symptoms (hot flashes and night sweats) per week -Amenorhea >6 months and <10 years -FSH, >40 mIU/mL -HT not contraindicated -Able to give informed consent Exclusion criteria -Fewer than 35 vasomotor symptoms (HF+NS) per week -Last menstrual period > 10-y duration -Positive pregnancy test or breastfeeding -Obesity, BMI >38kg/m2 -Previous history of endometrial hyperplasia/neoplasia -Previous history of cancers of the breast or reproductive tract -History of presence of myocardial infarction or stroke -History of severe recurrent	optimal sample size (n) for the primary outcome was calculated to be 22 per arm, for a total number of 88 women across all four arms of the study. This study was powered only to compare each botanical to placebo. Intention to treat Yes Details Setting University of Illinois at Chicago/National Institutes of Health Center for Botanical Dietary Supplements Research in outpatient care facilities at the University of Illinois Medical Center and at the Northwestern University Feinberg School of Medicine Randomisation method A random, computer-generated code assigned two women in each cluster to each of four treatment arms. There were 11 clusters with	Not reported Safety outcomes -Discontinuation CEE/MPA: n=2 due to adverse events -Major adverse events Not reported -Minor adverse events CEE/MPA: n=1 for headache	comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
	depression, or severe psychiatric disturbance -History or presence of cerebrovascular accident, severe varicose veins, sickle cell anemia History of alcohol or drug abuse -Abnormal vaginal bleeding of undetermined cause -Untreated or uncontrolled hypertension defined as systolic blood pressure > 165 mm Hg or diastolic blood pressure > 95 mm Hg -Concurrent administration of medication containing estrogen, progestin, SERM, St. John's wort, bisphosphonates, or dietary phytoestrogens -History of migraine associated with hormone use -History or presence of deep vein thrombosis, thrombophlebitis or thromboembolic disorder -Current participation in any other clinical trial within 30 days of enrollment ->5 alcoholic drinks per week		eight women in each cluster. Thus, from the first set of eight participants, two each were assigned to black cohosh, red clover, placebo, and the CEE/MPA arms. This same process was repeated for all women enrolled in the study. The randomisation procedure was the same at both sites. Statistical methods For each treatment baseline data was subtracted from the data at months 3, 6, 9 and 12 to assess symptom reduction. One way analysis of variance was used to analyse all data. Fisher's Least Significant Difference Procedure was used for pairwise comparison of the treatment groups. Missing measurements were imputed using the last-observation-carried-forward			

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
	-Smoker -Diabetes -Abnormal transvaginal ultrasound defined as >7-mm thickness -Abnormal endometrial biopsy or mammogram -Vegans (vegetarians who tend to consume greater than average doses of phytoestrogens)		method. All data was summarised as mean (SD), and p values of less than 0.05 were considered statistically significant.			
Full citation Hachul,H., Bittencourt,L.R., Andersen,M.L., Haidar,M.A., Baracat,E.C., Tufik,S., Effects of hormone therapy with estrogen and/or progesterone on sleep pattern in postmenopausal women, International Journal of Gynaecology and Obstetrics, 103, 207-212, 2008 Ref Id 226616 Country/ies where the study was carried out Brazil Study type Single-center, prospective, placebo-controlled study Aim of the study To investigate the effect of estrogen and progesterone on	Sample size N = 33 CEE: 14 Placebo: 19 Characteristics Age (yrs) CEE: 57.8 (5.1) Placebo: 54.5 (3.4) Postmenopause (yrs) CEE: 10.5 (8.6) Placebo: 9.0 (11.5) Inclusion criteria - Postmenopausal women - Aged 50 - 65 - Mean BMI less than 30 - No previous exposure to exogenous hormones Exclusion criteria - Endometrial thickness greater than 5 mm on ultrasound / positive result to progesterone test	Interventions 0.625 mg / day CEE orally	Power calculation Not reported. Intention to treat Not reported. Details Setting Not reported Randomisation No details provided. Reported as: "randomisation was stratified to obtain an approximately equal number" in each group. Statistical analysis Comparisons between groups - Chi squared test or Fisher test when presumptions of Chi squared test not met. Comparisons of quantitive variables (values at each testing) -	Results Epworth Sleepiness Scale Difficulty falling asleep CEE Baseline: 42.8 Follow-up: 40.0 Placebo: Baseline: 52.6 Follow-up: 37.5 - Pairwise comparisone between 2 groups at baseline: NS - Pairwise comparisone between 2 groups at follow-up: NS Sleep Apnea CEE Baseline: 14.2 Follow-up: 0 * * statistical difference with baseline and between 2 groups Placebo: Baseline: 26.3 Follow-up: 25.0 ** - Pairwise comparisone between 2 groups at baseline: NS - Pairwise comparisone between 2 groups at follow-up: p = 0.01	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Unclear A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: High B Performance bias B1 - Did groups get same level of care - Unclear B2 - Were participants blinded to treatment	Main outcome classification Psychological Main interventions classification HRT

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
sleep in postmenopausal women. Study dates Not reported Source of funding AFIP, CNPq, FAPESP, CEPID			Friedman K test.	Anxiety Reported as prevalence CEE Baseline: 64.2 Follow-up: 60.0 Placebo: Baseline: 52.6 Follow-up: 68.7 - Pairwise comparisone between 2 groups at baseline: NS - Pairwise comparisone between 2 groups at follow-up: NS Depression Reported as prevalence CEE Baseline: 28.5 Follow-up: 22.2 Placebo: Baseline: 31.5 Follow-up: 37.5 - Pairwise comparisone between 2 groups at baseline: NS - Pairwise comparisone between 2 groups at follow-up: NS	allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Yes Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Unclear D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: low Indirectness Does the study match the review	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Haines,C., Yu,S.L., Hiemeyer,F., Schaefers,M., Micro-dose transdermal estradiol for relief of hot flushes in postmenopausal Asian women: a randomized controlled trial, Climacteric, 12, 419-426, 2009 Ref Id 226623 Country/ies where the study was carried out Thailand, the Philippines, Singapore, Hong Kong, Malaysia Study type Multicenter, double-blind, randomized, placebo-controlled study Aim of the study To compare the effect of micro-dose transdermal estradiol and placebo on the incidence and severity of menopausal symptoms and well-being in postmenopausal Asian women with vasomotor symptoms	Sample size 165 subjects randomised to estradiol 0.014 mg/day (E2) or placebo. 80 per group were included in the analysis. By study completion, 77 in E2 and 74 in placebo groups. Characteristics Age at baseline, mean (SD), years Estradiol: 52.6 (3.99) Placebo: 52.2 (4.73) Time since last menstruation, mean (SD), months Estradiol: 56 (60.3) Placebo: 65.3 (61.3) Hysterectomy, n (%) Estradiol: 27 (33.8) Placebo: 33 (41.3) Bilateral oophorectomy, n (%) Estradiol: 19 (23.8) Placebo: 22 (27.5) Inclusion criteria -Women aged between 40 and 65 years	Interventions Transdermal patch delivering micro-dose E2 (0.014mg/day) or placebo for 12 weeks (one patch/week)	Power calculation Not reported Intention to treat Not reported Details Setting Not reported Sample size calculation Not reported Randomisation method Done by a centrally provided computer-generated list Allocation concealment and blinding Not reported. The study was double-blinded. Statistical methods Relative change in frequency of hot flushes from baseline to week 12 was compared between treatment groups using a two-sided Wilcoxin rank-sum (Mann-Whitney) test. Full analysis set	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms Not reported Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not reported -Quality of life Physical MenQoL subscore reported in absolute changes (SD). Placebo group improved more than the E2 group. Placebo group: -0.9 (1.04) E2 group: -0.6 (1.03) Safety outcomes -Discontinuation E2: adverse event n=1, withdrawal of consent n=1 Placebo: withdrawal of consent n=2 -Major adverse events Not reported -Minor adverse events Only minor adverse events of interest that arise in the study are reported Bleeding n (%)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias	Main outcome classification Hot flushes Musculoskeletal quality of life Discontinuation Minor adverse events-bleeding Main interventions classification Oestrogen (patch) and placebo (patch)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Study dates Between June 2005 and November 2006 Source of funding Bayer Schering Pharma AG	-Undergone natural menopause (≥12 months' amenorrhea or 6 months' amenorrhea with serum follicle stimulating hormone > 40 mIU/ml) or bilateral oophorectomy (≥6 weeks postsurgery) -At least 24 hot flushes (of any severity) within a 7-day screening period Exclusion criteria -Recently used oestrogen-containing products -Abnormal cervical smear test -Endometrial thickness of ≥5.0 mm -Any condition that could interfere with study medication or intepretation of results -Concomitant use of inducers or inhibitors of CYP3A4 or drugs effective in treating hot flushes -Received anticoagulant treatment for the past 6 months -Known severe dyslipoproteinemia		with the last observation carried forward was used to analyze hot flushes frequency, and full analysis set used for quality of life. Follow-up 12 weeks	Estradiol: 3 (3.8) Placebo: 1 (1.3)	C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: some	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					Other information Indirect to the UK population as Asian women were used in the study.	
Full citation Kalay,A.E., Demir,B., Haberal,A., Kalay,M., Kandemir,O., Efficacy of citalopram on climacteric symptoms, Menopause, 14, 223-229, 2007 Ref Id 226744 Country/ies where the study was carried out Turkey Study type Single-blind randomised control study, with particpants blinded to which medication they were taking Aim of the study To evaluate the efficacy of citalopram for climacteric symptoms and to assess the combined effect of citalopram and hormone therapy (HT) on climacteric symptoms in women inadequately responsive to HT alone Study dates Not reported Source of funding	Sample size Citalopram n=25 Placebo n=25 Characteristics Citalopram / Placebo Mean age, year (SD): 53.5 (5.3) / 51.7 (4.6) Surgical menopause n (%): 6 (24) / 6 (24) Natural menopause n (%): 19 (76) / 19 (76) Inclusion criteria Natural or surgical menopause More than seven to eight hot flashes per day Normal thyroid function Exclusion criteria Psychotic disease Undergoing psychiatric therapy Taking herbal products, dopaminergic or antidopaminergic drugs, or narcotic analgesics	Interventions The initial dose of citalopram was 10 mg/day. After 1 week, the dose was increased to 20 mg/day. By 4th week, the citalopram dose was increased to 40 mg/day in cases where sufficient improvement was not observed. Insufficient improvement was defined as unchanged score for vasomotor symptoms (the scores remained at the level of moderate-severe). One placebo tablet per day was given. After starting the medication, follow-up visits took place during the fourth and eighth weeks of treatment.	Power calculation Twenty-five study group participants would allow greater than 87% power to detect a significant difference on the vasomotor score. Intention to treat Not reported Details Setting Ankara Etlik Maternity and Women's Health Teaching Research Hospital, Turkey Randomisation method Block randomization was done with a computer-generated program Statistical methods One-way analysis of variance was used to compare differences between the groups at baseline with normally distributed variables. The Kruskal-Wallis test	Results Frequency of hot flushes (including night sweats) Not reported Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported -Depression Not reported -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Reported as change from baseline levels of Menopause-Specific Quality of Life Questionnaire scales for psychosocial score, median (minimum-maximum) Citalopram / Placebo -1.9 (-3.2 to 0) / 0 (-1.2 to 0) Psychosocial complaints significantly decreased in all groups (P = 0.01) Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not reported -Quality of life Reported as change from baseline levels of	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- No Level of bias: Low C Attrition bias	Main outcome classification Quality of life-psychological (MENQOL) Quality of life-musculoskeletal (MENQOL) Main interventions classification SSRI-Citalopram Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Not reported			was used for variables with skewed distribution. Frequency differences between the groups were analyzed using a [chi]2 test. To compare differences between time points within each group, the Wilcoxon signed rank test was used. To compare differences between groups throughout the study, repeated-measures analysis of variance was used	Menopause-Specific Quality of Life Questionnaire scales for physical score, median (minimum-maximum) Citalopram / Placebo -1.0 (-3.0 to 0) / 0 (-2.0 to 0) Physical well-being significantly improved in citalopram group (P=0.001) Safety outcomes -Discontinuation Not reported -Major adverse events Not reported -Minor adverse events Not reported	C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - No D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: some, population	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					was Turkish women	
Full citation Lin,S.Q., Sun,L.Z., Lin,J.F., Yang,X., Zhang,L.J., Qiao,J., Wang,Z.H., Xu,Y.X., Xiong,Z.A., Zhou,Y.Z., Wang,M.L., Zhu,J., Chen,S.R., Su,H., Yang,C.S., Wang,S.H., Zhang,Y.Z., Dong,X.J., Estradiol 1 mg and drospirenone 2 mg as hormone replacement therapy in postmenopausal Chinese women, Climacteric, 14, 472-481, 2011 Ref Id 226855 Country/ies where the study was carried out China Study type Double-blind, multicenter randomised study Aim of the study To compare the efficacy, safety and tolerability of 2 mg drospirenone/1 mg oestradiol (DRSP/E2) versus placebo in Chinese postmenopausal women with moderate to severe vasomotor symptoms (VMS). Study dates	Sample size DRSP/E2 n=183 Placebo n=61 Characteristics DRSP/E2 / Placebo Mean age, year (SD): 52.0 (3.81) / 51.9 (3.56) Inclusion criteria -24 or more moderate to severe hot flushes over 7 consecutive days during the 3-week screening period -Intact uterus with endometrial thickness < 5 mm by transvaginal ultrasonography or normal endometrial biopsy if endometrial thickness was ≥ 5 mm -Last mentrual bleed ≥ 1 year before, or bilateral oophorectomy ≥ 6 weeks before, or last natural menstrual bleed ≥ 6 months (but <1 year) previously, with serum follicle stimulating hormone ≥ 40 mIU/ml -Negative urinary pregnancy test -Negative bilateral mammography result Exclusion criteria -History of	Interventions 2 mg drospirenone/1 mg estradiol (DRSP/E2) versus placebo taken daily orally for four 28-day cycles (16 weeks)	Power calculation Based on the results of the European Angeliq Study, a sample size of 36 patients per group was calculated to be required to obtain 90% power for the primary efficacy parameter Intention to treat Not reported Details Setting Multicentre study in 9 centres in China--study does not report types of centres Randomisation method Centralized block randomisation for patient allocation at a ratio of 3:1 to DRSP/E2 or placebo groups, respectively Statistical methods Descriptive statistics (means with SD) and post-hoc statistical tests	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported -Depression Reported as percentage of depression incidences DRSP/ E2 Baseline: 42.1% / 49.2% After treatment at 16 week: 4% / 12.5% Reported as percent reduction in depression incidences from baseline to end of 16 week treatment -DRSP/E2: 38.1% -Placebo: 36.7% Group differences did not reach statistical significance -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Not reported Musculoskeletal symptoms Not reported Safety outcomes -Discontinuation Discontinuation due to adverse events -DRSP/E2 n=7 -Placebo n=5 -Major adverse events Not reported	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups	Main outcome classification Depression-depression incidences Discontinuation Minor adverse events-headache, bleeding Main interventions classification Oestrogen combined with progesterone (oral) Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Between May 2006 to October 2007 Source of funding Bayer Schering Pharma AG	cardiovascular disease -Uncontrolled thyroid disorders -Clinical depression -Malignant or premalignant disease -Abnormal gynecologic findings -Hepatic disease -Adrenal insufficiency or renal failure -Abnormal glucose tolerance and severe or congenital hypertriglyceridemia -Abnormal baseline laboratory findings -History of alcohol/drug abuse or current smoking -Hormonal therapy during the 4 weeks preceding enrolment -Concurrent therapy with prescription medicines -Use of herbal/other medicines for climacteric disorders -Known hypersensitivity to the study medication or its excipients			-Minor adverse events Bleeding reported as vaginal hemorrhage n (%) DRSP/E2 / Placebo: 2 (1.1) / 0 Headache n (%) DRSP/E2 / Placebo: 5 (2.7%) / 2 (3.3%)	comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Unclear D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: some, this study used Chinese women	
Full citation Nielsen,T.F.,	Sample size N = 335:	Interventions Pulsed estrogen	Power calculation Not reported	Results QoL scores from WHQ	Limitations NICE guidelines	Main outcome classification

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Ravn,P., Pitkin,J., Christiansen,C., Pulsed estrogen therapy improves postmenopausal quality of life: a 2-year placebo-controlled study, Maturitas, 53, 184-190, 2006 Ref Id 227060 Country/ies where the study was carried out Denmark Study type Double-blind, randomised, controlled 2 year study Aim of the study To investigate the effect of pulsed estrogen therapy S21400 on different quality of life (QoL) dimensions in early postmenopausal women Study dates Not reported Source of funding Not reported	Intranasal 17B estradiol: 150 ug/day: N = 114 300 ug/day: N = 103 Placebo: N = 118 Characteristics Age Placebo (N = 118): 52.8 ± 2.0 150 ug (N = 114): 52.6 ± 1.6 300 ug (N = 103): 52.8 ± 1.8 Hysterectomy (%) Placebo: 7.8 150 ug: 4.7 300 ug: 4.7 Inclusion criteria - 40 - 65 yrs old - Menopause defined as amenorrhea for more than 12 months or > 6 months with comitant serum level of estradiol < 0.16 nmol/L + FSH > 42 IU/L - All women who had undergone systerectomy had menopause confirmed by determination of serum estradiol and FSH at least 2 months prior to study entry. - Surgical menopause, if performed at least 6 weeks before study entry - Osteopenic (BMD	therapy S21400 (intranasal 17B estradial): 150 ug/day and 300 ug/day or placebo - Women with intact uterus additionally received oral micronised progesterone 200 mg/day, 14 days out of 28	Intention to treat Yes Details Setting Two Danish centers. Randomisation method Not reported Statistical methods Between group differences in mean change scores were evaluated with a non-parametric covariance analysis.	Anxiety/depressed mood Placebo Scores at baseline (±SD): 81.0 ± 14.3 Mean changes in scores (±SD): -1.6 ± 10.8 150 ug/d Scores at baseline (±SD): 81.9 ± 13.8 Mean changes in scores (±SD): -0.5 ± 12.6 Estimated difference (95% CI): 1.3 (-1.7, 4.2) - not significant 300 ug/day Scores at baseline (±SD): 81.7 ± 17.4 Mean changes in scores (±SD): 1.9 ± 11.8 Estimated difference (95% CI): 3.7 (0.9, 6.5) - not significant Somatic symptoms Placebo Scores at baseline (±SD): 69.8 ± 18.9 Mean changes in scores (±SD): -1.9 ± 14.8 150 ug/d Scores at baseline (±SD): 70.0 ± 16.3 Mean changes in scores (±SD): 0.8 ± 14.3 Estimated difference (95% CI): 12.9 (-0.6, 6.4) - not significant 300 ug/day Scores at baseline (±SD): 71.0 ± 17.9 Mean changes in scores (±SD): 2.0 ± 12.1 Estimated difference (95% CI): 4.2 (0.9, 7.6) - significant: p-value = 0.012 Sleep problems Placebo Scores at baseline (±SD): 61.3 ± 25.8 Mean changes in scores (±SD): -1.9 ± 18.9 150 ug/d Scores at baseline (±SD): 56.1 ± 25.6 Mean changes in scores (±SD): 8.1 ± 21.2 Estimated difference (95% CI): 8.2 (3.5, 12.9) - sig: <0.001 300 ug/day	manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Not reported A2 - Was there adequate concealment - Not reporte A3 - Were groups comparable at baseline - Unclear - Placebo had greater % of ERT compared to groups Level of bias: high B Performance bias B1 - Did groups get same level of care - Unclear B2 - Were participants blinded to treatment allocation- yes B3 - Were individuals administering care blinded to treatment allocation- yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups	Psychological Muscoloskeletal Main interventions classification HRT

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
	T score < - 1) and no complaint of severe climacteric symptoms Exclusion criteria - None stated			Scores at baseline (±SD): 60.7 ± 25.8 Mean changes in scores (±SD): 8.2 ± 17.7 Estimated difference (95% CI): 9.9 (5.5, 14.4) - sig: <0.001	comparable for dropout - Yes C3 - Were groups comparable for missing data - Unclear Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Unclear D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear level of bias: medium Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information - Danish, white women - Women who complained of severe climecteric	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					changes excluded	
Full citation Nir,Y., Huang,M.I., Schnyer,R., Chen,B., Manber,R., Acupuncture for postmenopausal hot flashes, Maturitas, 56, 383-395, 2007 Ref Id 227067 Country/ies where the study was carried out USA Study type Randomised, placebo-controlled pilot study Aim of the study To determine whether individually tailored acupuncture is an effective treatment option for reducing postmenopausal hot flashes and improving quality of life Study dates Not reported Source of funding Not reported	Sample size Active acupuncture n=12 Placebo acupuncture n=17 Characteristics Active acupuncture/placebo acupuncture / p-value if statistically significant Mean age, years (SD): 56.92 (1.73)/ 53.71 (4.24) / p=0.02 Mean age (years, SD) at menopause: 50.18 (2.96) / 48.57 (6.77) History of hormone therapy: 83% / 76% Inclusion criteria -Aged 45-65 -Had not experienced a menstrual period for at least 6 months or were at least 6 weeks post-bilateral oophorectomy -Baseline oestradiol concentration of less than 50 pg/mL and a normal TSH level -Average of at least 7 moderate to severe hot flashes (including night sweats) per 24 hours or an average of at least 70 hot flashes per week during the screening phase Exclusion criteria	Interventions 7 weeks (nine treatment sessions, twice weekly during the first two weeks and once weekly for the remaining five weeks) of either active acupuncture or placebo acupuncture (placebo needles that did not penetrate the skin at sham acupuncture points)	Power calculation Not reported Intention to treat Yes Details Setting Community clinics in the San Francisco Bay Area Randomisation method Separate randomisation table for each acupuncturist was created by generating a random string of permutations of two elements (blocked randomisation) Statistical methods Test for group differences in baseline characteristics included chi-square and t-tests. Differential impacts of both treatments on MSQL subscales were tested with a series of four repeated measures of analyses of variance.	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported -Depression Not reported -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Reported as mean (SD) menopausal specific quality of life-psychological Active acupuncture / placebo acupuncture Baseline: 2.85 (1.41)/ 2.92 (1.20) After the last treatment: 2.20 (0.73) / 2.82 (1.66) No significant reduction in MSQL psychological subscale Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not reported -Quality of life Reported as mean (SD) menopausal specific quality of life-physical Active acupuncture / placebo acupuncture Baseline: 3.49 (0.91)/ 3.31 (1.31) After the last treatment: 2.94 (0.73) / 2.89 (0.99)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes, however, participants in the active group were significantly older than those in the placebo group (p=0.01) Level of bias: Moderate B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Unclear B3 - Were individuals administering care blinded to treatment	Main outcome classification Psychological quality of life Musculoskeletal quality of life Discontinuation Minor adverse events-bleeding Main interventions classification Acupuncture Sham acupuncture

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
	-Endocrine disorders -Known or suspected oestrogen-dependent neoplasia -Known psychiatric disorders -Abnormal results on a laboratory TSH test -Baseline oestrogen level higher than 50 pg/mL -Any treatment for hot flashes, including black cohosh, phytoestrogens, or acupuncture during the 6 weeks before the study -Any unstable medical conditions -Use of any medication known to affect vasomotor symptoms -Having received acupuncture within the past year			No significant reduction in MSQL physical subscale Safety outcomes -Discontinuation Active acupuncture: n= 2 (1 due to concurrent unstable medical condition and 1 due to dissatisfaction with treatment) Placebo acupuncture: n=4 (2 due to concurrent unstable medical condition and 2 due to dissatisfaction with treatment) -Major adverse events Not reported -Minor adverse events Bleeding/bruising during treatment Active acupuncture n=8 Placebo n=1	allocation- No Level of bias: Unclear C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Unclear D4 - Were investigators blinded to intervention - No D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Odmark,I.S., Backstrom,T., Jonsson,B., Bixo,M., Well-being at onset of hormone replacement therapy: comparison between two continuous combined regimens, Climacteric, 7, 92-102, 2004 Ref Id 227091 Country/ies where the study was carried out Sweden Study type Randomised, double-blind, 1 month trial Aim of the study To compare the effect on well-being of two continuous combined HRT in women starting treatment and women switching from mainly sequential HRT Study dates Not reported. Source of funding Wyeth-Ayerst Pharmaceutical, Swedish Council of Research and a grant from the EU Regional Fund.	Sample size N = 246 - CE/MPA: N = 123 - E2/NETA: N = 123 Characteristics Age (yrs) CE/MPA = 55.7 ± 0.27 E2/NETA = 56.0 ± 0.29 Time to menopause (yrs) CE/MPA = 5.6 ± 0.35 E2/NETA = 5.4 ± 0.27 Inclusion criteria - Healthywomen with an intact uterus, had climacteric symptoms or ongoing HRT - Aged 52 or over Exclusion criteria - Contraindications - Use of steriod hormones	Interventions - CE/MPA 0.625 mg/5 mg - E2/NETA 2 mg/1 mg	Power calculation Not reported. Intention to treat Yes Details Setting 14 gyneacological centers in Sweden Randomisation method List in blocks of four was computer generated by statistician. Statistical methods - Differences in baseline characteristics between groups: Mann-Whitney independent sample test - Changes within a group: Wilcoxon test	Results Cyclicity Diagnoser (CD) scale Depression CE/MPA Baseline: 2.0 ± 0.18 Endpoint: 1.8 ± 0.17 E2/NETA: Baseline: 1.9 ± 0.18 Endpoint: 2.0 ± 0.22 - Changes within CE/MPA group: p-value = not significant - Changes within E2/NETA group: p-value = not significant Insomnia CE/MPA Baseline: 2.4 ± 0.21 Endpoint: 2.0 ± 0.20 E2/NETA: Baseline: 2.5 ± 0.25 Endpoint: 2.1 ± 0.19 - Changes within CE/MPA group: p-value = not significant - Changes within E2/NETA group: p-value = < 0.001 (deterioration by 16%) Discontinuation due to adverse events Headache: 3	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes - double dummy technique with dark coated tablet A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Unclear B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: low	Main outcome classification Psychological Main interventions classification HRT

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Yes Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Unclear D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes - validated scoring system D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Yes - participants recorded confounding factors in diary Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					Intervention: yes Outcomes: yes Indirectness: no	
Full citation Purdie,D.W., Empson,J.A., Crichton,C., Macdonald,L., Hormone replacement therapy, sleep quality and psychological wellbeing, British Journal of Obstetrics and Gynaecology, 102, 735-739, 1995 Ref Id 227189 Country/ies where the study was carried out UK Study type Randomised, single-blind, placebo-controlled trial Aim of the study To examine the effect of hormone replacement therapy upon sleep quality and duration in postmenopausal women. Study dates Not reported. Source of funding Wyeth Laboratories plc supplied HRT	Sample size N = 33 HRT: 17 Placebo: 16 Characteristics Mean age of HRT group: 54.3 yrs (range 49 - 60) Mean age of Placebo group: 53.6 yrs (range 50 - 59) Inclusion criteria - Amenorrheoic for at least 6 months - VSM symptoms - No HRT within past 6 months - Normotensive Exclusion criteria - Not reported.	Interventions HRT - 0.625mg conjugated equine oestrogen (orally), progestogen norgestrel 0.15 mg taken from days 17 - 28	Power calculation Sample size of 16 patients per group would be sufficient to detect a difference of 0.35 in waking episodes per hour of cumulative sleep, with 90% power using a two-sided test and placebo group over course of study. Intention to treat Not reported. Details Setting Princess Royal Hospital, Hull Randomisation method Randomisation schedule carried out in blocks of 4 Statistical methods ANCOVA	Results Sleep Quality - Stanford Sleepiness Questionnaire Arousals (number of shifts from deeper sleep to stage I sleep to wakefulness) HRT - Mean (SD) Baseline (First night): 13.94 (5.18) Endpoint (night 8): 10.88 Placebo Baseline (First night): 16.76 (5.60) Endpoint (night 8): 12.41 (5.66) - No significat difference attributable to HRT or placebo - Significant reduction in arousals in both groups during course of study ($p < 0.005$) Wakefulness (minutes) HRT Baseline (First night): 9.88 (9.34) Endpoint (night 8): 10.06 (13.44) Placebo Baseline (First night): 20.53 (15.87) Endpoint (night 8): 15.18 (12.47) - No significant difference between groups - Significant reduction in both groups: $p < 0.05$. Crown - Crisp experiential index Free floating Anxiety HRT Baseline: 7.06 (4.06) Endpoint (week 9 - 12): 4.63 (3.83) Placebo	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - No A3 - Were groups comparable at baseline - Unclear Level of bias: High B Performance bias B1 - Did groups get same level of care - Unclear B2 - Were participants blinded to treatment allocation- No - after bleeding occured, allocation became known to participants B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: High	Main outcome classification Psychological Main interventions classification HRT

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
				Baseline: 7.06 (3.70) Endpoint (week 9 - 12): 6.53 (3.56) - HRT group showed dsignificantly greater improvement between baseline and the mid and late periods (11th week) - p < 0.01 Somatic anxiety HRT Baseline: 6.13 (3.00) Endpoint (week 9 - 12): 3.94 (2.35) Placebo Baseline: 7.29 (3.31) Endpoint (week 9 - 12): 6.71 (2.69) - HRT group showed dsignificantly greater improvement between baseline and the mid and late periods (11th week) - p < 0.02 Depression HRT Baseline: 5.32 (1.92) Endpoint (week 9 - 12): 4.25 (2.24) Placebo Baseline: 5.82 (2.10) Endpoint (week 9 - 12): 5.64 (1.22) - HRT group showed dsignificantly greater improvement between baseline and the mid and late periods (11th week) - p < 0.025	C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: High D Detection bias D1 - Was follow-up appropriate length - Unclear D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Ross,L.A.,	Sample size Tibolone n=18	Interventions Oral conjugated	Power calculation A minimum of 26	Results Frequency of hot flushes (including night sweats)	Limitations NICE guidelines	Main outcome classification

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Alder,E.M., Cawood,E.H., Brown,J., Gebbie,A.E., Psychological effects of hormone replacement therapy: a comparison of tibolone and a sequential estrogen therapy, Journal of Psychosomatic Obstetrics and Gynecology, 20, 88-96, 1999 Ref Id 227235 Country/ies where the study was carried out Scotland Study type Randomised, initially double-blind, controlled trial Aim of the study To compare the psychological effects of two regimens of HRT in perimenopausal women Study dates Not reported Source of funding Organon Laboratories Ltd, UK	Sequential oestrogen (conjugated equine oestrogen plus progestogen) n=18 Characteristics Tibolone / sequential oestrogen / p-value Age, years (study does not report if mean or median age was used): 52.2 / 52.0 / 0.89 Inclusion criteria -Climacteric symptoms -At least 45 years of age -Intact uterus -Amenorrhea for at least 3 months -No past psychotic history nor current use of antidepressants or psychotherapeutic agents -No contraindications to oestrogen therapy Exclusion criteria Not reported	equine estrogen 0.625 mg daily plus progestogen (norgestrel) 150 micrograms for the last 12 days of each 28 day cycle, or tibolone 2.5 mg/day for 28 days for three months of the trial	patients would be required, 13 in each group to detect a 40% difference with 80% power between scores of depression on the Women's Health Questionnaire for the two drugs Intention to treat Yes Details Setting Queen Margaret College, Edinburgh, Edinburgh Healthcare NHS Trust, Family Planning and Well Woman Services, Edinburgh, Scotland Randomisation method Randomisation was made by pre-generated sequential randomisation lists with a block size of ten, and each packet was given a code number. Copies of the code were kept by Organon Laboratories and in the Department Office at Queen Margaret College in opaque sealed envelopes.	Not reported Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported -Depression Not reported -Cognitive function Reported as median change scores from baseline in Women's Health Questionnaire memory problems scale Tibolone (n) / Sequential oestrogens (n) / Significance Month 1: 0 (16) / 0.09 (15) / 0.03 Month 2: 0.08 (15) / 0.39 (13) / 0.006 Month 3: 0.01 (15) / 0.39 (12) / 0.05 For the first month, women taking sequential oestrogen improved slightly compared with the tibolone group. After 2 and 3 months, small difference in memory problems remained. There was no significant differences in any changes from baseline between the two groups. -Sleep disturbance Not reported -Quality of life Not reported Musculoskeletal symptoms Not reported Safety outcomes -Discontinuation Reported as withdrawal due to side effects Tibolone n=2 Sequential oestrogen n=3 -Major adverse events Not reported -Minor adverse events Not reported	manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for	Cognitive function-WHQ memory problems Discontinuation Main interventions classification Oestrogen combined with progestogen (oral conjugated equine estrogen 0.625 mg daily plus progestogen) Tibolone

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
			Statistical methods Mean values for 3 weeks baseline (before medication) and first, second and third months of HRT were analysed. Drugs were compared using a Mann-Whitney U test to measure for differences between changes from baseline between the two groups. Wilcoxon rank sum tests were used to test whether changes from baseline were significant within each group.		missing data - Unclear Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Rotem,C., Kaplan,B., Phyto-Female Complex for the relief of hot flushes, night sweats and quality of sleep: randomized, controlled, double-blind pilot study,	Sample size 25 randomised to Phyto-Female Complex group with 21 analysed. 25 randomised to placebo group with 23 analysed. 5 in placebo and 2 in study group	Interventions Oral Phyto-Female Complex (standardised extracts of black cohosh, dong quai, milk thistle, red clover, American ginseng, chaste-tree berry) or	Power calculation NR Intention to treat NR Details Setting Five community gynaecological clinics of major health	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there	Main outcome classification Sleep - sleep quality score Discontinuation Main interventions classification Herbal preparations Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Gynecological Endocrinology, 23, 117-122, 2007 Ref Id 227240 Country/ies where the study was carried out Israel Study type Randomized, double-blind, placebo-controlled trial Aim of the study To determine the efficacy and safety of the herbal formula Phyto-Female Complex (SupHerb, Netanya, Israel; ingredients: standardized extracts of black cohosh, dong quai, milk thistle, red clover, American ginseng, chaste-tree berry) for the relief of menopausal symptoms. Study dates Not reported (NR) Source of funding Not reported	dropped out during the first four weeks and 2 in placebo group during weeks 4-8 owing to lack of compliance or deciding voluntarily to discontinue participation. Characteristics Phyto-Female Complex- mean age (SD) 55.3±5.4, years in menopause: 6.88±4.77 Placebo- mean age (SD) 59.0±7.3, years in menopause: 8.95±6.44 Inclusion criteria -Amenorrhoea for at least 6 months, with hot flushes and/or night sweats at least three times daily -Healthy peri (study called perimenopausal premenopausal) and postmenopausal women, aged 44-65 years Exclusion criteria Not reported	matched placebo twice daily for 3 months	maintenance organisation in Israel Randomisation method Not reported Statistical methods A structured questionnaire on the frequency and intensity of menopausal symptoms was administered weekly from one week before throughout the 3-month treatment period, followed by biochemical tests, breast check, and transvaginal ultrasonography. Sleep quality was subjectively assessed on a scale of 1 to 5, with 1 meaning 'good sleeper'. Data were compared between groups and within groups, before treatment and at the end of treatment, using Student's paired two-tailed t test.	-Depression Not reported -Cognitive function Not reported -Sleep disturbance Reported as mean sleep quality score, SD Phyto-Female Complex / Placebo/ p-value -Baseline: 3.58 (1.14) / 2.57 (1.53) / NS -End of treatment at 3 months: 1.06 (1.04) / 2.05 (1.17) / 0.001 -Quality of life Not reported Musculoskeletal symptoms Not reported Safety outcomes -Discontinuation 7 women in the placebo group felt aggravation of or no change in symptoms and decided to stop the treatment -Major adverse events Not reported -Minor adverse events Not reported	appropriate randomisation - Unclear, method of randomisation was not reported A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: Unclear B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - No, reliability and validity of sleep quality score measure was not reported and the measur was self-rated D4 - Were investigators blinded to intervention - Unclear D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: some-the study used Israeli women Other information The first author is the scientific	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					consultant for the product tested in this study and SubHerb donated the Phyto-Female (herbal) capsules used in the study	
Full citation Rudolph,I., Palombo-Kinne,E., Kirsch,B., Mellinger,U., Breitbarth,H., Graser,T., Influence of a continuous combined HRT (2 mg estradiol valerate and 2 mg dienogest) on postmenopausal depression, Climacteric, 7, 301-311, 2004 Ref Id 227254 Country/ies where the study was carried out Germany Study type Randomised, double-blind, placebo-controlled Aim of the study To investigate the effects of continuous combined hrt with 2 mg estradiol valerate and 2 mg dienogest over 24 weeks on postmenopausal depression Study dates Not reported Source of funding Jenapharm GmbH & Co. KG.	Sample size N = 129 Characteristics EV + DNG (N = 65): Age (yrs): 55.3 + 5.1 Last menstrual period (months): 109.3 + 97.60 Placebo (N = 64): Age (yrs): 56.9 + 5.0 Last menstrual period (months): 123.3 + 95.2 Inclusion criteria - Healthy postmenopausal women - 48 - 65 yrs - Mild to moderate depressive epidode according to ICD10 and HAMD > 16 Exclusion criteria - Any contraindications for HRT wit estradiol - A severe depressive episode and acute stressful life events	Interventions - 2 mg Estradiol valerate (EV) + 2 mg Dienogest (DNG) per day	Power calculation Not reported. Intention to treat Yes Details Setting Two large practices Randomisation method Randomisation code produced using random number generator to select random permuted blocks. Statisticam methods Descriptive statistics and repeated analysis of variances (ANOVA, GLM, SAS). ANCOVA used in vsm and sleep disturbance	Results Depression (HAMD) Placebo (mean + SD) Baseline (n = 64): 18.8 + 3.9 Final (n = 38): 12.8 + 8.5 Mean difference (final - baseline): -6.4 + 7.7 EV + DNG Baseline (n = 65): 18.9 + 3.1 Final (n = 51): 8.9 + 6.4 Mean difference (final - baseline): -9.7 + 6.2 Depression severity Placebo (mean + SD) Baseline: 18.8 + 3.9 Final: 15.0 + 7.7 EV + DNG Baseline: 18.9 + 3.1 Final: 10.8 + 7.2 ANOVA Main effect treatment: p = 0.0044 Time by treatment interaction: p < 0.0001 Sleep disturbances (WHQ) ANCOVA (between-subject effects): Treatment p-value: 0.0475 Placebo (mean + SD) Baseline (n = 64): 18.8 + 3.9 Final (n = 38): 12.8 + 8.5	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes A3 - Were groups comparable at baseline - Yes Level of bias: low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: low	Main outcome classification Psychological Main interventions classification HRT

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Yes Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Unclear D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Schmidt,P.J., Nieman,L.,	Sample size 34 female subjects, 16 received	Interventions Placebo skin patch for 3 weeks.	Power calculation Not reported Intention to treat	Results Frequency of hot flushes (including night sweats) Not reported	Limitations NICE guidelines manual 2012:	Main outcome classification Depression

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Danaceau,M.A., Tobin,M.B., Roca,C.A., Murphy,J.H., Rubinow,D.R., Estrogen replacement in perimenopause-related depression: a preliminary report, American Journal of Obstetrics and Gynecology, 183, 414-420, 2000 Ref Id 227287 Country/ies where the study was carried out USA Study type Double-blind parallel design with those in the placebo group crossed over to the treatment group Aim of the study Examine the efficacy of estrogen in the treatment of perimenopausal-related depression in women with and without hot flushes Study dates Not reported Source of funding Not reported	estradiol first and 18 received placebo first. Characteristics Age, mean year (SD) and range: 17β-estradiol: 48.3 (2.7), 44-52 Placebo: 50.1 (3.1), 44-55 Subjects without hot flushes (n) 17β-estradiol: 9 Placebo: 9 Subjects with current Research Diagnostic Criteria for minor depression (n) 17β-estradiol: 13 Placebo: 13 Subjects with current Diagnostic and Statistical Manual III Revised Criteria for major depression (n) 17β-estradiol: 3 Placebo: 5 Inclusion criteria -Self-report onset of depression associated with mentrual cycle irregularity of at least 6 months' duration but with ≤1 of amenorrhea	17β-estradiol estraderm skin patch (0.05 mg/day) for 3 weeks. Subsequently, women receiving estradiol during the first 3 weeks continued receiving estradiol for an additional 3 weeks, whereas women who had received placebo crossed over to estradiol for 3 weeks.	Not reported Details Setting Outpatient clinic within the National Insitutes of Health Clinical Center in the US Randomisation method All subjects were given 1 week of single-blind placebo. Placebo non-responders were then randomised in a double-blind manner to receive either estraderm or placebo skin patch for 3 weeks. Depressed women with and without hot flushes were randomised separately. Both groups were randomised by a pharmacist who was not a study investigator. Statistical methods Symptom rating scores were compared by analysis of variance for repeated measures. Number of depressed perimenopausal women who	Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Reported as visual analog scale ratings (mean, SD) which ranged from 0 (not present) to 100 (present in the extreme) Estradiol at baseline: 56.4 (15.2) Placebo at baseline: 56.7 (13.1) Estradiol at week 4: 33.2 (21.5), P<0.01, week 4 versus baseline Placebo at week 4: 59.3 (19.9) P<0.01, estradiol (week 4) versus placebo (week 4) -Depression Reported as visual analog scale ratings (mean, SD) which ranged from 0 (not present) to 100 (present in the extreme) Estradiol at baseline: 56.2 (12.5) Placebo at baseline: 54.6 (15.9) Estradiol at week 4: 25.9 (16.0), P<0.01, week 4 versus baseline Placebo at week 4: 55.2 (22.8) P<0.01, estradiol (week 4) versus placebo (week 4) Reported as Center for Epidemiologic Studies-Depression (mean, SD) Estradiol at baseline: 23.0 (6.4) Placebo at baseline: 23.0 (8.4) Estradiol at week 4: 10.6 (6.9), P<0.01, week 4 versus baseline Placebo at week 4: 20.6 (6.9) P<0.01, estradiol (week 4) versus placebo (week 4) Reported as Hamilton Rating Scale for Depression	Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Unclear A2 - Was there adequate concealment - Yes A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Unclear Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for	Anxiety Main interventions classification Oestrogen (patch) Placebo (patch)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
	-diagnosis of major or minor depression determined by a strucured diagnostic interview -scores on the Center for Epidemiologic Studies Depression Scale ≥10 during 3 of the 4 screening visits -plasma levels of follicle-stimulating hormone ≥20 IU/L on 3 of 4 screening visits Exclusion criteria -medical illness -taking medication -abnormal result of a gynecologic examination or a mammogram -medical contraindication to oestrogen replacement therapy -history of psychiatric illness during the 2 years before the reported onset of the current episode of depression		responded to oestrogen or placebo on the basis of the percentage decrease in the Center for Epidemiologic Studies-Depression Scale scores after 3 weeks of oestrogen or placebo relative to baseline was examined.	(mean, SD) Estradiol at baseline: 14.6 (3.9) Placebo at baseline: 17.2 (5.8) Estradiol at week 4: 6.8 (5.2), P<0.01, week 4 versus baseline Placebo at week 4: 13.9 (5.9) P<0.01, estradiol (week 4) versus placebo (week 4) Please note results before cross-over are reported here. Musculoskeletal symptoms Not reported Safety outcomes -Discontinuation Not reported -Major adverse events Not reported -Minor adverse events Not reported	missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Unclear D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Soares,C.N., Arsenio,H., Joffe,H., Bankier,B., Cassano,P., Petrillo,L.F., Cohen,L.S., Escitalopram versus	Sample size For ITT: Estrogen and progestogen therapy (EPT) n=16 Escitalopram (ESCIT) n=16 Characteristics	Interventions 8 week open trial with ESCIT (flexible dose, 10-20 mg/day; fixed dose, 10mg/day for the first 4 weeks) or estrogen plus	Power calculation Not reported Intention to treat Yes-analyses included subjects who completed at least one treatment visit	Results Vasomotor Frequency of hot flushes (including night sweats)-not reported Altered sexual function Frequency of sexual intercourse-not reported (NR)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials	Main outcome classification Depression Discontinuation Minor adverse events-headache, weight change Main interventions

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
ethinyl estradiol and norethindrone acetate for symptomatic peri- and postmenopausal women: impact on depression, vasomotor symptoms, sleep, and quality of life, Menopause, 13, 780-786, 2006 Ref Id 227369 Country/ies where the study was carried out USA Study type Randomised open-label trial Aim of the study To examine efficacy and tolerability of escitalopram (ESCIT) compared to oestrogen and progestogen therapy (EPT) for the treatment of symptomatic peri- and postmenopausal women. Study dates Study participants recruited between June 2001 and September 2003 Source of funding Study partially supported by a National Alliance for Research on Schizophrenia and Depression Award (Dr. Soares) and a research grant from	Most women were white, divorced, with partial or completed college education, working outside the home, and presenting with menopause-related symptoms, particularly hot flashes. The majority of women in both groups met criteria for major depressive disorder. EPT/ESCIT Median age (range): 49 (40-58) /50 (40-59) Inclusion criteria Perimenopausal and postmenopausal women, aged 40 to 60 years, who presented with depressive disorders and menopause-related symptoms Exclusion criteria Clinical contraindications to estrogen therapy, undiagnosed abnormal vaginal bleeding, history of or current thrombophlebitis or thromboembolic disorderes Carcinoma of the breast Estrogen-dependent tumors Hepatic dysfunction or disease	progestogen therapy (ethinyl estradiol 5 mcg/day plus norethindrone acetate 1 mg/day)	(intention-to-treat), with the last observation carried forward. Details Setting Boston, MA, USA Randomisation method Not reported other than 40 women with depressive disorders and menopause-related symptoms were randomly assigned to an 8-week open-label escitalopram (ESCIT) or estrogen and progestogen therapy (EPT). Statistical methods Severity of depressive symptoms was assessed with the Montgomery-Asberg Depression Rating Scale (MADRS). Depressive symptoms were assessed at baseline and at weeks 2, 4, and 8. Scores from baseline to study end were assessed within the treatment groups using Wilcoxon signed	Psychological symptoms Anxiety: NR Depression: Full remission of depression (score of <10 on the Montgomery-Asberg Depression Rating Scale) was observed in 75% (12/16) of subjects treated with ESCIT, compared to 25% (4/16) treated with EPT (p=0.01). Decrease in depressive symptoms was significantly greater in subjects treated with ESCIT (median decline = 19.2 [range, 10-34]) compared with that in subjects treated with EPT (median decline = 9.4 [range, -6 to 30]) (p=0.03). Cognitive function: NR Sleep disturbance: NR Quality of life measurement (psychological):NR Musculoskeletal symptoms Symptom relief (joint pain and muscular pain [with and without] stiffness): NR Muscle strength: NR [validated] Physical activity (Greene sub-scale data): Reported in graphical format only Patient satisfaction: NR Quality of life (musculoskeletal): Reported in graphical format only Safety outcomes collected across NMA and standard reviews Discontinuation: Subjects dropped out due to "unwillingness to stay on hormones" (one subject on EPT at week 1, one subject on EPT at week 4), nausea (one subject on EPT at week 1), headaches (two subjects on ESCIT at week 1), "lack of efficacy" (one subject on EPT at week 4, one subject on ESCIT at week 3) Major adverse events Breast cancer-NR Other cancer-NR Arterial disease (e.g. coronary heart disease, stroke)-NR Venous thromboembolic disease (VTE) (e.g. DVT, thromboembolism)-NR Fracture-NR Mortality-NR Minor adverse events Bleeding pattern-NR	A Selection bias A1 - Was there appropriate randomisation - Unclear A2 - Was there adequate concealment - No A3 - Were groups comparable at baseline - Yes Level of bias: High B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- No B3 - Were individuals administering care blinded to treatment allocation- No Level of bias: High C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: High D Detection bias D1 - Was follow-up appropriate	classification Oestrogen combined with progesterone SSRI-Escitalopram

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Forest Pharmaceuticals (Drs. Cohen and Soares)			rank tests. Chi-square methods for discrete measures (or Fisher's exact test for small samples) and Mann-Whitney tests for continuous measures were used to examine potential differences between the treatment groups.	Headache-two subjects on ESCIT at week 1 Depression/anxiety/mood/mental health-NR Weight change/gain-Median weight hange observed after treatment with EPT was 1.62lb, which did not represent a significant variation when compared to weight observed at study entry. Women treated with ESCIT had a median change of 0.43lb, also nonsignificant compared to weight at study entry.	length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Unclear D4 - Were investigators blinded to intervention - No D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information Small sample size (16 on ESCIT and 16 on EPT). Open-label trial so patients were not kept "blind" to treatment allocation.	
Full citation Somunkiran,A., Erel,C.T., Demirci,F., Senturk,M.L., The effect of tibolone versus 17beta-estradiol on climacteric	Sample size Tibolone n=20 17 beta-oestradiol n=20 Characteristics Tibolone /17 beta-oestradiol / p Mean age (years,	Interventions Tibolone 2.5 mg/day or 17β-estradiol 2 mg/day for 6 months After 3 weeks washout period, treatment protocols	Power calculation Not reported Intention to treat Not reported Details Setting Department of Obstetrics and	Results Frequency of hot flushes (including night sweats) Not reported Frequency of sexual intercourse Not reported Psychological symptoms	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials	Main outcome classification Anxiety Depression Quality of life-psychological Quality of life-musculoskeletal

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
symptoms in women with surgical menopause: a randomized, cross-over study, Maturitas, 56, 61-68, 2007 Ref Id 227374 Country/ies where the study was carried out Turkey Study type Randomised, single-blind, cross-over study Aim of the study To compare the effectiveness of tibolone and 17β-estradiol on climacteric symptoms in surgically menopausal women. Study dates Not reported Source of funding Not reported	SD) 47.95 ± 3.28 / 47.58 ± 3.20 /Non-statistically significant The time interval between the surgery and the study was 3 weeks Inclusion criteria -Hysterectomy and bilateral oophorectomy -Perimenopausal period before the operation Exclusion criteria -Hypertensive disorders (systolic BP > 170 mmHg and/or diastolic BP > 105 mmHg) -Active liver disease -Cerebrovascular or thromboembolic disorders -Diabetes mellitus -Thyroid disorders -Any malignancies and chronic disease which may affect the quality of life	were exchanged for another 6 months	Gynecology, Duzce School of Medicine, Turkey Randomisation method Computer-generated list of random number groups Statistical methods The mean score of each symptom is calculated by the sum of all individual scores divided by the number of subjects. The score of the clusters are given as the sum of the mean scores of the symptoms within that cluster. For comparisons between baseline, tibolone and 17β-estradiol the non-parametric Wilcoxon Sign Rank Test was used.	-Anxiety Reported as mean score ± S.D. of the symptoms clusters of the Greene Climacteric Anxiety Scale during treatment Tibolone / 17beta-estradiol/p-value for tibolone vs 17beta-oestradiol 0.39 (0.58)/ 0.87 (1.01) /.002 Lower scores indicate improvement Compared with baseline, all subscores improved in both groups during treatment -Depression Reported as mean score ± S.D. of the symptoms clusters of the Greene Climacteric Depression Scale during treatment Tibolone / 17beta-estradiol/p-value for tibolone vs 17beta-oestradiol 0.25 (0.70)/ 1.25 (1.53) /reported as .000 Compared with baseline, all subscores improved in both groups during treatment -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Reported as mean score ± S.D. of the symptoms clusters of the Greene Climacteric Psychological Scale during treatment Tibolone / 17beta-estradiol/p-value for tibolone vs 17beta-oestradiol 0.64 (0.86)/ 2.12 (1.71) /reported as .000 Compared with baseline, all subscores improved in both groups during treatment Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not reported -Quality of life Reported as mean score ± S.D. of the symptoms clusters of the Greene Climacteric Somatic Scale	A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: Moderate B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation-Unclear B3 - Were individuals administering care blinded to treatment allocation-Unclear Level of bias: High C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: High	*All measured by Greene climacteric scale Main interventions classification Tibolone Oestrogen

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
				during treatment Tibolone / 17beta-estradiol/p-value for tibolone vs 17beta-oestradiol 0 / 0.43 (0.71) /.002 Compared with baseline, all subscores improved in both groups during treatment Safety outcomes -Discontinuation Not reported -Major adverse events Not reported -Minor adverse events Not reported	D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: some, study used Turkish women Other information This study was carried out among surgically menopausal women.	
Full citation Speroff,L., Efficacy and tolerability of a novel estradiol vaginal ring for relief	Sample size Vaginal ring delivering 50 mcg per day E2 (n = 113) or 100 mcg per	Interventions Vaginal ring delivering the equivalent of 50 mcg per day or 100	Power calculation Based on past unpublished studies of this E2 vaginal ring and	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse	Limitations NICE guidelines manual 2012: Appendix C: Methodology	Main outcome classification Anxiety Depression Quality of life-

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
of menopausal symptoms, Obstetrics and Gynecology, 102, 823-834, 2003 Ref Id 227387 Country/ies where the study was carried out USA Study type Double-blind, randomised, placebo-controlled trial Aim of the study To assess the efficacy, tolerability, and acceptance of a vaginal ring delivering the equivalent of 50 or 100 microg per day of estradiol (E2), compared with placebo, for relief of moderate to severe vasomotor symptoms and urogenital symptoms in postmenopausal women. Study dates Not reported Source of funding Warner Chilcott, a division of Galen Holdings PLC, which has developed this product	day E2 (n = 112), or a placebo vaginal ring (n = 108) for 13 weeks Characteristics Placebo/ Estradiol 50 mcg / Estradiol 100 mcg Mean age, year (SD): 50.7 (6.5) / 52.6 --- (8.3) / 51.8 (6.6) Hysterectomised, ovaries intact (%): 17 / 22 / 17 Inclusion criteria -At least 7 moderate to severe hot flushes per day or an average of at least 56 moderate to severe vasomotor symptoms per week for the 2 weeks before randomisation -Women with uterus were required to have had amenorrhea for more than 12 months before randomisation; if she had amenorrhea for less than 12 but at least 6 months, she was also required to have a FSH level of at least 40 IU and an E2 level of no mroe than 20 pg/mL -Women with hysterectomy must had bilateral oophorectomy	mcg per day of estradiol or a placebo vaginal ring for 13 weeks	assumptions of standard deviations, 80 women per group would be sufficient to detect a difference as small a 13 moderate to severe vasomotor symptoms per week, with a power of 0.80. Intention to treat Yes Details Setting The study reported the trial was conducted at 35 sites in the US with no indication of the setting type Randomisation method Randomisation schedule was generated with the SAS Proc Plan and women were randomised in blocks of six to 13 weeks of treatment Statistical methods Changes in Greene Climacteric Scale scores from baseline to weeks 4, 8, and 13 were analysed with analysis of variance and	Not reported Psychological symptoms -Anxiety Reported as mean change from baseline in Greene Climacteric Scale-Anxiety scores at week 13 50 mcg E2/ 100 mcg E2 / placebo Baseline: 4.85 / 4.87 / 5.78 Mean change from baseline at week 13: -2.56*/ -2.86*/ -1.94 * p < 0.002 versus placebo -Depression Reported as mean change from baseline in Greene Climacteric Scale-Depression scores at week 13 50 mcg E2/ 100 mcg E2 / placebo Baseline: 3.97 / 3.58 / 4.38 Mean change from baseline at week 13: -2.10*/ -1.88*/ -0.97 * p < 0.002 versus placebo -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Reported as mean change from baseline in Greene Climacteric Scale-Psychological scores at week 13 50 mcg E2/ 100 mcg E2 / placebo Baseline: 8.81 / 8.45 / 10.16 Mean change from baseline at week 13: -4.66*/ -4.74*/ -2.91 * p < 0.002 versus placebo Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Reported as mean change from baseline in Greene Climacteric Scale-somatic scores at week 13 50 mcg E2/ 100 mcg E2 / placebo Baseline: 3.40 / 3.39 / 4.39	checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: Unclear, as the study does not indicate where they recruited the subjects B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear	psychological Physical activity All measured by Greene Climacteric Scale Main interventions classification Oestrogen (depot)-oestradiol vaginal ring Placebo vaginal ring

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
	performed more than 6 weeks before randomisation; if they did not have bilateral oophorectomy must had a FSH level of at least 40 IU and an E2 level of no more than 20 pg/mL Exclusion criteria -Past or current thromoembolic disorder or cerebrovascular accident -Endometriosis -Allergy or intolerance to previous ERT or HRT, including disabling breakthrough bleeding -Past or current oestrogen-dependent neoplasia -Abnormal uninvestigated vaginal bleeding within 6 months of randomisation -Known or suspected pregnancy -Treatment with oestrogen, progestogen, androgen, or systemic corticosteroids by the oral route within 8 weeks of screening, by transdermal or buccal delivery		analysis of covariance	Mean change from baseline at week 13: -1.21*/ -1.38*/ -0.70 * p < 0.002 versus placebo -Quality of life Not reported Safety outcomes -Discontinuation Not reported -Major adverse events Not reported -Minor adverse events Not reported	C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
	within 4 weeks of screening, or by injection within 6 months of screening, hormone pellets or implants inserted within the previous 5 years or an implant removed within the past 3 months -Unopposed ERT for 6 months or more in women with an intact uterus or selective oestrogen receptor modulators within 8 weeks of screening					
Full citation Thomson,J., Oswald,I., Effect of oestrogen on the sleep, mood, and anxiety of menopausal women, British Medical Journal, 2, 1317-1319, 1977 Ref Id 227452 Country/ies where the study was carried out Scotland Study type Double-blind controlled study Aim of the study To investigate the effect of oestrogen therapy on sleep, mood, anxiety, and hot flushes in perimenopausal women. Study dates	Sample size Oestrogen n=17 Placebo n=17 Characteristics Mean age only reported Oestrogen: 49.7 Placebo: 48.5 Inclusion criteria -Aged 45-55 -Amenorrhoea for at least three months -Symptoms of insomnia, depression, anxiety, and hot flushes Exclusion criteria Not reported	Interventions In the first six weeks all patients received a placebo. In the remaining eight weeks one of each pair received piperazine oestrone sulphate in a dose of 1.5 mg twice daily while the other remained on placebo.	Power calculation Not reported Intention to treat Not reported Details Setting Patients were referred by local general practitioners in Scotland. Randomisation method Not reported Statistical methods Intragroup changes in the different periods of the experiment were compared by t tests for paired observations. The changes between the baseline period and first	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Measured by Hamilton anxiety score (SE) Oestrogen/placebo Start of study: 17.2 (1.8) / 20.1 (2.1) End of baseline period: 9.7 (1.3)/ 11.4 (1.3) End of first treatment month: 7.7 (1.2)/ 6.5 (1.1) End of second treatment month: 5.6 (1.4)/ 5.4 (0.7) No significant differences between the two groups. In both groups the difference in values between the start of the study and the end of the baseline period was significant (oestronegroup: P < 0.001; placebo group: P < 0.001). The decrease from the end of the baseline period to the end of the first treatment month was significant for the placebo group (P < 0.001) but not for the oestrone group, and the decrease from the end of the baseline period to the end of the study was significant in both groups (oestrone group: P < 0.01; placebo group: P <0.001).	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Unclear A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Unclear Level of bias: High B Performance bias B1 - Did groups get same level of care - Yes B2 - Were	Main outcome classification Anxiety-Hamilton anxiety score Depression-Hamilton depression score Sleep disturbance-mean duration of sleep, time awake that intervenes between periods of sleep, number of arousals from sleep to wakefulness Main interventions classification Oestrogen Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Not reported Source of funding Not reported			treatment month and between the baseline and second treatment month were also examined for each group, and the magnitude of change in the two groups was then compared using Student's t test. A one-tailed test was used for intervening wakefulness and frequency of arousals, which we had predicted would decrease with oestrogen treatment, and a two-tailed test in all other cases.	-Depression Measured by Hamilton depression score (SE) Oestrogen/placebo Start of study: 16.3 (1.9) / 18.2 (2.0) End of baseline period: 7.9 (1.2)/ 10.1 (1.5) End of first treatment month: 7.3 (1.3)/ 6.2 (1.3) End of second treatment month: 5.9 (1.8)/ 4.5 (0.7) In both groups the difference in values between the start and end of the baseline period was significant (oestrone group: P < 0.001; placebo group: P < 0.001). In the placebo group there was a significant decrease from the end of the baseline period to the end of the first treatment month (P < 0.02) and to the end of the second treatment month (P <0.01), but in the oestrone group these changes did not reach significance. There were no significant differences between the two groups. -Cognitive function Not reported -Sleep disturbance Measured by mean duration of sleep (SE) The duration of sleep increased in both groups. In the oestrogen group mean sleep duration increased from a baseline value of 423.2 (8.2) minutes to 442.2 (7.7) minutes in the first treatment month (P<0.01) and rose to 446.5 (7.2) minutes in the second treatment month (P <0.01). In the placebo group the increase from the baseline duration of 418.2 (7.2) minutes to 424.3 (8.2) minutes in the first treatment month was not significant, but the increase from the baseline value to 429.4 (7.2) minutes in the second treatment month was significant (P <0.02). The difference between the two groups was not significant. Measured by minutes (SE) awake that intervenes between periods of sleep Oestrogen/placebo/ p-value significance Change from baseline at first treatment month: -14.4 (5.1)/ -4.7 (4.5)/ not significant (p-value not reported) Change from baseline at second treatment month: -15.8 (5.8)/ 2.1 (2.2)/ significant difference between the two groups (p< 0.025) End of second treatment month: 446.5 (7.2)/ 4.5 (0.7)	participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Unclear D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
				Negative minutes denote decrease in the amount of intervening wakefulness Measured by mean number (SE) of arousals from sleep to wakefulness The oestrone-treated group woke less often. In the second treatment month they showed a decrease in the number of arousals from sleep to wakefulness of 0.9 (0.4) compared with the baseline period, whereas the placebo group showed a small mean increase of 0.1 (0.4). The difference between the two groups was significant (P<0.05). -Quality of life Not reported Musculoskeletal symptoms Not reported Safety outcomes -Discontinuation Not reported -Major adverse events Not reported -Minor adverse events Not reported	factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information Study does not report randomisation	
Full citation Tice,J.A., Ettinger,B., Ensrud,K., Wallace,R., Blackwell,T., Cummings,S.R., Phytoestrogen supplements for the treatment of hot flashes: the Isoflavone Clover Extract (ICE) Study: a randomized controlled trial, JAMA, 290, 207-214, 2003 Ref Id 227456 Country/ies where	Sample size Promensil n=84 assigned and analysed Rimostil n=83 assigned and analysed Placebo n=85 assigned and analysed Characteristics Promensil / Rimostil / Placebo Mean age, year (SD): 52.3 (2.8) / 52.3 (3.0) / 52.3 (3.4) Surgical menopause n (%): 6 (7) / 4 (5) /	Interventions -Promensil (82 mg of total isoflavones per day) -Rimostil (57 mg of total isoflavones per day) -Identical placebo contained less than 0.04 mg of total isoflavones per tablet -Participants were instructed to take 2 tablets once daily for 12 weeks	Power calculation The study was designed to have 90% power to detect at least a 15% greater reduction in hot flash frequency in the active treatment arms compared with the placebo arm. Intention to treat Yes Details Setting 3 academic clinical research sites located in	Results There were significant improvements from baseline in all 3 groups, but there were no statistically significant differences between groups on any of the Greene scales Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Reported as change in mean Greene Climacteric anxiety subscale (95% CI) from randomisation to the end of study Promensil / Promensil versus Placebo P value: -1.1 (-1.6 to 0.6) / .33	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes A3 - Were groups comparable at	Main outcome classification All effectiveness outcomes measured by Greene Climacteric Scale Anxiety Depression Quality of life-psychological Quality of life-musculoskeletal Discontinuation Minor adverse events-headache Main interventions classification Phytoestrogens Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
the study was carried out USA Study type Randomised, double-blind, placebo-controlled trial Aim of the study To compare the efficacy and safety of 2 dietary supplements derived from red clover with placebo in symptomatic menopausal women Study dates Between November 1999 and March 2001 Source of funding Novogen Inc	6 (7) Inclusion criteria -45 to 60 years -Experiencing at least 35 hot flashes per week -Had a follicle-stimulating hormone (FSH) level of 30 mIU/mL -Had either documented bilateral oophorectomy or at least 2 consecutive months of amenorrhea prior to enrollment with at least 6 months of amenorrhea in the year prior to entry Exclusion criteria -Vegetarian -Consumed soy products more than once per week -Took medications affecting isoflavone absorption (antibiotics, antacids) or hormonal preparations during the 3 months prior to enrollment -Had significant gastrointestinal disease -Drank more than 2 alcoholic beverages per day -Were allergic to red clover -Were regular users of dietary supplements containing		Oakland, California; Minneapolis, Minnesota; and Iowa City, Iowa. The study was administered through a coordinating center at the University of California, San Francisco. Randomisation method By the central pharmacy using computer-generated randomisation in blocks of 6, stratified by clinical site. Statistical methods Scores for the subscales of the Greene Climacteric Scale were calculated using the standard method described by Greene. Data are reported using the last observation carried forward.	Rimostil / Rimostil versus Placebo P value: -0.8 (-1.3 to 0.3) / .80 Placebo: -0.7 (-1.3 to 0.2) -Depression Reported as change in mean Greene Climacteric depression subscale (95% CI) from randomisation to the end of study Promensil / Promensil versus Placebo P value: -0.7 (-1.1 to 0.2) / .23 Rimostil / Rimostil versus Placebo P value: -0.4 (-0.8 to -0.2) / .79 Placebo: -0.3 (-0.7 to -0.2) -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Reported as change in mean Greene Climacteric psychological subscale (95% CI) from randomisation to the end of study Promensil / Promensil versus Placebo P value: -1.8 (-2.6 to 0.9) / .23 Rimostil / Rimostil versus Placebo P value: -1.2 (-2.0 to 0.3) / .77 Placebo: -1.0 (-1.9 to 0.1) Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not directly reported, although the study used Greene somatic scale, reported below	baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
	isoflavones, or consumed less than 80% of the expected study tablets during the 2-week placebo run-in period			-Quality of life Reported as change in mean Greene Climacteric somatic subscale (95% CI) from randomisation to the end of study Promensil / Promensil versus Placebo P value: -0.4 (-0.8 to -0.03) / .60 Rimostil / Rimostil versus Placebo P value: -0.6 (-1.1 to 0.2) / .82 Placebo: -0.6 (-1.0 to 0.1) Safety outcomes -Discontinuation 1 discontinued due to adverse event in Rimostil group -Major adverse events Not reported -Minor adverse events Reported as number and percentage of participants Promensil / Rimostil / Placebo / P value Headache: 5 (6) / 4 (5) / 11 (13) / .13	D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information	
Full citation Utian,W., Yu,H., Bobula,J., Mirkin,S., Olivier,S., Pickar,J.H., Bazedoxifene/conjugated estrogens and quality of life in postmenopausal women, Maturitas, 63, 329-335, 2009 Ref Id 227488 Country/ies where the study was carried out USA Study type Multicenter, double-	Sample size BZA 20 mg/CE 0.45 mg (n = 127) BZA 20 mg/CE 0.625 mg (n = 128) Placebo (n = 63) Characteristics BZA 20 mg/CE 0.45 mg / BZA 20 mg/CE 0.625 mg / Placebo /p-value Mean Age (SD): 53.57 (4.82) / 53.09 (4.41) / 53.62 (5.31) / 0.666 Inclusion criteria Postmenopausal women (aged 40–65 years) who had an	Interventions BZA 20 mg/CE 0.45 mg, BZA 20 mg/CE 0.625 mg, or placebo for 12 weeks	Power calculation Not reported Intention to treat Not reported Details Setting 43 sites in the United States (no further details) Randomisation method Not reported Statistical methods Changes from baseline in sleep scale and	Results Frequency of hot flushes (including night sweats) Not reported Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported -Depression Not reported -Cognitive function Reported as percentages of subjects reporting ability to concentrate per Menopause Symptoms Treatment Satisfaction Questionnaire (MS-TSQ) BZA 20 mg/CE 0.45 mg / BZA 20 mg/CE 0.625 mg /	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Unclear, randomisation methods not reported A2 - Was there adequate concealment -	Main outcome classification Cognitive function (ability to concentrate-MS-TSQ) Sleep disturbance (MOS sleep disturbance scale) Quality of life-psychological (MENQOL psychosocial) Quality of life-musculoskeletal (MENQOL physical) Main interventions classification Tissue selective oestrogen complexes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
blind, placebo-controlled study Aim of the study To assess the effects of bazedoxifene/conjugated estrogens (BZA/CE) on sleep parameters and health-related quality of life (HR-QOL) Study dates Not reported Source of funding Wyeth Research, Collegeville, PA, USA.	intact uterus and endometrial biopsy results at screening At least 7 moderate-to-severe hot flushes per day (or at least 50 per week) Exclusion criteria Uncontrolled hypertension (i.e., systolic blood pressure >140 mmHg or diastolic blood pressure >90 mmHg that was untreated) or controlled hypertension using greater than 2 antihypertensive medications prior to randomization Fasting total cholesterol >300 mg/dL or triglycerides >300 mg/dL Fasting blood glucose >125 mg/dL ECG findings suggestive of ischemia		MENQOL scores were analyzed using an analysis of covariance (ANCOVA), with treatment and study site as factors and baseline value as a covariate	Placebo 52.2* / 56.4 / 40.7 * Subjects receiving BZA 20 mg/CE 0.45 mg versus placebo reported significantly greater satisfaction with the ability to concentrate (P < 0.05) -Sleep disturbance Reported as mean (SD) baseline Medical Outcomes Study (MOS) sleep scale measures-sleep disturbance BZA 20 mg/CE 0.45 mg / BZA 20 mg/CE 0.625 mg / Placebo / p-value 47.0 (25.3) / 45.2 (22.5) / 46.4 (21.2) / 0.828 Mean (SE) change from baseline in MOS sleep scale-sleep disturbance measures at Week 12 BZA 20 mg/CE 0.45 mg / BZA 20 mg/CE 0.625 mg / Placebo −19.95 (1.93)*/ −21.41 (2.06)* / −5.90 (2.69) *P < 0.001 vs placebo Sleep scale measured on 6-point scale, ranges from 1 = “all of the time” to 6 = “none of the time”) At Week 12, both doses of BZA/CE showed significant improvements (P < 0.001) in scores for sleep disturbance compared with placebo Reported as effect size (95% CI) for MOS sleep measures-sleep disturbance at Week 12 BZA 20 mg/CE 0.45 mg / BZA 20 mg/CE 0.625 mg −0.65 (−0.98 to −0.31) / −0.75 (−1.08 to −0.41) The treatment effect sizes with BZA 20 mg/CE 0.45 and 0.625 mg were medium to large for sleep disturbance (−0.65 and −0.75) and the corresponding 95% CIs showed that these effect sizes were significant. -Quality of life Reported as mean (SD) baseline Menopause-Specific Quality of Life (MENQOL)-psychosocial function BZA 20 mg/CE 0.45 mg / BZA 20 mg/CE 0.625 mg / Placebo / p-value 3.66 (1.83) / 3.51 (1.66) / 3.68 (1.70) / 0.733 Reported as mean change from baseline in MENQOL psychosocial function scores at Week 12 BZA 20 mg/CE 0.45 mg / BZA 20 mg/CE 0.625 mg /	Unclear A3 - Were groups comparable at baseline - Yes Level of bias: Unclear B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation-Unclear B3 - Were individuals administering care blinded to treatment allocation-Unclear Level of bias: Unclear, as method of blinding not reported C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Unclear Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - N/A	(BZA 20 mg/CE 0.45 mg, BZA 20 mg/CE 0.625 mg) Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
				Placebo -0.9 / -1.2* / -0.7 *p < 0.05 vs placebo Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not reported -Quality of life Reported as mean (SD) baseline Menopause-Specific Quality of Life (MENQOL)-physical function BZA 20 mg/CE 0.45 mg / BZA 20 mg/CE 0.625 mg / Placebo / p-value 3.92 (1.51) / 3.68 (1.36) / 3.63 (1.38) / 0.308 Reported as mean change from baseline in MENQOL physical function scores at Week 12 BZA 20 mg/CE 0.45 mg / BZA 20 mg/CE 0.625 mg -1.1 / -1.3* / -0.8 *p < 0.01 vs placebo Safety outcomes -Discontinuation Not reported -Major adverse events Not reported -Minor adverse events Not reported	D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Unclear D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information	
Full citation Veerus,P., Fischer,K., Hovi,S.L., Karro,H., Rahu,M., Hemminki,E., Symptom reporting and quality of life in the Estonian Postmenopausal Hormone Therapy Trial, BMC Women's	Sample size N = 1823: Blind HT arm: 415 Placebo: N = 381 Non-blind HT arm: N = 503 Non-treatment arm: N = 524 Characteristics Mean Age (yrs)	Interventions - 0.625 mg CEE (regardless of hysterectomy status) + 2.5 mg MPA or: - 0.625 mg CEE and 5 mg MPA if they were within 3 years from their last period	Power calculation Not reported. Intention to treat Yes Details Setting Clinical centers in Estonia Randomisation method	Results % of participants reporting EuroQoL (EQ - 5D) scores Trouble sleeping (%) Non-blind HT Baseline: 31.4 Final: 34.1 Non-treatment:	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there	Main outcome classification Psychological Musculoskeletal Main interventions classification HRT

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Health, 8, 5-, 2008 Ref Id 227513 Country/ies where the study was carried out Estonia Study type Open-label Aim of the study To determine the effect of postmenopausal hormone therapy on women's symptom reporting and quality of life. Study dates 1999 - 2004 Source of funding Not reported.	All: 58.2 (4.0) Postmenopausal: 8.0 (4.0) years Inclusion criteria - Aged 50 - 64 - Estonian speaking in 2 areas (Tallinn and Tartu) and in 2 counties surrounding these towns Exclusion criteria Not reported.		Not reported Statistical method Mixed effects logistics regression with random subject specific intercepts, using a penalized quasi-likelihood metho.	Baseline: 30.3 Final: 36.2 Blind HT Baseline: 30.2 Final: 31.3 Placebo: Baseline: 34.2 Final: 33.3 95% OR = 0.66 (0.52 - 0.84) Depression Non-blind HT Baseline: 27.1 Final: 21.6 Non-treatment: Baseline: 27.2 Final: 23.6 Blind HT Baseline: 23.4 Final: 18.9 Placebo: Baseline: 21.0 Final: 19.3 95% CI: 0.81 (060 - 1.08) Anxiety Non-blind HT Baseline: 34.4 Final: 27.3 Non-treatment: Baseline: 36.1 Final: 29.5 Blind HT Baseline: 34.6 Final: 25.2 Placebo: Baseline: 33.2	appropriate randomisation - Method of randomisation not reported A2 - Was there adequate concealment - No A3 - Were groups comparable at baseline - Yes Level of bias: High B Performance bias B1 - Did groups get same level of care - Unclear B2 - Were participants blinded to treatment allocation- No - open label B3 - Were individuals administering care blinded to treatment allocation- No Level of bias: High C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Yes Level of bias: Low D Detection bias D1 - Was follow-up appropriate	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
				Final: 25.2 95% CI: 0.93 (0.73 - 1.19) Stiffness/aches in joints Non-blind HT Baseline: 57.5 Final: 57.5 Non-treatment: Baseline: 54.5 Final: 56.5 Blind HT Baseline: 56.3 Final: 54.4 Placebo: Baseline: 54.2 Final: 56.5 95% CI: 0.97 (0.82 - 1.15) - No difference between treatment and non-treatment arms in reporting any symptoms	length - Unclear D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes - EQ-5D D4 - Were investigators blinded to intervention - No D5 - Were investigators blinded to confounding factors - Unclear Level of bias: High Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Wiklund,I.K., Mattsson,L.A., Lindgren,R., Limoni,C., Effects of a standardized ginseng extract on quality of life and physiological parameters in symptomatic postmenopausal women: a double-blind, placebo-controlled trial. Swedish Alternative Medicine Group, International Journal	Sample size N = 384 Placebo = 191 Ginseng = 193 Characteristics Age yrs mean, (SD) Ginseng = 53.3 (4.0) Placebo = 53.6 (4.0) Weight kg (SD) Ginseng = 71.1 (11.6) Placebo = 69.9 (11.5) Inclusion criteria - Aged 45 - 65, without HRT for previous 2 months	Interventions Ginseng	Power calculation Estimated maximum placebo effect size 50% for a clinically relevant difference and an alpha value of 0.05, power of 80% subjects per treatment group. Sample size identified as 182 subjects per arm. Intention to treat Yes Details Setting	Results VSM Reported in seperate evidence table Quality of Life: Psychological General Well-Being (PGWB) score Anxiety Ginseng (N= 193) Baseline = 22.8 (4.3) After 16 weeks = 24.2 (4.3) Mean change = 1.4 (4.1) p value = 0.0001 Placebo (N = 191) Baseline = 22.9 (4.3) After 16 weeks = 24.2 (4.1) Mean change = 1.3 (3.9)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Unclear A2 - Was there adequate concealment - Yes A3 - Were groups	Main outcome classification Qulaity of life Psychological Sexual function Musculoskeletal Main interventions classification Non pharmaceutical treatment

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
of Clinical Pharmacology Research, 19, 89-99, 1999 Ref Id 227562 Country/ies where the study was carried out Sweden Study type Randomised, multicenter, double-blind, placebo-controlled parallel group study. Aim of the study To compare the effect of a 16 week treatment with ginseng or placebo in postmenopausal women with climacteric symptoms. Study dates Not reported. Source of funding Pharmaton S.A	and with no bleeding during previous 6 months Exclusion criteria - Women taking concomitant medication		Not reported Randomisation method Not reported Statistical method Student's t-test for independent samples used to analyse difference between groups. Frequency of adverse events compared using Chi-squared statistics and Fisher's exact test.	p value = 0.0001 Ginseng - placebo treatment difference = 0.1 (4.0), p-value = not significant Depression Ginseng Baseline = 15.2 (2.6) After 16 weeks = 16.0 (2.3) Mean change = 0.7 (2.4) p value = 0.0001 Placebo Baseline = 15.7 (2.1) After 16 weeks = 15.9 (2.3) Mean change = 0.2 (2.2) p value = not significant Ginseng-placebo treatment difference = 0.5 (2.3), p-value = 0.04 Quality of life - Women's Health Questionnaire (WHQ) Somatic symptoms Ginseng Baseline = 13.5 (4.0) After 16 weeks = 12.0 (3.5) Mean change = -1.5 (3.4) p value = 0.0001 Placebo Baseline = 13.3 (3.9) After 16 weeks = 12.4 (3.8) Mean change = -1.0 (3.3) p value = 0.001 Ginsent - placebo treatment difference = -0.5 (3.4), p-value = not significant Anxiety Ginseng Baseline = 6.3 (2.1) After 16 weeks = 5.6 (1.7) Mean change = -0.8 (1.8) p value = 0.0001 Placebo Baseline = 6.2 (2.0) After 16 weeks = 5.7 (1.8) Mean change = - 0.5 (1.6) p value = 0.001 Ginseng - placebo treatment difference = - 0.2 (1.7), p-value = not significant	comparable at baseline - Yes Level of bias: medium B Performance bias B1 - Did groups get same level of care - Unclear B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Yes Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Unclear D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
				Depression Ginseng Baseline = 12.9 (3.8) After 16 weeks = 11.5 (3.7) Mean change = -1.3 (3.4) p value = 0.0001 Placebo Baseline = 12.5 (3.7) After 16 weeks = 11.6 (3.7) Mean change = - 0.9 (3.4) p value = 0.001 Ginseng - placebo treatment difference = - 0.4 (3.4), p-value= not significant Sexual function Ginseng Baseline = 6.3 (2.5) After 16 weeks = 5.6 (1.7) Mean change = -0.1 (1.8) p value = not significant Placebo Baseline = 6.2 (2.3) After 16 weeks = 6.0 (2.3) Mean change = - 0.2 (1.9) p value = not significant Ginseng - placebo treatment difference = 0.1 (1.8), p-value= not significant Sleep problems Ginseng Baseline = 6.8 (2.3) After 16 weeks = 5.8 (2.3) Mean change = -1.0 (1.9) p value = 0.0001 Placebo Baseline = 6.7 (2.2) After 16 weeks = 6.0 (2.2) Mean change = - 0.7 (1.8) p value = 0.001 Ginseng - placebo treatment difference = - 0.2 (1.9), p-value= not significant	D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Wu,M.H., Pan,H.A., Wang,S.T., Hsu,C.C., Chang,F.M.,	Sample size 48 randomised 36 subjects completed 3 months of treatment and	Interventions Tibolone 2.5mg/day CEE 0.625 mg/day plus MPA 5mg/day Treatments were for	Power calculation Not reported Intention to treat Not reported Details	Results Frequency of hot flushes (including night sweats) Not reported Frequency of sexual intercourse	Limitations NICE guidelines manual 2012: Appendix C: Methodology	Main outcome classification Anxiety Depression Quality of life-

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Huang,K.E., Quality of life and sexuality changes in postmenopausal women receiving tibolone therapy, Climacteric, 4, 314-319, 2001 Ref Id 227582 Country/ies where the study was carried out Taiwan Study type Prospective, randomised, single-blind trial Aim of the study To investigate the effects of hormone replacement therapy (HRT) and tibolone on the sexuality and quality of life of Taiwanese postmenopausal women. Study dates Not reported Source of funding Organon Taiwan Ltd	thus analysed (analysis exclude those who did not complete the treatment) Tibolone n=24 randomised, 6 did not complete Continuous combined HRT (CEE plus MPA) n=24 randomised, 6 did not complete Characteristics Tibolone / CEE-MPA Mean age, year (SD): 51.22 (4.26) / 52.28 (2.85) Menopause age, year (SD): 49.39 (4.09) / 50.50 (2.62) Time since menopause, year (SD): 1.94 (0.94) / 1.83 (0.79) Inclusion criteria 12-36 months postmenopausal At least one climacteric symptom according to the Greene Climacteric Scale Exclusion criteria Patients who missed more than 3 days of assigned treatment per month were disqualified and excluded from the analysis	3 months	Setting Department of Obstetrics and Gynecology and Public Health, College of Medicine, National Cheng-Kung University, Tainan, Taiwan; Department of Obstetrics and Gynecology, Chang Gung Memorial Hospital, Kaoshiung, Taiwan Randomisation method Not reported Statistical methods Differences within and between groups were analysed using paired and unpaired student t tests	Not reported Psychological symptoms -Anxiety Reported as self-rated changed of Greene Climacteric Anxiety Scale, mean (SD) Pretreatment / post-treatment Tibolone: 6.61 (3.29) / 1.72 (1.23) CEE-MPA: 6.39 (3.52) / 2.11 (1.45) Within-group comparisons all showed statistically significant differences in all items post-treatment -Depression Reported as self-rated changed of Greene Climacteric Depression Scale, mean (SD) Pretreatment / post-treatment Tibolone: 5.06 (2.99) / 1.44 (0.92) CEE-MPA: 5.28 (3.23) / 2.22 (1.90) Within-group comparisons all showed statistically significant differences in all items post-treatment -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Reported as self-rated changed of Greene Climacteric Psychological Factor Scale, mean (SD) Pretreatment / post-treatment Tibolone: 11.72 (5.48) / 3.17 (1.76) CEE-MPA: 11.67 (6.33) / 4.39 (3.05) Within-group comparisons all showed statistically significant differences in all items post-treatment Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not reported	checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Unclear A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: High B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Unclear B3 - Were individuals administering care blinded to treatment allocation- Unclear Level of bias: High C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data -	psychological Quality of life-musculoskeletal Discontinuation Minor adverse events-bleeding *All measured by Greene Climacteric Scale Main interventions classification Tibolone Oestrogen combined with progesterone (CEE+MPA)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
				-Quality of life Reported as self-rated changed of Greene Climacteric Somatic Factor Scale, mean (SD) Pretreatment / post-treatment Tibolone: 8.5 (3.39) / 2.78 (1.7) CEE-MPA: 9.22 (4.72) / 3.78 (2.10) Within-group comparisons all showed statistically significant differences in all items post-treatment Safety outcomes -Discontinuation Reported as dropping out due to body discomfort Tibolone n=3 CEE-MPA n=4 -Major adverse events Not reported -Minor adverse events Reported as vaginal bleeding % 1 month: -CEE-MPA: 31% (5/16) -Tibolone: none 3 months: -CEE-MPA: 37% (6/16) -Tibolone: 12% (2/16)	Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Unclear D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: some, the study used Taiwanese women	
Full citation Amsterdam,J.D., Yao,Y., Mao,J.J., Soeller,I., Rockwell,K., Shults,J.,	Sample size N = 34 Black cohosh extract n = 15 Placebo n = 13 Characteristics	Interventions Black Cohosh (2 x 32 mg capsules daily) Placebo (2 x 100% rice powder daily)	Power calculation 25 participants per arm had 90% power to detect effect size of 0.94 and 80% power to	Results Frequency of hot flushes (including night sweats) Not reported	Limitations NICE guidelines manual 2012: Appendix C: Methodology	Main outcome classification Anxiety-Hamilton, Beck, GCS Depression-GCS Quality of life-

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Randomized, double-blind, placebo-controlled trial of Cimicifuga racemosa (black cohosh) in women with anxiety disorder due to menopause, Journal of Clinical Psychopharmacology, 29, 478-483, 2009 Ref Id 227637 Country/ies where the study was carried out USA Study type Randomised, double-blind, placebo controlled, parallel group RCT Aim of the study To examine the anxiolytic efficacy of a specific black cohosh extract preparation in reducing the symptoms of Anxiety Disorder due to menopause. Study dates Not reported Source of funding National Institute of Health/National Center for Complementary and Alternative medicine.	Black cohosh (N = 15) Age (years): 56.7 (6.53) / 50 - 76 Age at onset of Generalised Anxiety Disorder (GAD): 43.6 (8.6) / 19 - 53 Placebo (N = 13) Age (years): 44.9 (11.4)/ 10 - 55 Age at onset of Generalised Anxiety Disorder (GAD): 44.9 (11.4)/ 10 - 55 Inclusion criteria - Women who were either postmenopausal for ≥ 12 months or peri menopausal (with amenorrhea lasting to 2 to 11 months in the proceeding year) - Perimenopausal women were ≥ 40 years old and had no other demonstrable reason for their amenorrhea - Women with prior hysterectomy and uncertain menopausal status had a serum FSH level of ≥ 40 mIU/ml -Had a DSM IV Axis I diagnosis of Anxiety Disorder due to menopause that was ascertained via	Both for 12 weeks	detect effect size of 0.81, using 2-group t-test with a 0.05 significance level Intention to treat Yes Details Setting Depression Research Unit, University of Pennsylvania Randomisation method Performed using blocked randomisation with varying block sized. Group numbers were randomly permuted within each block. Random numbers generated and permuted within each block Statistical methods - Generalised estimating equations (GEE) and quasi-least squares (QLS) with 2-sided tests of hypothesis via the xtgls procedure for STATA.	Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Baseline scores: Mean (SD)/range Hamilton Anxiety Rating Scale (HAM-A) Score Black Cohosh (n=15): 16.9 (3.8)/10-22 Placebo (n=13): 15.9 (3.5)/9-22 p value = 0.39 Beck Anxiety Inventory (BAI) Black Cohosh: 11.8 (6.7)/3-26 Placebo: 14.1 (8.6)/5-36 p-value= 0.66 Estimated values in overall change for treatment groups using regression model HAM-A Est change difference, Black Cohosh: -2.56 Est change difference, Placebo: -4.90 Effect size: 0.72 p-value: 0.29 BAI: Est change difference, Black Cohosh: -1.17 Est change difference, Placebo: -4.46 Effect size: 0.34 p-value: 0.578 GCS anxiety Est change difference, Black Cohosh: 0.0084 Est change difference, Placebo: -1.93 Effect size: 0.55 p-value: 0.121 -Depression GCS Depression Est change difference, Black Cohosh: -0.19	checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes A3 - Were groups comparable at baseline - Yes Level of bias: low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Yes Level of bias: Low	Psychological-GCS Discontinuation Minor adverse events-bleeding, anxiety Main interventions classification Herbal preparations (black cohosh) Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
	the Structured Diagnostic Interview for DSM IV Exclusion criteria - Axis I diagnosis of Major Depressive Disorder, Bipolar disorder and other psychological disorders. - Co-morbidities and contraindications to menopause			Est change difference, Placebo: -0.98 Effect size: 0.54 p-value: 0.148 -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Greene Climatic Score (GCS) Psychology Est change difference, Black Cohosh: -0.30 Est change difference, Placebo: -2.80 Effect size: 0.61 p-value: 0.063 Musculoskeletal symptoms Not reported Safety outcomes -Discontinuation One patient (6.7%) on black cohosh discontinued treatment due to adverse events -Major adverse events Not reported -Minor adverse events Reported as menstrual flow, spotting and vaginal bleeding Black cohosh n = 1 Placebo n = 3 Reported as increased anxiety Black cohosh n = 1 Placebo n = 0	D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: yes Other information	
Full citation Barton,D.L., LaVasseur,B.I., Sloan,J.A., Stawis,A.N.,	Sample size Started treatment: 10 mg citalopram/placebo: n=54 / n=28	Interventions Citalopram at target doses of 10, 20, or 30 mg/d versus placebo for 6	Power calculation Multiple comparisons for the primary end point compared	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse	Limitations NICE guidelines manual 2012: Appendix C: Methodology	Main outcome classification Depression and anxiety (measured by POMS)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Flynn,K.A., Dyar,M., Johnson,D.B., Atherton,P.J., Diekmann,B., Loprinzi,C.L., Phase III, placebo-controlled trial of three doses of citalopram for the treatment of hot flashes: NCCTG trial N05C9, Journal of Clinical Oncology, 28, 3278-3283, 2010 Ref Id 227654 Country/ies where the study was carried out USA Study type Randomised, double-blind trial Aim of the study To identify effective nonhormonal options for hot flash relief Study dates November 2006 to April 2007 Source of funding Public Health Service grants	20 mg citalopram/placebo: n=56 / n=27 30 mg citalopram/placebo: n=55 / n=28 Evaluable for endpoint: 10 mg citalopram/placebo: n=44 / n=22 20 mg citalopram/placebo: n=44 / n=21 30 mg citalopram/placebo: n=44 / n=21 Characteristics Placebo/10 mg/20mg/30 mg Mean age (SD), years: 56.2 (9)/55.2 (7)/55.8 (9)/55.2 (8) Breast cancer history (%): 31/35/37/35 Current tamoxifen (%): 6/11/9/7 Inclusion criteria Postmenopausal and reported to be bothered with at least 14 hot flashes per week for at least the past month Exclusion criteria Not reported	weeks. Treatment for all participants was titrated to their assigned dose beginning with one tablet (10 mg/placebo) and increasing by one tablet per week (10 mg/placebo) up to their target dose, the largest of which was three tablets (30 mg/placebo) daily.	each of the three active arms with placebo, giving rise to three pairwise comparisons. This led to the adjustment of the P value to .05/3 = .0168. Therefore, each two-sided multiple comparison of the primary end point with 50 patients per treatment group at the end of 6 weeks of treatment had 80% power and 5% type I error rate to detect a difference of 0.82 standard deviations or 1.64 hot flashes per day, 4.10 units of hot flash score or a drop of 29% from the baseline score. This is considered a large effect size and is based on previous data with hot flash trials. Intention to treat Not reported Details Setting Collaborative trial of the North Central Cancer Treatment Group and Mayo Clinic Randomisation	Not reported Psychological symptoms -Anxiety Reported as mean changes in Profile of Mood States tension/anxiety subscale at end point Placebo/10 mg/20 mg/30 mg: 3.3/ 5.8/ 12.9*/ 4.1 * ANOVA P < 0.01, compared with the placebo arm -Depression Reported as mean changes in Profile of Mood States depression/dejection subscale at end point Placebo/10 mg/20 mg/30 mg: -0.1/ 6.0/ 5.2/ 6.5 -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Not reported Musculoskeletal symptoms Not reported Safety outcomes -Discontinuation Not reported -Major adverse events Not reported -Minor adverse events Not reported	checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Unclear A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: Unclear B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Unclear Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for	Main interventions classification SSRI-citalopram Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
			method Not reported Statistical methods Main statistical tests not reported, but measurements used were reported. Anxiety and depression were measured by the Profile of Mood States (POMS) and rated on a 0- to 100-point scale where 0 is as bad as can be and 100 is as good as can be.		missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Unclear D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Butt,D.A., Lock,M., Lewis,J.E., Ross,S., Moineddin,R., Gabapentin for the treatment of menopausal hot flashes: a randomized	Sample size Gabapentin n=99 assigned, n=95 included in intention-to-treat analysis Placebo n=98 assigned, n=98 included in	Interventions Gabapentin 300mg oral capsules or placebo 3 times daily for 4 weeks	Power calculation To accommodate conservative estimates, the reduction in mean hot flash score for the gabapentin group was	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias	Main outcome classification Psychology quality of life-MENQOL psychosocial Musculoskeletal quality of life-MENQOL physical Discontinuation

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
controlled trial, Menopause, 15, 310-318, 2008 Ref Id 227675 Country/ies where the study was carried out Canada Study type Randomised, double-blind, placebo-controlled trial Aim of the study To compare the effectiveness and tolerability of gabapentin with placebo for the treatment of hot flashes in women who enter menopause naturally. Study dates March 2004 to April 2006 Source of funding This study was funded by the Physicians` Services Incorporated Foundation (grant 03-19) and the University of Toronto, Faculty of Medicine Dean`s Fund (New Staff Grant). The gabapentin capsules were donated by Pfizer Inc. Neither funding source nor Pfizer had any role in study design; collection, analysis,	intention-to-treat analysis Characteristics Gabapentin/ placebo Mean age (SD), years: 55.9 (4.7) / 56.5 (4.4) Months since last menstrual period, mean (SD): 70.3 (67.3)/ 82.9 (78.5) Inclusion criteria -Postmenopausal women, defined as those who had experienced natural cessation of menses for 1 year -Between the ages of 45 and 65 years -At least 14 hot flashes per week Exclusion criteria -Use of HT, tamoxifen, raloxifene, SSRIs, SNRIs, or antiseizure medications -Present or planned antineoplastic or radiation therapy -Bilateral oophorectomy -Serum creatinine level greater than the laboratory normal range or creatinine clearance less than 30 mL/minute -Neurologic conditions -Hypothalamic dysfunction -Known		estimated to be 50% compared with the placebo group. Thus, a sample of 100 women in each group was required to detect an absolute 30% difference between groups with 85% power at the 5% significance level, allowing for 10% attrition, based on a similar study. Intention to treat Yes Details Setting Community practices associated with the North Toronto Primary Care Research Network and Greater Toronto area Randomisation method Random permutation schedule created by the study statistician. The drug packages were prepared and randomly assigned off-site by the central research pharmacy, which was not involved in the study design or	Not reported -Depression Not reported -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Reported as mean change in psychosocial MENQOL scores (95% CI) Gabapentin/placebo/ p-value between groups -0.6 (-0.8 to -0.4) / -0.4 (-0.6 to -0.1) / 0.12 Reported as baseline mean psychosocial MENQOL scores (SD) Gabapentin/placebo 3.0 (1.5)/3.1 (1.6) Reported as mean psychosocial MENQOL scores (SD) at week 4 Gabapentin/placebo 2.4 (1.3) / 2.7 (1.6) Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not reported -Quality of life Reported as mean change in physical MENQOL scores (95% CI) Gabapentin/placebo/ p-value between groups -0.7 (-0.9 to -0.4) / -0.3 (-0.5 to -0.2) / 0.03 Reported as baseline mean physical MENQOL scores (SD) Gabapentin/placebo 3.3 (1.4)/3.3 (1.4)	A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-	Minor adverse events-headache Main interventions classification Gabapentin Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
or interpretation of data; or the writing of this report.	hypersensitivity to gabapentin and its components -Inability to complete questionnaires		participant monitoring. The research nurse distributed the drug package to each woman in sequential order at randomization. Statistical methods Summary statistics, means and SDs for continuous measures, and percentages for categorical measures were calculated. For nonnormal continuous measurements, Wilcoxon rank sum or Mann-Whitney tests were used. Chi-square and t tests were used for comparing baseline characteristics and other measures between treatment groups. The secondary outcome of MENQOL change scores was compared between the groups using an unpaired t test for each domain.	Reported as mean physical MENQOL scores (SD) at week 4 Gabapentin/placebo 2.6 (1.2) / 3.0 (1.3) Safety outcomes -Discontinuation Gabapentin n=10 due to adverse events Placebo n=6 due to adverse events -Major adverse events Not reported -Minor adverse events Headache n (%): Gabapentin/placebo/p-value 2 (2)/ 5 (5)/ 0.44	up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information	
Full citation Grady,D., Cohen,B.,	Sample size Randomised/comple	Interventions Daily oral sertraline	Power calculation Total sample size	Results Frequency of hot flushes (including night sweats)	Limitations NICE guidelines	Main outcome classification

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Tice,J., Kristof,M., Olyaie,A., Sawaya,G.F., Ineffectiveness of sertraline for treatment of menopausal hot flushes: a randomized controlled trial, Obstetrics and Gynecology, 109, 823-830, 2007 Ref Id 227740 Country/ies where the study was carried out USA Study type Randomised, blinded, placebo-controlled trial Aim of the study To estimate the effect of the selective serotonin reuptake inhibitor sertraline on hot flush frequency and severity in perimenopausal and postmenopausal women. Study dates Women were screened for eligibility between February 2004 and October 2005 Source of funding Partial funding from Pfizer, rest of funding not reported	ted study Sertraline: 50 / 45 Placebo: 49 / 44 Characteristics Sertraline/ placebo Mean age (SD), year: 50.5 (5.0) / 52.6 (4.2) White (%): 46/ 67.3 African American (%): 38 /14.3 Time since menopause (year, SD): 3.9 (5.2) / 3.1 (3.6) Hysterectomy (%): 16/ 14.3 Bilateral oophorectomy (%): 0 /2 Inclusion criteria -Aged 40-60 -At least 14 hot flushes per week Exclusion criteria -History of breast or ovarian cancer -Depression -Chronic kidney or liver disease -Bipolar affective disorder -Seizures -Known hypersensitivity to sertraline or to SSRI	(50 mg) or identical placebo for 2 weeks. If no substantial side effects were noted, the dose was increased to two tablets daily (100 mg sertraline or placebo) and continued for an additional 4 weeks.	of 100 was calculated to provide 80% power to with two-tailed alpha .05 to detect a between-group difference of 20 percentage points in the percent change in hot flush frequency from baseline to 6 weeks. Intention to treat Yes Details Setting Women's Health Clinical Research Center of the University of California, San Francisco (UCSF) Randomisation method Treatment was assigned by a UCSF pharmacist in randomly permuted blocks of randomly varied size 2 to 4 in a 1:1 ratio within time since last mentrual period strata (1 year or less compared with more than 1 year). Statistical methods Mean percent changes were compared using t	Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported -Depression Not reported -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Reported as SF-36 Quality of Life Scale-Standardised Mental component (mean change at 6 weeks, SD) Score range (worst-best): 0-100 Sertraline / placebo / p-value 0.1 (9.1) / -0.3 (6.3) / .79 Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not reported -Quality of life Reported as SF-36 Quality of Life Scale-Standardised Physical component (mean change at 6 weeks, SD) Score range (worst-best): 0-100 Sertraline / placebo / p-value -2.3 (8.1) / 0.8 (6.4) / .05 Compared to placebo, treatment with sertraline resulted in greater worsening of scores on the Short Form 36 standardised physical component, but this is not statistically significant. Safety outcomes	manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes A3 - Were groups comparable at baseline - No Level of bias: Moderate as analysis adjusted for baseline characteristics B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups	Psychological quality of life-SF 36 Musculoskeletal quality of life-SF 36 Minor adverse events-headache, mood Main interventions classification SSRI-sertraline Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
			tests for primary analysis. For secondary analysis was restricted to sample of women in each group who were at least 80% adherent to treatment as assessed by pill count. Linear regression analyses were conducted to adjust between-group comparisons for baseline variables including age, race, or ethnicity, education, and years since menopause that were imperfectly balanced at baseline.	-Discontinuation Not reported -Major adverse events Not reported -Minor adverse events Sertraline / Placebo / Relative Risk (Sertraline compared to placebo) / p-value Headache n (%): 11 (22) / 11 (22.4) / 0.98 (0.47-2.85) / .96 Mood change n (%): 7 (14) / 4 (8.2) / 1.72 (0.54-5.49) / .3	comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Kim,D.I., Jeong,J.C., Kim,K.H., Rho,J.J., Choi,M.S., Yoon,S.H.,	Sample size Real acupuncture group n=27 Sham acupuncture group n=27	Interventions The real acupuncture group received 11 acupuncture	Power calculation This study was based on the results of a previous study in	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse	Limitations NICE guidelines manual 2012: Appendix C: Methodology	Main outcome classification Quality of life-psychological Quality of life-

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Choi,S.M., Kang,K.W., Ahn,H.Y., Lee,M.S., Acupuncture for hot flushes in perimenopausal and postmenopausal women: a randomised, sham-controlled trial, Acupuncture in Medicine, 29, 249-256, 2011 Ref Id 227776 Country/ies where the study was carried out South Korea Study type Randomised, sham-controlled trial Aim of the study To determine the effect of acupuncture in treating hot flushes in perimenopausal or postmenopausal women. Study dates April 2007 to October 2007 Source of funding Korean Institute of Oriental Medicine	Characteristics Real acupuncture group / Sham acupuncture group / p-value -Age, years, mean (SD): 50.4 (3.2) / 52.5 (3.5) / 0.0255 -Perimenopausal status n: 15 / 9 / 0.1003 -Postmenopausal status n: 12/ 18 / not reported Inclusion criteria -Perimenopausal and postmenopausal women (perimenopausal status defined as ≥3 months of self-reported menstrual irregularity; postmenopausal status was defined as amenorrhea for ≥12 months) with moderate or severe hot flushes -45–60 years of age; desire to receive treatment for hot flushes Exclusion criteria - Total hysterectomy or anticancer treatment due to malignancy -History of cancer within 5 years -Metallic allergy -Hyperthyroidism -Known psychiatric disorders -Any conventional medication (eg,	treatments for 7 weeks, and the control group underwent sham acupuncture on non-acupuncture points during the same period.	2006. The score differences of the hot flush Visual Analogue Scale (ranging 0–100) were 15, and the SDs of the study and control groups were 3.9 and 3.8, respectively. According to this result, 20.4 patients would be required in each group to detect significant differences (p=0.05, power=0.8). Assuming a 20% dropout rate, it was necessary to have at least 27 patients in each group. Intention to treat Yes Details Setting Dongguk University Ilsan Korean Medicine Hospital Randomisation method Random allocation software V.1.0 (Department of Anaesthesia, Isfanhan University of Medical Science) was used to randomise	Not reported Psychological symptoms -Anxiety Not reported -Depression Not reported -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Measured by Menopause Rating Scale-psychological (mean changes and SD at week 7 from baseline) Acupuncture: -3.1 (3.5) Sham: -1.1 (3.1) p= 0.8233, for mean changes of MRS psychological scale between real and sham acupuncture from baseline Measured by Menopause Rating Scale-psychological (mean, SD at baseline) Acupuncture: 8.2 (3.8) Sham: 5.0 (2.7) p= 0.0026, for comparing baseline values of MRS psychological scale between real and sham acupuncture Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not reported -Quality of life Measured by Menopause Rating Scale-somatic(mean changes and SD at week 7 from baseline) Acupuncture: -2.6 (1.9) Sham: -1.3 (2.5)	checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes A3 - Were groups comparable at baseline - Yes, however sham acupuncture group slightly older than the treatment group Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Unclear Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for	musculoskeletal Minor adverse event-bleeding Main interventions classification Acupuncture Sham acupuncture

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
	HRT or SSRIs) for hot flushes within the 8 weeks prior to the study -Medical conditions not appropriate for this study (eg, thromboembolic disease, heart disease, uncontrolled hypertension, diabetes mellitus or vaginal bleeding of unknown origin within 6 months)		patients into two groups. A block size of 4 was used. The allocation of each patient was concealed by placing each random code in an opaque, sealed envelope. Statistical methods For primary and secondary outcomes, the mean intergroup differences from baseline to each time point were assessed by using two-sample t tests or Wilcoxon rank sum tests.	p= 0.2962, for mean changes of MRS somatic scale between real and sham acupuncture from baseline Measured by Menopause Rating Scale-somatic (mean, SD at baseline) Acupuncture: 7.4 (2.6) Sham: 5.7 (2.4) p= 0.0048, for comparing baseline values of MRS somatic scale between real and sham acupuncture Safety outcomes -Discontinuation Not reported -Major adverse events Not reported -Minor adverse events Bleeding n=1 only in sham acupuncture group	dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes, but participants are Korean Intervention: yes Outcomes: yes Indirectness: no	
Full citation Painovich,J.M., Shufelt,C.L., Azziz,R., Yang,Y.,	Sample size N (total enrolled) = 60 N (total completed)=	Interventions -Traditional acupuncture: three treatments per week	Power calculation Mean MENQOL vasomotor domain core was 5.68	Results Frequency of hot flushes (including night sweats) Not reported	Limitations NICE guidelines manual 2012: Appendix C:	Main outcome classification Psychological quality of life

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Goodarzi,M.O., Braunstein,G.D., Karlan,B.Y., Stewart,P.M., Merz,C.N., A pilot randomized, single-blind, placebo-controlled trial of traditional acupuncture for vasomotor symptoms and mechanistic pathways of menopause, Menopause, 19, 54-61, 2012 Ref Id 227850 Country/ies where the study was carried out USA Study type Pilot randomised, single-blind, placebo-controlled trial Aim of the study A pilot study for the feasibility of planning a definitive clinical trial comparing traditional acpuncture (TA) with sham acupuncture (SA) and waiting control (WC) in relieving vasomotor symptoms (VMS), quality of life, and the hypothalamic-pituitary-adrenal axis in perimenopausal and postmenopausal women.	33 TA n = 12 SA n = 12 WC n = 9 Characteristics TA / SA / WA / p Mean age (SD) in years: 57.2±5.2 / 56.8±6.5 / 54.9±6.4 / p=0.43 Mean BMI (SD): 26.9±3.6 / 31.4±4.5 / 31.2±9.8 / p=0.13 Mean alcoholic drinks per week (SD): 2.1±4.5 / 3.6±3.8 / 2.3±2.5 / p=0.15 Mean years (SD) since menopause: 6.1±4.5 / 8.4±5.5 / 5.1±9.9 / p=0.2 Baseline VMS frequency: 8.3±4.4 / 9±3.8 / 9.9±4.6 / p=0.48 Inclusion criteria -Older than 40 with menopause-related VMS -At least 7 hot flushes per day -At least one missed menstrual cycle or spontaneous or medically-induced menopause Exclusion criteria -Concomittant illness with reasonable likelihood of limiting survival to <1 year. -Current substance abuse -Known, suspected or planned	for 12 weeks, 11 front points and 7 back points. Needles were inserted 0.5 - 1.5 inches, adhesive tape holding the plastic tubing in place, manually stimulated and left for 30 minutes. -Sham acupuncture: three treatments per week for 12 weeks, sham points, manipulated without skin penetration and secured with adhesive tape. -Waiting control: received no treatment for 3 months, underwent exit testing and subsequently had the option of 1 month (12 sessions) of complimentary TA.	with a standard deviation 1.3 among all study participants. With a sample size of 72 patients in each group, there would be adequate power (more than 95%) to detect a minimum 15% difference between SA (or TA) and WC groups at the significant level of 0.025. Intention to treat Not reported Details Setting Women who lived within a 5-mile radius and those who had access to the Cedars-Sinai Medical Center intranet. Randomisation method Participants were allocated to one of three study arms with equal probability using a randomized block design after signing the consent form. Appropriate statistical analyses that took the blocking into account were employed.	Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported -Depression Not reported -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Reported as mean (SD) psychosocial MENQOL Baseline TA / SA / WC / p-value: 2.8±1.6 / 3.5±1.8 / 3.2±1.8 / 0.68 Change from baseline at endpoint (12 weeks) TA / SA / WC / p-value: -0.5±1.4 / -0.9±1.7 / 1.0±1.6 / 0.16 Negative change denotes improvement Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not reported -Quality of life Reported as mean (SD) physical MENQOL Baseline TA / SA / WC / p-value: 3.4±1.3 / 3.7±1.3 / 3.9±1.1 / 0.58 Change from baseline at endpoint (12 weeks) TA / SA / WC / p-value: -0.5±1.6 / -1.1±1.4 / 0.3±0.9 / 0.17 Negative change denotes improvement Safety outcomes	Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - No A3 - Were groups comparable at baseline - Yes Level of bias: High B Performance bias B1 - Did groups get same level of care - No B2 - Were participants blinded to treatment allocation- Some B3 - Were individuals administering care blinded to treatment allocation- No Level of bias: High C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of	Musculoskeletal quality of life Main interventions classification Traditional acupuncture Sham acupuncture Waiting list

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Study dates Not stated Source of funding Not stated	pregnancy in next year -Concomittant menopause treatment -Participating in acupuncture treatment or psychological stress management within last year -Participating in another form of VMS treatment -HIV -Hepatitis -Blood-borne illness		Statistical methods Data are presented in tables as means and SD or SE for all continuous variables. Analyses were performed by applying non-parametric statistics. Comparing the demographic and symptom variables at baseline, the Kruskal-Wallis test was employed. Kruskal-Wallis test was applied for comparing the median in the three groups or the Wilcoxon rank sum test for comparing two related groups. All tests of hypotheses were two-sided with Type I error rate of 0.05. A $p < 0.05$ was considered statistically significant.	-Discontinuation Not reported -Major adverse events Not reported -Minor adverse events Not reported	bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - No D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: unclear Other information Subjects are likely to be employees of the centre conducting the study as they either lived close to the centre or could access the intranet and the study does not	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					indicate racial groups of subjects. TA and SA were blinded, however WC knew status and had a higher proportion of drop out due to not receiving acupuncture. The N value was fairly low.	
Full citation Pandya,K.J., Morrow,G.R., Roscoe,J.A., Zhao,H., Hickok,J.T., Pajon,E., Sweeney,T.J., Banerjee,T.K., Flynn,P.J., Gabapentin for hot flashes in 420 women with breast cancer: a randomised double-blind placebo-controlled trial, Lancet, 366, 818-824, 2005 Ref Id 227853 Country/ies where the study was carried out USA Study type Randomised double-blind placebo-controlled trial Aim of the study To assess the efficacy of gabapentin in controlling hot flashes in women	Sample size Placebo n=137 assigned, n=119 at week 4, n=113 at week 8 300 mg gabapentin n=139 assigned, n=123 at week 4, n=114 at week 8 900 mg gabapentin n=144 assigned, n=129 at week 4, n=120 at week 8 Characteristics Placebo / 300 mg gabapentin / 900 mg gabapentin Mean (SD) age, years: 54 (7) / 55 (9) / 55 (9) Currently taking tamoxifen (%): 103 (75) / 95 (68) / 100 (69) Inclusion criteria Aged 18 years or older who had breast cancer and were having an average of two or more hot flashes per day Exclusion criteria -Taking venlafaxine,	Interventions Placebo, gabapentin 100 mg, or gabapentin 300 mg, each to be taken by mouth three times a day, for 8 weeks	Power calculation In authors' previous research on clonidine, the SD of the percentage change from baseline in hot-flash frequency was about 35%. A sample of 114 evaluable participants per group would give 80% power to detect a 15% difference between any pair of groups. To allow for up to 16% dropout by 8 weeks, they planned to enrol 136 participants per group. Intention to treat Yes Details Setting Multicentre clinical trial at 18 geographically diverse member sites of the	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported -Depression Not reported -Cognitive function Reported as patient-report symptom inventory for memory Placebo/ gabapentin 300 mg / gabapentin 900 mg / p-value Change (95% CI) in memory symptoms from baseline to week 4: -0.33 (-0.73 to 0.07) / -0.38 (-0.70 to -0.06) / -0.31 (-0.62 to 0) / 0.209 Change (95% CI) in memory symptoms from baseline to week 8: -0.73 (-1.12 to -0.34) / -0.04 (-0.36 to 0.44) / -0.20 (-0.56 to 0.16) / 0.386 -Sleep disturbance Reported as patient-report symptom inventory for sleep disturbance Placebo/ gabapentin 300 mg / gabapentin 900 mg / p-value Change (95% CI) in sleep symptoms from baseline	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals	Main outcome classification Cognitive function (memory) Sleep disturbance Discontinuation Main interventions classification Placebo Gabapentin 300 mg and 900 mg

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
with breast cancer Study dates Between June 2001 and July 2003 Source of funding US National Cancer Institute	clonidine, or anticonvulsants -Pregnancy -Breastfeeding -Use of steroidal contraception -Coronary insufficiency -Recent history of myocardial infarction, symptomatic cardiac disease, peripheral or cerebrovascular disease, stroke, syncope, or symptomatic hypotension -Hepatic dysfunction (aspartate aminotransferase concentration above twice the upper limit of normal, or bilirubin concentration above the upper limit of normal, as defined at each institution) -Renal dysfunction (serum creatinine concentration above 1.25 times the upper limit of normal) -Known allergy to gabapentin		University of Rochester Community Clinical Oncology Program, New York Randomisation method Treatment assignment was done by use of a randomisation table created in SAS computer program (version 8) and was stratified by the Community Clinical OncologyProgram site and by the duration of hot flashes (<9 months or ≥9 months). A block size of three was used to ensure that the treatment assignment was balanced after every three participants within each stratum. Statistical methods For purposes of comparison, analyses were done on change scores and percentage change scores at week 4 and week 8 separately, by ANCOVA.	to week 4: -0.83 (-1.35 to -0.31) / -1.02 (-1.55 to -0.49) / -1.27 (-1.74 to -0.80) / 0.065 Change (95% CI) in sleep symptoms from baseline to week 8: -1.26 (-1.78 to -0.74) / -1.18 (-1.73 to -0.63) / -1.39 (-1.84 to -0.94) / 0.378 -Quality of life Not reported Musculoskeletal symptoms Not reported Safety outcomes -Discontinuation Due to side effects: -Placebo n=6 by week 4 -300 mg gabapentin n=3 by week 4, n=3 by week 8 -900 mg gabapentin n=8 by week 4, n=2 by week 8 -Major adverse events Not reported -Minor adverse events Not reported	administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Unclear D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation van,Die,M.D., Burger,H.G., Bone,K.M., Cohen,M.M., Teede,H.J., Hypericum perforatum with Vitex agnus-castus in menopausal symptoms: a randomized, controlled trial, Menopause, 16, 156-163, 2009 Ref Id 227916 Country/ies where the study was carried out Australia Study type Double-blind, randomized, placebo-controlled, parallel trial Aim of the study To evaluate the effectiveness of a phytotherapeutic intervention comprising a combination of St John's Wort (Hypercum) and Chaste tree/berry (Vitax) in the management of menopausal symptoms.	Sample size N = 93 total. - St John's Wort and Chaste: N = 50 - Placebo: N = 50 Characteristics Age (yrs): mean (SD) Placebo: 52.5 (3.8) Treatment: 51.9 (4.3) Perimenopausal Placebo: N = 16 Treatment: N = 17 Postmenopausal Placebo: N = 24 Treatment: N = 25 Hysterectomy Placebo: N = 9 Treatment: N = 8 Inclusion criteria - 40 - 60 yrs, postmenipausal or perimenopausal, experiencing a minimum of 5 hot flushes/sweating episones per day and scoring 20 + on Greene Climacteric Scale. - Hysterectomized women over 53 and FSH > 25 IU/L. Exclusion criteria - Using formulations or concomitant	Interventions St John's Wort (H. perforatum) and Chaste tree/berry (V. agnus-castus).	Power calculation Anticipating placebo effect of 30% for hot flush symptoms based on phytotherapeutic menopause RCTs and 30% for depression: calculated sample size of 102 would permit 0.8 power for the detection of moderate effects (d = 0.5), alpha level = 0.05. Intention to treat Yes Details Setting Royal Melbourne Institue of Technology and Jean Hailes Foundation for Women's Health. Randomisation method Computer generated random number table and labeled with code numbers. Statistical methods A mixed model,	Results Greene Climacteric Scale: Anxiety: mean score (SD), 95% CI Placebo Baseline: 6.36 (0.41), 5.59 - 7.14 Endpoint: 3.71 (0.41), 2.90 - 4.52 Mean change: 2.65 (0.57), 1.53 - 3.77 Treatment Baseline: 6.33 (0.39), 5.56 - 7.11 Endpoint: 4.60 (0.41), 3.80 - 5.40 Mean change: 1.73 (0.57), 0.62 - 2.85 - Difference between two groups at enpoint: p = 0.13 Depression Placebo Baseline: 5.12 (0.37), 4.40 - 5.84 Endpoint: 3.02 (0.39), 2.27 - 3.78 Mean change: 2.10 (0.53), 1.05 - 3.77 Treatment Baseline: 5.40 (0.37), 4.68 - 6.12 Endpoint: 3.89 (0.38), 3.15 - 4.64 Mean change: 1.51 (0.52), 0.47 - 2.55 - Difference between groups at endpoint: p = 0.11 Somatic Placebo: Baseline: 4.94 (0.35), 4.26 - 5.62 Endpoint: 2.83 (0.36), 2.12 - 3.54 Mean change: 2.11 (0.50), 1.14 - 3.10 Treatment:	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: low	Main outcome classification Psychological Musculoskeletal Main interventions classification Non pharmocological

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Study dates Not reported. Source of funding - MediHerb Australia Pty Ltd - active and placebo formulations - Australian College of Phytotherapy and Jean Hailes Foundation for Women's Health	therapies for menopausal/psychological symptoms - Pre-existing illness - Medically or surgically induced menopause		treating group as the between subject factor and phase as the within-subject factor.	Baseline: 4.64 (0.35), 3.96 - 5.32 Endpoint: 3.13 (0.36), 2.43 - 3.83 Mean change: 1.51 (0.52), 0.53 - 2.49 - Difference between groups at endpoint: p = 0.55 Sleep: Placebo: Baseline: 1.80 (0.13), 1.55 - 2.05 Endpoint: 1.26 (0.13), 1.00 - 1.52 Mean change: 0.54 (0.18), 0.18 - 0.90 Treatment: Baseline:1.85 (0.13), 1.65 - 2.15 Endpoint: 1.31 (1.13), 1.11 - 1.62 Mean change: 0.54 (0.18), 0.18 - 0.90 - Difference between groups at endpoint: p = 0.59 Hamilton Depression Inventory Placebo Baseline: 14.30 (0.75), 12.83 - 15.77 Endpoint: 8.40 (0.78), 6.87 - 9.93 Mean change: 5.90 (1.08) 3.78 - 8.02 Treatment: Baseline:14.76 (0.75), 13.29 - 16.23 Endpoint: 9.29 (0.77), 7.78 - 10.80 Mean change: 5.47 (1.07), 3.37 - 7.58 - Difference between groups at endpoint: p = 0.42 Utian Quality of Life Scale Placebo Baseline: 77.80 (1.85), 74.15 - 81.45 Endpoint: 77.22 (1.93), 73.41 - 81.02 Mean change: - 0.58 (2.67), -5.86 - 4.69 Treatment: Baseline: 79.04 (1.85), 75.39 - 82.69 Endpoint: 81.15 (1.93), 77.35 - 84.96 Mean change: 2.11 (2.67), -3.16 - 7.38 - Difference between groups at endpoint: p = 0.15	C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Yes Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Unclear D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Yes Level of bias: low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Full citation Yang,H.M., Liao,M.F., Zhu,S.Y., Liao,M.N., Rohdewald,P., A randomised, double-blind, placebo-controlled trial on the effect of Pycnogenol on the climacteric syndrome in peri-menopausal women, Acta Obstetricia et Gynecologica Scandinavica, 86, 978-985, 2007 Ref Id 227932 Country/ies where the study was carried out Taiwan Study type Double-blind, placebo-controlled study Aim of the study Investigae the effects of Pycnogenol on the complex peri-menopausal syndrome Study dates Jan 2002 - July 2005 Source of funding	Sample size N = 200 perimenopausal women Pycnogenol (N = 80) Placebo (N = 75) Characteristics Age (mean + SD) Pycnogenol (N = 80) = 46.73 (5.09) Placebo (N = 75) = 47.02 (4.220 Inclusion criteria - No menopausal cycle for 3 - 11 months but normal cycles appeared again (perimenopausal) - Hormone level FSH > 30 IU and estrogen E2 < 20 pg/l Exclusion criteria - Systematic or acute diseases, hormone therapy, contraceptive medication, hormone substitution, oophrectomy, illiteracy - Hysterectomy	Interventions - Pycnogenol 100 mg	Power calculation Not reported. Intention to treat Not reported. Details Setting Not reported. Randomisation method Not reported. Statistical methods Differences in baseline performance between 2 groups tested with one-way ANOVA. A teo-way ANOVA was performed with peri-menopausal symptom scores.	Results Somatic Problems (WHQ) Pycnogenol (mean (SD) Baseline: 2.61 (0.97) Endpoint: 3.21 (0.41) - p < 0.001 Placebo: Basline: 2.57 (1.00) Endpoint: 2.69 (0.87) - not significant Depressed (WHQ) Pycnogenol Baseline: 2.89 (0.91) Endpoint: 3.29 (0.46) - p < 0.001 Placebo Baseline: 2.91 (0.89) Endpoint: 2.89 (0.89) - not sig Anxiety (WHQ) Pycnogenol Baseline: 2.85 (0.91) Endpoint: 3.27 (0.44) - p < 0.001 Placebo Baseline: 2.91 (0.88) Endpoint: 2.92 (0.88) - not sig Sleep (WHQ) Pycnogenol Baseline: 2.55 (0.88) Endpoint: 3.22 (0.50) - p < 0.001 Placebo Baseline: 2.51 (0.91) Endpoint: 2.56 (0.90) - not sig	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Not reported A2 - Was there adequate concealment - Yes A3 - Were groups comparable at baseline - Yes Level of bias: High B Performance bias B1 - Did groups get same level of care - Unclear B2 - Were participants blinded to treatment allocation- Unclear - only reports that investigator was blinded B3 - Were individuals administering care blinded to treatment allocation- Unclear Level of bias: High C Attrition bias	Main outcome classification - Psychological - Musculoskeletal Main interventions classification non-pharmaceutical

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Yes Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Unclear D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes - WHQ questionnaire D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation	Sample size	Interventions	Power calculation	Results	Limitations	Main outcome

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Yurcheshen,M.E., Guttuso,T.,Jr., McDermott,M., Holloway,R.G., Perlis,M., Effects of gabapentin on sleep in menopausal women with hot flashes as measured by a Pittsburgh Sleep Quality Index factor scoring model, Journal of Women's Health, 18, 1355-1360, 2009 Ref Id 227936 Country/ies where the study was carried out USA Study type Secondary analysis of data from a cohort of menopausal women participating in a randomized, double-blind, placebo-controlled trial Aim of the study To analyze gabapentin's effect on Pittsburgh Sleep Quality Index (PSQI) scores in menopausal women Study dates Not reported Source of funding Not reported	Gabapentin n=30 Placebo n=29 Characteristics Gabapentin/Placebo Age, mean year (SD): 52.7 (3.6)/ 53.0 (3.1) White (%): 93.3%/ 93.1% Daily hot flush frequency, mean (SD): 10.8 (4.1)/ 10.3 (3.7) Duration of amenorrhea, mean months (SD): 67.8 (81.1)/ 44.8 (39.0) Inclusion criteria -Postmenopausal women -Experienced 7-20 daily hot flashes Exclusion criteria Not reported	Gabapentin (escalating to 300mg) or matching placebo three times daily for 12 weeks	Not reported Intention to treat Yes Details Setting Not reported Randomisation method Not reported Statistical methods The PSQI global and factor scores were analysed using a repeated-measures analysis of variance (ANOVA) model that included terms for treatment groups (gabapentin, placebo), week (categorical), and the interaction between treatment group and week.	Frequency of hot flushes (including night sweats) Not reported Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported -Depression Not reported -Cognitive function Not reported -Sleep disturbance Reported as mean PSQI factor scores (SD) Gabapentin/Placebo Baseline sleep quality score: 3.8 (2.1)/ 3.6 (1.9) Mean change from baseline to week 4 / p-value: -1.5 / -0.33 / $p < 0.05$ Mean change from baseline to week 12 / p-value: -1.27 / -0.28 / $p < 0.05$ Baseline sleep efficiency score: 2.5 (1.6)/ 2.4 (1.6) Mean change from baseline to week 4 / p-value: -1.03 / -0.15 / $p < 0.05$ Mean change from baseline to week 12 / p-value: 0.94 / 0.39 / not statistically significant Baseline daily disturbance score: 3.0 (1.0)/ 2.7 (0.9) Mean change from baseline to week 4 / p-value: -0.7 / -0.32 / not statistically significant Mean change from baseline to week 12 / p-value: -0.6 / -0.57 / not statistically significant Negative scores denote improvement -Quality of life Not reported Musculoskeletal symptoms Not reported Safety outcomes	NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Unclear A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Unclear, the study did not use significance tests to determine if differences between two groups' baseline characteristics are statistically significant Level of bias: Unclear B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment	classification Psychological-sleep disturbance Discontinuation Minor adverse events-bleeding Main interventions classification Gabapentin Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
				-Discontinuation Gabapentin: 4 subjects (13.3%), one each because of dizziness, rash, heart palpitations, and peripheral edema Placebo: 1 subject (3.4%) due to diarrhea -Major adverse events Not reported -Minor adverse events Onset of menses was more common in the placebo group (10.3%) than in the gabapentin group (6.7%)	allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Davis,S.R., Briganti,E.M., Chen,R.Q., Dalais,F.S., Bailey,M., Burger,H.G., The effects of Chinese medicinal herbs on postmenopausal vasomotor symptoms of Australian women: A randomised controlled trial, Medical Journal of Australia, 174, 68-71, 2001 Ref Id 255855 Country/ies where the study was carried out Australia Study type Randomised control trial-double blind Aim of the study To evaluate the effects of a defined formula of Chinese medicinal herbs (CMH) on menopausal symptoms (frequency of vasomotor symptoms (VMS). Study dates August 1998 - April 1999 Source of funding The Australian	Sample size N = 78 randomised n = 28 CMH completed n = 27 placebo completed Characteristics Means or percentages at baseline with 95% CI: Placebo / CMH / P Number: 27 / 28 / 0.07 Age: 54.1(52.6, 55.5) / 56.3(54.3,58.3) / 0.75 BMI: 26.1(24.3,27.9) / 25.7(23.9, 27.5) / 0.75 Duration of amenorrhea: 4.6(3, 6.2) / 5.8(3.9, 7.7) / 0.34 Previous use of HRT: 44.4% / 53.6% / 0.50 Previous use of natural therapies: 37% / 35.7% / 0.92 Frequency of hot flushes/night sweats per week: 46.6(35.4,57.8) / 46.2(38.75,53.7) / 0.94 MENQOL vasomotor domain: 4(3.3,4.8) / 3.8(3.1,4.5) / 0.6	Interventions Chinese medicinal herbs (CMH) which included the following formula: Rehmannia glutinosa Cornus officinalis Dioscorea opposita Alisma orientalis Paeonia suffruticosa Poria cocos Citrus reticulata Lycium chinensis Albizzia julibrissin Zizyphus jujuba Elipta prostrata Ligustrum lucidum Placebo Corn starch Placebo with bitter taste Both interventions were granules soluble in 200ml of water taken twice a day, and dispensed every 4 weeks. All packaging was identical. All herbs were listed with the Australian therapeutic Goods Administration, and administered in standard measures. They were screened for heavy metal contamination by two separate agencies.	Power calculation A clinically relevant effect of treatment is considered to be at least a 40% reduction in vasomotor events. Anticipating a 30% placebo response, for power of 80% and a significance level of 5%, a sample size of 28 subjects in each treatment group was required. This sample size was also adequate to determine a clinically relevant change of score of one point in the MENQOL domains. Intention to treat Not reported Details Setting Urban population in Australia recruited through the Jean Hailes Foundation Newsletter, newspapers, radio station interviews and the Medical Unit of the Jean Hailes Foundation Randomisation	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported -Depression Not reported -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life reported as psychosexual domain of MENQOL Mean values (95% CI) Placebo: 3.9 (3.3, 4.6) CMH: 3.6 (3.0, 4.2) P=0.45 Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not reported -Quality of life reported as physical domain of MENQOL Mean values (95% CI) Placebo: 5.6 (4.9, 6.2)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both	Main outcome classification Psychological-quality of life Musculoskeletal-quality of life Minor adverse events Main interventions classification Herbal preparations Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Menopause Society grant. 'Cathay Herbal' of Sydney donated the herbal preparations.	Inclusion criteria Non-Asian women, aged 45 to 70, resident in Australia for at least 10 years. >12 months amenorrhea due to menopause. FSH >25 IU/L >13 hot flushes/night sweats per week. Exclusion criteria Previous use of HRT, CMH or other natural therapies (including over-the-counter and complimentary medicine) >8 weeks pre baseline. Pre-existing gastrointestinal, renal or live disease, diabetes, uncontrolled hypertension, undiagnosed vaginal bleeding, systemic glucocorticosteroid use or cancer therapy. High phytoestrogen diet for 4 weeks pre baseline.		method Subjects were randomised to CMH or placebo using a randomisation chart constructed by randomising numbers 1 to 88 into two groups using Microsoft Excel Statistical method Frequency of hot flushes/night sweats was self-recorded during 4 week baseline period, and during the 12 weeks of study. The trial was powered based on the outcome of vasomotor frequency, with at least 40% reduction in VMS and MENQOL score considered effective. Analysis of variance was used to analyse the effects of treatment within and between groups over the study period. Analysis of covariance determined the effect of baseline characteristics on the average	CMH: 5.5 (5.2, 6.5) P=0.57 Safety outcomes -Discontinuation Not reported -Major adverse events Not reported -Minor adverse events Fifteen women (placebo, 9; CMH, 6) reported headache, joint pain or dizziness. Numbers not reported separately for each adverse event.	groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Unclear Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information Baseline characteristics of	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
			percentage of change in vasomotor symptoms and on the difference in scores for each domain of the MENQOL Questionnaire.		those who withdrew and those who completed the study were similar, except for the previous use of natural therapies for menopausal symptoms, which was more frequent in those who withdrew.	
Full citation Davis,S.R., Moreau,M., Kroll,R., Bouchard,C., Panay,N., Gass,M., Braunstein,G.D., Hirschberg,A.L., Rodenberg,C., Pack,S., Koch,H., Moufarege,A., Studd,J., APHRODITE Study Team., Testosterone for low libido in postmenopausal women not taking estrogen, New England Journal of Medicine, 359, 2005-2017, 2008 Ref Id 255862 Country/ies where the study was carried out UK, US, Canada, Australia, Sweden Study type Double-blind, placebo-controlled RCT Aim of the study To determine the	Sample size N = 814 Characteristics Age Placebo (N = 277): 54.4 ± 5.82 Testosterone 150 ug/Day (N = 267): 54.1 ± 5.37 Testosterone 300 ug/day (N = 267): 54.3 ± 6.53 Hysterectomy Placebo: 119 (43%) Testosterone 150 ug/Day: 117 (43.8%) Testosterone 300 ug/day: 122 (45.7%) Inclusion criteria - Surgical menopausal women: 20 - 70 yrs and postmenopausal for at least 12 months - natural menopause: 40 - 70 yrs and postmenopausal for	Interventions HRT: Testosterone 150 ug/Day, Testosterone 300 ug/day	Power calculation Two-sided, alpha level 0.05 Intention to treat Yes Details Setting 65 centers in US, UK, Canada, Australia, UK & Sweden Randomisation method Unclear Statistical methods ANCOVA adjusted for menopause type. ANOVA used to analyse secondary efficacy endpoints.	Results Baseline No. of satisfying sexual episodes over 4 week period Placebo (N = 277): 2.5 ± 2.7 Testosterone 150 ug/Day (N = 267): 2.9 ± 3.87 Testosterone 300 ug/day (N = 267): 2.5 ± 2.85 Increase in 4 week frequency of satisfying sexual events at week 24 Placebo (N = 265): 0.7 Testosterone 150 ug/Day (N = 252): 1.2 Testosterone 300 ug/day (N = 254): 2.1 (p<0.001) Subgroup with natural menopause: Placebo (N = 196): 0.5 Testosterone 150 ug/Day (N = 187): 1.2 Testosterone 300 ug/day (N = 189): 2.0 (p<0.001) Subgroup with surgically induced menopause: Placebo (N = 69): 1.5 Testosterone 150 ug/Day (N = 65): 1.1 Testosterone 300 ug/day (N = 65): 2.5 Adverse event All Placebo (N = 277): 243 Testosterone 150 ug/Day (N = 267): 225	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Unclear A2 - Was there adequate concealment - Yes A3 - Were groups comparable at baseline - Yes Level of bias: Medium B Performance bias B1 - Did groups get same level of care - unclear B2 - Were participants blinded to treatment	Main outcome classification Sexual Function Main interventions classification HRT: Testosterone patch

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
efficacy and safety of a testosterone patch (Intrinsa, Procter & Gamble Pharmaceuticals) for the treatment of hypoactive sexual desire disorder in women with natural or surgically induced menopause who were not receiving estrogen or estrogen plus progestin. Study dates July 2004 - February 2006 Source of funding Procter & Gamble Pharmaceuticals	at least 2 years Exclusion criteria - Use of systemic estrogen or estrogen plus progestin during previous 3 months (7 months for implantable testosterone)			Testosterone 300 ug/day (N = 267): 234 Serious Breast Cancer Placebo (N = 277): 0 Testosterone 150 ug/Day (N = 267): 1 - Ivasive ductal cancer grade II, diagnosed at 4 mo of treatment Testosterone 300 ug/day (N = 267): 1 - Intermediate - grade ductal carcinoma in situ, diagnosed at 7 month of treatment (patient had bloody nipple discharge before study entry) 1 - Estrogen- receptor-positive invasive breast cancer, diagnosed at 12 month of treatment	allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Yes Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Yes D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Unclear D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Yes Level of bias: Low Indirectness Does the study match the review	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation de Sousa-Munoz,R.L., Filizola,R.G., Efficacy of soy isoflavones for depressive symptoms of the climacteric syndrome, Maturitas, 63, 89-93, 2009 Ref Id 255875 Country/ies where the study was carried out Brazil Study type Placebo-controlled double-blind randomised study Aim of the study To evaluate the efficacy of soy isoflavones extract (SIE) in the treatment of depressive symptoms in women with climacteric syndrome. Study dates Not reported Source of funding Not reported	Sample size Daily dose of 120 mg of soy isoflavones extract (EG=experimental group) n=42 Two daily doses of Placebo made of starch (CG=control group) n=42 Characteristics No baseline characteristics data reported for each treatment group. Only overall characteristics reported. The age of the 84 patients in the sample ranged from 45 to 60 years (85.7% were from 50 to 60 years old), with an average of 53.35 (±3.62) years. Fifty-four women (64.3%) were married and 44 (52.3%) were brown or black, 61 (72.6%) had formal education from primary and complete intermediate levels; 73 (86.9%) belonged to middle-middle class and middle-lower economic classes	Interventions -The experimental group (EG) received the daily dose of 120 mg isoflavones divided into two oral doses of 60 mg -Control group received two daily doses of placebo (starch) The study does not reported how long the partipants took the capsules, however, it can be assumed the treatment was for 16 weeks as the final post-treatment visit was 16 weeks after initial treatment visit. VT1-initial treatment visit at baseline VT2-first follow-up visit eight weeks after the beginning of the treatment VT3-final post-treatment visit 16 weeks after VT1	Power calculation The sample size was calculated on 84 patients, based on the assumption that the treatment of depressive symptoms would be considered effective if the outcome was the reduction of 50% in the pre-treatment scores of a self-evaluation scale of these symptoms, considering a difference of 20% between experimental and control group as relevant, with statistical significance of 5% (p = 0.05) in a hypothesis test and 80% of statistical power. Intention to treat Not reported Details Setting Climacteric Clinic of the Lauro Wanderley University Hospital (HULW), Paraiba University Federal (UFPB),	Results Frequency of hot flushes (including night sweats) Not reported Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported -Depression The CES-D scores in the EG reduced from 12.5 (±4.2) in VT1 to 9.9 (±3.6) in VT2 (VT2 < VT1, p = 0.001) and 8.2 (±3.8) in VT3 (VT3 < VT2, p = 0.007), while the CG, reduced from 13.0 (±4.8) in VT1 to 10.1 (±4.1) in VT2 (VT2 < VT1, p = 0.001) and 9.4 (±4.1) in VT3 (VT2 = VT3, p > 0.05). In the outcome of the 16-week treatment (VT1–VT3), reduction of the CES-D scores did not reach statistical significance between groups. The ANOVA test for repeated measurements showed reduction statistically significant in scores between groups in relation to all evaluations (VT1–VT2–VT3) for measures of depressive symptoms according to CES-D (p = 0.001). -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Not reported Musculoskeletal symptoms Not reported Safety outcomes -Discontinuation In the EG, one patient dropped due to adverse event in the 2nd week (headache). No	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Unclear A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Unclear Level of bias: High B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Unclear B3 - Were individuals administering care blinded to treatment allocation- Unclear Level of bias: High	Main outcome classification Depression-CES-D Minor adverse events-headache Discontinuation Main interventions classification Phytoestrogen (soy isoflavones extract) Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
	and 43 (51.2%) performed no paid activity. EG and CG were homogeneous in relation to the distribution of these socio-demographic variables. Inclusion criteria -Age from 45 to 60 years -One year or more of amenorrhea for non-hysterectomized women -The presence of vasomotor and depression symptoms clinically detectable -Follicle-stimulating hormone (FSH) plasma levels greater than or equal to 25 IU/L -Minimum instruction necessary for understanding the questionnaire -Written agreement in participating in the study Exclusion criteria -Zero scores in the depressive symptoms assessment scale (Depression Scale of Center of Epidemiologic Studies of Depression, CES-D) -Use of psychoactive drugs		Joao Pessoa, Paraiba (PB), Brazil Randomisation method Systematic random allocation with no further details Statistical methods The primary efficacy measure was the comparison of the percentage reduction in the CES-D scores from VT3 between experimental (experimental and control groups) through the Student's t-test for independent samples. The calculation of percentage variation (Δ%) of the CES-D scores between VT1 and VT3 was made, using the following formula Δ% = (score of VT1 − score of VT3)/(score of VT1) × 100, considering the number of patients who completed the 16-week study (per protocol analysis).	discontinuation due to adverse events in the CG. -Major adverse events Not reported -Minor adverse events Reported as frequency of adverse events Headache EG frequency=2 CG frequency=2	C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Unclear Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes, though the study used the Brazilian version of CES-D D4 - Were investigators blinded to intervention - Unclear D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
	during the month before the beginning of the study -Treatment with oestrogens, phytoestrogens and selective synthetic modulators of oestrogen receptors in the six months before the beginning of the study -Diagnosis of gynaecological cancer, intestinal, liver, thyroid and/or renal diseases in activity -Mood disturbances -Ongoing psychotherapy -Use of oral antibiotics in the last two months, regular consumption of alcoholic drinks and exclusive vegetarian food		The comparison of average scores between evaluations in each group was also performed through the analysis of variance (ANOVA) for repeated measures, considering the mean scores obtained in the three visits (VT1, VT2, VT3). The Fisher exact test was used to compare the distribution of categorical variables.		Population: yes Intervention: yes Outcomes: yes Indirectness: some, the study used Brazilian women Other information	
Full citation De,NovaesSoaresC, Almeida,O.P., Joffe,H., Cohen,L.S., Efficacy of estradiol for the treatment of depressive disorders in perimenopausal women: A double-blind, randomized, placebo-controlled trial, Archives of General Psychiatry, 58, 529-534, 2001 Ref Id 255882 Country/ies where the study was carried out	Sample size Oestradiol group n=25 Placebo group n=25 Characteristics Oestradiol / Placebo / p-value Mean age, year (SD): 49.3 (3.8) / 50.3 (3.4) / .34 Duration of amenorrhea, d (SD): 165 (123) / 137 (133) / .44 Major depressive disorder (MDD) n (%): 15 (60) / 11 (44) / .47 Dysthymic disorder	Interventions Transdermal patches of 17β-estradiol (100 µg) or placebo for 12-week	Power calculation Not reported Intention to treat Yes Details Setting Institute of Psychiatry of the University of São Paulo, Brazil Randomisation method The randomisation scheme was externally controlled and based on a list of	Results Frequency of hot flushes (including night sweats) Not reported Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported -Depression Reported as mean Montgomery-Åsberg Depression Rating Scale scores (SD) Oestradiol/Placebo/Oestradiol vs placebo p-value Baseline: 24.6 (6.69) / 21.84 (4.43) / P=0.02 Week 4: 16.04 (4.83) / 18.12 (5.49) / n.s Week 8: 12.32 (4.71) / 17.44 (5.55) / n.s Week 12: 8.6 (5.02)* / 16.34 (6.29)* / P <.01	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at	Main outcome classification Depression - MADRS Discontinuation Minor adverse events-headache, bleeding Main interventions classification Oestrogen (patch)-17β-estradiol (100 µg) Placebo (patch)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Brazil Study type Double-blind, randomized, placebo-controlled trial Aim of the study To investigate the efficacy of 17beta-estradiol for the treatment of clinically significant depressive disorders in endocrinologically confirmed perimenopausal women Study dates Patients recruited between October 1996 and June 1998 Source of funding Grant 96/05105-8 from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)– São Paulo Research Foundation, São Paulo, Brazil.	n (%): 4 (16) / 7 (28) / .47 Minor depressive disorder n (%): 6 (24) / 7 (28) / .47 Inclusion criteria (1) age between 40 and 55 years (2) history of menstrual cycle irregularity or amenorrhea for less than 12 months (3) serum level of FSH greater than 25 IU/L (to document the gonadotropins' attempt to stimulate the declining ovarian function and, therefore, to confirm the perimenopausal status as the cause of menstrual irregularities) (4) diagnoses of MDD, dysthymic disorder, or minor depressive disorder, according to DSM-IV Exclusion criteria -Medical illness (assessed by general practitioners or gynaecologists at the study entry) -Use of hormone replacement therapy and/or psychoactive drugs in the 3 months prior to assessment -Contraindication to oestrogen therapy -Presence of		random numbers generated by computer Statistical methods Frequencies of categorical data were analysed using the Pearson χ^2 test or Fisher exact test, when appropriate. The independent t test (2-tailed) was used for between-group comparisons. A paired t test (2-tailed) was used for within-group comparisons.	*p <0.05 for within-group baseline vs week 12 -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Not reported Musculoskeletal symptoms Not reported Safety outcomes -Discontinuation 2 subjects randomised to placebo patches dropped out of the study due to patch-related skin irritation (n = 1) and nausea (n = 1). One subject treated with oestradiol dropped out because of adverse effects (headaches and nausea). -Major adverse events Not reported -Minor adverse events -Headaches n=1 in oestradiol group -Headaches n=3 (6%) in placebo group -Bleeding was reported by 4 (16%) of 25 subjects receiving oestradiol and by 2 (8%) of 25 subjects receiving placebo, during the treatment phase (12 weeks)	baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Unclear B3 - Were individuals administering care blinded to treatment allocation- Unclear Level of bias: Unclear C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
	psychotic features, suicidality, or severe aggressive behavior				method used to assess outcome - Unclear D4 - Were investigators blinded to intervention - Unclear D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: some, as this study used Brazilian women	
Full citation Frisk,J., Kallstrom,A.C., Wall,N., Fredrikson,M., Hammar,M., Acupuncture improves health-related quality-of-life (HRQoL) and sleep in women with breast cancer and hot flushes, Supportive Care in Cancer, 20, 715-724, 2012 Ref Id 256049 Country/ies where the study was	Sample size Electro-acupuncture (EA) n = 27 randomised, 26 analysed Hormone therapy (HT) n = 18 randomised and analysed Characteristics EA/HT/p-value Mean age (years), range: 54.1 (47-69) / 53.4 (43-67) / not significant Ongoing tamoxifen (yes/no):	Interventions -Electro-acupuncture treatment given by physiotherapist for 12 weeks -Hormone therapy group was treated with sequential or continuous combined oestrogen/progestagen therapy for 24 months	Power calculation Not reported Intention to treat Not reported Details Setting Three participating centres in southeast Sweden for an international, multi centre prospective study (HABITS) Randomisation method Computer generated	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported -Depression Not reported -Cognitive function Not reported -Sleep disturbance Reported as median times woken up/night (IQR 25th-75th pct): p-value based on pair-wise	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at	Main outcome classification Sleep- times woken up/night and WHQ sleep score Main interventions classification Acupuncture Oestrogen combined with progestogen

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
carried out Sweden Study type Multi-centre, randomised, prospective study Aim of the study Evaluate effects of electro-acupuncture (EA) and hormone therapy (HT) on health-related quality-of-life (HRQoL) and sleep in breast cancer survivors with vasomotor symptoms. Study dates Between 1998 and 2002 Source of funding Medical Research Council of South-East of Sweden, The Swedish Medical Research Council, and The County Council of Ostergotland	6/20 / 4/14 / not significant Inclusion criteria -Completed treatment for breast cancer in situ, T1 or T2 tumours with maximum four metastatic lymph nodes, T3 tumours without metastasis and vasomotor symptoms needing treatment according to the woman -Vasomotor symptoms Exclusion criteria -Ongoing treatment for breast cancer other than tamoxifen/torimefen, other malignancies, heredity or history of thromboembolic, cerebrovascular or liver disease, or porphyria and active cardiovascular disease		randomisation at the University of Uppsala and stratified for participating centre, previous HT use and ongoing treatment with tamoxifen Statistical methods Changes were analysed within and between both groups using the analysis of variance (ANOVA) for repeated measures and the Wilcoxon's signed rank-sum test was used for paired comparisons within each group	comparisons with baseline -EA group Baseline: 3.4 (2.3-4.3) 3 months: 2.0 (1-3): 0.01 6 months: 1.6 (0.8-2.9): 0.003 9 months: 1.6 (1.0-2.7): 0.03 12 months: 1.5 (1-2): 0.003 18 months: 1.4 (0.75-3.2): 0.03 24 months: 1.2 (1.2-1.3): 0.03 -HT group Baseline: 2.3 (0.8-3.0) 3 months: 1.3 (0.9-1.6): 0.01 6 months: 1.1 (0.3-1.6): 0.003 9 months: 1.2 (0.6-1.9): 0.02 12 months: 1.2 (0.5-1.5): 0.01 18 months: 0.9 (0.3-2.0): 0.01 24 months: 1.0 (0.3-1.4): 0.01 Reported as median WHQ sleep score (IQR 25th-75th pct): p-value based on pair-wise comparisons with baseline -EA group Baseline: 0.5 (0-0.75) 3 months: 0.33 (0-0.67): 0.05 6 months: 0.67 (0-0.67): 0.04 9 months: 0.33 (0-0.67): 0.01 12 months: 0 (0-0.67): 0.03 18 months: 0.33 (0.08-0.67): 0.14 24 months: 0.33 (0-0.33): 0.02 -HT group Baseline: 0.33 (0-0.67) 3 months: 0 (0-0.33): 0.01 6 months: 0 (0-0.33): 0.02 9 months: 0.16 (0-0.33): 0.07 12 months: 0 (0-0.5): 0.07 18 months: 0 (0-0.67): 0.65 24 months: 0 (0-0.67): 1.00	baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - No, different length of treatment B2 - Were participants blinded to treatment allocation- No B3 - Were individuals administering care blinded to treatment allocation- No Level of bias: High C Attrition bias C1 - Was follow-up equal for both groups - No C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
				-Quality of life Not reported Musculoskeletal symptoms Not reported Safety outcomes -Discontinuation Not reported -Major adverse events Not reported -Minor adverse events Not reported	assess outcome - Unclear D4 - Were investigators blinded to intervention - Unclear D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Guttuso,Jr, Kurlan,R., McDermott,M.P., Kieburtz,K., Gabapentin's effects on hot flashes in postmenopausal women: A randomized controlled trial, Obstetrics and Gynecology, 101, 337-345, 2003 Ref Id 256163 Country/ies where the study was carried out USA Study type Randomised, double-blind,	Sample size Gabapentin n=30 assigned and analysed Placebo n=29 assigned and analysed Characteristics Gabapentin / Placebo Mean age, year (SD): 52.7 (3.6) / 53 (3.1) Surgical menopause, n (%): 8 (26.7) / 6 (20.7) Inclusion criteria -An average of seven or more hot flashes per day accompanied by sweating -At least one	Interventions Gabapentin 900 mg per day or identically appearing placebo for 12 weeks	Power calculation Given the study's inclusion criterion of 7–20 hot flashes per day, the authors assumed a mean daily hot flash frequency at baseline of approximately 12 in each group. They also estimated a standard deviation of the change from baseline to 12 weeks in daily hot flash frequency of 4. Under these assumptions, a sample size of 22	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Reported as mean (SD) Profile of Mood States Tension/Anxiety Subscale Gabapentin / Placebo Baseline: 10.1 (8.1) / 8.1 (6.0) Absolute change from baseline to week 12 Gabapentin/Placebo/Treatment effect (gabapentin-placebo) / 95% CI / P -3.9 (6.4)/ -2.2 (3.5) / 0.0 / (-3.0, 2.0) / .77 Decreased value indicates improvement in this measure -Depression	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance	Main outcome classification Anxiety-Profile of Mood States Tension/Anxiety Subscale Quality of life-psychological-SF-36 Quality of life-musculoskeletal-SF-36 Discontinuation Minor adverse events-bleeding Main interventions classification Gabapentin Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
placebo-controlled trial Aim of the study To evaluate whether treatment with the anticonvulsant gabapentin may be effective in reducing hot flash frequency and severity. Study dates From July 2000 to March 2001 Source of funding General Clinical Research Center grant, 5 M01 RR00044 from the National Center for Research Resources, National Institutes of Health (NIH); an Experimental Therapeutics in Neurological Disease NIH Grant #5 T32 NS07338-12; and University of Rochester institutional research funds	daytime hot flash per day -Amenorrhea for more than 12 months or amenorrhea for 6–12 months with a serum follicle-stimulating hormone level greater than 40 mIU/mL and oestrogen less than 20 pg/mL or status post-bilateral oophorectomy for 2 months -An estimated creatinine clearance of 60 or more mL per minute -No oestrogen, progestin, leuprolide, or tamoxifen therapy within the past 2 months -No change in dose of raloxifene, clonidine, or any antidepressant therapy within the past month and no plan to change the dose in the future -No calcium channel antagonist or gabapentin therapy within the past 2 weeks -No previous allergic reaction to gabapentin Exclusion criteria -More than 50% of a patient's hot flashes associated with occurrence of		subjects per group was chosen to provide 90% power to detect a 33% reduction (from 12 to 8) in mean daily hot flash frequency with gabapentin, using a two-tailed t test at the 5% level of significance. Since some subjects would not complete the trial, they increased the sample size to 30 subjects per group (60 total). Intention to treat Yes Details Setting General Clinical Research Center at Strong Memorial Hospital, Rochester, New York Randomisation method The Office of Investigational Drug Services in the Department of Pharmacy at the University of Rochester prepared all study capsules and performed the randomisation via a random number table. The	Not reported -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Reported as mean (SD) SF-36 Mental Health Component Summary Gabapentin / Placebo Baseline: 49.4 (12.4) / 50.7 (11.2) Absolute change from baseline to week 12 Gabapentin/Placebo/Treatment effect (gabapentin-placebo) / 95% CI / P 4.4 (10.2)/ 2.2 (6.8) / 1.2 / (-1.7, 5.3) / .41 *Study does not report how to interpret SF-36 so an online search found higher SF-36 scores indicate less disability Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not reported -Quality of life Reported as mean (SD) SF-36 Physical Health Component Summary Gabapentin / Placebo Baseline: 49.2 (10.2) / 52.7 (6.6) Absolute change from baseline to week 12 Gabapentin/Placebo/Treatment effect (gabapentin-placebo) / 95% CI / P -1.1 (3.7)/ -0.3 (5.6) / -0.6 / (-3.0, 1.7) / .42 *Study does not report how to interpret SF-36 so an online search found higher SF-36 scores indicate less disability Safety outcomes -Discontinuation Reported as withdrawals due to adverse events	bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Unclear Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
	migraine headaches or ingestion of particular foods or beverages		randomisation was stratified by surgical menopause status. Statistical methods The Wilcoxon rank sum test was used to compare the treatment groups regarding all outcomes, except a $\chi 2$ test was used to compare the percentages of patients having a greater than 50% reduction in hot flash composite score from baseline to Week 12. Treatment effects were estimated using the Hodges–Lehmann estimate of the group difference in population medians and its associated 95% confidence interval.	Gabapentin n=4 Placebo n=1 -Major adverse events Not reported -Minor adverse events Reported as number of patients with onset of menses Gabapentin n=2 (6.7%) Placebo n=3 (10.3%)	intervention - Unclear D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information	
Full citation Kimmick,G.G., Lovato,J., McQuellon,R., Robinson,E., Muss,H.B., Randomized, double-blind, placebo-controlled, crossover study of sertraline (Zoloft) for	Sample size Sertraline n=33 assigned, 25 analysed Placebo n=29 assigned, 22 analysed Characteristics Placebo/Sertraline Median age, years (range): 52.3 (41.1-	Interventions 6 weeks of sertraline (50 mg each morning) versus placebo	Power calculation A targeted acrrual of 62 women with hot flashes provided at least 90% power to detect a 50% difference in the proportion of women still experiencing hot	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate	Main outcome classification Depression-CESD Discontinuation Minor adverse events-headache, anxiety Main interventions classification SSRI-sertraline Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
the treatment of hot flashes in women with early stage breast cancer taking tamoxifen, Breast Journal, 12, 114-122, 2006 Ref Id 256418 Country/ies where the study was carried out USA Study type Randomized, double-blind, placebo-controlled, crossover study Aim of the study To assess the effect of sertraline on the frequency and severity of hot flashes, mood status, and health-related quality of life Study dates Between October 1996 and June 2000 Source of funding Pfizer Pharmaceuticals	77.1) / 56.7 (36.6-77.0) Inclusion criteria -Aged 18 and older with localised breast cancer and receiving adjuvant tamoxifen therapy -Had at least one hot flash per day Exclusion criteria -Pregnant or breast-feeding -History of seizure disorder or hepatic or renal insufficiency -Concurrent or planned therapy with oestrogen, progestational agents, corticosteroids, androgens, or other anti-depressant therapy		flashes at 6 weeks (90% versus 45%) Intention to treat Yes Details Setting Wake Forest University School of Medicine Randomisation method Randomly assigned, in a double-blind fashion Statistical methods T-tests were used to compare treatment groups on mean daily hot flash frequency, mean hot flash score, and quality of life measures	-Depression Reported as CESD mean (SD) Placebo / sertraline / p Baseline: 11.5 (7.9) / 11.2 (9.2) / 0.49 6 weeks: 9.4 (7.4) / 8.9 (8.3) / 0.68 -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Not reported Musculoskeletal symptoms Not reported Safety outcomes -Discontinuation Reported as withdrawal by week 6 due to adverse events Sertraline n=3 Placebo n =2 -Major adverse events Not reported -Minor adverse events Reported as number of patients Headache: Placebo n=1 Sertraline n=1 Anxiety/nervousness: Placebo n=0 Sertraline n=3	randomisation - Unclear A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: Unclear B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- No B3 - Were individuals administering care blinded to treatment allocation- No Level of bias: High C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Unclear D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information No wash-out period reported	
Full citation Mann,E., Smith,M.J., Hellier,J., Balabanovic,J.A., Hamed,H., Grunfeld,E.A., Hunter,M.S., Cognitive behavioural treatment for women who have menopausal symptoms after	Sample size Usual care n=49 randomised, 45 analysed CBT n=47 randomised, 43 analysed Characteristics CBT / usual care Mean age, year (SD): 53.16 (8.10) / 54.05 (7.76) Time since breast	Interventions -Usual care-followed up every 6 months by an oncologist or clinical nurse specialist, with additional appointments as needed. Additionally, those treated in UK National Health Service hospitals in	Power calculation A sample size of 96 women was needed to provide 90% power to detect a two-point difference (SD 2.4; standardised effect size 0.8) in mean HFNS problem rating for the comparison of CBT to usual care	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Reported as WHQ anxiety or fears (higher scores indicate poorer wellbeing) CBT (mean, SD) / Usual care (mean, SD) / Adjusted mean difference (SE) /95% CI	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes	Main outcome classification Anxiety-WHQ anxiety or fears Depression-WHQ depressed mood Cognitive function-WHQ memory and concentration Sleep disturbance-WHQ sleep problems Quality of life-psychological- SF-36

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
breast cancer treatment (MENOS 1): a randomised controlled trial, Lancet Oncology, 13, 309-318, 2012 Ref Id 256621 Country/ies where the study was carried out UK Study type Randomised controlled trial Aim of the study Whether cognitive behavioural therapy (CBT) can help breast cancer survivors to effectively manage hot flushes and night sweats (HFNS) Study dates Between March 2009 to March 2011 Source of funding Cancer Research UK	cancer diagnosis, months, mean (SD): 47.75 (53.38) / 31.08 (30.63) Inclusion criteria -At least ten problematic HFNS per week (confirmed by a 2-week diary and a screening interview) for a duration of 2 months or more -Had completed medical treatment for breast cancer (surgery, radiotherapy, or chemotherapy), and had no evidence of other cancers or metastases -Women taking adjuvant endocrine treatment were eligible Exclusion criteria -Unable to attend sessions or who were seeking treatment for mood disorders rather than for HFNS were not eligible	southeast London were offered telephone support as part of the cancer survivorship programme. Women were sent an information leaflet produced by Breast Cancer Care and offered telephoned support every 2 weeks (average seven telephone calls, maximum ten). Nurses gave information about HFNS, advised on treatment options and practical ways of symptom management, and offered instructions for paced breathing and relaxation. -Group CBT comprised one 90 minute session a week for 6 weeks, and included psycho-education, paced breathing, and cognitive and behavioural strategies to manage HFNS. All participants received usual care—they had access to clinical specialists and cancer support services, either through routine follow-up appointments or as	at 9 weeks after randomisation. Intention to treat Analyses were based on modified intention-to-treat sample (excluding those who contributed no data) Details Setting Breast or oncology clinics in southeast London, UK Randomisation method Randomisation was done in blocks of 12–20 participants, allocating participants in a one-to-one ratio, stratifying by age (younger than 50 years, 50 years or older), and was done with a computer-generated sequence. Statistical methods Secondary outcomes were analysed wit h mixed linear regression models with random participant and cohort group intercepts and a time-by-treatment	Baseline: 0.34 (0.25) / 0.45 (0.30) / - / - 9 weeks: 0.23 (0.27) /0.40 (0.33)/-0.12 (0.06)* / -0.24 to -0.01 26 weeks:0.24 (0.31)/ 0.39 (0.31) / -0.10 (0.06)/ -0.21 to 0.01 *p<0.05 -Depression Reported as WHQ depressed mood (higher scores indicate poorer wellbeing) CBT (mean, SD) / Usual care (mean, SD) / Adjusted mean difference (SE) /95% CI Baseline: 0.23 (0.26)/ 0.31 (0.27)/ - / - 9 weeks: 0.13 (0.16)/0.28 (0.24)/-0.14 (0.05)*/ -0.23 to -0.06 26 weeks:0.13 (0.19)/0.28 (0.26)/-0.13 (0.05)*/-0.22 to -0.05 * p< 0.01 -Cognitive function Reported as WHQ memory and concentration (higher scores indicate poorer wellbeing) CBT (mean, SD) / Usual care (mean, SD) / Adjusted mean difference (SE) /95% CI Baseline: 0.75 (0.34) / 0.72 (0.36)/ - / - 9 weeks: 0.59 (0.36)/0.70 (0.32)/-0.14 (0.06)*/ -0.27 to -0.02 26 weeks: 0.51 (0.37)/0.62 (0.36)/-0.14 (0.06)*/-0.26 to -0.02 * p< 0.05 -Sleep disturbance Reported as WHQ sleep problems (higher scores indicate poorer wellbeing) CBT (mean, SD) / Usual care (mean, SD) / Adjusted mean difference (SE) /95% CI Baseline: 0.63 (0.30)/ 0.72 (0.29)/-/- 9 weeks: 0.37 (0.31)/ 0.65 (0.32)/ -0.26 (0.07)**/ -0.39 to -0.12 26 weeks: 0.43 (0.37)/ 0.61 (0.34)/ -0.16 (0.07)*/ -0.29 to -0.02 **p<0.0001 * p< 0.05	A2 - Was there adequate concealment - No A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- No B3 - Were individuals administering care blinded to treatment allocation- No Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes	mental health Symptom relief-SF-36 bodily pain Quality of life-musculoskeletal-WHQ somatic symptoms, SF-36 physical functioning, SF-36 physical role limitation Main interventions classification Cognitive behavioural therpy Usual care

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
		part of a breast cancer survivorship programme in southeast London.	interaction term; covariates in the model were treatment group, baseline value of outcome, the stratification factor age, and time. Results from all analyses were summarised at 9 weeks and 26 weeks with two-sided 95% CIs	-Quality of life Reported as SF-36 mental health, a higher score indicates better health CBT (mean, SD) / Usual care (mean, SD) / Adjusted mean difference (SE) /95% CI Baseline: 67.57 (17.89)/ 62.52 (17.37)/-/- 9 weeks: 74.63 (14.22)/ 66.46 (14.20)/ 6.03 (2.95)*/0.24 to 11.81 26 weeks: 70.70 (19.24)/ 64.5 (16.06)/3.86 (2.96)/ -1.94 to 9.65 * p< 0.05 Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Reported as SF-36 bodily pain, a higher score indicates better health CBT (mean, SD) / Usual care (mean, SD) / Adjusted mean difference (SE) /95% CI Baseline: 46.15 (22.73)/52.99 (21.64)/-/- 9 weeks: 53.68 (23.98)/52.16 (22.57)/ 6.35 (4.20)/-1.89 to 14.59 26 weeks: 51.00 (22.50)/46.58 (22.18)/ 9.85 (4.20)*/1.61 to 18.09 * p< 0.05 -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not reported -Quality of life Reported as WHQ somatic symptoms (higher scores indicate poorer wellbeing) CBT (mean, SD) / Usual care (mean, SD) / Adjusted mean difference (SE) /95% CI Baseline: 0.56 (0.26)/0.55 (0.25)/-/- 9 weeks: 0.44 (0.24)/0.46 (0.24)/-0.08 (0.06)/-0.21 to 0.04 26 weeks: 0.45 (0.23)/0.53 (0.23)/-0.03 (0.06)/-0.16 to 0.09 Reported as SF-36 physical functioning, a higher	D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Unclear D5 - Were investigators blinded to confounding factors - No Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
				score indicates better health CBT (mean, SD) / Usual care (mean, SD) / Adjusted mean difference (SE) /95% CI Baseline: 66.17 (22.89)/ 74.89 (22.27)/-/- 9 weeks: 75.38 (24.24)/79.23 (21.96)/4.76 (3.47)/-2.03 to 11.56 26 weeks: 74.13 (24.96)/73.88 (27.37)/8.86 (3.46)*/2.09 to 15.64 * p< 0.05 Reported as SF-36 physical role limitation, a higher score indicates better health CBT (mean, SD) / Usual care (mean, SD) / Adjusted mean difference (SE) /95% CI Baseline: 53.72 (43.29)/49.46 (40.31)/-/- 9 weeks: 60.00 (40.35)/60.90 (39.65)/-1.09 (8.14)/-17.03 to 14.85 26 weeks:55.77 (43.10)/51.92 (44.20)/2.63 (8.17)/-13.39 to 18.65 Safety outcomes -Discontinuation Not reported -Major adverse events Not reported -Minor adverse events Not reported		
Full citation Morrison,M.F., Kallan,M.J., Ten,Have T., Katz,I., Tweedy,K., Battistini,M., Lack of efficacy of estradiol for depression in postmenopausal women: a randomized, controlled trial, Biological Psychiatry, 55, 406-412, 2004 Ref Id	Sample size After 2 weeks of single-blind placebo treatment in 87 patients, 57 were randomly assigned to receive 8 weeks of treatment with oestradiol (.1 mg/day; n = 31) or placebo (n = 26). Characteristics Age, mean (SD) 61.8 (9.4) Placebo: 62.8 (9.5)	Interventions 8 weeks of treatment with estradiol (.1 mg/day) or placebo. All patients were then treated with medroxyprogesterone 10 mg/day for 2 weeks combined with the study patch.	Power calculation Not reported Intention to treat Not reported Details Setting Outpatient clinic of the Hospital of the University of Pennsylvania Randomisation method A study pharmacist, who was not an investigator,	Results Frequency of hot flushes (including night sweats) Not reported Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported -Depression Reported as Hamilton Depression Rating Scale Estradiol, baseline, mean (SD): 14.5 (2.6) Estradiol change from baseline at 8 weeks (95%	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Unclear A2 - Was there adequate concealment -	Main outcome classification Depression Discontinuation Minor adverse events-bleeding Main interventions classification Oestrogen (patch) Placebo (patch)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
256749 Country/ies where the study was carried out USA Study type Double-blind randomised, placebo-controlled trial Aim of the study Whether oestrogen therapy is effective in treating depressive disorders in older postmenopausal women and to determine whether progestins are associated with a deterioration of mood Study dates 1996-1999 Source of funding National Institute of Mental Health. Berlex provided study patches without charge.	Time since last mentrual periods, years (SD) Oestradiol: 16.6 (10.9) Placebo: 17.7 (13.0) Natural menopause (%) Oestradiol: 51.6 Placebo: 65.4 Inclusion criteria -50-90 years of age -postmenopausal at least 1 year with follicular stimulating hormone ≥ 40 mIU/mL for those within 5 years of menopause -Score ≥10 on the Center for Epidemiologic Studies Depression Scale and 8-20 on the Hamilton Depression Scale -Meet DSM-IV criteria for major depression, dysthymia, or minor depression Exclusion criteria -Use of hormonal medications within 3 months -Medical conditions that rendered a patient ineligible for oestrogen therapy -Structural disease of the central nervous system -Cognitive		randomly assigned subjects to 8 weeks of double-blind treatment with either 0.1mg/day estradiol skin patch or a placebo patch. Statistical methods Mixed effects piecewise linear regression was used to evaluate treatment effects. Baseline variables were compared using means with student's t-test or Pearson chi-square test.	CI): -2.8 (-4.5, -1.1), p=0.002 Placebo, baseline, mean (SD): 14.5 (3.1) Placebo change from baseline at 8 weeks (95% CI): -5.2 (-6.8, -3.5), p<0.001 Difference between estradiol and placebo at 8 weeks (95% CI): 2.4 (0, 4.7), p=0.05 Reported as Center for Epidemiological Studies Depression Scale Estradiol, baseline, mean (SD): 27.0 (8.8) Estradiol change from baseline at 8 weeks (95% CI): -3.5 (-6.0, -.9), p=0.01 Placebo, baseline, mean (SD): 29.8 (11.1) Placebo change from baseline at 8 weeks (95% CI): -5.9 (-8.4, -3.3), p<0.001 Difference between estradiol and placebo at 8 weeks (95% CI): 2.4 (-1.2, 6.0), p=0.19 -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Not reported Musculoskeletal symptoms Not reported Safety outcomes -Discontinuation 1 withdrew in estradiol group due to breast tenderness 1 withdrew in placebo group to seek conventional depression treatment -Major adverse events Not reported -Minor adverse events 4 women in oestradiol group developed bleeding after a mean of 4.75 weeks on oestradiol.	Unclear A3 - Were groups comparable at baseline - No Level of bias: High B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
	imparment as defined by a score of < 24 on the Mini-Mental Status Exam -Treatment for depression in previous 3 months -Alcohol or drug abuse or dependence during the previous 6 months -Serious medical problems resulting in a high probability of death within a year -Schizophrenia, bipoloar disorder or early-onset dysthymic disorder -Inability to comprehend English				and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: some Other information Populations in the oestradiol group had more African American than Caucasian (51.6% versus 41.9%), whereas placebo group is roughly the same (42.3% versus 46.1%). Greater proportions of people in placebo group had major depressive dsorder (past and current), and greater proportions in estradiol group	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					had minor depressive disorder.	
Full citation Nathorst-Boos,J., Floter,A., Jarkander-Rolff,M., Carlstrom,K., Schoultz,Bv, Treatment with percutanous testosterone gel in postmenopausal women with decreased libido--effects on sexuality and psychological general well-being, Maturitas, 53, 11-18, 2006 Ref Id 254534 Country/ies where the study was carried out Sweden Study type Double blind, randomised, crossover design Aim of the study To elucidate if percutanous treatment with 10mg testosterone per day could enhance sexuality and psychological well-being in postmenopausal women presenting problems with low libido Study dates Not reported Source of funding Swedish research	Sample size Testosterone n=30 allocated, 3 discontinued Placebo n=30 allocated, 4 discontinued Characteristics Women characteristics are reported as a whole rather than per treatment group. Mean ± S.D. age, weight and BMI for the 53 women completing the study were 55.4 ± 3.5 years, 65.4 ± 7.8 kg and 23.6 ± 2.8 kg/m2 Inclusion criteria -Between 50 and 65 years of age and complaining of total loss or significant decrease of libido during the postmenopausal period Exclusion criteria -Women who had experienced libido problems already before the menopause	Interventions As a complement to their already on-going HRT (combined oestrogen and progesterone), 10 mg of a testosterone gel (Testogel, Besins–Iscovesco) or placebo was administered to the subjects. Treatment continued for three months before cross over.	Power calculation Not reported Intention to treat Not reported Details Setting Karolinska Hospital, Sweden Randomisation method Randomisation was performed in blocks of eight and the code was kept in the local hospital pharmacy Statistical methods Differences in scores from baseline were compared among groups. Differences between the biological variables were examined by ANOVA.	Results Frequency of hot flushes (including night sweats) Not reported Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Reported as median value of Psychological general well being (PGWB) score- anxiety Placebo/ Testosterone/ p-value 24/ 27 / <0.001 -Depression Reported as median value of Psychological general well being (PGWB) score- depressed mood Placebo/ Testosterone/ p-value 15 /16 / 0.382 -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Not reported Musculoskeletal symptoms Not reported Safety outcomes -Discontinuation Not reported -Major adverse events Not reported -Minor adverse events Not reported separately	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Unclear, study did not report baseline characteristics per group Level of bias: Unclear B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes	Main outcome classification Anxiety (PGWB) Depression (PGWB) Main interventions classification Testosterone Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
council, the Karolinska Institute and Basins-Iscovesco					Level of bias: low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Unclear D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					Intervention: yes Outcomes: yes Indirectness: no	
Full citation Nijland,E.A., Weijmar Schultz,W.C., Nathorst-Boos,J., Helmond,F.A., van Lunsen,R.H., Palacios,S., Norman,R.J., Mulder,R.J., Davis,S.R., LISA,study investigators, Tibolone and transdermal E2/NETA for the treatment of female sexual dysfunction in naturally menopausal women: results of a randomized active-controlled trial, Journal of Sexual Medicine, 5, 646-656, 2008 Ref Id 254554 Country/ies where the study was carried out 6 European countries, US and Australia Study type RCT: Multicenter, double blind, randomized, clinical trial Aim of the study To compare the efficacy on sexual function of tribolone 2.5mg to continuous	Sample size N = 403 Tibolone N=199 Transdermal E2/NETA N=201 Characteristics Age Total mean = 56 yrs Transdermal E2/NETA = 55.8 yrs (n= 201) Tibolone = 55.8 yrs (N= 199) BMI Transdermal E2/NETA = 24.7 Tibolone = 25.0 Gynaecological surgery: Transdermal: 19% Tibolone: 18% Inclusion criteria - Aged between 48 - 68 years - Undergone natural menopause, had intact uterus - Reported that prior to menopause, their sex life was satisfying but since menopause they experienced decline in satisfaction with sexual activity that was associated with personal distress as measured by Female Sexual Distress Scale (FSDS ≥ 15).	Interventions - E2 (50 ug)/NETA (140 ug) in the form of a twice weekly patch plus a daily placebo tablet - Tribolone 2.5 mg as a daily tablet with a twice weekly placebo patch.	Power calculation Assumed a two-sided test, at the 0.05 alpha level, it was estimate that a maximum of 286 subjects would be required to provide 80% power to detect a standardized difference in treatment effect of 20% on the composite score (CS) of the Female Sexual Function Index (FSFI) between both groups. Intention to treat Yes Details Setting 29 study centers in 6 European countries, the US and Australia. Randomisation method Eligible women allocated in a 1:1 ratio using a computerized automatic interactive voice response system to treatment with either E2 ug)/NETA (140 ug) Allocation concealment and blinding	Results Reported as total sexual events in the 4-week frequency measured by a daily diary Tibolone (N=137) Baseline mean: 5.7 Mean change from baseline: 0.66 % change from baseline: 12% E2/NETA Baseline mean: 5.6 Mean change from baseline: 0.75 % change from baseline: 13% Within group p=0.02 Between group p= not significant Total satisfying sexual events Tibolone Baseline: 3.3 Mean change from baseline: 1.44 % change from baseline: 44% E2/NETA Baseline: 3.1 Mean change from baseline: 1.48 % change from baseline: 48% Within group p<0.001 Between group p= not significant Discontinuation Discontinued due to adverse events E2/NETA: n=41 Tibolone: n=23 Major adverse events Not reported Minor adverse events: Reported as vaginal hemorrhage Tibolone n=0 E2/NETA: n=22	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: Moderate B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-	Main outcome classification Altered sexual function Discontinuation Minor adverse events-bleeding Main interventions classification HRT: Tibolone vs combined oestrogen/progestero ne (estradiol/noresthister one acetate -NETA)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
combined transdermal estradion (E2)/norethisterone acetate (NETA) (50 ug/140 ug) in naturally postmenopausal women with sexual dysfunction. Study dates June 2004 - November 2005 Source of funding Not stated.	Exclusion criteria - Women who had other conditionsthat could have an impact on sexual function, including dyspareunia. - Were taking medication known to affect sexual function such as antidepressents, narcotics and antipsychotics. - Had a history or presense of liver or renal disease, breast cancer or estrogen dependent tumours, CVD, cerebrovascular disease or thromboembolic events or major gynaecologic surgery in the preceeding 3 months. - Previous unsuccessful use of testosterone/testosterone combinations or compounds known to enhance androgenic activity such as Tibolone, DHEA or transdermal estrogen-norethistorone therapy.		Not clear. Reported: "the investigators, study site personnel and participants remained blinded until after the database was locked". Statistical methods T-test. If the assumption for normality were violated, the Wilcoxon rank sum test. Sexual function assessed at baseline, week 12, and 24.		up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Unclear D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Polisseni,A.F.,	Sample size N = 174	Interventions - 2.5 mg Tribolone	Power calculation Sample size	Results Overall QoL (Women's Health Questionnaire):	Limitations NICE guidelines	Main outcome classification

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Andrade,A.T., Ribeiro,L.C., Castro,I.Q., Brandao,M., Polisseni,F., Guerra,Mde O., Effects of a continuous-combined regimen of low-dose hormone therapy (oestradiol and norethindrone acetate) and tibolone on the quality of life in symptomatic postmenopausal women: a double-blind, randomised study, Maturitas, 74, 172-178, 2013 Ref Id 254689 Country/ies where the study was carried out Brazil Study type Prospective, randomised, double-blind, compartive trial (RCT) Aim of the study To compare the effects of a combined, continuous, low-dose hormone therapy (LD-HT) with the effects of tibolone and a control group on the QoL of in the symptomatic postmenopausal women. Study dates June 2009 - June 2011	Characteristics Age (yrs) Tibolone (N = 42): 51.24 ± 3.48 E2 + NETA (N = 44): 52.98 ± 3.39 Control (Ca + Vit D3) (N = 44): 53.18 ± 4.06 Inclusion criteria - Between 45 - 60, postmenopausal with moderate - pronounced VSM symptoms & Blatt-Kupperman Menopausal index (BKMI) equal to or greater than 20 Menopause characterised by the absence of menstruation for at least 12 months & confirmed by increase of FSH Exclusion criteria - Outside age range - Had no or mild VSM symptoms, used HRT, herbal, isoflavone therapy or soy-based foods in last 6 months - Underwent surgery for breast cancer or had any comorbities	- 1mg ostradiol + 0.5 mg norethindrone acetate - Control: 50 mg Calcium carbonate + 200 UI vitamine D3	calculated using GraphPad StateMate version2. Parameters: alpha: 5%, beta = 20% (80% power) Intention to treat Not reported. Details Setting University Hospital of Federal University of Juiz de Fora, Minas Gerais, Brazil Randomisation method Computer generated list of random numbers used to allocate participants to group Statistical methods Wilcoxon signed-rank test assessed the significance of overall QoL in each domainfor each group. Comparisons between groups at all times for overall QoL for each domain were performed using Kruskal-Wallis test.	Baseline Tibolone (N = 42): 80.12 ± 14.04 E2 + NETA (N = 44): 77.73 ± 15.32 Control (Ca + Vit D3) (N = 44): 77.45 ± 15.42 Follow-up Tibolone (N = 42): 57.00 ± 15.50 - p<0.05 compared to baseline E2 + NETA (N = 44): 55.70 ± 16.67 - p<0.05 compared to baseline Control (Ca + Vit D3) (N = 44): 58.39 ± 12.6 - p<0.05 compared to baseline Qol - Depressed mood (WHQ) Baseline Tibolone (N = 42): 15.52 ± 4.46 E2 + NETA (N = 44): 15.16 ± 4.99 Control (Ca + Vit D3) (N = 44): 14.89 ± 5.49 Follow-up Tibolone (N = 42): 11.40 ± 3.83 - p<0.05 compared to baseline E2 + NETA (N = 44): 11.39 ± 4.81 - p<0.05 compared to baseline Control (Ca + Vit D3) (N = 44): 11.82 ± 4.66 - p<0.05 compared to baseline Somatic Symptoms (WHQ) Baseline Tibolone (N = 42): 18.17 ± 4.12 E2 + NETA (N = 44): 17.23 ± 4.61 Control (Ca + Vit D3) (N = 44): 17.36 ± 4.51 Follow-up Tibolone (N = 42): 14.33 ± 5.03 - p<0.05 compared to baseline E2 + NETA (N = 44): 12.70 ± 3.91 - p<0.05 compared to baseline Control (Ca + Vit D3) (N = 44): 13.41 ± 3.51 - p<0.05 compared to baseline QoL - Anxiety (WHQ) Baseline Tibolone (N = 42): 10.05 ± 2.95 E2 + NETA (N = 44): 8.82 ± 3.27 Control (Ca + Vit D3) (N = 44): 8.68 ± 3.00	manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes A3 - Were groups comparable at baseline - Yes Level of bias: low B Performance bias B1 - Did groups get same level of care - Unclear B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes - only pharmacist handlingg capsules knew contents Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups	Psychological outcomes Musculoskeletal symptoms Main interventions classification HRT

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Source of funding Cavalieri Dispensing Chemists Ltd				Follow-up Tibolone (N = 42): 6.76 ± 2.53 - p<0.05 compared to baseline E2 + NETA (N = 44): 6.66 ± 2.95 - p<0.05 compared to baseline Control (Ca + Vit D3) (N = 44): 6.70 ± 2.55 - p<0.05 compared to baseline Sleep problems (WHQ) Baseline Tibolone (N = 42): 8.05 ± 1.96 E2 + NETA (N = 44):7.95 ± 2.15 Control (Ca + Vit D3) (N = 44): 7.52 ± 2.04 Follow-up Tibolone (N = 42):5.83 ± 1.79 - p<0.05 compared to baseline E2 + NETA (N = 44): 5.91 ± 2.13- p<0.05 compared to baseline Control (Ca + Vit D3) (N = 44): 5.84 ± 1.93 - p<0.05 compared to baseline Baseline Tibolone (N = 42): 18.17 ± 4.12 E2 + NETA (N = 44): 17.23 ± 4.61 Control (Ca + Vit D3) (N = 44): 17.36 ± 4.51 Follow-up Tibolone (N = 42): 14.33 ± 5.03 - p<0.05 compared to baseline E2 + NETA (N = 44): 12.70 ± 3.91 - p<0.05 compared to baseline Control (Ca + Vit D3) (N = 44): 13.41 ± 3.51 - p<0.05 compared to baseline	comparable for dropout - Yes C3 - Were groups comparable for missing data - Yes Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Unclear D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes (WHQ) D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: - participants had to have 'moderate VSM' symptoms - BKMI = 20 or more)	
Full citation Qu,F., Cai,X., Gu,Y., Zhou,J., Zhang,R.,	Sample size N = 47 (total): GNL: N = 21	Interventions - GNL (200ml, oral) - control - Livial	Power calculation - Not reported Intention to treat	Results HAMD scores	Limitations NICE guidelines	Main outcome classification Psychological

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Burrows,E., Huang,H., Chinese medicinal herbs in relieving perimenopausal depression: a randomized, controlled trial, Journal of Alternative and Complementary Medicine, 15, 93-100, 2009 Ref Id 254731 Country/ies where the study was carried out China Study type RCT Aim of the study To explore the effects of GengNianLe (GNL, also called perimenopausal relieving formula), a defined formulaof Chinese medicinal herbs in relieving perimenopausal depression in Chinese women. Study dates Sept 2004 - April 2004 Source of funding National Natural Science Foundation of China	Control (tibolone): N = 26 Characteristics Age: GNL: 48.7 + 8.1 Control: 50.4 + 26 Duration of perimenopausal depression (months): GNL: 2.6 + 0.7 Control: 2.9 + 1.0 Inclusion criteria - Aged 40 - 60 with at least 6 consecutive months of amenorrhea with serum estradiol level < 20 pg/mL and FSH > 40 mIU/mL - minimum of 1 month of low mood, total HAMD score > 20 Exclusion criteria - Hormonal medication within past 3 months - medical conditions / contraindications	(Tibolone)	- Not reported Details Setting Zheijang University Randomisation methods Microsoft Excel randomised numbers into 2 groups Statistical analysis Mann Whitney tests used to analyse the inter and intra group differences of HAMD cores.	Depressed mood GNL: Baseline: 3.4 + 1.2 Post-treatment: 1.9 + 0.5 $p < 0.05$ compared to baseline Control: Baseline: 3.8 + 1.2 Post-treatment: 2.2 + 0.6 $p < 0.05$ compared to baseline Anxiety (Psychological) GNL Baseline: 3.3 + 1.3 Post-treatment: 2.3 + 0.5 $p < 0.05$ compared to baseline Control: Baseline: 3.2 + 0.7 Post-treatment: 2.5 + 0.5 $p < 0.05$ compared to baseline Anxiety (somatic) GNL Baseline: 3.9 + 0.9 Post-treatment: 3.3 + 0.6 $p < 0.05$ compared to baseline Control: Baseline: 3.7 + 1.0 Post-treatment: 3.5 + 0.5 - not significant	manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes A3 - Were groups comparable at baseline - Yes Level of bias: low B Performance bias B1 - Did groups get same level of care - Unclear B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Unclear Level of bias: low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups	Main interventions classification Non - pharmaceutical

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					comparable for missing data - Yes Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Unclear D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes (HAMD - validated) D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Simon,J., Braunstein,G., Nachtigall,L., Utian,W., Katz,M., Miller,S., Waldbaum,A., Bouchard,C., Derzko,C., Buch,A., Rodenberg,C.,	Sample size Placebo n=279 Testosterone n=283 Characteristics Women aged 26-70 years with hypoactive sexual desire disorder after bilateral salpingo-oophorectomy who	Interventions Testosterone (300 mcg/d) or placebo patches applied twice weekly for 24 weeks	Power calculation 230 patients/arm were estimated to be necessary to provide approximately 90% power to detect a difference between treatment groups	Results Frequency of hot flushes (including night sweats) Not reported Frequency of sexual intercourse Reported as mean frequency (SE) of total satisfying sexual activity over a 4 week period at 24 week, using a weekly diary, the sexual activity log (SAL) Placebo/Testosterone/Treatment difference (95% CI) / p	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there	Main outcome classification Sexual function Discontinuation Adverse events-headache Main interventions classification Testosterone Placebo

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Lucas,J., Davis,S., Testosterone patch increases sexual activity and desire in surgically menopausal women with hypoactive sexual desire disorder, Journal of Clinical Endocrinology and Metabolism, 90, 5226-5233, 2005 Ref Id 254964 Country/ies where the study was carried out USA, Canada, Australia Study type RCT Aim of the study Evaluate the efficacy and safety of a testosterone patch in surgically menopausal women with hypoactive sexual desire disorder (HSDD) Study dates Not reported Source of funding Procter & Gamble Pharmaceuticals, Inc.	were receiving concomitant oestrogen therapy. All women were in a stable, monogamous relationship with a partner who was sexually functional. Placebo / Testosterone Mean age (SD): 48.9 (7.4) / 49.2 (7.7) Mean time since oophorectomy (year): 8.2 (6.6) / 8.7 (7.0) Inclusion criteria 20-70 year of age, in good health, have a normal mammogram if age 40 year or older, have a normal Pap smear, have undergone bilateral salpingo-oophorectomy and hysterectomy at least 6 months before screening, and have no physical impediment to sexual function. Need to report having a satisfying sex life before oophorectomy and a meaningful loss of sexual desire and decrease in sexual activity after surgery and being bothered or concerned about this decrease in desire for sexual		of 0.34 satisfying sexual activities/week. Intention to treat Yes, with all patients who received at least one application of study medication included in the analyses. A last observation carried forward approach was used to account for patients who did not complete the study. Details Setting Multi-centre study in the US, Canada, and Australia Randomisation method All women were receiving a stable dose of oestrogen therapy (oral or transdermal patch) for at least 3 months before screening. Women were stratified by route of concomitant oestrogen therapy(transdermal or oral) and were then randomly assigned in a 1:1 ratio to receive placebo or 300 mcg testosterone	Baseline: 2.94 (0.19)/ 2.82 (0.15) / -0.12 (-0.60, 0.36) / 0.615 Value at wk 24: 3.93 (0.27) / 4.92 (0.30) / 0.99 (0.20, 1.79) / 0.015 Change from baseline: 0.98 (0.19) / 2.10 (0.25) / 1.11 (0.5, 1.73) / 0.0003 Reported as mean frequency (SE) of total sexual activity over a 4 week period at 24 week, using a weekly diary, the sexual activity log (SAL) Placebo/Testosterone/Treatment difference (95% CI) / p Baseline: 4.94 (0.28)/ 4.98 (0.24) / 0.04 (-0.69, 0.78) / 0.906 Value at wk 24: 5.39 (0.33) / 6.27 (0.33) / 0.88 (-0.04, 1.81) / 0.0602 Change from baseline: 0.45 (0.19) / 1.29 (0.23) / 0.84 (0.25, 1.43) / 0.0036 Psychological symptoms -Anxiety Not reported -Depression Not reported -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Not reported Musculoskeletal symptoms Not reported Safety outcomes -Discontinuation Patients who withdrew from study due to adverse events 19 in placebo, 24 in testosterone -Major adverse events Not reported -Minor adverse events Headache events Placebo n=21	appropriate randomisation - Yes A2 - Was there adequate concealment - Not reported A3 - Were groups comparable at baseline - Yes Level of bias: Moderate B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
	activity. Exclusion criteria Other conditions that could impact sexual function, including dysparenuia; major life change interfering with sexual function; a psychiatric disorder, including depression; or drug or alcohol dependency, or were taking medications known to affect sexual function, including androgens, phytoestrogens, selective serotonin reuptake inhibitors, systemic beta-blockers, raloxifene, tamoxifen, and sildenafil; had a history of breast cancer or oestrogen-dependent neoplasia, active gall bladder disease, diabetes, history of cerebrovascular disease or thromboembolic disorders, or abnormal levels of TSH, serum creatinine, or liver enzymes.		daily for 24 weeks in the form of a twice weekly patch worn on the abdomen. Patients and all study personnel were blinded to treatment assignments. Statistical methods All hypothesis tests were two-sided, and treatment differences were assessed at the 0.05 significance level. The primary efficacy end point was the change from baseline in the 4-wk frequency of total satisfying episodes during week 21–24. Treatment groups were compared using an analysis of covariance model, adjusting for route of administration of concomitant oestrogen therapy, baseline rate of activity, age, and pooled centre.	Testosterone n=28	D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Unclear D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information	
Full citation Soares,C.N., Thase,M.E., Clayton,A., Guico-	Sample size N = 607 Acute Desvenlafaxine: 224	Interventions SNRI: desvenlafaxine 100-200 mg/day	Power calculation Alpha level 5%, power of approx 90% = min of 250	Results HAM-D (MMRM analysis) Raw change from baseline, mean (SD) Desvenlafaxine (N = 110): -18.82 (5.51)	Limitations NICE guidelines manual 2012:	Main outcome classification Psychological Main interventions

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Pabia,C.J., Focht,K., Jiang,Q., Kornstein,S.G., Ninan,P., Kane,C.P., Cohen,L.S., Desvenlafaxine and escitalopram for the treatment of postmenopausal women with major depressive disorder, Menopause, 17, 700-711, 2010 Ref Id 255000 Country/ies where the study was carried out Argentina, Chile, Columbia, Mexico and US Study type Randomised, double-blind Aim of the study To assess the efficacy, safety and tolerability of the serotonin-norepinephrine reuptake inhibitor desvenlafaxine and the SSRI escitalopram for major depressive disorder (MDD) in postmenopausal women. Study dates Dec 2006 - Sept 2008 Source of funding Wyeth Research, acquired by Pfizer Inc	Escitalopram: 237 Continuation Phase Desvenlafaxine: 137 Escitalopram: 160 Characteristics Age Acute Desvenlafaxine: 56 (6) Escitalopram: 56 (6) Continuation Phase Desvenlafaxine: 56 (6) Escitalopram: 56 (6) Inclusion criteria - Postmenopausal, between 40 - 70 yrs with primary diagnosis of MDD - Depressive symptoms for at least 30 days before screening vidit and MADRS total score of 22 or higher Exclusion criteria - Ever previously received treatment or had known hypersensitivity to vanlafaxine, citapram, escitalopram - Had significant risk of suicide	SSRI: excitalopram 10-20 mg/d	women Intention to treat Yes Details Setting 72 centers Randomisation Method Wyeth's computerised randomisation and assignment system (CORE) Statistical analysis ANOVA, Mixed effects model for repeated measures (MMRM) analysis, Last observation carried forward (LOCF).	Escitalopram (N = 124): -17.88 (4.96) Difference in adjusted mean (95% CI) -0.70 (-1.82 - 0.43) p = 0.224 HAM-D (LOCF analysis) Raw change from baseline, mean (SD) Desvenlafaxine (N = 137): -16.44 (6.65) Escitalopram (N = 160): -15.68 (6.30) Difference in adjusted mean (95% CI) -0.48 (-1.79 - 0.83) p = 0.474 HAM-A (MMRM analysis) Raw change from baseline, mean (SD) Desvenlafaxine (N = 110): -15.10 (7.86) Escitalopram (N = 124): -15.02 (6.46) Difference in adjusted mean (95% CI) -0.35 (-1.51 - 0.81) p = 0.549 MADRS (MMRM analysis) Raw change from baseline, mean (SD) Desvenlafaxine (N = 110): -26.65 (6.29) Escitalopram (N = 124): -25.56 (6.32) Difference in adjusted mean (95% CI) -1.10 (-2.59 - 0.39) p = 0.333	Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - No - continuation phase had both blind and open-label A3 - Were groups comparable at baseline - Yes Level of bias: Medium B Performance bias B1 - Did groups get same level of care - Unclear B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- No Level of bias: High C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for	classification Non-hormonal pharmacological (SSRI & SNRI) non-hormonal pharmaceutical treatments

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					dropout - Yes C3 - Were groups comparable for missing data - Yes Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Unclear D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - No - continuation phase open label and blinded D5 - Were investigators blinded to confounding factors - Unclear Level of bias: High Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Uebelhack,R., Blohmer,J.U., Graubaum,H.J., Busch,R., Gruenwald,J.,	Sample size N = 301 (total) Treatment (Black Cohosh): 151 Placebo: 143	Interventions - Black Cohosh 1 mg triterpene glycosides and St John's Wort extract (0.25 mg total	Power calculation Not reported. Intention to treat Yes Details Setting	Results HAMD Treatment (N = 151) Baseline: 18.9 + 2.2 Endpoint: 11.0 + 3.8	Limitations NICE guidelines manual 2012: Appendix C: Methodology	Main outcome classification Psychological Main interventions classification Non - pharmaceutical

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Wernecke,K.D., Black cohosh and St. John's wort for climacteric complaints: a randomized trial, Obstetrics and Gynecology, 107, 247-255, 2006 Ref Id 255137 Country/ies where the study was carried out Germany Study type Double-blind, randomised placebo controlled Aim of the study To investigate the efficacy of the fixed combination of black cohosh and St John's wort extracts inwomen with climacteric complaints with a pronounced psychological component Study dates Oct 2003 - June 2004 Source of funding Schaper & Brummer GmbH & Co KG, Germany	Characteristics Mean Age (yrs) Treatment: 52.4 + 4.5 Placebo: 51.9 + 4.0 Number of gynaecological surgeries: Hysterectomy/unilateral oohorectomy/others Treatment: 25/9/49 Placebo: 21/14/59 Time since last menses (months) Trearment: 88 (9.5%) > 12 months Placebo: 97 (67.3%) > 12 months Inclusion criteria - 45 - 60 yrs, experiences climacteric complaints with pronounced psychological component for at least 3 months, left untreated for at least 2 months - HAMD total score 15 - 23 points Exclusion criteria - Treatment with hormones, nonhormonal climacteric drugs or any other treatment - Psychological therapy / therapy or depressive symptoms - Contraindications	hypericine) - Placebo 2 tablets orally twice per day (week 1 - 8) and 1 tablet orally twice per day (weeks 9 - 16)	Not reported Randomisation method Medication prenumbered using a 1:1 randomisation withblock size of 4. Statistical methods Mann-Whitney U test	Change from baseline: -7.9 + 4.0 p < 0.001 Placebo (N = 143) Baseline: 18.9 + 2.1 Endpoint: 16.5 + 4.3 Change from baseline: -2.4 + 4.3 p < 0.001 Adverse events (any) Treatment: 35 (23.2 %) Placebo: 32 (21.3%) - no discontinuation due to adverse events	checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Not reported A3 - Were groups comparable at baseline - Yes Level of bias: Medium B Performance bias B1 - Did groups get same level of care - Unclear B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Yes Level of bias: Low	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					D Detection bias D1 - Was follow-up appropriate length - Unclear D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes - HAMD scores D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Veerus,P., Hovi,S.L., Sevon,T., Hunter,M., Hemminki,E., The effect of hormone therapy on women's quality of life in the first year of the Estonian Postmenopausal Hormone Therapy trial, BMC Research Notes, 5, 176-, 2012	Sample size N = 1395 Non-HT arm (placebo and non-treatment arms): N = 673 HT arm (blind and non-blind HT arms): N = 686 N = 1395:	Interventions - 0.625 mg CEE (regardless of hysterectomy status) + 2.5 mg MPA or: - 0.625 mg CEE and 5 mg MPA if they were within 3 years from their last period	Power calculation Not reported. Intention to treat Yes Details Setting Clinical centres in Estonia Randomisation method Not reported	Results WHQ scale Depressed mood (mean (SE)) Non-HT: 0.22 (0.01) HT: 0.21 (0.01) Between group p-value*: 0.308 Between group p-value**: 0.539 Anxiety/fear (mean (SE))	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation -	Main outcome classification Psychological Main interventions classification HRT

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Ref Id 255171 Country/ies where the study was carried out Estonia Study type Randomised (both blind and open label) Randomised (both blind and open label) Aim of the study To analyse the impact of the HT on different aspects of symptom experience on QOL during a randomised trial. Study dates 1999 - 2001 Source of funding Academy of Finland, STAKES and Estonian Ministry of Education and Research	Non-HT arm (placebo and non-treatment arms): N = 673 HT arm (blind and non-blind HT arms): N = 686 Characteristics Mean Age (yrs) Non-HT: 60.1 (4.0) HT: 59.5 (4.0) Inclusion criteria - Aged 50 - 64 - Estonian speaking in 2 areas (Tallinn and Tartu) Exclusion criteria Not reported.		Statistical method Between group significants: t-test, Chi squared, Wilcoxon rank test Setting Clinical centres in Estonia Randomisation method Not reported Statistical method Between group significants: t-test, Chi squared, Wilcoxon rank test	Non-HT: 0.27 (0.01) HT: 0.27 (0.01) Between group p-value*: 0.519 Between group p-value**: 0.642 Sleep problems (mean (SE)) Non-HT: 0.39 (0.01) HT: 0.34 (0.01) Between group p-value*: 0.005 Between group p-value**: 0.005 * = Wilcoxon rank sum test ** = t-test	Not reported A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: High B Performance bias B1 - Did groups get same level of care - Unclear B2 - Were participants blinded to treatment allocation- No - some arms open label B3 - Were individuals administering care blinded to treatment allocation- No Level of bias: High C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Yes Level of bias: Low D Detection bias D1 - Was follow-	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
					up appropriate length - Unclear D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes - WHQ D4 - Were investigators blinded to intervention - No D5 - Were investigators blinded to confounding factors - Unclear Level of bias: High Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Wang,C.C., Cheng,K.F., Lo,W.M., Law,C., Li,L., Leung,P.C., Chung,T.K., Haines,C.J., A randomized, double-blind, multiple-dose escalation study of a Chinese herbal medicine preparation (Dang Gui Buxue Tang) for moderate to severe	Sample size 1.5g/day DBT n =20 randomised, 17 analysed 3.0g/day DBT n =20 randomised, 19 analysed 6.0g/day DBT n =20 randomised, 16 analysed Characteristics 1.5g / 3.0g / 6.0g / p-value Mean age, year (SD): 51.79 (3.73) /	Interventions Chinese herbal medicine preparation, Dang Gui Buxue Tang (DBT) given orally daily at 1.5, 3.0, or 6.0 g/day for 12 weeks	Power calculation A sample size of 20 per dose group was calculated to provide 80% power at the 5% significance level, with an anticipated mean difference (SD) of 10.3 (15.1), to show the difference in menopausal symptoms	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Not reported -Depression Not reported -Cognitive function Not reported	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate	Main outcome classification Quality of life-psychological: GCS, MENQOL Quality of life-musculoskeletal: GCS, MENQOL Discontinuation Main interventions classification Herbal preparations-Chinese herbal preparations in 3 different dosages

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
menopausal symptoms and quality of life in postmenopausal women, Menopause, 20, 223-231, 2013 Ref Id 255207 Country/ies where the study was carried out Hong Kong Study type A randomized, double-blind, multiple-dose escalation study Aim of the study To investigate the dose-response relationship of a Chinese herbal medicine preparation, Dang Gui Buxue Tang (DBT), with short-term menopausal symptoms and quality of life in local postmenopausal women Study dates Not reported Source of funding Area of Excellence Grant of the University Grants Committee in Hong Kong	51.84 (3.54) / 52.07 (3.16) / 0.96 Mean years since menopause (SD): 2.42 (1.03) / 3.99 (1.79) / 2.85 (1.71) / 0.439 Inclusion criteria -At least 3 moderate to severe hot flashes per day or at least 21 moderate or severe hot flashes per week -Amenorrhea for at least 12 months -Serum follicle-stimulating hormone concentrations higher than 18 IU/L -Luteinzing hormone concentrations higher than 12.6 IU/L -17 beta-oestradiol concentrations lower than 361 pmol/L at screening Exclusion criteria -Usage of any Chinese medicine, herbal medicinal products, or hormone therapy before the study -Serious underlying medical disorders or undiagnosed vaginal bleeding		between DBT and placebo from baseline to week 12, as shown in the authors' phase I clinical trial. Intention to treat Yes Details Setting Chinese University of Hong Kong Randomisation method Each participant was randomised and allocated to one of three dose groups according to a computer-generated randomisation code list in a 1:1:1 ratio using a block size of six. The DBT preparations were prepared and packed in capsule form and provided in an envelope with the randomisation code. The randomisation code was not broken for anyone during the study. Statistical methods Only those participants who completed all the visits and measurements	-Sleep disturbance Not reported -Quality of life Reported as mean Greene Climacteric Scale-Psychological (SD) 1.5g / 3.0g / 6.0g / p-value for difference between dose groups Baseline (1 to 4 weeks before intervention): 0.13 (1.11) / 0.13 (1.37) / 0.12 (0.94) / 0.06 0th week: 0.12 (1.11) / 0.14 (1.33) / 0.13 (0.90) / 0.086 4th week: 0.15 (1.00) / 0.15 (1.12)*^ / 0.11 (0.63)*^ / 0.046 12th week: 0.09 (0.89)* / 0.17 (1.23)^ / 0.10 (0.61)*^ / 0.006 Reported as mean MENQOL-Psychosocial scores (SD) 1.5g / 3.0g / 6.0g / p-value for difference between dose groups Baseline (1 to 4 weeks before intervention): 2.65 (1.00) / 3.34 (1.06) / 2.52 (1.15) / 0.061 0th week: 2.53 (1.06) / 3.37 (1.29) / 2.50 (1.07) / 0.051 4th week: 2.55 (0.97) / 3.02 (1.33)*^ / 1.84 (1.01)*^ / 0.021 12th week: 2.32 (0.75) / 2.93 (1.11)* / 2.04 (1.24) / 0.046 *p< 0.05 compared with baseline ^ p< 0.05 compared with other doses Reduction in scores indicate improvement Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported	concealment - Yes A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear C3 - Were groups comparable for missing data - Unclear Level of bias: Unclear D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
			were included for analysis. Repeated-measures ANOVA was performed to test the significant dose x time effects of DBT on quality of life scores. Paired t test was used to analyse within-group differences.	-Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) The study reported Greene somatic scale as quality of life-see below -Quality of life Reported as mean Greene Climacteric Scale-Somatic (SD) 1.5g / 3.0g / 6.0g / p-value for difference between dose groups Baseline (1 to 4 weeks before intervention): 0.14 (0.96) / 0.15 (1.20) / 0.12 (0.92) / 0.281 0th week: 0.13 (1.05) / 0.16 (1.23) / 0.13 (0.95) / 0.376 4th week: 0.13 (0.92) / 0.14 (1.04) / 0.10 (0.63)* / 0.067 12th week: 0.11 (0.90) / 0.16 (1.10) / 0.11 (0.68)* / 0.092 Reported as mean MENQOL-Physical scores (SD) 1.5g / 3.0g / 6.0g / p-value for difference between dose groups Baseline (1 to 4 weeks before intervention): 3.05 (0.84) / 3.60 (0.89) / 2.85 (0.84) / 0.365 0th week: 2.92 (0.95) / 3.68 (0.99)^ / 2.84 (0.79)^ / 0.015 4th week: 2.76 (1.06) / 3.29 (1.17)^ / 3.21 (0.46)*^ / 0.046 12th week: 2.84 (1.04) / 3.19 (0.94)*^ / 2.06 (0.98)*^ / 0.005 *p< 0.05 compared with baseline ^ p< 0.05 compared with other doses Reduction in scores indicate improvement Safety outcomes -Discontinuation Reported as discontinuation due to treatment-emergent adverse event 1.5g n=1 at week 4 6.0g n=1 at week 0 -Major adverse events Not reported	and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: some, the study used Chinese women Other information No placebo control was included in the study	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
				-Minor adverse events Not reported		
Full citation Xia,Y., Zhao,Y., Ren,M., Zhang,J., Wang,Y., Chang,Y., Fu,S., Fan,G., Zhu,Y., Huang,Y., Gao,X., A randomized double-blind placebo-controlled trial of a Chinese herbal medicine preparation (Jiawei Qing'e Fang) for hot flashes and quality of life in perimenopausal women, Menopause, 19, 234-244, 2012 Ref Id 255270 Country/ies where the study was carried out China Study type Randomised, double-blind placebo-controlled RCT Aim of the study To evaluate the effictiveness and safety of a Chinese herbal medicine preperation, Jiawei Qing'e Fang (JQF), on menopausal symptoms in perimenopausal women. Study dates August 2009. Source of funding National Science &	Sample size N = 72 perimenopausal women * JQF: N = 32 Placebo: N = 32 * perimenopausal defined as menstrual irregularity or amenorrhea for a period of 3 to 11 months. Characteristics Age JQF (N=36) = 50.69 ± 3.45 Placebo (N = 36) = 50.39 ± 2.46 BMI JQF (N=36) = 25.38 ± 2.62 Placebo (N = 36) = 24.38 ± 2.62 Inclusion criteria - Aged 45 - 55 yrs, perimenopausal who reported 14 or more hot flushes per week Exclusion criteria - Hyperplasia, abnormal bleeding - Surgical menopause - known hypersensitivity to drugs and contraindications.	Interventions Jiawei Qing'e Fang (JQF) herbal medicine Placebo	Power calculation Unclear Intention to treat Unclear Details Setting Second Affiliated Hospital of Tianjin University of Traditional Chinese Medicine Randomisation method Predefined computer-generated randomisation list with a balaced 1:1 randomisation using a block size of 4. Statistical methods Continuous variables - means compared used independent t test for normally distrubed and Wilcoxon test for skewed distribution. Categorical variables compared using chi squared test.	Results Menopause specific quality of life (MENQOL) scores VSM Reported in seperate table Psychosocial (score, mean ± SD) Placebo (N = 32) Baseline = 3.15 ± 1.25 4 weeks = 3.06 ± 0.95 8 weeks = 3.00 ± 1.28 12 weeks = 3.07 ± 1.14 % reduction from baseline 4 weeks = 3.97 8 weeks = 4.54 12 weeks = 2.41 JQF (N = 32) Baseline = 3.56 ± 1.31 4 weeks = 3.18 ± 1.13 8 weeks = 2.95 ± 1.15 12 weeks = 3.00 ± 1.10 % reduction from baseline 4 weeks = 10.41 8 weeks = 17.19 12 weeks = 15.81 * p = 0.055 Physical Baseline = 3.17 ± 1.02 4 weeks = 3.06 ± 0.95 8 weeks = 3.02 ± 0.88 12 weeks = 2.98 ± 0.82 % reduction from baseline 4 weeks = 3.57 8 weeks = 4.74 12 weeks = 6.04 JQF Baseline = 3.29 ± 1.32 4 weeks = 2.90 ± 1.13 8 weeks = 2.66 ± 1.06 12 weeks = 2.85 ± 1.04 % reduction from baseline 4 weeks = 11.65	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: low C Attrition bias C1 - Was follow-up equal for both groups - Yes	Main outcome classification Psychological Musculoskeletal Sexual Main interventions classification non-pharmaceutical treatments

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
technology Pillar Programme, International Cooperative Project of the Science and Technology Ministry, Programme for the Changjiang Scholars and Innovative Research Team in Tianjin.				8 weeks = 18.97 12 weeks = 13.14 * P = 0.034 Sexual Baseline = 3.16 ± 1.79 4 weeks = 3.19 ± 1.63 8 weeks = 3.02 ± 1.59 12 weeks = 3.17 ± 1.55 % reduction from baseline 4 weeks = - 1.32 8 weeks = 4.29 12 weeks = - 0.33 JQF Baseline = 3.21 ± 1.63 4 weeks = 3.05 ± 1.50 8 weeks = 2.90 ± 1.41 12 weeks = 2.88 ± 1.41 % reduction from baseline 4 weeks = 4.97 8 weeks = 9.74 12 weeks = 0.39 * p = 0.249	C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Yes Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Unclear D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: low Indirectness Does the study match the review protocol in terms of Population: No Intervention: yes Outcomes: yes Indirectness: no	
Full citation Bao,T., Cai,L., Snyder,C., Betts,K., Tarpinian,K., Gould,J., Jeter,S., Medeiros,M., Chumsri,S.,	Sample size Acupuncture n=25, analyzed n=24 Sham acupuncture n=26, analyzed n=23 Characteristics	Interventions Sham acupuncture and Acupuncture weekly for 8 weeks	Power calculation Not reported Intention to treat Yes Details Setting John Hopkins and	Results Frequency of hot flushes (including night sweats) Reported in separate evidence table Frequency of sexual intercourse Not reported	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised	Main outcome classification Hot flashes Depression Main interventions classification Acupuncture vs sham

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Bardia,A., Tan,M., Singh,H., Tkaczuk,K.H., Stearns,V., Patient-reported outcomes in women with breast cancer enrolled in a dual-center, double-blind, randomized controlled trial assessing the effect of acupuncture in reducing aromatase inhibitor-induced musculoskeletal symptoms, Cancer, 120, 381-389, 2014 Ref Id 328293 Country/ies where the study was carried out USA Study type Dual-center, double-blind, randomized controlled trial Aim of the study Assess whether real acupuncture (RA), compared with sham acupuncture (SA), improves patient-reported outcomes (PROs) in patients with breast cancer who are receiving an adjuvant AI. Study dates Not reported Source of funding American Society of Clinical Oncology Foundation Young Investigator's Award, Susan Komen Postdoctoral	Sham acupuncture/Acupuncture Median age, year (range): 61 (44-82) / 61 (45-85) Duration of aromatase inhibitors: median (range),d: 426 (137-1561)/389 (109-1738) Inclusion criteria -Postmenopausal -Stage 0-3 hormone receptor-positive breast cancer who had been receiving AI therapy for greater than or equal to 1 month -Reported AI-associated musculoskeletal symptoms -Had not received acupuncture within the past 12 months Exclusion criteria Not reported		University of Maryland Cancer Center Randomisation method Generated by trial statistician using specialised randomisation software before the start of the trial. Randomisation assignments were provided to center acupuncturists. Randomisation sequence was not concealed Statistical methods -Comparison between treatment in change from baseline to week 8 used Wilcoxon signed-rank test -ANCOVA	Psychological symptoms -Anxiety Not reported -Depression Reported as CESD median (IQR) Sham Acupuncture/Acupuncture Baseline: 10.5 (10) / 16 (9) Week 12: 7.5 (11.75) / 10 (10.5) P-value for change from baseline between group: 0.442 -Cognitive function Not reported -Sleep disturbance Not reported -Quality of life Not reported Musculoskeletal symptoms -Symptom relief (joint pain and muscular pain [with and without] stiffness) Not reported -Muscle strength Not reported -[validated] Physical activity (Greene sub-scale data) Not reported -Quality of life Not reported Safety outcomes -Discontinuation Not reported -Major adverse events Not reported -Minor adverse events Not reported	controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - No A3 - Were groups comparable at baseline - Yes Level of bias: Moderate B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- No Level of bias: Moderate C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Unclear Level of bias: Moderate	acupuncture

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
Fellowship Award, Breast Cancer Research Foundation, Komen for the Cure					D Detection bias D1 - Was follow-up appropriate length - N/A D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Unclear D5 - Were investigators blinded to confounding factors - Unclear Level of bias: High Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no	
Full citation Zheng,T.P., Sun,A.J., Xue,W., Wang,Y.P., Jiang,Y., Zhang,Y., Lang,J.H., Efficacy and safety of Cimicifuga foetida extract on menopausal syndrome in Chinese women, Chinese Medical Journal,	Sample size N=96 participated in study Group A: Cimicifuga rhizome extract, n=32 (n=31 completed treatment) Group B: Oestradiol valerate +progesterone, n=32 (n=30	Interventions Group A: Cimicifuga foetida extract (three tablets) every day for three months Group B: Oestradiol valerate (one tablet) for 30 days each cycle, from the 19th day, also took two capsules of	Power calculation Not reported Intention to treat Not reported Details Setting Department of Peking Union Medical College Hospital, China Randomisation	Results Frequency of hot flushes (including night sweats) Not reported Frequency of sexual intercourse Not reported Psychological symptoms -Anxiety Reported as scores of the Hospital Anxiety and Depression score (HADS) (mean, SD) Group A/Group B/Group C Baseline: 5.23 (3.39)/6.43 (2.81)/5.71 (3.84) After 3 months (final): 4.42 (3.16)/5.00 (3.13)/4.79	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation -	Main outcome classification Anxiety Depression Vaginal bleeding Main interventions classification Non-pharmaceutical treatments: Herbal preparation- black cohosh Hormonal

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
126, 2034-2038, 2013 Ref Id 288683 Country/ies where the study was carried out China Study type Prospective randomised controlled trial Aim of the study To compare the clinical effects of different regimens of three-month course on climacteric symptoms, so as to evaluate the efficacy and safety of black cohosh extract Study dates Recruitment: from July 2009 to July 2010 Source of funding Not reported	completed treatment) Group C: Oestradiol valerate +medroxyprogesterone acetate (MPA), n=32 (n=28 completed treatment) Characteristics Age (mean, years, SD): Group A: 53.4 (3.0) Group B: 52.7 (3.6) Group C: 52.1 (3.2) Amenorrhea (mean, months (duration), SD): Group A: 27.0 (14.1) Group B: 28.5 (16.4) Group C: 29.5 (15.0) Height (mean, cm, SD): Group A: 159.29 (4.82) Group B: 161.40 (3.70) Group C: 159.46 (4.68) Weight (mean, kg, SD): Group A: 64.65 (9.21) Group B: 59.00 (7.07) Group C: 60.09 (9.08) Inclusion criteria Women aged 40 to 60 years, early menopausal, going through climacteric symptoms Early menopause was defined as going through	progesterone for 12 days (for three cycles) Group C: Oestradiol valerate (one tablet) for 30 days each cycle, from the 19th day, two tablets of MPA added to treatment for 12 days (for three cycles)	method 96 participants randomly and equally assigned to group A, B, or C in 16 blocks, generated by SAS software according to magnitude of random number Statistical methods Two-tailed tests were performed with a significant level of 0.05. Quantitative data meeting normal distribution were presented as mean (SD). Intra-group comparison was carried out between before and after treatment, paired-samples t test was used if data was of normal distribution, otherwise Wilcoxon W test was preferred. ANOVA was chosen for comparisons among groups if data was of normal distribution and equal variance, and P<0.05, LSD was chosen for post hoc multiple	(3.11) P value: 0.015/0.003/0.282 Quality of life reported as MENQOL scores (mean, SD) Group A/Group B/Group C Baseline: 4.33 (1.27)/4.69 (1.40)/4.40 (1.33) After 3 months (final): 3.72 (1.20)/3.40 (1.19)/3.39 (1.64) P value: 0.01/<0.001/0.001 -Depression Reported as scores of the Hospital Anxiety and Depression score (HADS) (mean, SD) Group A/Group B/Group C Baseline: 5.19 (2.94)/5.90 (3.92)/5.93 (4.02) After 3 months (final): 5.13 (3.22)/5.00 (3.17)/5.75 (3.80) P value: 0.7/0.1/0.9 Cognitive function Not reported Sleep disturbance Not reported Musculoskeletal symptoms Quality of life reported as MENQOL scores (mean, SD) Group A/Group B/Group C Baseline: 4.58 (1.07)/4.63 (1.10)/4.58 (1.37) After treatment (endpoint):3.79 (0.98)/3.20 (0.98)/3.54 (1.27) P value: <0.001/<0.001/<0.001 Muscle strength Not reported Physical activity Not reported	Yes A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: High B Performance bias B1 - Did groups get same level of care - yes B2 - Were participants blinded to treatment allocation- Unclear B3 - Were individuals administering care blinded to treatment allocation- Unclear Level of bias: High C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - No. Group C had 12.5% drop out C3 - Were groups comparable for missing data - unclear Level of bias: high D Detection bias D1 - Was follow-	pharmaceutical treatments: oestrogen combined with progesterone

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments	Identifiers
	amenorrhea above 6 months and within 5 years, serum E2 concentration <30pg/ml, and serum follicle stimulating hormone (FSH) concentration >40 IU/L Exclusion criteria Uterine fibroid (fibroid diameter ≥5cm or the size of uterus ≥8 gestational weeks), history of diabetes or hypertension, history of thromboembolism, severe endometriosis, epilepsy, asthma, hyperprolactinaemia, first degree relative having a history of breast cancer, receiving HRT in the past three months, and endometrial thickness ≥0.5 cm after withdrawal bleeding		comparisons. Kruskal-Wallis H test was used for data not fitting normal distribution. Enumeration data were reported as frequencies and rates, and X2 test (Fisher's exact test) was used for rate comparison.		up appropriate length - Yes D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Unclear D5 - Were investigators blinded to confounding factors - Unclear Level of bias: High	

H.5 Urogenital atrophy

H.5.1 Local oestrogens for short-term treatment

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Full citation Karp,D.R., Jean-Michel,M., Johnston,Y., Suciu,G., Aguilar,V.C., Davila,G.W., A randomized clinical trial of	Sample size N = 65 E-string = 22 Placebo (PLA) = 21 Control (CON) = 22	Interventions Women were randomised to either an estradiol-	Details 1. Standardised history and vaginal health assessmnets were performed at baseline and at 6 and 12 weeks after	Results Efficacy endpoints 1. Change in maturation value 2. Vaginal pH 3. Vaginal atrophy	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
the impact of local estrogen on postoperative tissue quality after vaginal reconstructive surgery, Female Pelvic Medicine and Reconstructive Surgery, 18, 211-215, 2012 Ref Id 226751 Country/ies where the study was carried out United States Study type Randomised controlled trial Aim of the study To evaluate the use and effect of early administration of vaginal oestrogen in the immediate post-operative period via a continuous low-dose estradiol vaginal ring in a placebo-controlled trial. Study dates October 2008 to January 2010 Source of funding No funding reported and Pfizer supplied the placebo vaginal rings	Characteristics Age (years) - Mean (SD) E-string = 65 (7.4) PLA = 66 (7.9) CON = 65 (7.8) Time since last period (years) - Median (Range) E-string = 14.5 (3 - 30) PLA = 17 (4 - 29) CON = 15 (3 - 35) Ethnicity White - n (%) Not reported Dyspareunia - n (%) Not reported Vaginal Dryness - n (%) Not reported Inclusion criteria 1. Inclusion criteria were postmenopausal women at least 2 years after spontaneous or sugical menopause with symptomatic urogenital atrophy and pelvic organ prolapse and had opted to undergo reconstructive vaginal surgery. 2. Eligible candidates had to have at least one symptom (vaginal dryness, vulvar pruritus, dyspareunia, dysuria, or urinary urgency) and/or sign (vaginal pallor, petechiae, friability) of atrophic vaginitis. Exclusion criteria Women were excluded if they had contra-indications to oestrogen use (vaginal bleeding, oestrogen-dependent cancers, hepatic or	releasing vaginal ring placed immediately after surgery, a placebo ring of identical size and shape or a control group who did not have any vaginal ring.	surgery. The women were asked to complete symptom and severity questionnaires in which the presence and severity of vaginal dryness, pruritus, dyspareunia, dysuria and urinary urgency were recorded by the patient. 2. Specimens for maturation value, microscopic inflammation and vaginal pH were collected at 6 and 12 weeks. For vaginal cytology, vaginal smears were taken from the upper right or left lateral vaginal walls with a plastic spatula, spread on a slide and immediately fixed with fixative spray. 3. Presence and severity of vaginal pallor, petechiae, friability, and dryness were noted at 6 and 12 weeks post-operatively and were assessd on a scale of 0 (none) to 4 (severe) 4. Maturation value (MV) = number of superficial cell + [0.5 x (number of intermediate cells)] + [0 x (number of parabasal cells)] divided by 2. A value of 0 to 49 indicated low oestrogen effect, 50 to 64 indicated moderate oestrogen effect and 65 to 100 indicated high oestrogen effect	Safety endpoints Not objectively evaluated Acceptability endpoints Withdrawal due to adverse events Quality of life endpoints Not evaluated EFFICACY Maturation value, mean percentage change at week 12 E-string = 27.1 PLA = -34.7 CON = -15.4 P < 0.01 Vaginal pH, number (%) of participants with pH less than 5.5 E-string = 12 (54.5) PLA = 0 (0) CON = 2 (9.1) Mean percentage difference in overall objective atrophy E-string = -63 PLA = +13 CON = +2.4 ACCEPTABILITY Withdrawal due to adverse events E-string = 2 PLA = 2 CON = 0	A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	thrombotic disease), allergies to silicone and/or vaginal pH of less than or equal to 4.0, or use of vaginal or systemic oestrogen in the previous 6 months.				comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - See results C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information Data from vaginal ring and placebo ring groups only used in guideline review.
Full citation Griesser,H., Skonietzki,S., Fischer,T., Fielder,K., Suesskind,M., Low dose estriol pessaries for the treatment of vaginal atrophy: a double-blind placebo-controlled trial investigating the efficacy of pessaries containing 0.2mg and 0.03mg estriol, Maturitas, 71, 360-368, 2012 Ref Id 226600 Country/ies where the study was carried out Germany Study type Randomised controlled trial Aim of the study To confirm the superior	Sample size N = 436 Estriol 0.2mg (0.2 ES) = 142 Estriol 0.03mg (0.03 ES) = 147 Placebo (PLA) = 147 Characteristics Age (years) - Mean (SD) 0.2 ES = 64.9 (8.1) O.03 ES = 65.4 (7.3) PLA = 64.8 (7.8) Time since last period (years) - Median (Range) Not reported Ethnicity White - n (%) Not reported Dyspareunia - n (%)	Interventions 1. The women were randomly assigned in a 1:1:1 ratio to receive either 0.2mg estriol, 0.03mg estriol or placebo. 2. The treatment duration was 12 weeks with once-daily applications for 20 days, followed by twice weekly administration for a further 9 weeks as a maintenance	Details 1. Primary efficacy endpoints were the rise (increase) in the vaginal maturation index, the normalisation (decrease of the vaginal pH value, and the improvement (decrease) in intensity of the subjective most bothersome symptom of vaginal atrophy after 12 weeks. 2. Secondary efficacy variables comprised the time course of the vaginal maturation index, of vaginal pH, and the most bothersome symptom, the physician's evaluation of effcacy and the rate of responders (meeting	Results Efficacy endpoints 1. Change in maturation index (increase) 2. Vaginal pH (decrease) 4. Subjective assessment of severity of most bothersome symptom of vaginal atrophy (decrease) Safety endpoints Treatment related adverse events Acceptability endpoints 1. Withdrawal due to adverse events 2. Subjective assessment of accepatbility to treatment Quality of life endpoints Not evaluated EFFICACY Maturation index, mean (SD) change at week	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Unclear A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
efficacy of pessaries with 0.03 mg and/or 0.2 mg estriol compared to pessaries without an active substance in the treatment of vaginal atrophy. Study dates October 2008 to January 2011 Source of funding Study was sponsored by Dr. Kade Pharmazeutische Fabrik gmbH	Not reported Vaginal Dryness - n (%) Not reported Inclusion criteria 1. Postmenopausal women (last menstrual period more than 12 months ago or having undergone bilateral ovariectomy) aged 18 years or older with a clinical diagnosis of vaginal atrophy, a vaginal maturation index > 40% and a vaginal pH value > 5. 2. At least one subjective symptom of vaginal atrophy (dryness, pain/burning sensation, pruritus, discharge, dyspareunia) had to be rated at a score of ≥ on a visual analogue scale. Exclusion criteria Hormone replacement therapy; therapy with phytoestrogens or local vaginal hormonal therapy during the 12 weeks preceding baseline as well as current or suspected estrogen-dependent malignant tumor; a pap smear ≥ grade III; endometrial thickness > 5mm; current or suspected vaginal infection; current symptomatic urinary tract infection; existing or previous breast cancer or suspicion thereof; undiagnosed bleeding in the genital area; current venous thromboembolic disease; known severe	therapy.	simultaneously the criteria of vaginal maturation index ≥ 55%, vaginal pH ≤ 5 and most bothersome symptom ≤ 35 on the visual analogue scale). 3. Maturation value was calculated as follows: number of superficial cells + [0.5 x (number of intermediate cells)] + [0 x (number of parabasal cells)].	12 (pairwise comparisons) 0.2 ES = 46.3 (17.0) PLA = 23.9 (21.5) 0.03 ES = 38.4 (19.4) PLA = 23.9 (21.5) Vaginal pH, mean (SD) change at week 12 (pairwise comparisons) 0.2 ES = -1.6 (0.8) PLA = -0.6 (0.8) 0.03 ES = -1.4 (0.9) PLA = -0.6 (0.8) Severity of most bothersome symptom score, mean (SD) change at week 12 (pairwise comparisons) 0.2 ES = -52.2 (23.7) PLA = -31.8 (26.3) 0.03 ES = -47.1 (23.4) PLA = -31.8 (26.3) SAFETY Treatment related adverse events, n (%) 0.2 ES = 34 (23.9) 0.03 ES = 32 (21.8) PLA = 38 (25.9) ACCEPTABILITY Withdrawal due to adverse events 0.2 ES = 5/142 0.03 ES = 7/147 PLA = 5/147 Percentage reporting 'very good' or 'good' tolerability 0.2 ES = 94.6 0.03 ES = 88.9 PLA = 80.5	or treatment allocation) - Unclear A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Unclear risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - See results C2b. The groups were comparable for treatment completion (that is, there

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	renal insufficiency or hypersensitivity to estriol or any excipients (hard fat and emulsifiers) of the study medication.				were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Unclear D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Unclear risk of bias Indirectness

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious
Full citation Bachmann,G., Bouchard,C., Hoppe,D., Ranganath,R., Altomare,C., Vieweg,A., Graepel,J., Helzner,E., Efficacy and safety of low-dose regimens of conjugated estrogens cream administered vaginally, Menopause, 16, 719-727, 2009 Ref Id 226127 Country/ies where the study was carried out Canada & United States Study type Randomised controlled trial Aim of the study To evaluate the efficacy and safety of low dose conjugated oestrogen cream 0.3mg (equivalent to Premarin Vaginal Cream 0.5g) for the treatment of vulvovaginal atrophy Study dates Not reported Source of funding The study was supported by Wyeth Research, Collegeville, PA	Sample size N = 423 Conjugated oestrogen cream daily for 3 weeks then 1 week off (CE 21/7) for 12 weeks = 143 Conjugated oestrogen cream twice weekly (CE 2/W) for 12 weeks = 72 Placebo daily for 3 weeks then 1 week off (PLA 21/7) for 12 weeks = 140 Placebo twice weekly (PLA 2/W) for 12 weeks = 68 Characteristics Age (years) - Mean (SD) CE 21/7 = 57.7 (±5.8) CE 2/W = 57.5 (±5.5) PLA 21/7 = 58.0 (±5.8) PLA 2/W = 58.7 (±5.8) Time since last period (years) - Mean (SD) CE 21/7 = 8.9 (±6.0) CE 2/W = 7.9 (±5.8) PLA 21/7 = 9.7 (±6.6) PLA 2/W = 9.9 (±6.7) Ethnicity White - n (%) CE 21/7 = 134 (93.7) CE 2/W = 127 (90.7) PLA 21/7 = 63 (87.5) PLA 2/W = 60 (97.1) Dyspareunia - n (%) CE 21/7 = 88 63.8) CE 2/W = 83 (60.6) PLA 21/7 = 33 (47.1) PLA 2/W = 37 (55.2)	Interventions Women were treated with either conjugated oestrogen cream daily for 3 weeks then 1 week off, conjugated oestrogen cream twice weekly, placebo daily for 3 weeks then 1 week off, or placebo twice weekly for a period of 12 weeks. All women went on to receive open-label treatment with conjugated oestrogen cream for the next 40 weeks using the same regimen to which they were assigned during the initial 12 week phase.	Details 1. Primary endpoints were changes from baseline in vaginal maturation indices, vaginal pH and the severity of pateint-reported most bothersome symptom at 12 weeks. 2. Vaginal pH and the percentage of superficial and parabasal cells (on vaginal cytologic smear) were measured at baseline, 4, 6, 12 and 52 weeks or the time of study discontinuation. 3. The severity of each symptom was recorded daily on a daily diary card and the weekly score derived from an average of daily scores during that week. 4. A secondary endpoint was the GHCE perfomed at baseline, 4, 6, 12 and 52 weeks or the time of study discontinuation	Results Efficacy parameters 1. Change in vaginal maturation index (percentages of superficial and parabasal cells in vaginal smear) 2. Change in vaginal pH 4. Severity of most bothersome symptom of atrophic vaginitis: vaginal dryness, itching, burning, or dyspareuinia Safety parameters Treatment related adverse events Acceptability parameters Withdrawal due to adverse events Quality of life parameters Not evaluated EFFICACY Superficial cells, mean (SD) percentage change from baseline to week 12 CE 21/7 = 27.9 (±20.3) CE 2/W = 25.8 (±20.1) PLA 21/7 = 3.0 (±20.4) PLA 2/W = 1.0 (±19.8) P ≤ 0.001 Parabasal cells, mean (SD) percentage change from baseline to week 12 CE 21/7 = -60.9 (±20.3) CE 2/W = -58.2 (±26.0) PLA 21/7 = -21.5 (±25.5) PLA 2/W = -6.6 (±25.6) P ≤ 0.001 Vaginal pH, mean (SD) change from baseline to week 12 CE 21/7 = -1.6 (±1.2), 143 CE 2/W = -1.6 (±1.2), 140 PLA 21/7 = -0.4 (±0.8), 72	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Unclear A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Unclear A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Unclear risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Vaginal Dryness - n (%) CE 21/7 = 34 (24.6) CE 2/W = 22 (23.4) PLA 21/7 = 21 (30.0) PLA 2/W = 16 (23.9) Inclusion criteria Healthy postmenopausal women aged between 45 and 80 with an intact uterus and syl score of 15 or less on the Genital Health Clinical Evaluationotamptoms of moderate-to-severe vaginal atrophy defined as; a baseline composite score, at the screening visit, of at least 5 (1 = mild, 2 = moderate, 3 = severe) on the four symptoms (dyspareunia, vaginal dryness, vaginal itching and vaginal burning) at least one of these symptom said to be moderate or severe a total score of 15 or less on the Genital Health Clinical Evaluation (GHCE) vaginal pH of at least 5 a clinical diagnosis of atrophic vaginitis (defined as 0% to 5% superficial cells on vaginal cytologic smear) Additional criteria included a serum estradiol concentration of 30 pg/ml or less and a serum follicle-stimulayting hormone level greater than the lower limit of normal for postmenopausal women			PLA 2/W = - 0.3 (±0.8), 68 P ≤ 0.001 Mean change in severity score for most bothersome symptom reported CE 21/7 = -1.3 CE 2/W = -1.4 PLA 21/7 = -0.8 PLA 2/W = -0.7 P ≤ 0.001 SAFETY Treatment related adverse events, n (%) CE 21/7 = 95 (66.4) CE 2/W = 100 (71.4) PLA 21/7 = 46 (63.9) PLA 2/W = 47 (69.1) ACCEPTABILITY Withdrawal due to adverse events CE 21/7 = 6/143 CE 2/W = 8/140 PLA 21/7 = 3/72 PLA 2/W = 4/68	B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? See results C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	at the given laboratory Exclusion criteria 1. Use of an intrauterine device within 3 months of screening or the use of any oral, vaginal, or transdermal medication containing oestrogens, androgens or progestins within 8 weeks of screening. 2. Women who had used vaginal moisturizers, lubricants, jellies, ointments, douches, herbal medications, over-the-counter preparations, home remedies or natural oestrogen products for the treatment of menopausal symptoms agreed to refrain from using them for a minimum of 7 days before screening. 3. Women who currently used more than two antihypertensive medications, had used any investigational drug or device within 30 days of screening, or had urogynecologic surgery within 3 months of screening were also excluded				Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information 1. Standard deviation for results calculated from the standard error reported using the following formula: SD = SE x √N 2. Data for the CE 21/7 group used in the analysis as this is the recommended (labelled) regimen
Full citation Cano,A., Estevez,J., Usandizaga,R., Gallo,J.L., Guinot,M., Delgado,J.L.,	Sample size N = 167 Estriol gel (EST) 114 Placebo (PLA) = 53	Interventions Depending on the randomisation	Details 1. Efficacy was assessed by the evaluation of the cytological MV, vaginal pH,	Results Efficacy endpoints 1. Change in maturation value 2. Vaginal pH	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist:

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Castellanos,E., Moral,E., Nieto,C., del Prado,J.M., Ferrer,J., The therapeutic effect of a new ultra low concentration estriol gel formulation (0.005% estriol vaginal gel) on symptoms and signs of postmenopausal vaginal atrophy: results from a pivotal phase III study, Menopause, 19, 1130-1139, 2012 Ref Id 255650 Country/ies where the study was carried out Spain Study type Randomised controlled trial Aim of the study To evaluate the efficacy and safety of 0.005% estriol vaginal gel, delivering an ultra-low dose of estriol per application, for the local treatment of postmenopausal vaginal atrophy. Study dates Not reported Source of funding Study funded by Italfarmaco SA	Characteristics Age (years) - Mean (SD) EST = 56.5 (±5.72) PLA = 57.2 (±6.70) Time since last period (years) - Mean (SD) EST = 9.7 (±6.57) PLA = 10.2 (±6.68) Ethnicity - White n (%) EST = 114 (100) PLA = 53 (100) Dyspareunia - n (%) Not reported Vaginal Dryness - n (%) Not reported Inclusion criteria Women wre included if they were postmenopausal (at least 2 years of amenorrhea by either natural or sugical menopause (bilateral oophorectomy)). They also presented symptoms and signs of atrophy of the vaginal mucosa including as a minimum vaginal dryness and at least one sign of vaginal atrophy (a thinned vaginal mucosa, a mucosa with flattening of the folds or a dry, fragile or pale vaginal mucosa); and the presence of petechiae or any other alteration that the investigator considered indicative of vaginal atrophy were assessed by the investigators in gynecological examination. Exclusion criteria	schedule, women received either 1g of vaginal gel containing 50micrograms of estriol or 1g of placebo. The placebo formulation was a highly hydrating gel identical in appearance, aroma, and texture to the estriol formulation but with the exclusion of the hormone. Women were advised to administer the gel preferably at night. The gel was administered with an applicator inserted deep inside the vagina.	and symptoms and signs of vaginal atrophy at baseline and after 3 and 12 weeks of treatment. 2. Maturation value (MV) = number of superficial cell + [0.6 x (number of intermediate cells)] + [0.2 x (number of parabasal cells)] 3. Vaginal pH was assessed using a vaginal pH strip 4. A composite symptom score (Global Symptom Score) of - (none) tr 3 (severe) was used 5. Safety was assesed by evaluation of adverse effects, gynecological and physical examinations and vital signs.	4. Signs and symptoms of vaginal atrophy Safety endpoints Treatment related adverse events Acceptability endpoints 1. Withdrawal due to adverse events 2. Subjective assessment of acceptability Quality of life endpoints Not evaluated EFFICACY Maturation index, mean (SD) change from baseline to week 12 EST = 26.9 (±23.3) PLA = 3.2 (±16.5) Vaginal pH, mean (SD) change from baseline to week 12 EST = -1.2 (±1.4) PLA = - 0.4 (±1.2) Vaginal dryness, percentage of women cured/improved at week 12 EST = 88.2 PLA = 66.7 P = 0.001; RR=1.32 (1.08-1.62) Vaginal pruritus, burning, and dysuria Improved in estriol group but no significant differences detected. Dyspareunia, percentage of women cured/improved at week 12 EST = 86.5 PLA = 75.0 P = 0.095; RR=1.15 (0.96-1.39) SAFETY Treatment related adverse events, n (%) EST = 52 (45.6) PLA = 21 (39.6) ACCEPTABILITY Withdrawal due to adverse events EST = 1/114	randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Unclear A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Unclear A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Unclear risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	1. Women were excluded if they had a history of malignant or premalignant lesions of the breasts or endometrium; malignant colon or hepatic tumors; malignant melanoma; venous thromboembolic disorders or arterial thromboembolic disorders; peripheral arterial disease; mesenteric artery thrombosis; renal artery thrombosis or coagulopathies. 2. Women were also excluded if they had undiagnosed vaginal bleeding, grade II or higher uterovaginal prolapse or signs and symptoms suggestive of infection of the genital or urinary tract. 3. Women with endometrial thickness equal to or less than 4 mm measured by transvaginal ultrasound or who had received any type of vulvovaginal treatment with 15 days of study initiation, women who had received phytoestrogens with 1 month and women who had received hormonal therapy within 3 months of study start.			PLA = 0/53 Percentage of women rating the intervention as 'excellent' or 'good' EST = 73.6 PLA = 43.1	differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - See results C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious
Full citation Simon,J., Nachtigall,L., Gut,R., Lang,E., Archer,D.F., Utian,W., Effective treatment of vaginal atrophy with an ultra-low-dose estradiol vaginal tablet.[Erratum appears in Obstet Gynecol. 2008 Dec;112(6):1392], Obstetrics and Gynecology, 112, 1053-1060, 2008 Ref Id 227345 Country/ies where the study was carried out United States Study type Randomised controlled trial Aim of the study To evaluate the efficacy and safety of a new ultra-low dose 10-microgram E2 vaginal tablet in a placebo-controlled, 52-week, double	Sample size N = 309 Endogenous estradiol (E2) = 205 Placebo (PLA) = 104 Characteristics Age (years) - Mean (SD) E2 = 57.5 (±5.64) PLA = 57.7 (±5.27) Time since last period (years) - Mean (SD) E2 = 8.0 (±5.8) PLA = 8.2 (±5.3) Ethnicity White - n (%) E2 = 192 (93.7) PLA = 95 (91.3) Dyspareunia - n (%) Not reported Vaginal Dryness - n (%) Not reported	Interventions 1. Women were randomly assigned in a 2:1 ratio in blocks of 6 to receive vaginal tablets containing either 10 micrograms E2 (Novo-nordisk A/S) or placebo. 2. All vaginal tablets were identical in appearance. 3. Treatment instructions were to insert one vaginal tablet daily for 14 days and the subsequently	Details 1. The primary efficacy endpoints included the mean change form baseline to weeks 12 (Last observation carried forward = LOCF) in vaginal maturation index abd value, vaginal pH, and the mean score of the most bothersome moderate to severe symptom as identied by the woman. 2. For vaginal cytology, smears were taken form the upper third of the right lateral vaginal wall and the samples used to calculate the maturation index. 3. The maturation value was calculated according to the following formula = 1 x number of superficial cells + [0.5 x (number of intermediate cells)] + [0 x	Results Efficacy endpoints 1. Percentage of superficial cells on the vaginal smear 2. Percentage of parabasal cells on the vaginal smear 3. Percentage of intermediate cells on the vaginal smear 4. Maturation index 5. Vaginal pH 6. Mean score for most bothersome urogenital symptom (dyspareunia and vaginal dryness) [0 = none, 3 = severe] Safety endpoints Treatment related adverse events Acceptability endpoints Withdrawal due to adverse events Quality of life endpoints Not evaluated EFFICACY	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
blind clinical trial Study dates March 2005 to May 2006 Source of funding Supported by Novodisk A/S	Inclusion criteria 1. The study included nonhysterectomised, postmenopausal (2 or more years since final menstrual cycle or bilateral oophorectomy) women who were at least 45 years of age or older, with at least three urogenital symptoms (vaginal dryness, vaginal and/or vulvar irritation/itching, vaginal soreness, dysuria, or dyspareunia and vaginal bleeding associated with sexual activity), one of which had to be moderate in severity 2. All women were required to have serum E2 levels less than 20pg/ml, follicle stimulating hormone levels more than 40 milli-international units/ml, 5% or more superficial cells in vaginal cytology, vaginal pH more than 5.0, an endometrial thickness of less than 4.0mm as assessed by transvaginal ultrasonography, and a normal mammogram within the 6 months before study entry. Exclusion criteria 1. Known or suspected history of breast carcinoma, hormone-dependent tumor, genital bleeding of unknown cause, acute thrombophlebitis or thromboembolic disorder associated with oestrogen use, vaginal infection	one tablet twice per week. The women were instructed to insert the tablets at the same time each day.	(number of parabasal cells)] divided by 2.	Superficial cells, mean percentage change from baseline to week 12 10 E2 = 13 PLA = 4 P < 0.001 Intermediate cells, mean percentage change from baseline to week 12 10 E2 = 24 PLA = 5 P < 0.001 Parabasal cells, mean percentage change from baseline to week 12 10 E2 = -37 PLA = -9 P < 0.001 Maturation index, mean change from baseline to week 12 10 E2 = 25.0 PLA = 6.5 Vaginal pH, participants with pH less than 5.5 at week 12, n (%) 10 E2 = 145 (72) PLA = 37 (36) Change in mean score for most bothersome urogenital symptom at week 12 10 E2 = -1.23 PLA = -0.87 P = 0.003 SAFETY Treatment related adverse events, n (%) 10 E2 = 158 (77) PLA = 77 (75) ACCEPTABILITY Withdrawal due to adverse events, n (%) 10 E2 = 11 (5) PLA = 5 (5)	comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - See results C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	requiring treatment, allergy to the test drug or its constituents, or any serious disease or chronic condition that could interfere with study compliance. 2. The use of any investigational drug within the 30 days preceding screening, exogenous sex hormones within 3 months before study drug initiation, or current use of corticosteroids were prohibited.				those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Intervention: Yes Outcomes: Yes Indirectness: No serious Other information Study primary endpoint was 12 weeks. Continued till week 52 of which results are reported in long-term review question. Endometrial safety evaluated at week 52.
Full citation Bachmann,G., Lobo,R.A., Gut,R., Nachtigall,L., Notelovitz,M., Efficacy of low-dose estradiol vaginal tablets in the treatment of atrophic vaginitis: a randomized controlled trial, Obstetrics and Gynecology, 111, 67-76, 2008 Ref Id 226126 Country/ies where the study was carried out United States Study type Randomised controlled trial Aim of the study To evaluate and compare the efficacy of vaginal tablets containing 25mcg E2, 10mcg E2 and placebo for vaginal atrophy in post-menopausal women. Study dates Enrollment lasted from 1994 to 1996 Source of funding Supported by Novo Nordisk A/S	Sample size N = 230 25 mcg Estradiol (25 E2) = 91 10 mcg estradiol (10 E2) = 92 Placebo (PLA) = 47 Characteristics Age (years) - Mean (SD) 25 E2 = 58.3 (±7.4) 10 E2 = 57.7 (±6.5) PLA = 57.6 (±4.8) Time since last period (years) - Mean (SD) 25 E2 = 14.8 (±9.6) 10 E2 = 13.5 (±7.8) PLA = 13.6 (±8.1) Ethnicity - White n (%) 25 E2 = 88 (96.7) 10 E2 = 83 (90.2) PLA = 41 (87.2) Dyspareunia - n (%) Not reported Vaginal Dryness - n (%) Not reported Inclusion criteria 1. Women aged 45 years or older with moderate-to-severe vaginal dryness and soreness. 2. All women had serum	Interventions A low dose oestrogen vaginal tablet, containing 25 mcg estradiol or 10 mcg estradiol, in a hydrophilic cellulose-nased matrix were used in double-blind fashion for 12 weeks and compared with an identical-looking placebo. treatment instructions were to insert one vaginal tablet daily for 14 days and subsequently one tablet twice per week. The women werre instructed to insert the tablet at the same time each day.	Details 1. Evaluations for safety and efficacy occurred at weeks 2, 4, 7 and 12 in the double-blind phase and at 12, 26. 39 and 51 weeks in the open label phase. 2. The primary efficacy outcome was the change in the composite score of three vaginal symptoms (dryness, soreness and irritation). 3. Routine laboratory assessments included haematology, blood chemistry and urinalysis measured at screening at at weeks 12 and 52. 4. Physical examinations findings were recoded by the investigators.	Results Efficacy endpoints 1. Maturation index (percentage change in superficial and intermediate cells on the vaginal smear) 2. Change in vaginal pH 4. Change in composite score of three vaginal symptoms (dryness, soreness, and irritation) Safety endpoints 2. Endometrial histology 3. Treatment related adverse events Acceptability endpoints Withdrawal due to adverse events Quality of life endpoints Not evaluated EFFICACY Maturation value, mean (SD) percentage change from baseline to week 12 25 E2 = 11.5 (±13.3) 10 E2 = 13.1 (±13.3) PLA = 8.7 (±16.4) Significant increase in superficial and intermediate cells Vaginal pH, proportion of participants with pH less than 5 at week 12 25 E2 = 51% 10 E2 = 39% PLA = 21% Vaginal symptom composite score Significant reduction in scores for both E2	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	E2 concentrations of 20pg/ml or less, with 5% or less superficial vaginal cells. 3. Participants were also required to be at least 12 months post-menopausal, with an endometrial thickness of 5mm or less as determined by transvaginal ultrasonography Exclusion criteria Known or suspected history of breast carcinoma; hormone dependent tumor; genital bleeding of unknown cause; acute thrombophlebitis or thromboembolic disorder associated with oestrogen use; vaginal infection requiring treatment; allergy to the test drug or its constituents; or any serious disease or chronic condition that could interfere with study compliance. The use of any investigational drug within 30 days preceding screening. Any homeopathic preparation with the 7 days preceding study drug administration, and any exogenous corticosteroid or sex hormones within the 8 weeks preceding study drug initiation was prohibited.			groups compared to placebo SAFETY Endometrial histology One case of hyperplasia in the 25 mcg E2 group Treatment related adverse events No apparent trends reported ACCEPTABILITY Withdrawal due to adverse events 25 E2 = 4/91 10 E2 = 6/92 PLA = 1/47	B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? See results C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information Standard deviation for results calculated from the standard error reported using the following formula: SD = SE x √N *Data from 25 E2 and 10 E2 group combined for the analysis as both doses are

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					recommended in the BNF
Full citation Dessole,Salvatore, Rubattu,Giovanni, Ambrosini,Guido, Gallo,Omar, Capobianco,Giampiero, Cherchi,Pier Luigi, Marci,Roberto, Cosmi,Erich, Efficacy of low-dose intravaginal estriol on urogenital aging in postmenopausal women, Menopause (New York, N.Y.), 11, 49-56, 2004 Ref Id 319335 Country/ies where the study was carried out Italy (City of Sassari) Study type Propective, randomized, double-blind placebo-controlled study Aim of the study To assess the efficacy and safety of intravaginal estriol administration on urinary incontinence, urogenital atrophy, and recurrent urinary tract infections in postmenopausal women Study dates May 1999 to April 2002 Source of funding Not reported	Sample size Total = 88 Intravaginal estriol ovule group=44 Placebo group=44 Characteristics Postmenopausal women between 55 and 70 years of age Treatment and control groups were homogenous for age and urogenital aging symptoms Age (years) Intravaginal estriol ovule group=58 (4) Placebo group=56 (5) BMI (kg/m²) Intravaginal estriol ovule group=21.8 (4.5) Placebo group=22.4 (4.9) Race Intravaginal estriol ovule group=99% Placebo group=98% Vaginal parity Intravaginal estriol ovule group=2.9 (1.8) Placebo group=2.6 (1.2) Duration of menopause (years) Intravaginal estriol ovule group=7.5 (5.2) Placebo group=7.0 (4.8) Duration of urogenital atrophy symptoms (years) Intravaginal estriol ovule group=4.8 (5.0) Placebo group=5.0 (5.2)	Interventions Intravaginal estriol ovule group: Intravaginal estriol ovules: 1 ovule (1 mg) once daily for 2 weeks and then 2 ovules once weekly as maintenance therapy for a total of 6 months. Placebo group: Inert placebo vaginal suppositories in a similar regimen All were identical in appearance	Details Sample size calculated on the basis of prevalence of urinary incontinence, urogenital atrophy, and recurrent urinary tract infections in postmenopausal women. Determination of vaginal pH, colposcopic examination, vaginal and urethral smeras, and urodynamic examination performed at baseline and after 6 months of treatment. Randmization used sets of sequenced, sealed, opaque envelopes, each containing the bottle number to be given to each participant. Vaginl dryness and dyspareunia were classified as: none, moderate, or severe Degree of urogenital atrophy visually assessed and classified as none, moderate, or severe; taking into account pallor, petechiae, friability, and vaginal dryness (yes or no) Vaginal pH measured using an indicator strip	Results Efficacy endpoints 1. Vaginal dryness 2. Dyspareunia 3. Urogenital atrophy (n) 4. Vaginal pH Safety endpoints Treatment related adverse events Acceptability endpoints Withdrawal due to adverse events Quality of life endpoints Not evaluated EFFICACY Number with vaginal dryness Intravaginal estriol ovule group: Before treatment - 44/44 After treatment - 14/44 Control group: Before treatment - 44/44 After treatment - 37/44 P<0.001 Number with dyspareunia Intravaginal estriol ovule group: Before treatment - 38/44 After treatment - 9/44 Control group: Before treatment - 37/44 After treatment - 38/44 P<0.001 Number with urogenital atrophy Intravaginal estriol ovule group: Before treatment - 44/44 After treatment - 12/44 Control group: Before treatment - 44/44 After treatment - 41/44 P<0.01 Vaginal pH, mean (SD) Intravaginal estriol ovule group: Before treatment - 5.65 (0.97) After treatment - 4.12 (0.96) Control group: Before treatment - 5.47 (0.93) After treatment - 5.30 (0.75) P<0.05	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Inclusion criteria Postmenopausal women with urogenital aging symptoms (symptoms and signs of urinary stress incontinence, vaginal atrophy symptoms including vaginal dryness and dyspareunia, and histories of recurrent urinary tract infections. None had received estrogen therapy before the study. Exclusion criteria Anatomical lesions of the urogenital tract, such as uterovaginal prolapse, cystocele, and rectocele of grade I or II, presence of severe systemic disorders, thromboembolic diseases, biliary lithiasis, previous breast or uterine cancer, abnormal uterine bleeding, and body mass index of 25 kg/m² or higher. Wome with detrusor over activity and abnormal maximal cystometric capacity were also excluded.			SAFETY Treatment related adverse events Intravaginal estriol ovule group: 4 Control group: 3 ACCEPTABILITY Withdrawal due to adverse events Intravaginal estriol ovule group: 4 Control group: 7	kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - See results C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Yes Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious
Full citation Eriksen,P.S., Rasmussen,H., Low-dose 17 beta-estradiol vaginal tablets in the treatment of atrophic vaginitis: a double-blind placebo controlled study, European Journal of Obstetrics, Gynecology, and Reproductive Biology, 44, 137-144, 1992 Ref Id 226455 Country/ies where the study was carried out Denmark Study type Double-blind randomized placebo controlled trial	Sample size N=164 Treatment group: 81 Placebo group: 83 Characteristics Women between 45 and 70 years of age No statistical significant difference between the two groups concerning all baseline variables Age (years) Treatment group: 58.1 (6.0) Placebo group: 58.6 (6.0) Weight (kg) Treatment group: 63.2 (11.5)	Interventions Treatment group: Vaginal tablet contaiing 25 µg micronized 17ß-estradiol in a hydrophilic matrix system. One vaginal tablet daily for the first 2 weeks and then one tablet twice a week for the last 10 weeks Placebo group: Tablets using the same	Details Women interviwed about degree of vaginal dryness, burning and itching, dyspareunia related to the vagina at each visit. Gynecological examination to establish the degree of atrophy, signs of inflammation, pallor, petechiae and thickness of mucosa. Degree of atrophy assessed at 2 and 12 weeks.	Results Moderate to severe atrophy of vaginal mucosa (%) Treatment group: Before treatment - 78.8%; After 2 weeks treatment - 14.3%; After 12 weeks treatment - 10.7% Placebo group: Before treatment - 81.9%; After 2 weeks treatment - 35.4%; After 12 weeks treatment - 29.9% P-value at 2 weeks < 0.001 P-value at 12 weeks < 0.001 Vaginal dryness (%) Treatment group: Before treatment - 70.0%; After 12 weeks treatment - 14.7% Placebo group: Before treatment - 65.1%; After 12 weeks treatment - 28.2% No difference after 2 weeks P-value at 12 weeks < 0.002	Limitations Method of randomisation, treatment allocation not reported.

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Aim of the study To investigate the effect of 25 µg 17ß-estradiol administered as a small vaginal tablet for 12 weeks on the symptoms of the vagina related to atrophy. Study dates May 1989 to April 1990 Source of funding Not reported	Placebo group: 64.6 (9.9) Systolic blood pressure (mmHg) Treatment group: 141 (21) Placebo group: 142 (21) Inclusion criteria Women suffering from vaginal symptoms related to postmenopausal atrophy and not subjected to any estrogen treatment for the duration of at least 1 month before participation. Exclusion criteria Past history of acncer or thromboembolic episodes, vaginal bleeding of unknown origin, or if pregnant.	applicator		Vaginal burning and itching (%) Treatment group: Before treatment - 46.3%; After 12 weeks treatment - 10.6% Placebo group: Before treatment - 38.6%; After 12 weeks treatment - 25.6% No difference after 2 weeks P-value at 12 weeks < 0.088 Vaginal dyspareunia (%) Treatment group: Before treatment - 42.5%; After 2 weeks treatment - 14.2; After 12 weeks treatment - 8.0% Placebo group: Before treatment - 45.8%; After 2 weeks treatment - 25.9; After 12 weeks treatment - 24.4% P-value at 2 weeks < 0.003 P-value at 12 weeks < 0.002 Dropouts due to several reasons (n) Treatment group: 6 Placebo group: 4	
Full citation Casper,F., Petri,E., Local treatment of urogenital atrophy with an estradiol-releasing vaginal ring: a comparative and a placebo-controlled multicenter study. Vaginal Ring Study Group, International Urogynecology Journal, 10, 171-176, 1999 Ref Id 255671 Country/ies where the study was carried out Germany Study type Double-blind placebo-controlled study Aim of the study To detect differences between the efficacy and safety of the low-dose estradiol-releasing silicone vaginal ring compared to a placebo ring in the relief of	Sample size N=84 Number in each treatment arm not reported, but 67 reported to have completed 24-week treatment. Estradiol vaginal ring group: 33 Placebo group: 34 Characteristics Postmenopausal women recruited from 10 clinical sites No clinically significant differences found between the two treatment groups. Inclusion criteria At least 2 years post spontaneous or surgical menopause presenting with one or more of the following signs and symptoms of atrophic vaginitis due to estrogen	Interventions Low-dose estradiol-releasing vaginal ring - has a core containing 2 mg of 17β-estradiol within a silicone vaginal ring Placebo ring	Details Physical and gynecological examinations, including vaginal sonography, vaginal smear and pH measurement were performed at inclusion visit. Efficacy analyses conducted on a per-protocol analyses Safety analyses conducted on an intention-to-treat analyses	Results EFFICACY endpoints 1. Epithelial maturation values estimated as MV=(1.0 X % superficial cells) + (0.6 x % intermediate cells) + (0.2 x % parabasal cells) 2. Vaginal pH 3. Physician assessment of epithelial atrophy (vaginal pallor, petechiae, friability, and dryness) 4. Symptoms of estrogen deficiency - vaginal dryness, pruritus, dyspareunia, dysuria, and urinary urgency SAFETY endpoints 1. Endometrial thickness 2. Treatment-related adverse events ACCEPTABILITY endpoints Not evaluated QUALITY OF LIFE endpoints Not evaluated EFFICACY	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Unclear A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Unclear A3. The groups were comparable at baseline

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
symptoms of estrogen deficiency and the reduction of urogenital atrophy (vaginal pH an epithelial maturation values) in postmenopausal women. Study dates Not reported. Study published in 1999. Source of funding Not reported.	deficiency: 1. Pruritus vulvae, dyspareunia, dysuria, urinary urgency 2. Petechiae, friability or vaginal dryness on examination by a gynecologist Exclusion criteria Women who had received sex hormone therapy within the previous 3 months, or who had severe hepatic or renal diseases, estrogen-dependent neoplasms and urinary tract infections despite antibiotic treatment, or presented an endometrial thickness > 5mm or a vaginal ulceration, irritation, or bleeding from causes other than epithelial atrophy.			Maturation value Mean maturation value in estradiol group significantly higher than in placebo group at week 24 (P = 0.004) Vaginal pH Estradiol ring group: decrease in vaginal pH from 6.7 to 5.3 Placebo group: decrease in vaginal pH from 6.8 to 6.2 P = 0.0006 Relief of dyspareunia, % Estradiol ring group: 90 Placebo group: 45 P=0.028 Free of vaginal dryness, n (%) Estradiol ring group: 32 (69) Placebo group: 33 (73) P = not significant SAFETY Mean endometrial thickness, mm Estradiol ring group: 3.1 at baseline to 3.4 at 24 weeks Placebo group: 3.0 at baseline to 2.8 at 24 weeks Adverse effects No significant difference in adverse effects between the two groups	including all major confounding and prognostic factors - Yes Unclear risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - 67 of 84 completed treatment C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Yes Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Intervention: Yes Outcomes: Yes Indirectness: No serious
Full citation Bachmann,G.A., Komi,J.O., Ospemifene Study Group., Ospemifene effectively treats vulvovaginal atrophy in postmenopausal women: results from a pivotal phase 3 study, Menopause, 17, 480-486, 2010 Ref Id 226136 Country/ies where the study was carried out 76 centers in the United States Study type Randomized, double-blind phase 3 study Aim of the study To evaluate the efficacy and safety of ospemifene in the treatment of vulvovaginal atrophy (VVA) in postmenopausal women for 12-weeks. Study dates Not reported. Source of funding QuatRx Pharmaceuticals Company	Sample size N = 826 Ospemifene 30 mg/day: 282 Ospemifene 60 mg/day: 276 Placebo: 268 Characteristics Ninety percent of women in all groups were white. Age, mean (SD) years Ospemifene 30 mg/day: 58.4 (6.3) Ospemifene 60 mg/day: 58.6 (6.3) Placebo: 58.9 (6.1) BMI, mean (SD) kg/m² Ospemifene 30 mg/day: 26.4 (4.5) Ospemifene 60 mg/day: 26.0 (4.4) Placebo: 26.1 (4.4) Inclusion criteria Postmenopausal women aged 40 to 80 years, with the following criteria of VVA: 5% or less superficial cells on the vaginal smear (maturation index), vaginal pH greater than 5.0, and at least one moderate or severe symptom of VVA. Exclusion criteria 1. Endometrial thickness of 4mm or greater on centrally read transvaginal ultrasound 2. Pathological findings on endometrial biopsy or Papanicolaou test 3. Any other clinical	Interventions 30 or 60 mg/day of ospemifene or placebo. Study medication taken in the morning. All women were provided with a nonhormonal luubricant for use as needed throughout treatment period.	Details Participants randomized in a 1:1:1 ratio Tablets and packaging were identical in appearance.	Results EFFICACY endpoints 1. Percentage of superficial cells on the vaginal smear at week 12 2. Percentage of parabasal cells on the vaginal smear at week 12 3. Vaginal pH at week 12 4. Self-assessed symptoms of dyspareunia at week 12 SAFETY endpoints 1. Endometrial thickness 2. Endometrial histology 3. Treatment emergent adverse events ACCEPTABILITY endpoints Withdrawal due to adverse events QUALITY OF LIFE endpoints Not evaluated EFFICACY Superficial cells, percentage change from baseline to week 12 Ospemifene 30 mg/day: 7.8 Ospemifene 60 mg/day: 10.8 Placebo: 2.2 P < 0.001 Parabasal cells, percentage change from baseline to week 12 Ospemifene 30 mg/day: -21.9 Ospemifene 60 mg/day: -30.1 Placebo: 3.98 P < 0.001 Maturation index Significant improvement in maturation index for both ospemifene groups after 4 weeks of treatment P < 0.001 Vaginal pH, change from baseline to week 12 Ospemifene 30 mg/day: -0.67	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	significant gynaecological abnormality other than VVA (eg. uterine bleeding of unknown origin) 4. Body mass index of 37 kg/m² or greater 5. Systolic blood pressure of 180 mmHg or diastolic blood pressure of 100 mmHg or higher 6. Abnormal breast examination or mammogram results 7. Suspicion of malignancy or history of any malignancy within 10 years 8. Current or past thromboembolic or blood coagulation disorder 9. Women who consumed more than 14 drinks of alcohol per week 10. Women currently using itraconazole, ketoconazole, or digitalis alkaloids 11. Use of any HT (unless the woman had a sufficient washout period before any procedures (eg. 14 days for vaginal estrogens and 60 days for oral/transdermal therapy)			Ospemifene 60 mg/day: -1.01 Placebo: -0.10 P < 0.001 Vaginal dryness, change in symptom score at 12 weeks Ospemifene 30 mg/day: -1.22 Ospemifene 60 mg/day: -1.26 Placebo: -0.84 Significant for both ospemifene groups compared to placebo Dyspareunia, change in symptom score at 12 weeks Ospemifene 30 mg/day: -1.02 Ospemifene 60 mg/day: -1.19 Placebo: -0.89 Only significant for the 60 mg ospemifene compared to placebo SAFETY Endometrial thickness, mean (SD) change from baseline, mm Ospemifene 30 mg/day: 0.42 (1.35) Ospemifene 60 mg/day: 0.72 (1.59) Placebo: -0.02 (1.03) Endometrial hyperplasia or carcinoma No cases reported Treatment emergent adverse events Incidence of adverse events similar across treatment groups ACCEPTABILITY Withdrawal due to adverse events 5% in each group discontinued the study because of adverse events	B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - 5% of participants in each treatment group C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Yes Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information Used results for the 60 mg dosage of Ospemifene as the standard deviation of the means were reported by the previous review.
Full citation Goldstein,S.R., Bachmann,G.A., Koninckx,P.R., Lin,V.H., Portman,D.J., Ylikorkala,O., Ospemifene Study Group., Ospemifene 12-month safety and efficacy in postmenopausal women with vulvar and vaginal atrophy,	Sample size N = 426 Ospemifene 60 mg/day: 363 Placebo: 63 Characteristics Postmenopausal women 40-80 years of age, with vulvar and vaginal atrophy, defined as having	Interventions 60 mg ospemifene (or matching placebo) taken orally each morning with food.	Details Women randomized in a 6:1 ratio to ospemifene or matching placebo by sequential allocation of randomization number. Randomization stratified by study center.	Results EFFICACY endpoints 1. Percentage of superficial cells in the maturation index on the vaginal smear 2. Percentage of parabasal cells in the maturation index on the vaginal smear 3. Vaginal pH SAFETY endpoints Endometrial thickness	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Climacteric, 17, 173-182, 2014 Ref Id 319531 Country/ies where the study was carried out 23 sites in Europe Study type Randomized double-blind placebo-controlled parallel-group study Aim of the study Assessment of 12-month safety of ospemifene 60 mg/daily for the treatment of postmenopausal women with vulvar and vaginal atrophy. Study dates October 2007 to July 2009 Source of funding Hormos Medical Ltd, subsidiary of QuatRx Pharmaceuticals. Shionogi Inc.	a proportion of superficial cells ≤ 5% in the vaginal smear and a vaginal pH > 5. Age, mean (SD) years Ospemifene 60 mg/day: 61.7 (6.2) Placebo: 62.9 (6.5) BMI, mean (SD) kg/m² Ospemifene 60 mg/day: 24.7 (2.9) Placebo: 24.1 (2.9) Inclusion criteria Intact uterus and normal findings (except for atrophic vaginal signs) on pelvic examination, breast palpation, and recent mammogram. Subjects were not enrolled based on symptoms (ie. vaginal dryness or dyspareunia). Exclusion criteria Abnormal endometrial histology other than atrophy based on baseline biopsy, uterine bleeding of unknown origin or clinically significant abnormal gynaecological findings.			ACCEPTABILITY endpoints Not evaluated for 12 weeks. QUALITY OF LIFE endpoints Not evaluated EFFICACY Maturation index Superficial cells, median (range) percentage / mean (SD) change from baseline to week 12 Ospemifene 60 mg/day: 5 (-5, 60.0) / 5 (10.8) Placebo: 0 (-5, 28) / 0 (8.25) P < 0.0001 Parabasal cells, median (range) percentage / mean (SD) change from baseline to week 12 Ospemifene 60 mg/day: -40 (-100, 75) / -40 (29.2) Placebo: 0 (-90, 98) / 0 (47) P < 0.0001 Vaginal pH, mean (SD) change from baseline to week 12 Ospemifene 60 mg/day: -1.21 (0.912) Placebo: -0.16 (0.945) P < 0.0001 SAFETY Endometrial thickness, mean (SD) change from baseline to week 12, mm Ospemifene 60 mg/day: 0.44 (1.7) Placebo: 0.31 (1.5)	of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all majorconfounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? 96.1% and 98.4% completed treatment at week 12. C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information Was a 52 week RCT but efficacy outcomes were reported at 12-weeks. Long-term outcomes have been reported in long-term review question.
Full citation Karoussos,K.E., Studer,S., Wyss,H.J., The treatment of atrophic vaginal conditions with Ortho-Gynest A pilot study, Journal of International Medical Research, 7, 569-572, 1979 Ref Id 291535 Country/ies where the study was carried out Switzerland Study type Open pilot study. Observational study (pre and post intervention study). Aim of the study To evaluate the efficacy and	Sample size N=24 Characteristics Postmenopausal women with atrophic vaginal changes. Age range: 50-72 years; Mean: 61.1 years Onset of menopause: 1-23 years; Mean: 10.9 years Inclusion criteria 1. Normal physiological postmenopausal state with atrophic vaginal epithelial changes. 2. Post-operative postmenopausal state with atrophic vaginal epithelial changes.	Interventions Ortho-Gynest suppositories (contains 0.5 mg oestriol per suppository).	Details Study duration: 3 months Tests performed prior to commencing treatment 1. Cytological smear of the fornix. 2. Cervical smear. 3. Iodine test for glycogen content. 4. Examination of vulva and vagina. Schedule of treatment 1. 1 supp per day for first 7 days 2. 2 supp per week from day 7 to week 4 3. 2 supp per week from	Results EFFICACY endpoints 1. Dyspareunia 2. Pruritus 3. Vaginal cytological index 4. Appearance of vagina SAFETY endpoints Treatment-related adverse events ACCEPTABILITY endpoints Withdrawal due to treatment related adverse events QUALITY OF LIFE endpoints Not evaluated EFFICACY Vaginal cytological index	Limitations Other information NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study):

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
incidence of side-eefects associated with the use of Ortho-Gynest vaginal suppositories. Study dates Not reported. Study published in 1979. Source of funding Not reported.	3. Combination of inflammatory vaginal epithelial changes and other postmenopausal signs. Exclusion criteria 1. Suspected or diagnosed pregnancy. 2. Suspected or established estrogen-dependent neoplasia. 3. Suspected or confirmed carcinoma of the breast. 4. Blood-stained discharge per vaginam without any evident reason.		week 4 to month3	Increase in vaginal index Clinical evaluation of the appearance of the vagina 1. No change in thickness of vulval epithelium. 2. Narrowing of vagina improved. 3. Improvement of atrophic changes. SAFETY Treatment related adverse events 4 complained of side-effects: Unpleasant burning sensation, lower abdominal sensation, nausea and malaise, pruritus, spotting. ACCEPTABILITY Withdrawal due to treatment related adverse effects 2 patients withdrew because of side-effects 17 patients completed follow-up	N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: N/A A3. The groups were comparable at baseline, including all major confounding and prognostic factors: N/A Level of risk: Unclear risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: N/A B3. Individuals administering care were kept 'blind' to treatment allocation: N/A Level of risk: Unlear risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): N/A C2a. How many participants did not

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					complete treatment in each group? 7/24 did not complete followup. C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): Unclear C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: Unclear risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					confounding and prognostic factors: N/A Level of bias: Low risk of bias
Full citation Portman,D., Palacios,S., Nappi,R.E., Mueck,A.O., Ospemifene, a non-oestrogen selective oestrogen receptor modulator for the treatment of vaginal dryness associated with postmenopausal vulvar and vaginal atrophy: A randomised, placebo-controlled, phase III trial, Maturitas, 78, 91-98, 2014 Ref Id 319560 Country/ies where the study was carried out USA Study type Randomised, double-blind, parallel-group, multicentre phase III 12-week study Aim of the study To evaluate the efficacy and safety of ospemifene in the treatment of vaginal dryness in postmenopausal women with vulvovaginal atrophy Study dates July 2008 to August 2009 Source of funding QuatRx Pharmaceuticals Company	Sample size N = 314 Ospemifene 60 mg/day = 160 Placebo = 154 Characteristics Womem aged 40-80 years with diagnosed vulvovaginal atrophy and moderate or severe symptoms of vaginal dryness Age, mean (SD) years Ospemifene 60 mg/day - 59.9 (6.7) Placebo - 59.3 (7.0) BMI, mean (SD), kg/m² Ospemifene 60 mg/day - 27.2 (4.6) Placebo - 26.5 (4.6) Inclusion criteria Naturally or surgically menopausal Moderate or severe symptoms of vaginal atrophy 5% or fewer superficial cells in maturation index of vaginal smear Vaginal pH greater than 5.0 Self-reported most bothersome symptom of vaginal dryness or vaginal pain associated with sexual activity, with a severity of moderate or severe at randomization Exclusion criteria BMI ≥ 37 kg/m², the	Interventions One daily 60 mg ospemifene or placebo that were identical in appearance.	Details Participants took a one-daily dose of study medication with food in the morning for 12 weeks. Participants seen on weeks 4 and 12 for completion of VVA symptom questionnaire, assessment of vaginal pH, vaginal smear, and visual examination of vagina. Transvaginal ultrasound and endometrial biopsy conducted on week 12.	Results EFFICACY endpoints 1. Percentage of superficial cells in the maturation index on the vaginal smear 2. Percentage of parabasal cells in the maturation index on the vaginal smear 3. Vaginal pH 4. Severity of vaginal dryness SAFETY endpoints 1. Endometrial thickness 2. Endometrial histology 3. Treatment-related adverse events ACCEPTABILITY endpoints Withdrawal due to adverse events QUALITY OF LIFE endpoints Not evaluated EFFICACY Superficial cells, mean percentage (SD) change from baseline to week 12 Ospemifene 60 mg/day: 7.0 (11.5) Placebo: 0.0 (11.3) P < 0.001 Parabasal cells, mean percentage (SD) change from baseline to week 12 Ospemifene 60 mg/day: -31.7 (26.7) Placebo: -3.9 (27.1) P < 0.001 Vaginal pH, mean (SD) change from baseline to week 12 Ospemifene 60 mg/day: -0.95 (0.847) Placebo: -0.25 (0.844) P < 0.001 Severity of vaginal dryness, mean (SD) change in severity score from baseline to week 12 Ospemifene 60 mg/day: -1.3 (1.08)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Unclear A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	presence of clinically sugnificant abnormaol gynaecological findings other than signs of vaginal atrophy and concomitant hormonal medications, SERMs, or products expected to have oestrogenic and/or antioestogenic effects.			Placebo: -1.1 (1.02) P = 0.08 SAFETY Endometrial thickness, mean (SD) change from baseline to week 12, mm Ospemifene 60 mg/day: 0.82 (1.68) Placebo: -0.11 (1.20) *Assessed in only patients with an intact uterus Endometrial hyperplasia or carcinoma No cases reported Treatment related adverse events, n (%) Ospemifene 60 mg/day: 43 (26.9) Placebo: 18 (11.7) ACCEPTABILITY Withdrawal due to treatment related adverse events, n (%) Ospemifene 60 mg/day: 12 (7.5) Placebo: 5 (3.2)	treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - See results C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Yes Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information Two sets of analyses undertaken: Primary analyses: Intent-to-treat population Subsidiary analyses: Per-protocol population - consisted of all participants who had completed at least 10 weeks of treatment and had taken 85% or more of study medication. Efficacy and safety of ospemifene demonstrated using ITT analyses.

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Full citation Portman,D.J., Bachmann,G.A., Simon,J.A., Ospemifene Study Group., Ospemifene, a novel selective estrogen receptor modulator for treating dyspareunia associated with postmenopausal vulvar and vaginal atrophy, Menopause, 20, 623-630, 2013 Ref Id 254703 Country/ies where the study was carried out 110 sites in the United States Study type Multicenter phase 3 randomized, double-blind, parallel-group design study Aim of the study To compare the efficacy, safety, and tolerability of ospemifene 60 mg/day versus placebo in the treatment of moderate to severe dyspareunia in postmenopausal women with vulvar and vaginal atrophy (VVA). Study dates July 2008 to August 2009 Source of funding QuatRx Pharmaceuticals Company	Sample size N= 605 Ospemifene 60 mg/day = 303 Placebo = 302 Characteristics Most participants were white (90.6%) aged 40 to 79 years and had BMI values ranging from 16.7 to 37.1 kg/m² Inclusion criteria 1. Postmenopausal women aged 40 to 80 years who reported having moderate or severe vaginal pain (dyspareunia) with sexual activity as their most bothersome symptom. 2. Having VVA, defined as 5% or less superficial cells in the maturation index of the vaginal smear and a vaginal pH higher than 5. 3. Either hysterectomized or had an intact uterus with a double-layer endometrial thickness less than 4 mm and had no evidence of hyperplasia, cancer, or other pathology. 4. Negative Papanicolaou test result or lacked an intact cervix. 5. Negative mammogram result 9 months or less before randomization. 6. Normal breast examination result at screening. 7. Provided written informed consent. Exclusion criteria 1. BMI of 37 kg/m² or higher 2. SBP of 180 mmHg or	Interventions 60 mg/daily ospemifene or placebo with food in the morning for 12 weeks.	Details Ospemifene and placebo supplied as tablets identical in appearance. Nonhormonal vaginal lubricant provided to all participants and used as needed. Participants seen on weeks 4 and 12 for assesment. Participants underwent transvaginal ultrasound and endometrial biopsy on week 12.	Results EFFICACY endpoints 1. Percentage of superficial cells in the maturation index on the vaginal smear 2. Percentage of parabasal cells in the maturation index on the vaginal smear 3. Vaginal pH 4. Severity of dyspareunia associated with sexual intercourse SAFETY endpoints 1. Endometrial thickness 2. Endometrial histology 3. Treatment-related adverse events ACCEPTABILITY endpoints Withdrawal due to treatment-related adverse events QUALITY OF LIFE endpoints Not evaluated EFFICACY Superficial cells, mean percentage (SD) change from baseline to week 12 Ospemifene 60 mg/day: 12.3 (14.8) Placebo: 1.7 (6.9) P < 0.0001 Parabasal cells, mean percentage (SD) change from baseline to week 12 Ospemifene 60 mg/day: -40.2 (38.8) Placebo: 0.0 (30.0) P < 0.0001 Vaginal pH, mean (SD) change from baseline to week 12 Ospemifene 60 mg/day: -0.94 (1.0) Placebo: -0.07 (0.8) P < 0.0001 Dyspareunia, mean (SD) change in severity score from baseline to week 12 Ospemifene 60 mg/day: -1.5 (1.1) Placebo: -1.2 (1.1) P < 0.0001	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Unclear A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	DBP of 100 mgHg or higher 3. Clinically significant abnormal gynaecological findings. 4. Other signs of vaginal atrophy such as: uterine bleeding of unkown origin, uterine polyps or symptomatic and/or large uterine fibroids (> 3 cm), or vaginal infection requiring medication. 5. Significant abnormal findings on physical examination, mammography, ECG, safety lab tests, or liver function screening. 6. More than 14 alcoholic drinks per week. 7. Took heparin, digitalis alkaloids, or strong cytochrome P450 3A4 inhibitors 8. Used any hormonal medications, SERMs, or products expected to have estrogenic and/or antoestrogenic effects within prespecified time frames before study screening. 9. Used ospemifene before study screening. 10. Women who were positive for factor V Leiden mutation or had current or past cerebrovascular incidents, thromboembolic disorders, blood coagulation disorders, severe hepatic or renal impairment, or suspicion of malignancy on mammography within 10 years.			Percentage of participants reporting no vaginal pain after sexual activity on week 12 Ospemifene 60 mg/day: 38.0 Placebo: 28.1 *Ospemifene demonstrated statistically significant efficacy compared to placebo for all 4 efficacy parameters. SAFETY Endometrial thickness, mean (SD) change from baseline to week 12, mm Ospemifene 60 mg/day: 0.40 (1.25) Placebo: 0.10 (1.29) *Ospemifene caused a slight increase in endometrial thickness Endometrial hyperplasia or carcinoma No cases reported Adverse events, n (%) Ospemifene 60 mg/day: 79 (26.1) Placebo: 44 (14.6) ACCEPTABILITY Withdrawal due to treatment related adverse events, n (%) Ospemifene 60 mg/day: 10 (3.3) Placebo: 4 (1.3)	allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - 4.6% in ospemifene group and 3.3% in placebo group C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Yes Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information Two sets of analyses undertaken: Primary analyses: Intent-to-treat population Subsidiary analyses: Per-protocol population - consisted of all participants who had completed at least 10 weeks of treatment and had taken 85% or more of study medication. Efficacy and safety of ospemifene demonstrated using ITT analyses.
Full citation Rutanen,E.M., Heikkinen,J.,	Sample size N = 160	Interventions Three different	Details Participants had a washout	Results EFFICACY endpoints	Limitations NICE guidelines manual

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Halonen,K., Komi,J., Lammintausta,R., Ylikorkala,O., Effects of ospemifene, a novel SERM, on hormones, genital tract, climacteric symptoms, and quality of life in postmenopausal women: a double-blind, randomized trial, Menopause, 10, 433-439, 2003 Ref Id 227258 Country/ies where the study was carried out Finland Study type Double-blind randomised controlled study Aim of the study Effects of three different daily doses of ospemifene on hormone levels, genital tract organs, climacteric symptoms, and quality of life. Study dates Not reported. Source of funding Hormos Medical Corporation	Ospemifene 30 mg/day = 40 Ospemifene 60 mg/day = 40 Ospemifene 90 mg/day = 40 Placebo = 39 1 woman in placebo group did not start treatment at all. Characteristics No differences in baseline characteristics between treatment groups Age, mean (SD) Ospemifene 30 mg/day: 56.9 (4.5) Ospemifene 60 mg/day: 56.9 (4.7) Ospemifene 90 mg/day: 57.6 (4.3) Placebo: 58.2 (5.4) BMI, mean (SD) Ospemifene 30 mg/day: 24.4 (2.4) Ospemifene 60 mg/day: 25.0 (3.0) Ospemifene 90 mg/day: 25.1 (3.3) Placebo: 24.5 (2.7) Inclusion criteria 1. Healthy postmenopausal women aged 45 to 65 years 2. At least 12 months post last spontaneous menstrual bleed 3. FSH levels exceeding 40 IU/L and E2 levels below 0.11 nmol/L Exclusion criteria 1. BMI of 30 kg/m² or more 2. Blood pressure of 160/105 mmHg or higher 3. Pathological finding on gynaecological	doses (30, 60, or 90 mg daily) of ospemifene or placebo for 3 months.	period of 90 days for any systemic hormone medications or for 30 days for vaginal estrogen medication. Prestudy screening included clinical examination and laboratory assessments. Endometrial thickness measured by vaginal ultrasonography at screening and at 3 months.	1. Percentage of parabasal, intermediate, and superficial cells on the vaginal smear SAFETY endpoints 1. Endometrial thickness 2. Endometrial histology 3. Adverse events ACCEPTABILITY endpoints Withdrawal due to adverse events QUALITY OF LIFE endpoints Changes in Work Ability Index in depression, anxiety, or activity (self-confidence) EFFICACY Changes in parabasal, intermediate, and superficial cells during treatment period Clear difference between ospemifene and placebo groups in mean changes in these cells (P<0.05) Significant differences in pairwise comparisons SAFETY Endometrial thickness, mean (SD) change from baseline, mm Ospemifene 30 mg/day: 0.64 (1.14) P<0.05 Ospemifene 60 mg/day: 0.54 (1.01) P<0.05 Ospemifene 90 mg/day: 0.42 (0.82) P<0.05 Placebo: -0.01 (0.69) All ospemifene groups differed significantly from placebo. No differences in endometrial thickness were noticeable among the differing ospemifene dose levels Endometrial histology Endometrium remained atrophic after 3 months. Adverse events Frequency of participants reporting adverse events similar across treatment groups ACCEPTABILITY Withdrawal due to adverse events	2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Unclear A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Unclear A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Unclear risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	examination or pap smear 4. Endometrial thickness of 5mm or more 5. Uterine fibroids more than 5 cm in diameter 6. Known endometrial polyps or submucous fibroids 7. Current or history of any malignancy of the reproductive organs or breasts 8. Any other hormone-dependent malignancy 9. Any present drug therapy except thyroxin			Ospemifene 30 mg/day: 1 Ospemifene 60 mg/day: 3 Ospemifene 90 mg/day: 1 Placebo: 0 Side effects included: headache, facial numbness, nausea, dizziness, or ameba infection QUALITY OF LIFE No differences in quality of life indices at baseline or at 3 months.	C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - See results C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information Were not clear on whether adverse events were treatment related.
Full citation Voipio,S.K., Komi,J., Kangas,L., Halonen,K., DeGregorio,M.W., Erkkola,R.U., Effects of ospemifene (FC-1271a) on uterine endometrium, vaginal maturation index, and hormonal status in healthy postmenopausal women, Maturitas, 43, 207-214, 2002 Ref Id 227527 Country/ies where the study was carried out Finland Study type Double-blind, placebo-	Sample size N=40 25 mg ospemifene = 8 50 mg ospemifene = 8 100 mg ospemifene = 8 200 mg ospemifene = 8 Placebo = 8 Characteristics Healthy postmenopausal Caucasian females Age, mean (SD) years 25 mg ospemifene = 60 (4.0) 50 mg ospemifene = 62 (4.5) 100 mg ospemifene = 60 (4.6)	Interventions Oral doses of ospemifene 25 mg ospemifene; 50 mg ospemifene; 100 mg ospemifene; 200 mg ospemifene; or matching Placebo for 12 weeks.	Details Gynaecological examination, measurement of the double-layer thickness of the uterine endometrium, vaginal maturation index were performed and endometrial biopsy taken at baseline and at 12 weeks' visit. Estrogenic effects on vaginal epithelium estimated by routine maturation index. Visual analogue scale used to assess vaginal dryness.	Results EFFICACY endpoints 1. Percentage of parabasal cells in the maturation index on the vaginal smear 2. Percentage of intermediate cells in the maturation index on the vaginal smear 3. Percentage of superficial cells in the maturation index on the vaginal smear 4. Vaginal dryness SAFETY endpoints 1. Endometrial thickness 2. Endometrial histology 3. Treatment-related adverse events ACCEPTABILITY endpoints Withdrawal due to treatment related adverse events	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Unclear A2. There was adequate concealment of allocation

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
controlled phase I study Aim of the study To investigate the effects of ospemifene on the uterine endometrium, vaginal maturation index, and hormonal status in healthy postmenopausal women with an atrophic vaginal epithelium. Study dates Not reported. Source of funding Not reported.	200 mg ospemifene = 62 (5.1) Placebo = 62 (4.6) Inclusion criteria Postmenopausal, 55-75 years of age, body weight between 50-90 kg, in good general health, with an intact uterus. Exclusion criteria 1. Use of any hormonal medication (thyroxin allowed) during the 12 previous months 2. Strong susceptibility to allergic reactions 3. Participation in a drug study or blood donation within 60 days prior to the study 4. Evidence of clinically significant cardiovascular, renal, hepatic, hematological, gastrointestinal, pulmonary, metabolic, neurological or psychic disease or continuous medication to these diseases 5. Excessive use of alcohol			QUALITY OF LIFE endpoints Not evaluated EFFICACY Parabasal cells Decrease in percentage of cells for all ospemifene doses Intermediate cells Increase in percentage of cells for all ospemifene doses Superficial cells Increase in percentage of cells for all ospemifene doses Vaginal dryness No statistical significant difference between treatment groups. SAFETY Endometrial thickness, median (range) change from baseline, mm Treatment arm Baseline 12 weeks 25 mg ospemifene 2.38 (0.62) 1.65 (0.23) 50 mg ospemifene 2.40 (1.32) 3.48 (4.59) 100 mg ospemifene 2.38 (0.78) 2.38 (1.22) 200 mg ospemifene 1.40 (0.18) 2.20 (1.08) Placebo 2.38 (0.78) 1.93 (0.31) No clinically significant changes seen in endometrial thickness at any dose level Endometrial histology Weak effect of ospemifene on endometrial histology. No secretory changes or hyperplasia observed. Treatment-related adverse events Generally, ospemifene well tolerated ACCEPTABILITY	(such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Unclear A3. The groups were comparable at baseline including all majorconfounding and prognostic factors - Yes Unclear risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - 1 each in two

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				Withdrawal due to adverse effects, n 50 mg ospemifene: 1 due to gallstones and pancreatitis 200 mg ospemifene: 1 due to hot flushes, dizziness, and chest pain	treatment groups did not complete treatment C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious
Full citation Constantine, G. D., Goldstein, S. R., Archer, D. F., Endometrial safety of ospemifene: results of the phase 2/3 clinical development program, Menopause, 22, 36-43, 2015 Ref Id 338232 Country/ies where the study was carried out 23 sites in Europe Study type Six randomised, phase 2/3 double-blind, placebo controlled, parallel-group studies Aim of the study To assess the endometrial safety of ospemifene based on phase 2/3 clinical trials of postmenopausal women with up to 52 weeks of exposure to ospemifene 60 mg/day versus placebo Study dates Not reported Source of funding Shionogi Inc.	Sample size N=2166 women with 1863 completing the study. Ospemifene 60 mg/day: 1,242 women Placebo: 924 Number completed the study, n (%): Ospemifene 60 mg/day: 1061 (85.4) Placebo: 802 (86.8) Characteristics Postmenopausal women 40-80 years of age, with vulvar and vaginal atrophy, defined as having a proportion of superficial cells ≤ 5% in the vaginal smear and a vaginal pH > 5. Age, mean (SD) years Ospemifene 60 mg/day: 59.4 (6.49) Placebo: 58.9 (6.24) BMI, mean (SD) kg/m² Ospemifene 60 mg/day: 25.7 (4.03) Placebo: 26.0 (4.20) Women with intact uterus, n (%) Ospemifene 60 mg/day: 851 (68.5)	Interventions 60 mg ospemifene (or matching placebo) taken orally each morning with food	Details Participants were randomized 1:1 to ospemifene 60 mg/day or placebo in one 6-week trial and three 12-week trials; one of the 12-week trials had a 40-week extension study. In a separate 52-week trial, women were randomized 6:1 to ospemifene 60 mg/day or placebo by sequential allocation of randomization number. Randomization stratified by study center. Endometrial safety was assessed by endometrial histology (biopsy), transvaginal ultrasound, and gynecologic examination.	Results Short term outcomes at 12 weeks EFFICACY endpoints 1. Percentage of superficial cells in the maturation index on the vaginal smear 2. Percentage of parabasal cells in the maturation index on the vaginal smear 3. Vaginal pH 4. Vaginal atrophy 5. Vaginal dryness 6. Dyspareunia 7. Itching and discomfort SAFETY endpoints 1. Endometrial thickness 2. Breast pain/blood oestradiol levels 3. Treatment-emergent adverse events ACCEPTABILITY endpoints Not evaluated for 12 weeks. QUALITY OF LIFE endpoints Not evaluated EFFICACY Superficial cells, median (range) percentage / mean (SD) change from baseline to week 12 Not reported Parabasal cells, median (range) percentage / mean (SD) change from baseline to week 12 Not reported Vaginal pH, mean (SD) change from baseline to week 12	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Placebo: 543 (58.8) Inclusion criteria Postmenopausal women with vulvar and vaginal atrophy (5% or less superficial cells on vaginal smear (maturation index), vaginal pH higher than 5.0, and at least one moderate or severe symptom of VVA) In three of the studies, participants were required to have an intact uterus: One 12-week study (N = 79), the 40-week long-term extension study (N = 118), and the 52-week long term safety study (N = 426) required participants to have an intact uterus Exclusion criteria Abnormal endometrial histology other than atrophy based on baseline biopsy, uterine bleeding of unknown origin, clinically significant abnormal gynecologic findings, endometrial thickness of 4 mm or more on centrally read TVUS, pathologic findings on endometrial biopsy or Papanicolaou test, or clinically significant findings on physical examination			Not reported Vaginal atrophy Not reported Vaginal dryness Not reported Dyspareunia Not reported Itching and discomfort: Not reported SAFETY Endometrial thickness, mean (SD) change from baseline to week 12, mm Ospemifene 60 mg/day: 0.51 (1.5) Placebo: 0.06 (1.2) Breast pain/blood oestradiol levels Not reported Treatment-emergent adverse events Not reported	B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? 85.4% and 86.8% completed treatment in the ospemifene and placebo group respectively. C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information Long-term outcomes have been reported in long-term review question. This study consists of some data on women in

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Goldstein's 2014 study.

H.5.2 Local oestrogens for long-term treatment

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Full citation Iosif,C.S., Effects of protracted administration of estriol on the lower genito urinary tract in postmenopausal women, Archives of Gynecology and Obstetrics, 251, 115-120, 1992 Ref Id 226712 Country/ies where the study was carried out Sweden Study type Observational study Aim of the study To examine the effect of protracted administration of estriol in the lower genito-urinary tract symptoms Study dates 1980 to 1989 Source of funding Not reported	Sample size N = 48 Characteristics Age (years) - Mean (range) 59.2 (57 - 65) Time since last period (years) - Mean (range) 9.1 (5 - 15) Ethnicity White Not reported Dyspareunia - n (%) Not reported Vaginal Dryness - n (%) Not reported Inclusion criteria Women had symptoms of vaginal atrophy, urinary incontinence, or recurrent urinary tract infections Exclusion criteria Women with a proliferative endometrium	Interventions Women were given long-term treatment with vaginal suppositories containing 0.5 mg oestriol (Organon). Dose used was one vaginal suppository every evening for first two weeks and then one vaginal suppository twice a week for the remainder of the study. Were followed for 8-10 years	Details To exclude women with a proliferative endometrium, medroxy-progesterone 5mg was given once a day for 7 days two weeks before starting oestrogen treatment and no women entering the study had a withdrawal bleed. Endometrial samples were taken 8 - 10 years after starting treatment. The women had a gynecological examination prior to the treatment as weel as at 3 months, 6 months and once a year up to 10 years after starting treatment.	Results Efficacy parameters Symptoms of moderate to severe atrophic vaginitis Safety parameters 1. Endometrial histology 2. Treatment related adverse events EFFICACY Atrophic vaginitis (number symptom free at year 1) 31 of 32 SAFETY Endometrial histlogy, n (%) 7 (16.6) reported as proliferative endometrium over 8 - 10 years Treatment related adverse events 7 complained of vaginal pruritus 6 complained of local irritation and vaginal pain ACCEPTABILITY Withdrawal due to adverse events, n (%) Year 1: 9 (18.8) Year 2: 14 (19.2) Year 4: 16 (33.3)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: N/A A3. The groups were comparable at baseline, including all major confounding and prognostic factors: N/A Level of risk: Unclear risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): N/A C2a. How many participants did not complete treatment in each group? See results section C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: Unclear risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					to the intervention: Unclear D5. Investigators were kept 'blind' to other important confounding and prognostic factors: Unclear Level of bias: Low risk of bias Other information For the symptoms of atrophic vaginitis outcome, the paper reports that 98% of women were symptom free at 1 year so the NCC calculated the number from the women who had not dropped out at year 1 (48-16=32).
Full citation Ulrich,L.S., Naessen,T., Elia,D., Goldstein,J.A., Eugster-Hausmann,M., trial,investigators, Endometrial safety of ultra-low-dose Vagifem 10 microg in postmenopausal women with vaginal atrophy, Climacteric, 13, 228-237, 2010 Ref Id 227483 Country/ies where the study was carried out Denmark,Finland, France, Hungary,Norway, Sweden,Czech Republic Study type Observational study (non-comparative cohort study) Aim of the study To evaluate the endometrial safety of 10µg estradiol vaginal tablet in postmenopausal women with vaginal atrophy. Study dates January 2000 to November 2008 Source of funding Novo Nordisk A/S	Sample size N = 336 Characteristics Age (years) - Mean ± SD E = 59.5 ± 6.2 Time since last period (years) - Mean ± SD E = 9.4 ± 5.9 Ethnicity White - n (%) E = 296 (88.1%) Dyspareunia - n (%) Not reported Vaginal Dryness - n (%) Not reported Inclusion criteria Women were incldued if they were healthy, non-hysterectomized postmenopausal women aged 45 years or older at the time of screening, had their last menses or had a bilateral oophorectomy performed more than 2 years prior to the time of screening had one or more urogenital	Interventions Using the pre-loaded applicator, subjects inserted 10µg estradiol vaginal tablet once daily during the first 2 weeks of the study and in the remainder of the study subjects inserted one tablet twice weekly.	Details This was a 52 week open-label, multi-centre trial. Visits to screening centre: weeks 0, 8, 26, and 52. Phone consultations: weeks 16, 35 and 42. Endometrial biopsies used pipelle de Cornier preceded by transvaginal ultrasound at baseline and endpoint. Only women treated ≥3 months had endpoint biopsies.	Results Efficacy parameters Not evaluated Safety parameters 1. Endometrial thickness 2. Endometrial histology 3. Treatment related adverse events Acceptability parameters Withdrawal due to adverse events Quality of life parameters Not evaluated SAFETY Endometrial thickness, mean change from baseline, mm Decrease from 2.04 mm at study start to 1.94 mm after 52 weeks Endometrial hyperplasia or carcinoma No cases reported Treatment related adverse events, n(%)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: N/A A3. The groups were comparable at baseline, including all major confounding and prognostic factors: N/A Level of risk: Unclear risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	symptoms of moderate to severe intensity (as identified by the subject) including vaginal dryness, vaginal and/or vulvar irritation/itching, vaginal soreness, dysuria, dyspareunia, and vaginal bleeding associated with sexual activity. All women were required to have serum follicle stimulating hormone (FSH) levels 4 40 mIU/ml, serum estradiol520 pg/ml, 5% or fewer superficial cells in vaginal cytology, vaginal pH >5.0, endometrial thickness >4.0 mm as assessed by transvaginal ultrasound, and a normal mammogram within 6 months prior to enrolment into the trial. Exclusion criteria Women were excluded from the study if they had a known or suspected history of breast cancer or past estrogen-dependent neoplasia, endometrial hyperplasia or endometrial polyps diagnosed during the screening period, or abnormal genital bleeding of unknown etiology. Exposure to exogenous sex steroid hormone therapies within the past 3 months prior to the screening visit, hysterectomy or endometrial ablation, use of any vaginal or vulvar preparations 1 month prior to baseline, hot flushes requiring systemic hormonal therapy, active deep venous thrombosis or thromboembolic disorders,			186 (55.4) reported treat-emergent adverse events. None were judged to be related to study drug. ACCEPTABILITY Withdrawal due to adverse events, n (%) 18 (5.4%)	received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): N/A C2a. How many participants did not complete treatment in each group? 292 of 336 completed the study C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): Yes C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: Unclear risk of bias D. Detection bias (bias in how outcomes are ascertained,

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	active arterial thrombosis, known or suspected hepatic and/or renal impairment, porphyria, body mass index >35.0 kg/m2, Papanicolaou cervical smear test (Pap smear) presenting in Pap class >II, known or suspected vaginal infection requiring treatment, uterovaginal prolapse Grade II–IV POPQ (pelvic organ prolapse qualification scale), known diabetes mellitus, current use of steroid hormones				diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: Unclear D5. Investigators were kept 'blind' to other important confounding and prognostic factors: Unclear Level of bias: Unclear risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious
Full citation Simunic,V., Banovic,I., Ciglar,S., Jeren,L., Pavicic,Baldani D., Sprem,M., Local estrogen treatment in patients with urogenital symptoms, International Journal of Gynecology and Obstetrics, 82, 187-197, 2003 Ref Id 220302 Country/ies where the study was carried out Croatia Study type Randomised controlled trial Aim of the study To determine the efficacy and safety of low dose	Sample size N = 1612 17β-estraliol (E) = 828 PLacebo (P) = 784 Characteristics Age (years) - Mean ± SD E = 58.1 ± 6.9 P = 59.5 ± 7.1 Time since last period (years) - Mean ± SD E = 8.6 ± 3.5 P = 9.9 ± 3.8 Ethnicity White - n (%) Not reported Dyspareunia - n (%) E = 361 (43.6%) P = 298 (38.0%)	Interventions Women were randomised to receive either 25µg of micronized 17B-estradiol or placebo as vaginal tablets. The women were treated once a day over a 2 week period, and then twice a week for the remaining 12 months.	Details Assessments included a full history questionnaire, micturition diary, clincial (gynecologic) and cystometric examination, transvaginal ultrasound, and serum 17B-estradiol determination at the beginning, after 4 and 12 montsh of treatment	Results Efficacy parameters 1. Symptoms of vaginal atrophy (vaginal dryness, itching, burning, and dyspareunia) 2. Vaginal atrophy score index Safety parameters 1. Endometrial thickness 3. Treatment related adverse events Acceptability parameters 1. Withdrawal due to adverse events 2. Subjective assessment of acceptability by	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
(25µg) of micronized 17β-estradial administered vaginally in the management of patients with urogenital symptoms Study dates April 2000 to May 2001 Source of funding Not reported	Vaginal Dryness - n (%) E = 560 (67.6%) P = 504 (64.3%) Inclusion criteria Women with urogenital complains at least 1 year post-menopause Exclusion criteria Women were excluded if they had any hormone replacement therapy for at least six months any systemic disease or infection suspected or proven malignant disease unexplained uterine bleeding previous hysterectomy or surgical correction for genuine stress urinary incontinence acute gynecological infection			participants (Satisfaction rate) Quality of life parameters Not evaluated EFFICACY With symptoms of vaginal atrophy, n (%) Baseline E: 664 (84.8) P: 567 (77.3) P=0.412 After 12 months E: 121 (15.5) P: 430 (58.6) P=0.0013 Vaginal atrophy total score index, mean (SD) Baseline E: 1.95 (0.01) P: 2.19 (0.03) P=0.236 After 12 months E: 0.21 (0.02) P: 1.15 (0.04) P=0.026 SAFETY Endometrial thickness, mean (SD) mm Baseline E: 3.1 (0.4) P: 3.2 (0.3) P=0.432 After 12 months	A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - See results C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment.

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				E: 2.9 (0.5) P: 3.0 (0.4) P=0.324 Treatment related adverse events, n (%) E: 21 (2.7) P: 3.0 (0.4) No significant differences ACCEPTABILITY Withdrawal due to adverse events, n (%) E: 10 (1.3) P: Not reported No significant differences Satisfaction rate, % E: 84.5 P: 29.3	C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious
Full citation Gerbaldo,D., Ferraiolo,A., Croce,S., Truini,M., Capitanio,G.L., Endometrial morphology after 12 months of vaginal oestriol therapy in post-menopausal women, Maturitas, 13, 269-274, 1991	Sample size N = 23 Characteristics Age (years) - Mean ± SD 64.9 ± 9.2 Time since last period (years) - Mean ± SD Not reported	Interventions Women were given E3 Oestriol Vaginal cream 0.5mg (Colpogyn by Angelini Acraf) every day for the first 3 weeks and then 0.5mg twice weekly for 12 months	Details Prior to study, endometrial atrophy was assessed by hysteroscopy followed by endometrial biopsy. The same evaluation was repeated after weeks 6 and 12 of treatment.	Results Efficacy parameters Not evaluated Safety parameters 1. Endometrial thickness 2. Endometrial histology Acceptability parameters Not evaluated	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Ref Id 291560 Country/ies where the study was carried out Italy Study type Observational study (Non-comparative cohort study) Aim of the study To evaluate the endometrial response to long-term vaginal E3 treatment Study dates Not stated Source of funding Not stated	Ethnicity White - n (%) Not reported Dyspareunia - n (%) Not reported Vaginal Dryness - n (%) Not reported Inclusion criteria Non-obese, post-menopausal women complaining of urogenital atrophy Exclusion criteria Women were not included if the had receivec oestrogen therapy during year before study or if they were experiencing post-menopausal bleeding			Quality of life parameters Not evaluated SAFETY Endometrial thickness, mean change from baseline, mm Rsults not reported Endometrial histology Atrophic nature of endometrium confirmed	(that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: N/A A3. The groups were comparable at baseline, including all major confounding and prognostic factors: N/A Level of risk: Unclear risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): N/A C2a. How many participants did not complete treatment in each group? None C2b. The groups were comparable for treatment

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: Unclear risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Unclear D4. Investigators were kept 'blind' to participants' exposure to the intervention: Unclear D5. Investigators were kept 'blind' to other important confounding and prognostic factors: Unclear Level of bias: Unclear risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Full citation Simon,J., Nachtigall,L., Gut,R., Lang,E., Archer,D.F., Utian,W., Effective treatment of vaginal atrophy with an ultra-low-dose estradiol vaginal tablet.[Erratum appears in Obstet Gynecol. 2008 Dec;112(6):1392], Obstetrics and Gynecology, 112, 1053-1060, 2008 Ref Id 227345 Country/ies where the study was carried out Canada and United States Study type Randomised control trial Aim of the study To evaluate the efficacy and safety of ultra low dose 10microgram E2 oestradiol vaginal tablets in postmenopausal women with vaginal atrophy. Study dates March 2005 to May 2006 Source of funding Supported by Novo Nordisk A/S	Sample size N = 309 Estradiol (E) = 205 Placebo (P) = 104 Characteristics Age (years) - Mean ± SD E = 57.5 ± 5.64 P = 57.7 ± 5.27 Time since last period (years) - Mean ± SD E = 8.0 ± 5.8 P = 8.2 ± 5.3 Ethnicity White - n (%) E = 192 (93.7%) P = 95 (91.3%) Dyspareunia - n (%) Not reported Vaginal Dryness - n (%) Not reported Inclusion criteria Women were included if they were ≥45 years old. ≥2 years since last menses or oophorectomy. FSH >40 MI/mL ≥3 urogenital symptoms (including those of moderate to severe intensity). Serum E2 levels <20pg/mL ≤5% superficial cells in cytology test. Vaginal pH>5 Endometrial thickness <4mm Normal mammogram within 6 months of trial. Intact uterus Good general health with no significant illness. Exclusion criteria Women were excluded if they	Interventions Women were randomised (2:1) in blocks of 6 to either 10 micrograms E2 or placebo. All vaginal tables were identical in appearance.	Details All data reported at weeks 12 and 52 are from intent-to-treat analyses, with missing values for each individual imputed using last observation carried forward. The primary efficacy endpoints included mean change from baseline to week 12 in vaginal Maturation Index and Value, vaginal pH, and the mean score of most bothersome moderate to severe symptom as identified by the patient. The endometrial safety of the E2 tablet was evaluated through endometrial biopsies conducted at screening and at the end of the trial	Results Efficacy endpoints 1. Maturation index 2. Vaginal pH 6. Mean score for most bothersome urogenital symptom (dyspareunia and vaginal dryness) [0 = none, 3 = severe] Safety endpoints Treatment related adverse events Acceptability endpoints Withdrawal due to adverse events Quality of life endpoints Not evaluated EFFICACY Maturation index, mean change from baseline to week 52 10 E2 = 24.5 PLA = 5.9 Vaginal pH, participants with pH less than 5.5 at week 52, n (%) 10 E2 = 131 (64.8) PLA = 30 (29.4) Change in mean score for most bothersome urogenital symptom at week 52 10 E2 = -1.23 PLA = -0.87 P = 0.004 SAFETY Treatment related adverse events, n (%) 10 E2 = 158 (77) PLA = 77 (75)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	were allergic to treatment or its constituents. used of any investigational drug <30 days of treatment used exogenous sex hormones withi 3 months were using corticostedoids had a known or suspected history of breast carcinoma had genital bleeding of unknwon cause had acute thrombophlebitis or thromboembolic disorder associated with estrogen use had vaginal infection required treatment had any serious disease or condition that could interfere with study compliance			ACCEPTABILITY Withdrawal due to adverse events, n (%) 10 E2 = 11 (5) PLA = 5 (5)	to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - See results C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious
Full citation Goldstein,S.R., Bachmann,G.A., Koninckx,P.R., Lin,V.H., Portman,D.J., Ylikorkala,O., Ospemifene Study Group., Ospemifene 12-month safety and efficacy in postmenopausal women with vulvar and vaginal atrophy, Climacteric, 17, 173-182, 2014 Ref Id 319531 Country/ies where the study was carried out 23 sites in Europe Study type 52-week randomized double-blind placebo-controlled parallel-group study Aim of the study Assessment of 12-month safety of ospemifene 60 mg/daily for the treatment of postmenopausal women with vulvar and vaginal atrophy. Study dates October 2007 to July 2009 Source of funding Hormos Medical Ltd, subsidiary of QuatRx Pharmaceuticals.	Sample size N = 426 with 349 completing the study. Ospemifene 60 mg/day: 363 Placebo: 63 Characteristics Postmenopausal women 40-80 years of age, with vulvar and vaginal atrophy, defined as having a proportion of superficial cells ≤ 5% in the vaginal smear and a vaginal pH > 5. Age, mean (SD) years Ospemifene 60 mg/day: 61.7 (6.2) Placebo: 62.9 (6.5) BMI, mean (SD) kg/m² Ospemifene 60 mg/day: 24.7 (2.9) Placebo: 24.1 (2.9) Inclusion criteria Intact uterus and normal findings (except for atrophic vaginal signs) on pelvic examination, breast palpation, and recent mammogram. Subjects were not enrolled based on symptoms (ie. vaginal dryness or dyspareunia). Exclusion criteria Abnormal endometrial histology other than atrophy based on baseline biopsy, uterine bleeding of unknown origin or clinically	Interventions 60 mg ospemifene (or matching placebo) taken orally each morning with food.	Details Women randomized in a 6:1 ratio to ospemifene or matching placebo by sequential allocation of randomization number. Randomization stratified by study center.	Results EFFICACY endpoints 1. Vaginal dryness 2. Signs of vaginal atrophy SAFETY endpoints 1. Endometrial thickness 2. Endometrial histology 3. Treatment-emergent adverse events ACCEPTABILITY endpoints 1. Withdrawal due to treatment related adverse events 2. Compliance to treatment QUALITY OF LIFE endpoints Not evaluated EFFICACY Maturation index Vaginal dryness, percentage with no dryness at week 52 Ospemifene 60 mg/day: 81.5 Placebo: 32.1 P < 0.0001 Vaginal atrophy, percentage with no signs	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all majorconfounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Shionogi Inc.	significant abnormal gynecological findings.			of atrophy at week 52 Ospemifene 60 mg/day: 80 Placebo: 30 SAFETY Endometrial thickness, mean (SD) change from baseline to week 52, mm Ospemifene 60 mg/day: 0.75 (1.5) Placebo: 0.17 (1.3) Endometrial histological biopsy characteristics No tissue changes (hyperplasia or carcinoma) reported Treatment-emergent adverse events, n (%) Ospemifene 60 mg/day: 308 (84.6) Placebo: 47 (75.8) ACCEPTABILITY Withdrawals due to treatment related adverse events, n (%) Ospemifene 60 mg/day: 49 (13.5) Placebo: 6 (9.7) Compliance to treatment, % Ospemifene 60 mg/day: 95 Placebo: 99	B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? 81.0% and 87.3% completed treatment in the ospemifene and placebo group respectively. C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained,

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information Short-term outcomes of this study have been reported in short-term review question.
Full citation Simon,J.A., Lin,V.H., Radovich,C., Bachmann,G.A., Ospemifene Study Group., One-year long-term safety extension study of ospemifene for the treatment of vulvar and vaginal atrophy in postmenopausal women with a uterus, Menopause, 20, 418-427, 2013 Ref Id 319569 Country/ies where the study was carried out	Sample size N = 180 Ospemifene 30 mg/day = 62 Ospemifene 60 mg/day = 69 Placebo = 49 Characteristics Most participants were white aged 46 to 79 years with BMI values ranging from 15.7 to 36.8 kg/m² Inclusion criteria Postmenopausal women aged 40 to 80 years, with the following criteria of VVA: 5% or less superficial cells on the vaginal smear (maturation index), vaginal pH greater than 5.0, and at least	Interventions 30 or 60 mg/day of ospemifene or placebo for 40 additional weeks. Study medication taken in the morning.	Details 40-week safety extension of a 12-week, phase 3, efficacy and safety study. Blinding was according to the original blinding assignment for the 12-week study. Total duration was 52-weeks followed by a 4-week posttreatment follow-up period. Endometrial thickness assessed by transvaginal ultrasonography.	Results EFFICACY endpoints 1. Vaginal dryness SAFETY endpoints 1. Endometrial thickness 2. Endometrial histology 3. Adverse events ACCEPTABILITY endpoints 1. Withdrawal due to adverse events 2. Compliance to dosing schedules QUALITY OF LIFE	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
United States Study type Multicentre, randomized, double-blind 40-week extension study of a 12-week study (226136) Aim of the study To assess the safety of ospemifene for the treatment of vulvar and vaginal atrophy (VVA) in postmenopausal women with a uterus Study dates May 2006 to September 2008 Source of funding QuatRx Pharmaceuticals	one moderate or severe symptom of VVA. Exclusion criteria 1. Endometrial thickness of 4mm or greater on centrally read transvaginal ultrasound 2. Pathological findings on endometrial biopsy or Papanicolaou test 3. Any other clinical significant gynaecological abnormality other than VVA (eg. uterine bleeding of unknown origin) 4. Body mass index of 37 kg/m² or greater 5. Systolic blood pressure of 180 mmHg or diastolic blood pressure of 100 mmHg or higher 6. Abnormal breast examination or mammogram results 7. Suspicion of malignancy or history of any malignancy within 10 years 8. Current or past thromboembolic or blood coagulation disorder 9. Women who consumed more than 14 drinks of alcohol per week 10. Women currently using itraconazole, ketoconazole, or digitalis alkaloids 11. Use of any HT (unless the woman had a sufficient washout period before any procedures (eg. 14 days for vaginal estrogens and 60 days for oral/transdermal therapy)			endpoints Not evaluated EFFICACY Vaginal dryness Improvement in severity scores for vaginal dryness from baseline to both week 26 and 52 for both ospemifene doses compared to placebo SAFETY Endometrial thickness, mean (SD) change Ospemifene 60 mg/day: 1.14 (1.56) Placebo: -0.04 (1.15) Endometrial histology No hyperplasia or carcinoma reported Adverse events, n (%) Ospemifene 30 mg/day: 38 (61.3) Ospemifene 60 mg/day: 44 (63.8) Placebo: 22 (44.9) ACCEPTABILITY Withdrawal due to adverse events, n (%) Ospemifene 30 mg/day: 3 (4.8) Ospemifene 60 mg/day: 4 (5.8) Placebo: 1 (2.0) Compliance rates, mean % Ospemifene 30 mg/day: 85.5 Ospemifene 60 mg/day: 84.6 Placebo: 93.4	that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - See results C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Yes Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious
Full citation Constantine, G. D., Goldstein, S. R., Archer, D. F., Endometrial safety of ospemifene: results of the phase 2/3 clinical	Sample size N=2166 women with 1863 completing the study. Ospemifene 60 mg/day: 1,242 women Placebo: 924	Interventions 60 mg ospemifene (or matching placebo) taken orally each morning with food	Details Participants were randomized 1:1 to ospemifene 60 mg/day or placebo in one 6-week trial and three 12-week trials; one of the 12-week trials had a 40-week extension study. In a	Results Long term outcomes at 52 weeks EFFICACY endpoints 1. Vaginal dryness 2. Signs of vaginal	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
development program, Menopause, 22, 36-43, 2015 Ref Id 338232 Country/ies where the study was carried out 23 sites in Europe Study type Six randomised, phase 2/3 double-blind, placebo controlled, parallel-group studies Aim of the study To assess the endometrial safety of ospemifene based on phase 2/3 clinical trials of postmenopausal women with up to 52 weeks of exposure to ospemifene 60 mg/day versus placebo Study dates Not reported Source of funding Shionogi Inc.	Number completed the study, n (%): Ospemifene 60 mg/day: 1061 (85.4) Placebo: 802 (86.8) Characteristics Postmenopausal women 40-80 years of age, with vulvar and vaginal atrophy, defined as having a proportion of superficial cells ≤ 5% in the vaginal smear and a vaginal pH > 5. Age, mean (SD) years Ospemifene 60 mg/day: 59.4 (6.49) Placebo: 58.9 (6.24) BMI, mean (SD) kg/m² Ospemifene 60 mg/day: 25.7 (4.03) Placebo: 26.0 (4.20) Women with intact uterus, n (%) Ospemifene 60 mg/day: 851 (68.5) Placebo: 543 (58.8) Inclusion criteria Postmenopausal women with vulvar and vaginal atrophy (5% or less superficial cells on vaginal smear (maturation index), vaginal pH higher than 5.0, and at least one moderate or severe symptom of VVA) In three of the studies, participants were required to have an intact uterus: One 12-week study (N = 79), the 40-week long-term extension study (N = 118), and the 52-week long term safety study (N = 426) required participants to have an intact uterus		separate 52-week trial, women were randomized 6:1 to ospemifene 60 mg/day or placebo by sequential allocation of randomization number. Randomization stratified by study center. Endometrial safety was assessed by endometrial histology (biopsy), transvaginal ultrasound, and gynecologic examination.	atrophy 3. Dyspareunia 4. Itching and discomfort SAFETY endpoints 1. Endometrial thickness 2. Endometrial histology 3. Treatment-emergent adverse events ACCEPTABILITY endpoints 1. Withdrawal due to treatment related adverse events 2. Compliance to treatment QUALITY OF LIFE endpoints Not evaluated EFFICACY Vaginal dryness Not reported Vaginal atrophy Not reported Dyspareunia Not reported Itching and discomfort Not reported SAFETY Endometrial thickness, mean (SD) change from baseline to week 52, mm Ospemifene 60 mg/day: 0.81 (1.5) Placebo: 0.07 (1.2) Endometrial histological biopsy characteristics No tissue changes (hyperplasia with atypia	differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Exclusion criteria Abnormal endometrial histology other than atrophy based on baseline biopsy, uterine bleeding of unknown origin, clinically significant abnormal gynecologic findings, endometrial thickness of 4 mm or more on centrally read TVUS, pathologic findings on endometrial biopsy or Papanicolaou test, or clinically significant findings on physical examination			or carcinoma) reported Simple endometrial hyperplasia without atypia on biopsy 3 months after the last dose of the study drug was reported for one woman who received ospemifene 60 mg/d Treatment-emergent adverse events Not reported ACCEPTABILITY Withdrawals due to treatment related adverse events, n (%) Ospemifene 60 mg/day: 95 (7.6) Placebo: 34 (3.7) Compliance to treatment, n (%) Not reported	C2a. How many participants did not complete treatment in each group? 85.4% and 86.8% completed treatment in the ospemifene and placebo group respectively. C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Low risk of bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information Short-term outcomes of this study have been reported in short-term review question. This study consists of some data on women in Goldstein's 2014 study.

H.5.3 Short-term effectiveness of ospemifene

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Full citation Bachmann,G.A., Komi,J.O., Ospemifene Study Group., Ospemifene effectively treats vulvovaginal atrophy in postmenopausal women: results from a pivotal phase 3 study, Menopause, 17, 480-486, 2010 Ref Id 226136 Country/ies where the study was carried out 76 centers in the United States Study type Randomized, double-blind phase 3 study Aim of the study To evaluate the efficacy and safety of ospemifene in the treatment of vulvovaginal atrophy (VVA) in	Sample size N = 826 Ospemifene 30 mg/day: 282 Ospemifene 60 mg/day: 276 Placebo: 268 Characteristics Ninety percent of women in all groups were white. Age, mean (SD) years Ospemifene 30 mg/day: 58.4 (6.3) Ospemifene 60 mg/day: 58.6 (6.3) Placebo: 58.9 (6.1) BMI, mean (SD) kg/m² Ospemifene 30 mg/day: 26.4 (4.5) Ospemifene 60 mg/day: 26.0 (4.4) Placebo: 26.1 (4.4) Inclusion criteria Postmenopausal women aged 40 to 80 years, with the following criteria of VVA: 5% or less superficial cells on the vaginal smear	Interventions 30 or 60 mg/day of ospemifene or placebo. Study medication taken in the morning. All women were provided with a nonhormonal luubricant for use as needed throughout treatment period.	Details Participants randomized in a 1:1:1 ratio Tablets and packaging were identical in appearance.	Results EFFICACY endpoints 1. Percentage of superficial cells on the vaginal smear at week 12 2. Percentage of parabasal cells on the vaginal smear at week 12 3. Vaginal pH at week 12 4. Self-assessed symptoms of dyspareunia at week 12 SAFETY endpoints 1. Endometrial thickness 2. Endometrial histology 3. Treatment emergent adverse events ACCEPTABILITY endpoints Withdrawal due to adverse events QUALITY OF LIFE endpoints Not evaluated EFFICACY Superficial cells, percentage change from baseline to week 12 Ospemifene 30 mg/day: 7.8 Ospemifene 60 mg/day: 10.8 Placebo: 2.2 P < 0.001	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
postmenopausal women for 12-weeks. Study dates Not reported. Source of funding QuatRx Pharmaceuticals Company	(maturation index), vaginal pH greater than 5.0, and at least one moderate or severe symptom of VVA. Exclusion criteria 1. Endometrial thickness of 4mm or greater on centrally read transvaginal ultrasound 2. Pathological findings on endometrial biopsy or Papanicolaou test 3. Any other clinical significant gynaecological abnormality other than VVA (eg. uterine bleeding of unknown origin) 4. Body mass index of 37 kg/m² or greater 5. Systolic blood pressure of 180 mmHg or diastolic blood pressure of 100 mmHg or higher 6. Abnormal breast examination or mammogram results 7. Suspicion of malignancy or history of any malignancy within 10 years 8. Current or past thromboembolic or blood coagulation disorder 9. Women who consumed more than 14 drinks of alcohol per week 10. Women currently using itraconazole, ketoconazole, or digitalis alkaloids 11. Use of any HT (unless the woman had a sufficient washout period before any procedures (eg. 14 days for vaginal estrogens and 60 days for oral/transdermal therapy)			Parabasal cells, percentage change from baseline to week 12 Ospemifene 30 mg/day: -21.9 Ospemifene 60 mg/day: -30.1 Placebo: 3.98 P < 0.001 Maturation index Significant improvement in maturation index for both ospemifene groups after 4 weeks of treatment P < 0.001 Vaginal pH, change from baseline to week 12 Ospemifene 30 mg/day: -0.67 Ospemifene 60 mg/day: -1.01 Placebo: -0.10 P < 0.001 Vaginal dryness, change in symptom score at 12 weeks Ospemifene 30 mg/day: -1.22 Ospemifene 60 mg/day: -1.26 Placebo: -0.84 Significant for both ospemifene groups compared to placebo Dyspareunia, change in symptom score at 12 weeks Ospemifene 30 mg/day: -1.02 Ospemifene 60 mg/day: -1.19 Placebo: -0.89 Only significant for the 60 mg ospemifene compared to placebo SAFETY Endometrial thickness, mean (SD) change from baseline, mm Ospemifene 30 mg/day: 0.42 (1.35) Ospemifene 60 mg/day: 0.72 (1.59) Placebo: -0.02 (1.03) Endometrial hyperplasia or carcinoma No cases reported Treatment emergent adverse events	Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - 5% of participants in each treatment group C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				Incidence of adverse events similar across treatment groups ACCEPTABILITY Withdrawal due to adverse events 5% in each group discontinued the study because of adverse events	were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Yes Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Used results for the 60 mg dosage of Ospemifene as the standard deviation of the means were reported by the previous review.
Full citation Goldstein,S.R., Bachmann,G.A., Koninckx,P.R., Lin,V.H., Portman,D.J., Ylikorkala,O., Ospemifene Study Group., Ospemifene 12-month safety and efficacy in postmenopausal women with vulvar and vaginal atrophy, Climacteric, 17, 173-182, 2014 Ref Id 319531 Country/ies where the study was carried out 23 sites in Europe Study type Randomized double-blind placebo-controlled parallel-group study Aim of the study Assessment of 12-month safety of ospemifene 60 mg/daily for the treatment of postmenopausal women with vulvar and vaginal atrophy. Study dates October 2007 to July 2009 Source of funding Hormos Medical Ltd, subsidiary of QuatRx Pharmaceuticals. Shionogi Inc.	Sample size N = 426 Ospemifene 60 mg/day: 363 Placebo: 63 Characteristics Postmenopausal women 40-80 years of age, with vulvar and vaginal atrophy, defined as having a proportion of superficial cells ≤ 5% in the vaginal smear and a vaginal pH > 5. Age, mean (SD) years Ospemifene 60 mg/day: 61.7 (6.2) Placebo: 62.9 (6.5) BMI, mean (SD) kg/m² Ospemifene 60 mg/day: 24.7 (2.9) Placebo: 24.1 (2.9) Inclusion criteria Intact uterus and normal findings (except for atrophic vaginal signs) on pelvic examination, breast palpation, and recent mammogram. Subjects were not enrolled based on symptoms (ie. vaginal dryness or dyspareunia). Exclusion criteria Abnormal endometrial histology other than atrophy based on baseline biopsy, uterine bleeding of unknown origin or clinically significant abnormal gynaecological findings.	Interventions 60 mg ospemifene (or matching placebo) taken orally each morning with food.	Details Women randomized in a 6:1 ratio to ospemifene or matching placebo by sequential allocation of randomization number. Randomization stratified by study center.	Results EFFICACY endpoints 1. Percentage of superficial cells in the maturation index on the vaginal smear 2. Percentage of parabasal cells in the maturation index on the vaginal smear 3. Vaginal pH SAFETY endpoints Endometrial thickness ACCEPTABILITY endpoints Not evaluated for 12 weeks. QUALITY OF LIFE endpoints Not evaluated EFFICACY Maturation index Superficial cells, median (range) percentage / mean (SD) change from baseline to week 12 Ospemifene 60 mg/day: 5 (-5, 60.0) / 5 (10.8) Placebo: 0 (-5, 28) / 0 (8.25) P < 0.0001 Parabasal cells, median (range) percentage / mean (SD) change from baseline to week 12 Ospemifene 60 mg/day: -40 (-100, 75) / -40 (29.2) Placebo: 0 (-90, 98) / 0 (47) P < 0.0001 Vaginal pH, mean (SD) change from baseline to week 12 Ospemifene 60 mg/day: -1.21 (0.912) Placebo: -0.16 (0.945) P < 0.0001 SAFETY Endometrial thickness, mean (SD) change from baseline to week 12, mm Ospemifene 60 mg/day: 0.44 (1.7)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all majorconfounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				Placebo: 0.31 (1.5)	treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? 96.1% and 98.4% completed treatment at week 12. C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information Was a 52 week RCT but efficacy outcomes were reported at 12-weeks. Long-term outcomes have been reported in long-term review question.
Full citation Portman,D., Palacios,S., Nappi,R.E., Mueck,A.O., Ospemifene, a non-	Sample size N = 314 Ospemifene 60 mg/day = 160 Placebo = 154 Characteristics	Interventions One daily 60 mg ospemifene or placebo that were identical in appearance.	Details Participants took a one-daily dose of study medication with food in the morning for 12 weeks. Participants seen on weeks 4	Results EFFICACY endpoints 1. Percentage of superficial cells in the maturation index on the vaginal smear 2. Percentage of parabasal cells in the maturation index on the vaginal smear	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
oestrogen selective oestrogen receptor modulator for the treatment of vaginal dryness associated with postmenopausal vulvar and vaginal atrophy: A randomised, placebo-controlled, phase III trial, Maturitas, 78, 91-98, 2014 Ref Id 319560 Country/ies where the study was carried out USA Study type Randomised, double-blind, parallel-group, multicentre phase III 12-week study Aim of the study To evaluate the efficacy and safety of ospemifene in the treatment of vaginal dryness in postmenopausal women with vulvovaginal atrophy Study dates July 2008 to August 2009 Source of funding QuatRx Pharmaceuticals Company	Womem aged 40-80 years with diagnosed vulvovaginal atrophy and moderate or severe symptoms of vaginal dryness Age, mean (SD) years Ospemifene 60 mg/day - 59.9 (6.7) Placebo - 59.3 (7.0) BMI, mean (SD), kg/m² Ospemifene 60 mg/day - 27.2 (4.6) Placebo - 26.5 (4.6) Inclusion criteria Naturally or surgically menopausal Moderate or severe symptoms of vaginal atrophy 5% or fewer superficial cells in maturation index of vaginal smear Vaginal pH greater than 5.0 Self-reported most bothersome symptom of vaginal dryness or vaginal pain associated with sexual activity, with a severity of moderate or severe at randomization Exclusion criteria BMI ≥ 37 kg/m², the presence of clinically sugnificant abnormaol gynaecological findings other than signs of vaginal atrophy and concomitant hormonal medications, SERMs, or products expected to have oestrogenic and/or antioestogenic effects.		and 12 for completion of VVA symptom questionnaire, assessment of vaginal pH, vaginal smear, and visual examination of vagina. Transvaginal ultrasound and endometrial biopsy conducted on week 12.	3. Vaginal pH 4. Severity of vaginal dryness SAFETY endpoints 1. Endometrial thickness 2. Endometrial histology 3. Treatment-related adverse events ACCEPTABILITY endpoints Withdrawal due to adverse events QUALITY OF LIFE endpoints Not evaluated EFFICACY Superficial cells, mean percentage (SD) change from baseline to week 12 Ospemifene 60 mg/day: 7.0 (11.5) Placebo: 0.0 (11.3) P < 0.001 Parabasal cells, mean percentage (SD) change from baseline to week 12 Ospemifene 60 mg/day: -31.7 (26.7) Placebo: -3.9 (27.1) P < 0.001 Vaginal pH, mean (SD) change from baseline to week 12 Ospemifene 60 mg/day: -0.95 (0.847) Placebo: -0.25 (0.844) P < 0.001 Severity of vaginal dryness, mean (SD) change in severity score from baseline to week 12 Ospemifene 60 mg/day: -1.3 (1.08) Placebo: -1.1 (1.02) P = 0.08 SAFETY Endometrial thickness, mean (SD) change from baseline to week 12, mm Ospemifene 60 mg/day: 0.82 (1.68) Placebo: -0.11 (1.20) *Assessed in only patients with an intact uterus Endometrial hyperplasia or carcinoma	(systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Unclear A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				No cases reported Treatment related adverse events, n (%) Ospemifene 60 mg/day: 43 (26.9) Placebo: 18 (11.7) ACCEPTABILITY Withdrawal due to treatment related adverse events, n (%) Ospemifene 60 mg/day: 12 (7.5) Placebo: 5 (3.2)	participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - See results C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Yes Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information Two sets of analyses undertaken: Primary analyses: Intent-to-treat population Subsidiary analyses: Per-protocol population - consisted of all participants who had completed at least 10 weeks of treatment and had taken 85% or more of study medication. Efficacy and safety of ospemifene demonstrated using ITT analyses.
Full citation Portman,D.J., Bachmann,G.A., Simon,J.A., Ospemifene Study Group., Ospemifene, a novel selective estrogen receptor modulator for treating dyspareunia associated with postmenopausal vulvar	Sample size N= 605 Ospemifene 60 mg/day = 303 Placebo = 302 Characteristics Most participants were white (90.6%) aged 40 to 79 years and had BMI values ranging from 16.7 to 37.1 kg/m² Inclusion criteria 1. Postmenopausal women	Interventions 60 mg/daily ospemifene or placebo with food in the morning for 12 weeks.	Details Ospemifene and placebo supplied as tablets identical in appearance. Nonhormonal vaginal lubricant provided to all participants and used as needed. Participants seen on weeks 4 and 12 for assesment. Participants underwent transvaginal ultrasound and endometrial biopsy on week	Results EFFICACY endpoints 1. Percentage of superficial cells in the maturation index on the vaginal smear 2. Percentage of parabasal cells in the maturation index on the vaginal smear 3. Vaginal pH 4. Severity of dyspareunia associated with sexual intercourse SAFETY endpoints 1. Endometrial thickness	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
and vaginal atrophy, Menopause, 20, 623-630, 2013 Ref Id 254703 Country/ies where the study was carried out 110 sites in the United States Study type Multicenter phase 3 randomized, double-blind, parallel-group design study Aim of the study To compare the efficacy, safety, and tolerability of ospemifene 60 mg/day versus placebo in the treatment of moderate to severe dyspareunia in postmenopausal women with vulvar and vaginal atrophy (VVA). Study dates July 2008 to August 2009 Source of funding QuatRx Pharmaceuticals Company	aged 40 to 80 years who reported having moderate or severe vaginal pain (dyspareunia) with sexual activity as their most bothersome symptom. 2. Having VVA, defined as 5% or less superficial cells in the maturation index of the vaginal smear and a vaginal pH higher than 5. 3. Either hysterectomized or had an intact uterus with a double-layer endometrial thickness less than 4 mm and had no evidence of hyperplasia, cancer, or other pathology. 4. Negative Papanicolaou test result or lacked an intact cervix. 5. Negative mammogram result 9 months or less before randomization. 6. Normal breast examination result at screening. 7. Provided written informed consent. Exclusion criteria 1. BMI of 37 kg/m² or higher 2. SBP of 180 mmHg or DBP of 100 mgHg or higher 3. Clinically significant abnormal gynaecological findings. 4. Other signs of vaginal atrophy such as: uterine bleeding of unkown origin, uterine polyps or symptomatic and/or large uterine fibroids (> 3 cm), or vaginal infection requiring medication. 5. Significant abnormal findings on physical examination,		12.	2. Endometrial histology 3. Treatment-related adverse events ACCEPTABILITY endpoints Withdrawal due to treatment-related adverse events QUALITY OF LIFE endpoints Not evaluated EFFICACY Superficial cells, mean percentage (SD) change from baseline to week 12 Ospemifene 60 mg/day: 12.3 (14.8) Placebo: 1.7 (6.9) P < 0.0001 Parabasal cells, mean percentage (SD) change from baseline to week 12 Ospemifene 60 mg/day: -40.2 (38.8) Placebo: 0.0 (30.0) P < 0.0001 Vaginal pH, mean (SD) change from baseline to week 12 Ospemifene 60 mg/day: -0.94 (1.0) Placebo: -0.07 (0.8) P < 0.0001 Dyspareunia, mean (SD) change in severity score from baseline to week 12 Ospemifene 60 mg/day: -1.5 (1.1) Placebo: -1.2 (1.1) P < 0.0001 Percentage of participants reporting no vaginal pain after sexual activity on week 12 Ospemifene 60 mg/day: 38.0 Placebo: 28.1 *Ospemifene demonstrated statistically significant efficacy compared to placebo for all 4 efficacy parameters. SAFETY Endometrial thickness, mean (SD) change from baseline to week 12, mm	treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Unclear A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	mammography, ECG, safety lab tests, or liver function screening. 6. More than 14 alcoholic drinks per week. 7. Took heparin, digitalis alkaloids, or strong cytochrome P450 3A4 inhibitors 8. Used any hormonal medications, SERMs, or products expected to have estrogenic and/or antoestrogenic effects within prespecified time frames before study screening. 9. Used ospemifene before study screening. 10. Women who were positive for factor V Leiden mutation or had current or past cerebrovascular incidents, thromboembolic disorders, blood coagulation disorders, severe hepatic or renal impairment, or suspicion of malignancy on mammography within 10 years.			Ospemifene 60 mg/day: 0.40 (1.25) Placebo: 0.10 (1.29) *Ospemifene caused a slight increase in endometrial thickness Endometrial hyperplasia or carcinoma No cases reported Adverse events, n (%) Ospemifene 60 mg/day: 79 (26.1) Placebo: 44 (14.6) ACCEPTABILITY Withdrawal due to treatment related adverse events, n (%) Ospemifene 60 mg/day: 10 (3.3) Placebo: 4 (1.3)	follow-up) - Yes C2a. How many participants did not complete treatment in each group? - 4.6% in ospemifene group and 3.3% in placebo group C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants'

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Yes Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information Two sets of analyses undertaken: Primary analyses: Intent-to-treat population Subsidiary analyses: Per-protocol population - consisted of all participants who had completed at least 10 weeks of treatment and had taken 85% or more of study medication. Efficacy and safety of ospemifene demonstrated using ITT analyses.
Full citation Rutanen,E.M., Heikkinen,J., Halonen,K., Komi,J., Lammintausta,R., Ylikorkala,O., Effects of ospemifene, a novel SERM, on hormones, genital tract, climacteric symptoms, and quality of life in postmenopausal women: a double-blind, randomized trial, Menopause, 10, 433-439, 2003	Sample size N = 160 Ospemifene 30 mg/day = 40 Ospemifene 60 mg/day = 40 Ospemifene 90 mg/day = 40 Placebo = 39 1 woman in placebo group did not start treatment at all. Characteristics No differences in baseline characteristics between treatment groups Age, mean (SD) Ospemifene 30 mg/day: 56.9 (4.5) Ospemifene 60 mg/day:	Interventions Three different doses (30, 60, or 90 mg daily) of ospemifene or placebo for 3 months.	Details Participants had a washout period of 90 days for any systemic hormone medications or for 30 days for vaginal estrogen medication. Prestudy screening included clinical examination and laboratory assessments. Endometrial thickness measured by vaginal ultrasonography at screening and at 3 months.	Results EFFICACY endpoints 1. Percentage of parabasal, intermediate, and superficial cells on the vaginal smear SAFETY endpoints 1. Endometrial thickness 2. Endometrial histology 3. Adverse events ACCEPTABILITY endpoints Withdrawal due to adverse events QUALITY OF LIFE endpoints Changes in Work Ability Index in depression, anxiety, or activity (self-confidence)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Unclear

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Ref Id 227258 Country/ies where the study was carried out Finland Study type Double-blind randomised controlled study Aim of the study Effects of three different daily doses of ospemifene on hormone levels, genital tract organs, climacteric symptoms, and quality of life. Study dates Not reported. Source of funding Hormos Medical Corporation	56.9 (4.7) Ospemifene 90 mg/day: 57.6 (4.3) Placebo: 58.2 (5.4) BMI, mean (SD) Ospemifene 30 mg/day: 24.4 (2.4) Ospemifene 60 mg/day: 25.0 (3.0) Ospemifene 90 mg/day: 25.1 (3.3) Placebo: 24.5 (2.7) Inclusion criteria 1. Healthy postmenopausal women aged 45 to 65 years 2. At least 12 months post last spontaneous menstrual bleed 3. FSH levels exceeding 40 IU/L and E2 levels below 0.11 nmol/L Exclusion criteria 1. BMI of 30 kg/m² or more 2. Blood pressure of 160/105 mmHg or higher 3. Pathological finding on gynaecological examination or pap smear 4. Endometrial thickness of 5mm or more 5. Uterine fibroids more than 5 cm in diameter 6. Known endometrial polyps or submucous fibroids 7. Current or history of any malignancy of the reproductive organs or breasts 8. Any other hormone-dependent malignancy 9. Any present drug therapy except thyroxin			EFFICACY Changes in parabasal, intermediate, and superficial cells during treatment period Clear difference between ospemifene and placebo groups in mean changes in these cells (P<0.05) Significant differences in pairwise comparisons SAFETY Endometrial thickness, mean (SD) change from baseline, mm Ospemifene 30 mg/day: 0.64 (1.14) P<0.05 Ospemifene 60 mg/day: 0.54 (1.01) P<0.05 Ospemifene 90 mg/day: 0.42 (0.82) P<0.05 Placebo: -0.01 (0.69) All ospemifene groups differed significantly from placebo. No differences in endometrial thickness were noticeable among the differing ospemifene dose levels Endometrial histology Endometrium remained atrophic after 3 months. Adverse events Frequency of participants reporting adverse events similar across treatment groups ACCEPTABILITY Withdrawal due to adverse events Ospemifene 30 mg/day: 1 Ospemifene 60 mg/day: 3 Ospemifene 90 mg/day: 1 Placebo: 0 Side effects included: headache, facial numbness, nausea, dizziness, or ameba infection QUALITY OF LIFE No differences in quality of life indices at baseline or at 3 months.	A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Unclear A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Unclear risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - See results

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information Were not clear on whether adverse events were treatment related.
Full citation Voipio,S.K., Komi,J., Kangas,L., Halonen,K., DeGregorio,M.W., Erkkola,R.U., Effects of ospemifene (FC-1271a) on uterine endometrium, vaginal maturation index, and hormonal status in healthy postmenopausal women, Maturitas, 43, 207-214, 2002 Ref Id 227527 Country/ies where the study was carried out Finland Study type Double-blind, placebo-controlled phase I study Aim of the study To investigate the effects of ospemifene on the uterine endometrium, vaginal maturation index, and hormonal status in healthy postmenopausal women with an atrophic vaginal	Sample size N=40 25 mg ospemifene = 8 50 mg ospemifene = 8 100 mg ospemifene = 8 200 mg ospemifene = 8 Placebo = 8 Characteristics Healthy postmenopausal Caucasian females Age, mean (SD) years 25 mg ospemifene = 60 (4.0) 50 mg ospemifene = 62 (4.5) 100 mg ospemifene = 60 (4.6) 200 mg ospemifene = 62 (5.1) Placebo = 62 (4.6) Inclusion criteria Postmenopausal, 55-75 years of age, body weight between 50-90 kg, in good general health, with an intact uterus. Exclusion criteria 1. Use of any hormonal medication (thyroxin allowed) during the 12 previous months 2. Strong susceptibility to allergic reactions	Interventions Oral doses of ospemifene 25 mg ospemifene; 50 mg ospemifene; 100 mg ospemifene; 200 mg ospemifene; or matching Placebo for 12 weeks.	Details Gynaecological examination, measurement of the double-layer thickness of the uterine endometrium, vaginal maturation index were performed and endometrial biopsy taken at baseline and at 12 weeks' visit. Estrogenic effects on vaginal epithelium estimated by routine maturation index. Visual analogue scale used to assess vaginal dryness.	Results EFFICACY endpoints 1. Percentage of parabasal cells in the maturation index on the vaginal smear 2. Percentage of intermediate cells in the maturation index on the vaginal smear 3. Percentage of superficial cells in the maturation index on the vaginal smear 4. Vaginal dryness SAFETY endpoints 1. Endometrial thickness 2. Endometrial histology 3. Treatment-related adverse events ACCEPTABILITY endpoints Withdrawal due to treatment related adverse events QUALITY OF LIFE endpoints Not evaluated EFFICACY Parabasal cells Decrease in percentage of cells for all ospemifene doses Intermediate cells Increase in percentage of cells for all ospemifene doses Superficial cells Increase in percentage of cells for all ospemifene doses Vaginal dryness	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Unclear A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Unclear A3. The groups were comparable at baseline including all majorconfounding and prognostic factors - Yes Unclear risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
epithelium. Study dates Not reported. Source of funding Not reported.	3. Participation in a drug study or blood donation within 60 days prior to the study 4. Evidence of clinically significant cardiovascular, renal, hepatic, hematological, gastrointestinal, pulmonary, metabolic, neurological or psychic disease or continuous medication to these diseases 5. Excessive use of alcohol			No statistical significant difference between treatment groups. SAFETY Endometrial thickness, median (range) change from baseline, mm Treatment arm Baseline 12 weeks 25 mg ospemifene 2.38 (0.62) 1.65 (0.23) 50 mg ospemifene 2.40 (1.32) 3.48 (4.59) 100 mg ospemifene 2.38 (0.78) 2.38 (1.22) 200 mg ospemifene 1.40 (0.18) 2.20 (1.08) Placebo 2.38 (0.78) 1.93 (0.31) No clinically significant changes seen in endometrial thickness at any dose level Endometrial histology Weak effect of ospemifene on endometrial histology. No secretory changes or hyperplasia observed. Treatment-related adverse events Generally, ospemifene well tolerated ACCEPTABILITY Withdrawal due to adverse effects, n 50 mg ospemifene: 1 due to gallstones and pancreatitis 200 mg ospemifene: 1 due to hot flushes, dizziness, and chest pain	intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - 1 each in two treatment groups did not complete treatment C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious
Full citation Constantine, G. D., Goldstein, S. R., Archer, D. F., Endometrial safety of ospemifene: results of the phase 2/3 clinical	Sample size N=2166 women with 1863 completing the study. Ospemifene 60 mg/day: 1,242 women Placebo: 924 Number completed the	Interventions 60 mg ospemifene (or matching placebo) taken orally each morning with	Details Participants were randomized 1:1 to ospemifene 60 mg/day or placebo in one 6-week trial and three 12-week trials; one of the 12-week trials had a 40-week extension study. In a	Results Short term outcomes at 12 weeks EFFICACY endpoints 1. Percentage of superficial cells in the maturation index on the vaginal smear 2. Percentage of parabasal cells in the maturation index on the vaginal smear	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
development program, Menopause, 22, 36-43, 2015 Ref Id 338232 Country/ies where the study was carried out 23 sites in Europe Study type Six randomised, phase 2/3 double-blind, placebo controlled, parallel-group studies Aim of the study To assess the endometrial safety of ospemifene based on phase 2/3 clinical trials of postmenopausal women with up to 52 weeks of exposure to ospemifene 60 mg/day versus placebo Study dates Not reported Source of funding Shionogi Inc.	study, n (%): Ospemifene 60 mg/day: 1061 (85.4) Placebo: 802 (86.8) Characteristics Postmenopausal women 40-80 years of age, with vulvar and vaginal atrophy, defined as having a proportion of superficial cells ≤ 5% in the vaginal smear and a vaginal pH > 5. Age, mean (SD) years Ospemifene 60 mg/day: 59.4 (6.49) Placebo: 58.9 (6.24) BMI, mean (SD) kg/m² Ospemifene 60 mg/day: 25.7 (4.03) Placebo: 26.0 (4.20) Women with intact uterus, n (%) Ospemifene 60 mg/day: 851 (68.5) Placebo: 543 (58.8) Inclusion criteria Postmenopausal women with vulvar and vaginal atrophy (5% or less superficial cells on vaginal smear (maturation index), vaginal pH higher than 5.0, and at least one moderate or severe symptom of VVA) In three of the studies, participants were required to have an intact uterus: One 12-week study (N = 79), the 40-week long-term extension study (N = 118), and the 52-week long term safety study (N = 426) required participants to have	food	separate 52-week trial, women were randomized 6:1 to ospemifene 60 mg/day or placebo by sequential allocation of randomization number. Randomization stratified by study center. Endometrial safety was assessed by endometrial histology (biopsy), transvaginal ultrasound, and gynecologic examination.	3. Vaginal pH 4. Vaginal atrophy 5. Vaginal dryness 6. Dyspareunia 7. Itching and discomfort SAFETY endpoints 1. Endometrial thickness 2. Breast pain/blood oestradiol levels 3. Treatment-emergent adverse events ACCEPTABILITY endpoints Not evaluated for 12 weeks. QUALITY OF LIFE endpoints Not evaluated EFFICACY Superficial cells, median (range) percentage / mean (SD) change from baseline to week 12 Not reported Parabasal cells, median (range) percentage / mean (SD) change from baseline to week 12 Not reported Vaginal pH, mean (SD) change from baseline to week 12 Not reported Vaginal atrophy Not reported Vaginal dryness Not reported Dyspareunia Not reported Itching and discomfort: Not reported SAFETY Endometrial thickness, mean (SD) change from baseline to week 12, mm Ospemifene 60 mg/day: 0.51 (1.5) Placebo: 0.06 (1.2)	between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	an intact uterus Exclusion criteria Abnormal endometrial histology other than atrophy based on baseline biopsy, uterine bleeding of unknown origin, clinically significant abnormal gynecologic findings, endometrial thickness of 4 mm or more on centrally read TVUS, pathologic findings on endometrial biopsy or Papanicolaou test, or clinically significant findings on physical examination			Breast pain/blood oestradiol levels Not reported Treatment-emergent adverse events Not reported	C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? 85.4% and 86.8% completed treatment in the ospemifene and placebo group respectively. C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information Long-term outcomes have been reported in long-term review question. This study consists of some data on women in Goldstein's 2014 study.

H.5.4 Long-term effectiveness of ospemifene

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Full citation Goldstein,S.R., Bachmann,G.A., Koninckx,P.R., Lin,V.H., Portman,D.J., Ylikorkala,O., Ospemifene Study Group., Ospemifene 12-month safety and efficacy in postmenopausal women with vulvar and vaginal atrophy, Climacteric, 17, 173-182,	Sample size N = 426 with 349 completing the study. Ospemifene 60 mg/day: 363 Placebo: 63 Characteristics Postmenopausal women 40-80 years of age, with vulvar and vaginal atrophy, defined as having a proportion of superficial cells ≤ 5% in the vaginal smear and a vaginal pH > 5.	Interventions 60 mg ospemifene (or matching placebo) taken orally each morning with food.	Details Women randomized in a 6:1 ratio to ospemifene or matching placebo by sequential allocation of randomization number. Randomization stratified by study center.	Results EFFICACY endpoints 1. Vaginal dryness 2. Signs of vaginal atrophy SAFETY endpoints 1. Endometrial thickness 2. Endometrial histology 3. Treatment-emergent adverse events	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
2014 Ref Id 319531 Country/ies where the study was carried out 23 sites in Europe Study type 52-week randomized double-blind placebo-controlled parallel-group study Aim of the study Assessment of 12-month safety of ospemifene 60 mg/daily for the treatment of postmenopausal women with vulvar and vaginal atrophy. Study dates October 2007 to July 2009 Source of funding Hormos Medical Ltd, subsidiary of QuatRx Pharmaceuticals. Shionogi Inc.	Age, mean (SD) years Ospemifene 60 mg/day: 61.7 (6.2) Placebo: 62.9 (6.5) BMI, mean (SD) kg/m² Ospemifene 60 mg/day: 24.7 (2.9) Placebo: 24.1 (2.9) Inclusion criteria Intact uterus and normal findings (except for atrophic vaginal signs) on pelvic examination, breast palpation, and recent mammogram. Subjects were not enrolled based on symptoms (ie. vaginal dryness or dyspareunia). Exclusion criteria Abnormal endometrial histology other than atrophy based on baseline biopsy, uterine bleeding of unknown origin or clinically significant abnormal gynecological findings.			ACCEPTABILITY endpoints 1. Withdrawal due to treatment related adverse events 2. Compliance to treatment QUALITY OF LIFE endpoints Not evaluated EFFICACY Maturation index Vaginal dryness, percentage with no dryness at week 52 Ospemifene 60 mg/day: 81.5 Placebo: 32.1 P < 0.0001 Vaginal atrophy, percentage with no signs of atrophy at week 52 Ospemifene 60 mg/day: 80 Placebo: 30 SAFETY Endometrial thickness, mean (SD) change from baseline to week 52, mm Ospemifene 60 mg/day: 0.75 (1.5) Placebo: 0.17 (1.3) Endometrial histological biopsy characteristics No tissue changes (hyperplasia or carcinoma) reported Treatment-emergent adverse events, n (%) Ospemifene 60 mg/day: 308 (84.6)	balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all majorconfounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? 81.0% and 87.3% completed treatment in the ospemifene and placebo group respectively. C2b. The groups were

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				Placebo: 47 (75.8) ACCEPTABILITY Withdrawals due to treatment related adverse events, n (%) Ospemifene 60 mg/day: 49 (13.5) Placebo: 6 (9.7) Compliance to treatment, % Ospemifene 60 mg/day: 95 Placebo: 99	comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Indirectness: No serious Other information Short-term outcomes of this study have been reported in short-term review question.
Full citation Simon,J.A., Lin,V.H., Radovich,C., Bachmann,G.A., Ospemifene Study Group., One-year long-term safety extension study of ospemifene for the treatment of vulvar and vaginal atrophy in postmenopausal women with a uterus, Menopause, 20, 418-427, 2013 Ref Id 319569 Country/ies where the study was carried out United States Study type Multicentre, randomized, double-blind 40-week extension study of a 12-week study (226136) Aim of the study To assess the safety of ospemifene for the treatment of vulvar and vaginal atrophy (VVA) in postmenopausal women with a uterus Study dates May 2006 to September 2008 Source of funding QuatRx Pharmaceuticals	Sample size N = 180 Ospemifene 30 mg/day = 62 Ospemifene 60 mg/day = 69 Placebo = 49 Characteristics Most participants were white aged 46 to 79 years with BMI values ranging from 15.7 to 36.8 kg/m² Inclusion criteria Postmenopausal women aged 40 to 80 years, with the following criteria of VVA: 5% or less superficial cells on the vaginal smear (maturation index), vaginal pH greater than 5.0, and at least one moderate or severe symptom of VVA. Exclusion criteria 1. Endometrial thickness of 4mm or greater on centrally read transvaginal ultrasound 2. Pathological findings on endometrial biopsy or Papanicolaou test 3. Any other clinical significant gynaecological abnormality other than VVA (eg. uterine bleeding of unknown origin) 4. Body mass index of 37 kg/m² or greater 5. Systolic blood pressure of 180 mmHg or diastolic blood pressure of 100 mmHg or higher 6. Abnormal breast examination or mammogram results 7. Suspicion of malignancy or history of any malignancy within 10 years 8. Current or past thromboembolic	Interventions 30 or 60 mg/day of ospemifene or placebo for 40 additional weeks. Study medication taken in the morning.	Details 40-week safety extension of a 12-week, phase 3, efficacy and safety study. Blinding was according to the original blinding assignment for the 12-week study. Total duration was 52-weeks followed by a 4-week posttreatment follow-up period. Endometrial thickness assessed by transvaginal ultrasonography.	Results EFFICACY endpoints 1. Vaginal dryness SAFETY endpoints 1. Endometrial thickness 2. Endometrial histology 3. Adverse events ACCEPTABILITY endpoints 1. Withdrawal due to adverse events 2. Compliance to dosing schedules QUALITY OF LIFE endpoints Not evaluated EFFICACY Vaginal dryness Improvement in severity scores for vaginal dryness from baseline to both week 26 and 52 for both ospemifene doses compared to placebo SAFETY Endometrial thickness, mean (SD) change Ospemifene 60 mg/day: 1.14 (1.56) Placebo: -0.04 (1.15) Endometrial histology No hyperplasia or carcinoma reported	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	or blood coagulation disorder 9. Women who consumed more than 14 drinks of alcohol per week 10. Women currently using itraconazole, ketoconazole, or digitalis alkaloids 11. Use of any HT (unless the woman had a sufficient washout period before any procedures (eg. 14 days for vaginal estrogens and 60 days for oral/transdermal therapy)			Adverse events, n (%) Ospemifene 30 mg/day: 38 (61.3) Ospemifene 60 mg/day: 44 (63.8) Placebo: 22 (44.9) ACCEPTABILITY Withdrawal due to adverse events, n (%) Ospemifene 30 mg/day: 3 (4.8) Ospemifene 60 mg/day: 4 (5.8) Placebo: 1 (2.0) Compliance rates, mean % Ospemifene 30 mg/day: 85.5 Ospemifene 60 mg/day: 84.6 Placebo: 93.4	treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - See results C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Yes Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information
Full citation Constantine, G. D., Goldstein, S. R., Archer, D. F., Endometrial safety of ospemifene: results of the phase 2/3 clinical development program, Menopause, 22, 36-43, 2015 Ref Id 338232 Country/ies where the study was carried out 23 sites in Europe Study type Six randomised, phase 2/3 double-blind, placebo controlled, parallel-group studies Aim of the study To assess the endometrial safety of ospemifene based on phase 2/3 clinical trials of postmenopausal women with up to 52 weeks of exposure to ospemifene 60 mg/day versus placebo	Sample size N=2166 women with 1863 completing the study. Ospemifene 60 mg/day: 1,242 women Placebo: 924 Number completed the study, n (%): Ospemifene 60 mg/day: 1061 (85.4) Placebo: 802 (86.8) Characteristics Postmenopausal women 40-80 years of age, with vulvar and vaginal atrophy, defined as having a proportion of superficial cells ≤ 5% in the vaginal smear and a vaginal pH > 5. Age, mean (SD) years Ospemifene 60 mg/day: 59.4 (6.49) Placebo: 58.9 (6.24) BMI, mean (SD) kg/m² Ospemifene 60 mg/day: 25.7 (4.03)	Interventions 60 mg ospemifene (or matching placebo) taken orally each morning with food	Details Participants were randomized 1:1 to ospemifene 60 mg/day or placebo in one 6-week trial and three 12-week trials; one of the 12-week trials had a 40-week extension study. In a separate 52-week trial, women were randomized 6:1 to ospemifene 60 mg/day or placebo by sequential allocation of randomization number. Randomization stratified by study center. Endometrial safety was assessed by endometrial histology (biopsy), transvaginal ultrasound, and gynecologic examination.	Results Long term outcomes at 52 weeks EFFICACY endpoints 1. Vaginal dryness 2. Signs of vaginal atrophy 3. Dyspareunia 4. Itching and discomfort SAFETY endpoints 1. Endometrial thickness 2. Endometrial histology 3. Treatment-emergent adverse events ACCEPTABILITY endpoints 1. Withdrawal due to treatment related adverse events 2. Compliance to treatment QUALITY OF LIFE endpoints Not evaluated	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Study dates Not reported Source of funding Shionogi Inc.	Placebo: 26.0 (4.20) Women with intact uterus, n (%) Ospemifene 60 mg/day: 851 (68.5) Placebo: 543 (58.8) Inclusion criteria Postmenopausal women with vulvar and vaginal atrophy (5% or less superficial cells on vaginal smear (maturation index), vaginal pH higher than 5.0, and at least one moderate or severe symptom of VVA) In three of the studies, participants were required to have an intact uterus: One 12-week study (N = 79), the 40-week long-term extension study (N = 118), and the 52-week long term safety study (N = 426) required participants to have an intact uterus Exclusion criteria Abnormal endometrial histology other than atrophy based on baseline biopsy, uterine bleeding of unknown origin, clinically significant abnormal gynecologic findings, endometrial thickness of 4 mm or more on centrally read TVUS, pathologic findings on endometrial biopsy or Papanicolaou test, or clinically significant findings on physical examination			EFFICACY Vaginal dryness Not reported Vaginal atrophy Not reported Dyspareunia Not reported Itching and discomfort Not reported SAFETY Endometrial thickness, mean (SD) change from baseline to week 52, mm Ospemifene 60 mg/day: 0.81 (1.5) Placebo: 0.07 (1.2) Endometrial histological biopsy characteristics No tissue changes (hyperplasia with atypia or carcinoma) reported Simple endometrial hyperplasia without atypia on biopsy 3 months after the last dose of the study drug was reported for one woman who received ospemifene 60 mg/d Treatment-emergent adverse events Not reported ACCEPTABILITY Withdrawals due to treatment related adverse events, n (%) Ospemifene 60 mg/day: 95 (7.6)	B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? 85.4% and 86.8% completed treatment in the ospemifene and placebo group respectively. C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				Placebo: 34 (3.7) Compliance to treatment, n (%) Not reported	systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information Short-term outcomes of this study have been reported in short-term review question. This study consists of some data on women in Goldstein's 2014 study.

H.6 Review and referral

H.7 Starting and stopping HRT

<table>
<tr><th>Study details</th><th>Study Design</th><th>Intervention</th><th>Results</th><th>Quality checklist</th><th>Other information</th></tr>
<tr>
<td>Full citation
Lindh-Astrand,L., Bixo,M., Hirschberg,A.L., Sundstrom-Poromaa,I., Hammar,M., A randomized controlled study of taper-down or abrupt discontinuation of hormone therapy in women treated for vasomotor symptoms, Menopause, 17, 72-79, 2010
Ref Id
226863
Country/ies where the study was carried out
Sweden
Source of funding
The Research Council of Southeast of Sweden
Swedish Society of Obstetrics and Gynaecology.
Study dates
March 2005 to December 2007.</td>
<td>Study type
Randomized open-label controlled trial.
Inclusion criteria
Used HRT for between 3 and 11 years, used continuous estrogen-progestogen therapy or tibolone at least during the last year, had originally started HRT because of vasomotor symptoms and were suitable to try to discontinue HRT according to the gynaecologists and her own judgement.
Exclusion criteria
Unstable thyroid or other metabolic disease. Any indication to stop HRT rapidly (e.g. breast cancer). Recently started or changed medication for any psychiatric disorder. Undergoing other treatments for vasomotor symptoms. Having more than one hot flush per 24 hours according to the 2-week screening diary. Having had unsuccessful discontinuation of HRT during the last year. Undergoing HRT because of premenopausal hypogonadism.
Method of blinding
The randomization and block lengths were unknown to the investigators and nurses participating in the study. Participants were not blinded to their allocation.
Randomization
An independent statistician prepared a computer generated separate randomization list for each centre, and the randomization was carried out with blocks of four women.
Power calculation
The assumption was that tapering of HRT would lead to a mean recurrence of 2 hot flushes per 24 hours, and abrupt discontinuation would cause 20% more hot flushes per 24 hours (i.e. 2.4 flushes per 24 hours). 80% power to detect a significant difference at the 5% level would require 100 women in each arm.
An alternative power calculation was based on the assumption that 33% of women in the taper group and 66% of women in the abrupt group would have resumed HRT after 4 months. 80% power at the 5% level would require 35 women per arm.</td>
<td>Interventions
Tapering of HRT by taking usual dose every other day for a four week period, before stopping completely.
Comparator
Immediate discontinuation of HRT.
Symptom reporting
A manual hot flush diary was used during the 2-week screening period, 4-week tapering period, and 6 weeks after discontinuation.
Number and severity of hot flashes were registered daily after waking up and before bedtime.
Severity was rated with a scale ranging from 0 (not bothersome at all) to 10 (extremely bothersome) and comprised a summative rating of all hot flushes</td>
<td>Results
Results
<table>
<tr><th>Variable</th><th>Taper group</th><th>Abrupt discontinuation</th></tr>
<tr><td>Hot flash frequency at 6 weeks</td><td>3.4 (1.3 to 6.4)</td><td>4.0 (1.4 to 6.1)</td></tr>
<tr><td>Hot flash severity at 6 weeks</td><td>3.1 (0.7 to 7.4)</td><td>4.1 (1.0 to 7.0)</td></tr>
<tr><td>PGWB score</td><td>86 (70 to 96)</td><td>85 (75 to 92)</td></tr>
<tr><td>Resumption of HRT at 6 weeks</td><td>6/45 (13.3%)</td><td>5/36 (13.9%)</td></tr>
<tr><td>Resumption of HRT at 12 months</td><td>24/44 (55%)</td><td>14/36 (39%)</td></tr>
<tr><td>Adverse events*</td><td>39 (54%)</td><td>29 (48%)</td></tr>
</table>
*Numbers as reported in the article, but percentages do not equate to number in each group. Likely adverse events are reported as absolute number of events, but percentage represents percentage of participants who experienced at least one adverse event.</td>
<td>A1 - An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups)
Yes
A2 - There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation)
Yes
A3 - The groups were comparable at baseline, including all major confounding and prognostic factors
Yes
B1 - The comparison</td>
<td>Other information
Limitations
Open label study design. Whether investigators were blinded to other potential confounders (such as duration of HRT use) is unclear.
Baseline data for women lost to follow up are unknown, therefore unclear whether there may be systematic differences between these women and those who completed the trial.
Outcomes of menopausal symptom severity are only reported at 6 weeks. It is unclear whether this is an adequate length of follow up time.</td>
</tr>
</table>

Study details	Study Design	Intervention	Results	Quality checklist	Other information
	Sample size N = 87 • n = 46 taper-down group • n = 41 immediate discontinuation Characteristics **Variable (median and IQR unless otherwise stated)** / **Taper group** / **Abrupt discontinuation group** Age (years) / 58 (54 to 61) / 59 (57 to 61) Age at menopause (years) / 50 (48 to 52) / 49.5 (48 to 51.8) Duration of HRT (years) / 9.0 (5.3 to 10.0) / 9.5 (6.0 to 10.9) No. of hot flushes per 24 hours / 0 (0.00 to 0.07) / 0 (0.0 to 0.18) Reason for stopping HRT (n, %) / / / Fear of adverse effects / 14 (31) / 10 (28) Woman's decision / 23 (53) / 20 (56) Physician's advice / 7 (16) / 6 (17)	experienced. The baseline average number and severity of hot flushes per 24 hours were calculated from the 2-week screening period. The 6-week figure was calculated as an average of the 7 day period of the 6th week diary. For women who recommenced treatment with HRT during the 6-week follow up period (n=9) the mean number of frequency and severity from the last 7 days for the specific woman (before she resumed HRT) was carried forward to constitute her 6 week data. The PGWB form was used to assess health related quality of life at baseline and 6 weeks after discontinuation of HRT. It contains 22 items related to anxiety, depressed mood, well-being, self-control, general		groups received the same care apart from the intervention(s) studied Yes B2 - Participants receiving care were kept 'blind' to treatment allocation No B3 - Individuals administering care were kept 'blind' to treatment allocation No C1 - All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) Yes C2a - How many participants did not complete treatment in each group? Taper down group: 1 excluded due to protocol violation. Abrupt discontinuation group: 3 protocol violations, 1 withdrew consent.	

Study details	Study Design	Intervention	Results	Quality checklist	Other information
		health and vitality. Each item is graded between 0 (most negative opinion) and 5 (most positive opinion), with a total score of between 0 and 110.		C2b - The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) Unclear C3a - For how many participants in each group were no outcome data available? Taper down group, n= 6: 1 excluded due to protocol violation, 5 lost to follow up. Abrupt discontinuation group, n = 6: 3 protocol violations, 1 withdrew consent, 2 lost to follow up. C3b - The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences	

Study details	Study Design	Intervention	Results	Quality checklist	Other information
				between groups in terms of those for whom outcome data were not available). Yes D1 - The study had an appropriate length of follow-up Unclear D2 - The study used a precise definition of outcome Yes D3 - A valid and reliable method was used to determine the outcome Yes D4 - Investigators were kept 'blind' to participants' exposure to the intervention No D5 - Investigators were kept 'blind' to other important confounding and prognostic factors Unclear	
Full citation Cunha,E.P., Azevedo,L.H., Pompei,L.M., Strufaldi,R., Steiner,M.L.,	Study type Randomized, double-blind, placebo controlled trial. Inclusion criteria Postmenopausal women using estrogen-progestogen hormone therapy in full doses, defined as CEE 0.625mg/day (or equivalent) in association with	Interventions Tapering of HRT dose to low dose regimen (1mg estradiol plus 0.5mg	Results Scores at 2 months:	A1 - An appropriate method of randomisation was used to allocate	Other information Also presents data on outcomes at 2 months and 4 months. This

<table>
<tr><th>Study details</th><th>Study Design</th><th>Intervention</th><th>Results</th><th>Quality checklist</th><th>Other information</th></tr>
<tr>
<td>Ferreira,J.A., Peixoto,S., Fernandes,C.E., Effect of abrupt discontinuation versus gradual dose reduction of postmenopausal hormone therapy on hot flushes, Climacteric, 13, 362-367, 2010
Ref Id
226368
Country/ies where the study was carried out
Brazil
Source of funding
Medication provided by Biolab Sanus Farmacêutica Ltda (São Paulo, Brazil).
Study dates
Not reported.</td>
<td>medroxyprogesterone acetate 5.0mg (sequential scheme) or 2.5mg (continuous scheme) or equivalent of other progestogens.
In addition, they had to have been using HRT for at least 6 months, should wish to discontinue HRT for personal reasons (not due to adverse effects) and HRT must have been prescribed for the treatment of climacteric vasomotor symptoms.
Exclusion criteria
Use of medication or behavioural therapy for weight control. Use of any type of medication other than HRT that has recognised action of climacteric vasomotor symptoms. Medical indication for the immediate discontinuation of HRT. Presentation of severe liver failure, heart failure, previous thrombosis, uncontrolled thyroid disease, hyperplasia, endometrial polyps or thickening, or cancer in any organ. Discontinuation of HRT due to adverse effects.
Method of blinding
Placebo controlled.
Randomization
By means of RandomAllocation Software in blocks of 12 participants each.
Power calculation
80% power to detect an 80% reduction in symptoms (level of significance not reported, assumed 5%) would require 17 patients per group.
Sample size
N = 60
• n = 20 Group 1: immediate discontinuation of usual dose HRT
• n = 20 Group 2: 2 months low dose HRT followed by immediate discontinuation
• n = 20 Group 3: 4 months low dose HRT followed by immediate discontinuation
Characteristics
<table>
<tr><th>Variable (years, mean and SD unless otherwise stated)</th><th>Immediate discontinuation</th><th>Low dose treatment for 2 months</th><th>Low dose treatment for 4 months</th></tr>
<tr><td>Age</td><td>52.71 ± 4.19</td><td>52.61 ± 6.16</td><td>51.32 ± 4.63</td></tr>
<tr><td>Ethnicity</td><td></td><td></td><td></td></tr>
<tr><td>Caucasian</td><td>13 (76.5%)</td><td>14</td><td>13 (68.4%)</td></tr>
</table>
</td>
<td>norethisterone acetate daily) for either two months (group 2) or four months (group 3) prior to discontinuation.
Comparator
Immediate discontinuation of standard dose HRT.
Symptom reporting
Reported using the Blatt-Kupperman Menopausal Index at baseline (randomization) and again after 2, 4 and 6 months. The index comprises a numerical summation of 11 menopausal complaints, such as hot flushes, insomnia, palpitation, fatigue etc. Some symptoms are weighted more heavily than others, and each symptom is ranked according to its severity.</td>
<td>
<table>
<tr><th>Variable</th><th>Group 1 (placebo)</th><th>Group 2 (2 months low dose, then placebo)</th><th>Group 3 (4 months low dose, then placebo)</th></tr>
<tr><td>Mean total score for Blatt-Kupperman index (± SD)</td><td>11.8 ± 6.3</td><td>8.2 ± 5.3</td><td>8.1 ± 6.0</td></tr>
<tr><td>Mean score for hot flushes (± SD)</td><td>5.4 ± 4.2</td><td>0.4 ± 1.9</td><td>1.9 ± 3.6</td></tr>
</table>
No significant difference between any two groups for total score. Significantly lower scores in group 2 and group 3 when compared to group 1 for hot flushes.
Scores at 4 months:
<table>
<tr><th>Variable</th><th>Group 1 (placebo)</th><th>Group 2 (2 months low dose, then placebo)</th><th>Group 3 (4 months low dose, then placebo)</th></tr>
<tr><td>Mean total score for Blatt-Kupperman index (± SD)</td><td>14.0 ± 6.4</td><td>15.7 ± 8.9</td><td>9.7 ± 7.7</td></tr>
<tr><td>Mean score for hot flushes (± SD)</td><td>7.1 ± 4.8</td><td>6.0 ± 4.2</td><td>2.1 ± 3.6</td></tr>
</table>
No significant difference between any two groups for total score. Significantly lower scores
</td>
<td>participants to treatment groups (which would have balanced any confounding factors equally across groups)
Yes
A2 - There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation)
Yes
A3 - The groups were comparable at baseline, including all major confounding and prognostic factors
Yes
B1 - The comparison groups received the same care apart from the intervention(s) studied
Yes
B2 - Participants receiving care were kept 'blind' to treatment allocation
Yes</td>
<td>shows a significant difference in outcomes only between groups who were still taking and no longer taking HRT, not between any groups who had completed discontinuation.
Limitations
The trial was double-blind in design, but it is unclear whether individuals administering care to the participants (as opposed to the study investigators) were also blinded to treatment allocation.
It is unclear whether investigators were also blinded to other potential confounders, in addition to treatment allocation.
Follow up was at 6 months, when the abrupt discontinuation group had been without treatment for 6 months, and the tapered dose groups had been</td>
</tr>
</table>

Study details	Study Design				Intervention	Results	Quality checklist	Other information
			(77.8%)			in group 3 than group 1 or 2 for hot flushes. Scores at 6 months: **Variable** / **Group 1 (placebo)** / **Group 2 (2 months low dose, then placebo)** / **Group 3 (4 months low dose, then placebo)** Mean total score for Blatt-Kupperman index (± SD): 13.4 ± 7.7 / 17.1 ± 10.0 / 14.9 ± 7.5 Mean score for hot flushes (± SD): 6.4 ± 4.5 / 6.2 ± 4.2 / 6.1 ± 3.6 No significant difference between any two groups for either outcome.	B3 - Individuals administering care were kept 'blind' to treatment allocation Unclear C1 - All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) Yes C2a - How many participants did not complete treatment in each group? None. C2b - The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) Yes C3a - For how many participants in each group were no outcome data available? Group 1, n = 3 lost to follow up	off treatment for 2 and 4 months. It is unclear whether this is an appropriate length of follow up.
	Non-Caucasian	4(23.5%)	4 (22.2%)	6 (31.6%)				
	Marital status							
	Stable relationship	11 (64.7%)	15 (83.3%)	12 (63.2%)				
	Other	6 (35.3%)	3 (16.7%)	7 (36.8%)				
	Age at menopause	47.29 ± 3.58	45.78 ± 4.39	46.21 ± 5.13				
	Time since menopause	5.41 ± 2.37	6.83 ± 5.22	5.11 ± 2.94				
	Duration of HRT	4.94 ± 3.63	5.39 ± 3.57	4.11 ± 2.98				
	Body mass index (kg/m2)	23.0 ± 3.1	24.5 ±3.8	24.8 ± 4.7				

Study details	Study Design	Intervention	Results	Quality checklist	Other information
				Group 2, n = 2 lost to follow up Group 3, n = 1 lost to follow up C3b - The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available). Yes D1 - The study had an appropriate length of follow-up Unclear D2 - The study used a precise definition of outcome Yes D3 - A valid and reliable method was used to determine the outcome Yes D4 - Investigators were kept 'blind' to participants' exposure to the intervention	

Study details	Study Design	Intervention	Results	Quality checklist	Other information
				Yes D5 - Investigators were kept 'blind' to other important confounding and prognostic factors Unclear	
Full citation Haimov-Kochman,R., Barak-Glantz,E., Arbel,R., Leefsma,M., Brzezinski,A., Milwidsky,A., Hochner-Celnikier,D., Gradual discontinuation of hormone therapy does not prevent the reappearance of climacteric symptoms: a randomized prospective study, Menopause, 13, 370-376, 2006 Ref Id 226622 Country/ies where the study was carried out Israel Source of funding Not reported. Study dates May 2001 to April 2003.	Study type Open-label randomized controlled trial. Inclusion criteria Women treated with combined estrogen-progestogen therapy or estrogen-alone therapy for more than 3 years. Exclusion criteria Taking concomitant medication or over-the-counter supplementation that could affect their evaluation during the study. Women with the following conditions were excluded: smoking, alcoholism, severe liver or kidney disorders, active ischaemic heart disease, evidence of acute thrombosis and infectious diseases, abnormal Pap smear, vaginal bleeding of undiagnosed cause, endometrial hyperplasia, severe uncontrolled hypertension. Method of blinding Open label study. Randomization Randomization with SAS 8e package. Power calculation A sample size of 100 women was needed to give 90% power to detect a difference of 25% in reuptake of HRT rates between the two groups, at the 5% level (assumed 40% return to HRT in the abrupt discontinuation group and 15% in the gradual discontinuation group). Sample size N = 91 • n = 54 Group 1: abrupt discontinuation 4 withdrawals after randomization due to exclusion criteria, therefore n = 50 • n = 46 Group 2: gradual discontinuation 5 withdrawals after randomization due to exclusion criteria, therefore n = 41 Characteristics	Interventions Reduction of HRT by one tablet per week per month, so complete cessation took place after 6 months. Comparator Immediate discontinuation of HRT. Symptom reporting Symptoms were monitored with the Greene scale. 21 different symptoms clustered into 4 different subclasses are assessed: 11 psychological symptoms (6 anxiety and 5 depression), 7 somatic symptoms (e.g. headaches, muscle and joint pain), 2 vasomotor symptoms (hot	Results Total Greene Climacteric score during follow up: At 1 month: significantly lower scores in taper group than abrupt discontinuation (p=0.001) At 3 months: significantly lower scores in taper group than abrupt discontinuation (p=0.047) At 6, 9 and 12 months: no significant difference between the two groups. Vasomotor Greene Climacteric score during follow up: At 1 month: significantly lower scores in taper group than abrupt discontinuation (p=0.0001) At 3 months: significantly lower scores in taper group than abrupt discontinuation (p=0.001) At 6 months: significantly higher scores in taper group than abrupt discontinuation (p=0.001) At 9 and 12 months: no significant difference between the two groups. Resumption of HRT: 21/50 (42%) group 1 versus 15/41 (36.6%) group 2 (p = 0.67)	A1 - An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) Yes A2 - There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) Yes A3 - The groups were comparable at baseline, including all major confounding and prognostic	Other information Limitations The trial was open-label by design. Whether investigators were blinded to other potential confounding factors is not clear.

Study details	Study Design	Intervention	Results	Quality checklist	Other information
	Age, years (mean, SD) = 56.8 ± 4.2 Duration of HRT use, years (mean, SD) = 8.8 ± 3.8	flushes and night sweats) and a sexual symptom (loww of sexual interest). Each symptom score ranges from 0 ("not at all") to 3 ("quite a bit") compiling a Greene score range of 0 to 63. The questionnaire was completed at 1, 3, 6, 9 and 12 months by the physician at the time of patient visits, and by telephone questionnaire.		factors Yes B1 - The comparison groups received the same care apart from the intervention(s) studied Yes B2 - Participants receiving care were kept 'blind' to treatment allocation No B3 - Individuals administering care were kept 'blind' to treatment allocation No C1 - All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) Yes C2a - How many participants did not complete treatment in each group? None C2b - The groups were comparable for treatment completion (that is, there were no important or	

Study details	Study Design	Intervention	Results	Quality checklist	Other information
				systematic differences between groups in terms of those who did not complete treatment) Not applicable C3a - For how many participants in each group were no outcome data available? None C3b - The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available). Not applicable D1 - The study had an appropriate length of follow-up Yes D2 - The study used a precise definition of outcome Yes D3 - A valid and reliable method	

<table>
<tr><th>Study details</th><th>Study Design</th><th>Intervention</th><th>Results</th><th>Quality checklist</th><th>Other information</th></tr>
<tr><td></td><td></td><td></td><td></td><td>was used to determine the outcome
Yes
D4 - Investigators were kept 'blind' to participants' exposure to the intervention
No
D5 - Investigators were kept 'blind' to other important confounding and prognostic factors
Unclear</td><td></td></tr>
<tr><td>Full citation
Aslan,E., Bagis,T., Kilicdag,E.B., Tarim,E., Erkanli,S., Kuscu,E., How best is to discontinue postmenopausal hormone therapy: immediate or tapered?, Maturitas, 56, 78-83, 2007
Ref Id
226110
Country/ies where the study was carried out
Turkey
Source of funding
Not reported.
Study dates
Not reported.</td><td>Study type
Randomized controlled trial.
Inclusion criteria
Current HRT users choosing to discontinue their medication.
Exclusion criteria
Not reported.
Method of blinding
Not reported - assumed open label.
Randomization
"rank randomization" (not described).
Power calculation
Sample size of 64 patients would give 80% power to detect a change of 2 symptom scores (SD = 4) on the hot flush scoring system, at the 5% level.
Sample size
N = 72
2 withdrawals prior to commencing any discontinuation programme.
• n = 35 tapering
• n = 35 immediate discontinuation

Characteristics
<table><tr><th>Variable</th><th>Immediate discontinuation</th><th>Tapered discontinuation</th></tr></table></td><td>Interventions
Use of medication once every other day for 2 weeks, then discontinued.
Comparator
Immediate discontinuation.
Symptom reporting
Recording of vasomotor symptoms on a symptom scale.
Severity recorded as:
Mild: temporary warmth sensation, no sweating, does not interfere with daily activity.
Moderate: temporary warmth</td><td>Results
Hot flush score after 2 weeks:
Immediate discontinuation group (mean ± SEM) : 3.06 ± 0.87
Tapered discontinuation group (mean ± SEM) : 1.96 ± 0.65
p = 0.323
Hot flush score after 4 weeks:
Immediate discontinuation group (mean ± SEM) : 3.23 ± 1.10
Tapered discontinuation group (mean ± SEM) : 2.83 ± 1.04
p = 0.792
VMS severity
<table><tr><th>VMS severity after 2 weeks</th><th>Immediate discontinuation (n, %)</th><th>Tapered discontinuation (n, %)</th></tr><tr><td>None</td><td>17 (48)</td><td>19 (54.3)</td></tr><tr><td>Mild</td><td>15 (42.9)</td><td>13 (37.1)</td></tr><tr><td>Moderate</td><td>1 (2.9)</td><td>2 (5.7)</td></tr><tr><td>Severe</td><td>2 (5.7)</td><td>1 (2.9)</td></tr><tr><td>VMS severity after 4 weeks</td><td>Immediate discontinuation</td><td>Tapered discontinuation</td></tr></table></td><td>A1 - An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups)
Unclear
A2 - There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment</td><td>Other information
Limitations
Method of randomisation was not made clear in the article. Study was open label by design, but whether investigators were blinded to potential confounders (other than treatment allocation) is unclear. Follow up was for four weeks only (2 weeks after discontinuation in the tapering group) and it is unclear whether this is sufficiently</td></tr>
</table>

Study details	Study Design			Intervention	Results			Quality checklist	Other information
	Mean age (years; mean, SD)	53 ± 3.8	53.3 ± 4.6	sensation, sweating, interferes with daily activity to a lesser degree. Severe: temporary warmth sensation, sweating, interferes with daily activity severely. Any night sweats. Frequency was noted as average daily episodes of hot flushes in each severity group. Symptom scores were obtained using the severity and frequency of symptoms. One point was given for every mild hot flush, two for a moderate hot flush and three for a severe hot flush. The hot flush score was also grouped as none (0 point), mild (1-8 points), moderate (9-16 points) and severe (17 and higher points).		(n, %)	(n, %)	allocation) Yes A3 - The groups were comparable at baseline, including all major confounding and prognostic factors Yes B1 - The comparison groups received the same care apart from the intervention(s) studied Yes B2 - Participants receiving care were kept 'blind' to treatment allocation No B3 - Individuals administering care were kept 'blind' to treatment allocation No C1 - All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) Yes C2a - How many participants did not complete	long.
	Duration of menopause (years; mean, SD)	6.3 ± 0.68	5 ± 0.52		None	18 (51.4)	18 (51.4)		
	Duration of HRT use (years; mean, SD)	3.03 ± 0.31	3.31 ± 0.37		Mild	13 (37.1)	15 (42.9)		
	Presence of VMS before treatment (%)	77.1	80		Moderate	2 (5.7)	0 (0)		
					Severe	2 (5.7)	2 (5.7)		
					Adverse effects **Adverse effects** / **Immediate discontinuation (n, %)** / **Tapered discontinuation (n, %)** Vaginal bleeding: 3 (8.6) / 2 (5.7)				

Study details	Study Design	Intervention	Results	Quality checklist	Other information
				treatment in each group? None C2b - The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) Not applicable C3a - For how many participants in each group were no outcome data available? None C3b - The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available). Not applicable D1 - The study had an appropriate	

Study details	Study Design	Intervention	Results	Quality checklist	Other information
				length of follow-up Unclear D2 - The study used a precise definition of outcome Yes D3 - A valid and reliable method was used to determine the outcome Yes D4 - Investigators were kept 'blind' to participants' exposure to the intervention No D5 - Investigators were kept 'blind' to other important confounding and prognostic factors Unclear	

H.8 Long term risk and benefits of HRT

H.8.1 Venous thromboembolism

Study details	Design	Comparison	Results	Other
Full citation Eischer,L., Eichinger,S., Kyrle,P.A., The risk of recurrence in women with venous thromboembolism while using estrogens: a prospective cohort study, Journal of Thrombosis and Haemostasis, 12, 635-640, 2014 Ref Id 328803 Study type Prospective cohort study Source of funding Austrian National Bank Country/ies where the study was carried out Austria Study dates 1992-2012	Aim of the study To test the hypothesis that women who had a first VTE while using estrogen have a low risk of recurrence. Inclusion criteria Between 1992 and 2008 consecutive patients with a first distal and/or proximal deep vein thrombosis of the leg and/or pulmonary embolism (PE) who had been treated with anticoagulants for 3-18 months were included. Exclusion criteria -age younger than 18 years; -VTE associated with surgery, trauma, cancer, prolonged immobilization or pregnancy; -requirement for long-term antithrombotic treatment for reasons other than VTE	Interventions Estrogen Details Methods Setting: Hospital Methods: Ascertainment of estrogen use: at study entry, a detailed medical history, including a systematic documentation of estrogen use, was obtained. Ascertainment of VTE: recurrent symptomatic DVT was confirmed by venography of colour duplex songraphy Statistic methods: -categorical data were compared among groups using contingency-table analyses (chi-square test). -continuous data were compared by means of Mann-Whitney U-tests. -cox proportional-hazards models were used to analyse the association between estrogen use and the risk of recurrent VTE. Analyses were adjusted for age, presence or absence of FV leiden and site of VTE. Follow-up: averagely more than 5 years, losses to follow-up were 6.5% Sample size N=630 Estrogen users: n=333 [only 58 were menopausal hormone therapy (MHT) users, 275 were estrogen-containing contraceptives users] Non-users: n=297	Characteristics Age in years, mean (SD): non users: 55 (15) estrogen users: 38 (15) Observation time in months, mean (SD): non users: 61 (50) estrogen users: 76 (52) Factor V leiden, n(%): non users: 48 (16%) oestrogen users: 98 (28%) Results Risk of recurrent VTE in relation to estrogen use, n/N, adjusted RR (95% CI): Non users: 49/297, 1 (reference group) Estrogen (MHT) users: 8/58, 0.7 (0.3-1.5) -Analysis adjusted for age, site of VTE (distal deep vein thrombosis (DVT), proximal DVT, pulmonary embolism) and factor V Leiden.	Other information Limitations Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. No (participants were women with a confirmed first VTE) Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline, including all major confounding and prognostic factors. No, estrogen users were younger compared with non-users (mean 38 vs. 55), had longer duration of estrogen use (mean 76 months vs. 61 months) Level of risk: Low Performance bias The comparison groups received the same care apart from the intervention(s) studied. Unclear. Participants receiving care were kept 'blind' to treatment allocation. N/a Individuals administering care were kept 'blind' to treatment allocation. N/a Level of risk: Unclear Attrition bias

Study details	Design	Comparison	Results	Other
				All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). No, observation time for estrogen users was about 1 year (mean) longer but reason not reported How many participants did not complete treatment in each group? Not reported [just reported as a total losses to follow-up were low (6.5%)] The groups were comparable for treatment completion. Unclear For how many participants in each group were outcome data not available? Not reported The groups were comparable with respect to the availability of outcome data. Unclear Level of risk: High Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. N/A Investigators were kept 'blind' to other important confounding and prognostic factors. N/A

Study details	Design	Comparison	Results	Other
				Level of risk: Low Study quality
Full citation Benson,V.S., Canonico,M., Reeves,G.K., Abbott,S., Allen,N., Armstrong,M., Balkwill,A., Banks,E., Benson,V., Beral,V., Black,J., Brown,A., Bull,D., Cairns,B., Callaghan,K., Canfell,K., Canoy,D., Chivenga,J., Crossley,B., Crowe,F., Ewart,D., Ewart,S., Fletcher,L., Gathani,T., Gerrard,L., Goodill,A., Green,J., Guiver,L., Hilton,E., Kan,S.W., Keene,C., Kirichek,O., Kroll,M., Langston,N., Lingard,I., Liu,B., Luque,M.J., Pank,L., Pirie,K., Reeves,G., Roddam,A., Shaw,K., Sherman,E., Sherry-Starmer,E., Strange,H., Sweetland,S., Timadjer,A., Tipper,S., Travis,R., Wang,X., Watson,J., Wright,L., Yang,T., Young,H., Venous thromboembolism risk in relation to use of different types of postmenopausal hormone therapy in a large prospective study, Journal of Thrombosis and Haemostasis, 10, 2277-2286, 2012 Ref Id 310765 Study type Prospective cohort study. Source of funding UK Medical Research Council Cancer Research UK UK National Health Service Breast Screening Programme Country/ies where the study was carried out UK Study dates Recruitment from June 1996 to March 1998. Follow up for 1.9 to 3.9 years.	Aim of the study To assess the relationship between the type of hormone replacement therapy used and the incidence of VTE. Inclusion criteria Postmenopausal women aged 50 to 69 years. Exclusion criteria Premenopausal or perimenopausal women. Women with a history of cancer (except non-melanoma skin cancer) at recruitment. Previous history of VTE or treatment for blood clots at recruitment. Hospital record for VTE prior to recruitment, or surgery in the 12 weeks prior to recruitment. Unknown use of HRT.	Interventions Not applicable. Details Cox regression was used to estimate the relative risk of hospital admission or death for VTE in relation to use of HRT. Methods Women provided information on their use of HRT, socio-demographic and anthropometric factors, and medical and reproductive history at recruitment. A second questionnaire was sent to study participants 3 years later to update the information on HRT use and other factors (with a 65% response rate). Study participants were followed by record linkage using their NHS number for deaths, cancer registrations, emigration and NHS hostpial admissions. The main outcome measure for this analysis (VTE) was defined as the first diagnosis following recruitment into the study of pulmonary embolism or deep vein thrombosis as in inpatient/day-case hospital admisssion, or as the underlying cause of death. Records of VTE were validated using a sample of 1000 women with and without a record of VTE identified. 93% of hospital diagnoses were confirmed by the general practitioner. Only 3 women (0.3%) with no hospital record of VTE were reported by their general practitioner to have had a diagnosis of VTE during the follow up period. Sample size N = 1058259 n = 476711 never users of HRT n = 201515 past users of HRT	Characteristics For whole cohort Age, years† 56.7 (4.5) BMI, kg/m²† 26.1 (4.6) Current smokers 20.8% Number with VTE 2200 (0.2%) †mean (standard deviation) Results Relative risks (RR) are shown compared to never users of HRT and adjusted for geographical region, socioeconomic status and BMI. Use of any HRT preparation Current use of HRT RR (95% CI): 1.59 (1.45 to 1.75) Past use of HRT RR (95% CI): 0.95 (0.84 to 1.08) Different routes and HRT preparations Current use of transdermal oestrogen only HRT RR (95% CI): 0.82 (0.64 to 1.06) Current use of oral oestrogen only HRT RR (95% CI): 1.42 (1.22 to 1.66) Current use of oral oestrogen plus progestin HRT RR (95% CI): 2.07 (1.86 to 2.32) Age of user Current use of transdermal oestrogen only HRT in women < 50 years RR (95% CI): 0.80 (0.55 to 1.15) Current use of oral oestrogen only HRT in women < 50 years RR (95% CI): 1.45 (1.17 to 1.80) Current use of oral oestrogen plus progestin HRT in women < 50 years RR (95% CI): 1.87 (1.59 to 2.21)	Other information Limitations Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. Unclear. Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes (but other known risk factors, such as family history and thrombiphilia were not recorded nor controlled for in analysis) The groups were comparable at baseline, including all major confounding and prognostic factors. No - past and current users of HRT were younger, and more likely to have used oral contraceptives, than never users. Level of risk: High Performance bias The comparison groups received the same care apart from the intervention(s) studied. N/A Participants receiving care were kept ‘blind’ to treatment allocation. N/A Individuals administering care were kept ‘blind’ to treatment allocation. N/A Level of risk: unclear

Study details	Design	Comparison	Results	Other
		n = 380033 current users of HRT	Current use of transdermal oestrogen only HRT in women aged 50+ years RR (95% CI): 0.85 (0.61 to 1.20) Current use of oral oestrogen only HRT in women aged 50+ years RR (95% CI): 1.33 (1.06 to 1.65) Current use of oral oestrogen plus progestin HRT in women aged 50+ years RR (95% CI): 2.16 (1.90 to 2.45) Duration of use Current use of transdermal oestrogen only HRT commenced within the past 2 years RR (95% CI): 1.63 (0.41 to 6.53) Current use of oral oestrogen only HRT commenced within the past 2 years RR (95% CI): 3.83 (1.91 to 7.71) Current use of oral oestrogen plus progestin HRT commenced within the past 2 years RR (95% CI): 3.17 (2.10 to 4.78) Current use of transdermal oestrogen only HRT for <5 years RR (95% CI): 0.71 (0.42 to 1.18) Current use of oral oestrogen only HRT for <5 years RR (95% CI): 1.27 (0.94 to 1.71) Current use of oral oestrogen plus progestin HRT for <5 years RR (95% CI): 2.07 (1.77 to 2.42) Current use of transdermal oestrogen only HRT for 5+ years RR (95% CI): 0.85 (0.63 to 1.13) Current use of oral oestrogen only HRT for 5+ years RR (95% CI): 1.49 (1.24 to 1.77) Current use of oral oestrogen plus progestin HRT for 5+ years RR (95% CI): 2.05 (1.80 to 2.33) Different types and doses of	Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). No, the study reported that "many women in the UK ceased HRT use after publications of the first report of results from the WHI study in 2002", but did not report the data in detail. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Level of risk: High Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Unclear. Investigators were kept ‘blind’ to participants’ exposure to the intervention. N/A Investigators were kept ‘blind’ to other important confounding and

Study details	Design	Comparison	Results	Other
			oestrogen use in users of oestrogen-only HRT Current use of conjugated equine oestrogen RR (95% CI): 1.46 (1.23 to 1.75) Current use of ≤ 0.625mg conjugated equine oestrogen RR (95% CI): 1.30 (1.04 to 1.62) Current use of > 0.625mg conjugated equine oestrogen RR (95% CI): 1.82 (1.38 to 2.40) Current use of oestradiol RR (95% CI): 1.45 (1.06 to 1.98) Current use of ≤ 1mg oestradiol RR (95% CI): 1.71 (1.16 to 2.53) Current use of > 1mg oestradiol RR (95% CI): 1.26 (0.77 to 2.06) Different types of progestin use in users of oestrogen-progestin HRT Current use of norethisterone RR (95% CI): 1.82 (1.52 to 2.17) Current use of norgestrel RR (95% CI): 1.98 (1.71 to 2.29) Current use of medroxyprogesterone acetate RR (95% CI): 2.67 (2.25 to 3.17) Current use of continuous combined regimen RR (95% CI): 2.30 (1.99 to 2.67) Current use of sequential combined regimen RR (95% CI): 1.93 (1.69 to 2.21)	prognostic factors. N/A Level of risk: Unclear
Full citation Canonico,M., Fournier,A., Carcaillon,L., Olie,V., Plu-Bureau, Oger,E., Mesrine,S., Boutron-Ruault,M.C., Clavel-Chapelon,F., Scarabin,P.Y., Postmenopausal hormone therapy and risk of idiopathic venous thromboembolism: results from the E3N cohort study, Arteriosclerosis, Thrombosis and Vascular Biology, 30, 340-345, 2010	Aim of the study To investigate the impact of oestrogens by route of administration as well as the influence of concomitant progestogens on the risk of idiopathic venous thrombosis. Inclusion criteria Postmenopausal women born between 1925 and 1950, insured by a healthcare plan covering mostly teachers. Exclusion criteria Thrombotic event before the start of follow up. Personal history of cancer, other than	Interventions Not applicable. Details Cox proportional hazards models were used to estimate the hazard ratios for venous thromboembolism associated with HRT. Methods Participants completed biennial self-administered questionnaires which included items about anthropometric measurements,	Characteristics Only reported for the entire cohort Age, years† 54.0 (4.3) BMI, kg/m²† 22.6 (3.2) Current smokers 7095 (9.9%) †mean (standard deviation) Results Hazard ratios (HR) are reported as compared to never users of HRT unless otherwise stated, and adjusted for age, BMI, parity,	Other information -HRT use was self-reported and nondifferential misclassification regarding exposure might have occured during follow-up. Limitations Study quality Selection bias The method of allocation to treatment groups was

Study details	Design	Comparison	Results	Other
Ref Id 301085 Study type Prospective cohort study. Source of funding Mutuelle Générale de l'Education Nationale. Institut National de la Santé et de la recherché Médicale. Institut Gustave Roussy. 3M Company. Country/ies where the study was carried out France Study dates 1990 to July 2005.	basal cell carcinoma. Non-idiopathic thrombotic event or a VTE without information on predisposing factors. In addition, 68 women with a validated thrombotic event were censored at the point of cancer diagnosis, because of a validated cancer predating the thrombotic event.	medical history, menopausal status and a variety of lifestyle habits. Nonfatal VTE events were initially reported by women in the questionnaires. Participants who declared to have either a DVT or PE were then asked to complete a specific questionnaire and to send medical documentation relating to the event. To be validated, VTE events had to be diagnosed using an imaging procedure. Events were centrally validated by a medical committee blinded to HRT use. Cases of fatal pulmonary embolism were identified from death certificates. -15-yr follow-up time Sample size N = 80308 n = 549 cases with VTE n = 79759 controls without VTE (number using and not using HRT is not described)	educational level and time period. Different preparations of HRT Current use of oral oestrogens HR (95% CI): 1.7 (1.1 to 2.8) Current use of transdermal oestrogens HR (95% CI): 1.1 (0.8 to 1.8) Past use of HRT HR (95% CI): 1.1 (0.8 to 1.5) Current use of oral oestrogens compared to current use of transdermal oestrogens HR (95% CI): 1.5 (1.1 to 2.0) Different types of progestagen Current use of micronized progesterone HR (95% CI): 0.9 (0.6 to 1.5) Current use of pregnane derivatives HR (95% CI): 1.3 (0.9 to 2.0) Current use of norpregnane derivatives HR (95% CI): 1.8 (1.2 to 2.7) Current use of nortestosterone derivatives HR (95% CI): 1.4 (0.7 to 2.4)	unrelated to potential confounding factors. No, participants are mostly teachers with a health insurance Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes, but there could be other unknown risk factors not controlled for The groups were comparable at baseline, including all major confounding and prognostic factors. Unclear - data not reported separately for HRT users and non-users. Level of risk: High Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. N/A Individuals administering care were kept 'blind' to treatment allocation. N/A Level of risk: unclear Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported.

Study details	Design	Comparison	Results	Other
				The groups were comparable for treatment completion. Not applicable. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Not applicable. Level of risk: Unclear Detection bias The study had an appropriate length of follow up. Yes, 15-yr follow-up The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear. Level of risk: unclear
Full citation Cherry,N., Oestrogen therapy for prevention of reinfarction in postmenopausal women: A randomised placebo controlled trial, Lancet, 360, 2001-2008, 2002 Ref Id 295717 Study type Randomised, blinded, lacebo controlled trial. Source of funding UK National Health Service Research and Development	Aim of the study To assess the effect of unopposed oestradiol valerate on risk of another cardiac event or death in postmenopausal women who had just survived their first myocardial infarction. Inclusion criteria Women aged 50 to 69 years admitted to coronary care units or general medical wards with a diagnosis of myocardial infarction, in participating hospitals for the duration of the study. Discharged alive from hospital within 31 days of admission. Exclusion criteria	Interventions Women were randomly allocated to receive either 2mg oestradiol valerate or placebo, taken as one tablet daily for 2 years. Participants and investigators were blinded to treatment allocation. Details Number (percentage) of VTE events in the placebo group were compared to the events in the HRT group. Methods At recruitment, baseline information	Characteristics HRT group Age at admission to hospital, years†: 62.3 (5.2) BMI, kg/m²†: 26.8 (5.1) Placebo group Age at admission to hospital, years†: 62.9 (4.9) BMI, kg/m²†: 26.7 (5.3) †mean (standard deviation) Results Unadjusted relative risk (RR) for	Other information Limitations Power of study was less than planned. Known non-compliance was high. Non-compliance probably under-reported. Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Yes.

Study details	Design	Comparison	Results	Other
Programme on Cardiovascular Disease and Stroke. University of Manchester. Schering Health Care Ltd. Country/ies where the study was carried out England and Wales Study dates July 1996 and February 2000. Trial duration 2 years.	Previous myocardial infarction (prior to the index event). Use of HRT or vaginal bleeding in the 12 months prior to admission. History of breast, ovarian or endometrial carcinoma. Active thrombophlebitis, or a history of deep vein thrombosis or pulmonary embolus. Acute or chronic liver disease, Rotor syndrome, Dubin-Johnson syndrome or severe renal disease.	was collected from participants regarding height, weigh, smoking status, alcohol use, education, occupation, ethnic group, use of OCP or HRT, age at LMP, previous hysterectomy, history of agina, hypertension, stroke or diabetes, and fractures in the previous 10 years. Sample size N = 1017 n = 513 HRT n = 504 placebo	VTE are reported for HRt group as compared to placebo group. Risk of DVT RR (95% CI): 1.96 (0.18 to 21.60) Risk of PE RR (95% CI): 0.98 (0.20 to 4.84) Risk of any VTE RR (95% CI): 1.23 (0.33 to 4.55)† †Calculated by the NCC WCH technical team from data reported in the article.	There was adequate concealment of allocation. Yes. The groups were comparable at baseline. Yes. Level of risk: Low risk of bias Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Yes (was only disclosed if the information was required by patient's doctor. In such cases, patient withdrew from treeatment) Individuals administering care were kept 'blind' to treatment allocation. Yes. Level of risk: Low risk of bias Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 184 placebo, n = 294 HRT. The groups were comparable for treatment completion. No - more women in the HRT group did not comply with treatment, due to vaginal bleeding.

Study details	Design	Comparison	Results	Other
				For how many participants in each group were outcome data not available? None. The groups were comparable with respect to the availability of outcome data. No (high droput rate in HRT group) Level of risk: High risk of bias Detection bias The study had an appropriate length of follow up. Yes. (2-yr follow-up) The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Yes. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear. Level of risk: Low risk of bias
Full citation Grodstein,F., Stampfer,M.J., Goldhaber,S.Z., Manson,J.E., Colditz,G.A., Speizer,F.E., Willett,W.C., Hennekens,C.H., Prospective study of exogenous hormones and risk of pulmonary embolism in women, Lancet, 348, 983-987, 1996 Ref Id 229373 Study type Prospective cohort study. Source of funding	Aim of the study To assess the association between oral contraceptives and postmenopausal hormones with pulmonary embolism. Inclusion criteria Female registered nurses in 11 states. Exclusion criteria Women with a history of previous PE, cancer (except non-melanoma skin cancer), angina, myocardial infarction, stroke and other cardiovascular disease. Women who did not provide any information on exogenous hormone use.	Interventions Not applicable. Details Proportional hazards models were used to construct relative risks of PE associated with hormone use, adjusted for known or suspected risk factors. Methods Participants completed a detailed questionnaire at baseline that included items about their medical history and cardiovascular risk factors. Every two years, follow up	Characteristics Women's age at baseline: 30-55 years; No other data reported. Results Relative risks (RR) are reported for occurrence of pulmonary embolism in HRT users compared to non-users and are adjusted for age, BMI, diabetes, hypertension, hypercholesterolaemia, smoking status, parity and 2-year time period. Current postmenopausal HRT use	Other information -Information on HRT use was collected from the women themselvels, misclassification is possible. But in this study participants were registered nurses, acccuracy of self-reported HRT use should be high. Limitations Study quality Selection bias The method of allocation

Study details	Design	Comparison	Results	Other
Research grants from the National Institutes of Health. Country/ies where the study was carried out USA Study dates 1976 to 1992 (The Nurses Health Study).		questionnaires were sent so that information on risk factors could be kept up to date and newly diagnosed major illnesses could be recorded. The analysis of pulmonary embolism was restricted to cases that occurred between 1976 and June 1st 1992. PE was confirmed if supported by a high probability lung scan, a positive pulmonary arteriogram or necropsy. 16-year follow-up time Sample size N = 112593 (separate numbers for HRT use and no HRT use are not reported)	RR (95% CI): 2.1 (1.2 to 3.8) Past postmenopausal HRT use RR (95% CI): 1.3 (0.7 to 2.4) Duration of use Current use of HRT for up to 5 years RR (95% CI): 2.6 (1.2 to 5.2) Current use of HRT for over 5 years RR (95% CI): 1.9 (0.9 to 4.0) Dose of oestrogen Current use of 0.3 mg oestrogen daily RR (95% CI): 1.9 (0.5 to 8.3) Current use of 0.625 mg oestrogen daily RR (95% CI): 1.5 (0.6 to 3.7) Current use of ≥1.25 mg oestrogen daily RR (95% CI): 1.4 (0.4 to 5.0)	to treatment groups was unrelated to potential confounding factors. No, (participants were registered nurses) Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline, including all major confounding and prognostic factors. Unclear. Level of risk: High Performance bias The comparison groups received the same care apart from the intervention(s) studied. Unclear (nurses taking HRT might undergo more diagnostic procedures) Participants receiving care were kept 'blind' to treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Level of risk: High Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Not applicable.

Study details	Design	Comparison	Results	Other
				For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Not applicable. Level of risk: Unclear Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. No Investigators were kept 'blind' to other important confounding and prognostic factors. No Level of risk: Unclear
Full citation Hoibraaten,E., Qvigstad,E., Arnesen,H., Larsen,S., Wickstrom,E., Sandset,P.M., Increased risk of recurrent venous thromboembolism during hormone replacement therapy--results of the randomized, double-blind, placebo-controlled estrogen in venous thromboembolism trial (EVTET), Thrombosis and Haemostasis, 84, 961-967, 2000 Ref Id 300785 Study type Randomised controlled trial. Source of funding Novo-Nordisk Pharma.	Aim of the study To assess whetehr oestradiol treatment influences the risk of VTE. Inclusion criteria Postmenopausal women (no natural menstruation for at least 1 year) aged less than 70 years who had suffered previous DVT or PE. Previous VTE verified by objective means (venography or ultrasound for DVT, lung scan, helical CT or angiography for PE), or women without objective testing who had a typical history and were subsequently treated for VTE. Exclusion criteria Use of anti-coagulants within the last 3 months, familial antithrombin deficiency, any type of malignant disease, acute or chronic liver disease, history of liver disease in which	Interventions Women were randomly allocated to treatment with HRT containing 2mg oestradiol plus 1mg norethistereone acetate (Kliogest, Novo-Nordisk) or to placebo tablets with equivalent looking appearance. Details The study was stratified for age (< 60 or > 60 years of age) as this was considered the most important risk factor for VTE. Women were allocated to treatment by computer generated 1:1 block randomisation with fixed block sizes of 10 women. Methods At the initial visit, data were	Characteristics HRT group: Age, years† 55.8 (7.0) BMI, kg/m²† 26.8 (4.3) Current smoker 15 (21%) Family history of VTE 25 (35%) Placebo group: Age, years† 55.7 (5.9) BMI, kg/m²† 27.4 (4.0) Current smoker 20 (29%) Family history of VTE 18 (26%) † mean (standard deviation) Results Number of VTE events in placebo group n/N: 1/69 Number of VTE events in HRT	Other information Limitations All women were at high risk of VTE, due to their previous history. Small sample size. Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Yes. There was adequate concealment of allocation. Yes. The groups were comparable at baseline. Yes.

Study details	Design	Comparison	Results	Other
Research Forum, Ullevål University Hospital, Oslo. Country/ies where the study was carried out Norway Study dates February 1996 to February 1999. Trial duration 2 years.	liver function tests had failed to return to normal, porphyria, known drug abuse or alcoholism, life expectancy less than 2 years, or women who had taken part in other clinical trials within 12 weeks before study entry.	collected on demographic characteristics, reproductive and health history, risk factors for VTE and medication use. All women were given detailed instructions on symptoms and signs of DVT and PE and were advised to contact their own physician, local hospital, the investigator or a 24 hour telephone number if symptoms occurred. Scheduled follow up visits took place after 3 and 12 months, and an end of study visit at 24 months. Adverse events reported by the patient spontaneously, given in response to direct questioning, or observed on clinical examination were evaluated by the investigator. The major outcome was VTE as verified by objective tests (venography or ultrasound in the case of DVT, lung-scan, helical CT or angiography in the case of PE). All primary end points were independently and blindly confirmed by a radiologist and/or an internist/haematologist at the patient's local hospital. Sample size N = 140 n = 71 HRT group n = 69 placebo group	group n/N: 8/71 (includes one cerebral venous sinus thrombosis, in addition to DVT/PE outcomes) Relative risk of VTE in HRT group (95% CI): 8.63 (1.09 to 388.6)	Bias: Low risk of bias Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Yes. Individuals administering care were kept 'blind' to treatment allocation. Yes. Bias: Low risk of bias Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 23 HRT group, n = 14 placebo group The groups were comparable for treatment completion. Yes. For how many participants in each group were outcome data not available? None. The groups were comparable with respect to the availability of outcome data. Yes. Bias: Low risk of bias Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes.

Study details	Design	Comparison	Results	Other
				A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Yes Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear. Bias: Low risk of bias
Full citation Holmberg,L., Iversen,O.E., Rudenstam,C.M., Hammar,M., Kumpulainen,E., Jaskiewicz,J., Jassem,J., Dobaczewska,D., Fjosne,H.E., Peralta,O., Arriagada,R., Holmqvist,M., Maenpaa,J., Maenpa,J., HABITS Study Group, Increased risk of recurrence after hormone replacement therapy in breast cancer survivors, Journal of the National Cancer Institute, 100, 475-482, 2008 Ref Id 302449 Study type Randomised controlled trial. Source of funding Novo Nordic Pharma. Nordic Cancer Union. Swedish Cancer Society. Country/ies where the study was carried out Sweden. Study dates May 1997 until December 2003. Trial duration 2 years.	Aim of the study To evaluate whether HRT for menopausal symptoms is safe in women with previously treated breast cancer. Inclusion criteria Women who had previously completed primary treatment for breast cancer, including a complete removal of the tumour and axillary surgery, radiotherpay and chemotherapy as stipulated by local treatment guidelines. Treatment with tamoxifen was permitted. Tumour stage 0-2 with less than 4 involved axillary lymph nodes. Presence of menopausal symptoms that both the woman and her doctors felt needed treatment. Exclusion criteria Concomitant treatment with aromatase inhibitors. Four or more involved axillary lymph nodes or tumour stage > 2. Tumour recurrence, other history of malignancy or serious disease. Other contraindications to HRT treatment.	Interventions Women were randomly assigned to receive either HRT or best symptomatic treatment without hormones. Choice of the specific type of HRT was determined by local practice. If there was no preferred specific therapy in a particular centre then a sequential oestrgoen-progestagen regimen was prescribed for women with an intact uterus whose LMP was within the past 2 years. A continous combined regimen was prescribed for women 2 or more years past the menopause. The majority of centres prescribed a regimen of oestradiol hemihydrate and norethisterone acetate. Medium potency oestrogens alone were prescribed for women who had undergone hysterectomy. The majority of centres prescribed estradiol alone for these women. The study interventions were open label. Details The allocation scheme was computer generated in blocks of eight and stratified by participating centre, use of HRT before diagnosis of the original breast cancer, and treatment with	Characteristics Reported only for those women who were not lost to follow up. HRT group: Age, years† 55.6 (42 - 75) Follow up in years‡ 4.1 (0.01 to 7.8) Non-HRT group: Age, years† 54.8 (38 - 74) Follow up in years‡ 4.0 (0.2 to 7.7) †mean (range) ‡median (range) Results Occurrence of VTE in non-HRT group n/N: 2/224 Occurrence of VTE in HRT group n/N: 2/223 Relative risk of VTE in HRT group (95% CI): 1.00 (0.14 to 7.01)	Other information Limitations All women had previous breast cancer Open label trial therefore high risk of more vigorous follow-up in HRT group. Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Yes. There was adequate concealment of allocation. Yes. The groups were comparable at baseline. Yes. Bias: Low risk of bias Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No - open label trial. Individuals administering care were kept 'blind' to

Study details	Design	Comparison	Results	Other
		tamoxifen. Block size was unknown to the participating clinicians. Methods Participants were followed by a breast cancer specialist at least twice yearly for the first three years after assignment, and continue to be followed at least annually for a minimum of five years in total. It was recommended that participants receive mammograms every 12 to 24 months. Participants were also required to be seen by a gynaecologist every year. New breast cancer events, other new cancer, compliance and side effects of treatment were recorded prospectively. Sample size N = 447 n = 224 assigned to best symptomatic treatment without treatment n = 223 assigned to HRt		treatment allocation. No - open label trial. Bias: High risk of bias Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 11 HRT arm (never exposed to HRT), n = 43 non-HRT arm (drop-in to HRT group) The groups were comparable for treatment completion. No - more participants in the non-HRT arm actually were exposed to HRT during the trial. For how many participants in each group were outcome data not available? n = 2 HRT arm, n = 3 non-HRT arm. The groups were comparable with respect to the availability of outcome data. Yes. Bias: High risk of bias Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Unclear - patient reported side effects. Not described

Study details	Design	Comparison	Results	Other
				whether events were verified by scan. Investigators were kept 'blind' to participants' exposure to the intervention. No - open label trial. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear. Bias: High risk of bias
Full citation Laliberte,F., Dea,K., Duh,M.S., Kahler,K.H., Rolli,M., Lefebvre,P., Does the route of administration for estrogen hormone therapy impact the risk of venous thromboembolism? Estradiol transdermal system versus oral estrogen-only hormone therapy, Menopause, 18, 1052-1059, 2011 Ref Id 300451 Study type Retrospective cohort study. Source of funding Novartis Pharmaceuticals Corporation. Country/ies where the study was carried out Canada. Study dates January 2002 to October 2009.	Aim of the study To quantify the magnitude of risk reduction for VTE events associated with transdermal relative to oral oestrogen only HRT preparations in a real-world setting. Inclusion criteria Women aged 35 years or older at the date of first dispensing of HRT. To have a record of at least 2 dispensings of either transdermal or oral oestrogen only HRT. Continous health plan enrollment during the observation period and for 180 days before the index date (first dispensation). Exclusion criteria Receipt of any other oestrogen HRT agents during the 180 day baseline period (prior to the index date), or if they had been diagnosed with a VTE prior to the index date.	Interventions Not applicable. Details The risk of VTE among participants receiving transdermal as compared to oral oestrogen only preparations was evaluated using adjusted incidence rate ratios. Methods Health insurance claims from the Thomson Reuters MarketScan database were used to conduct the analysis. Participants receiving transdermal oestrogen were matched 1:1 with participants receiving oral oestrogen based on age (5 year intervals), baseline concomitant medication use (antihypertensive, antihyperlipidaemic, progestin and anticoagulant), Charlson comorbidity index, year of the index date, menopausal and postmenopausal disorders, hysterectomy, oophorectomy and risk factors for VTE (major surgery, hypertension and coagulation defect). Incidence of VTE was identified using ICD-9 codes. -7-year follow-up time Sample size N = 54036 n = 27018 transdermal HRT users	Characteristics Transdermal HRT users Age, years† 48.9 (7.1) Oral HRT users Age, years† 48.9 (7.1) †mean (standard deviation) Results Rate ratios (RR) compare use of transdermal HRT to oral HRT and are adjusted for baseline healthcare costs, census region, baseline oral contraceptive pill use, and binary variables for progestin and other oestrogen agents used concomitantly with the treatment of interest. Current use of transdermal HRT compared to oral HRT RR (95% CI): 0.67 (0.49 to 0.92)	Other information -Information on participants' weight and BMI was not available in the database therefore couldn't be controlled for in analysis. Limitations Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. Yes (while participants were all commercially insured) Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. (a matched-cohort design was used) The groups were comparable at baseline, including all major confounding and prognostic factors. Yes. Level of risk: Unclear Performance bias The comparison groups received the same care apart from the

Study details	Design	Comparison	Results	Other
		n = 27018 oral HRT users		intervention(s) studied. Unclear Participants receiving care were kept 'blind' to treatment allocation. No Individuals administering care were kept 'blind' to treatment allocation. No. Level of risk: Unclear Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Not applicable. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Not applicable. Level of risk: Unclear Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Unclear. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear.

Study details	Design	Comparison	Results	Other
				Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear. Level of risk: Unclear
Full citation Manson,J.E., Chlebowski,R.T., Stefanick,M.L., Aragaki,A.K., Rossouw,J.E., Prentice,R.L., Anderson,G., Howard,B.V., Thomson,C.A., LaCroix,A.Z., Wactawski-Wende,J., Jackson,R.D., Limacher,M., Margolis,K.L., Wassertheil-Smoller,S., Beresford,S.A., Cauley,J.A., Eaton,C.B., Gass,M., Hsia,J., Johnson,K.C., Kooperberg,C., Kuller,L.H., Lewis,C.E., Liu,S., Martin,L.W., Ockene,J.K., O'Sullivan,M.J., Powell,L.H., Simon,M.S., Van,Horn L., Vitolins,M.Z., Wallace,R.B., Menopausal hormone therapy and health outcomes during the intervention and extended poststopping phases of the Women's Health Initiative randomized trials, JAMA, 310, 1353-1368, 2013 Ref Id 294268 Study type Randomised controlled trial. After discontinuation of the trial, participants were followed up as an observational cohort study. Source of funding National Heart, Lung and Blood Institute, U.S. Department of Health and Human Services. Active study drug and placebo were supplied by Wyeth Ayerst. Country/ies where the study was carried out USA Study dates	Aim of the study To determine the benefits and risks of hormone replacement therapy when taken for chronic disease prevention by a group of predominantly healthy postmenopausal women. Inclusion criteria Oestrogen plus progesterone arm: Postmenopausal women with an intact uterus, aged 50 to 79 years at randomisation. Oestrogen alone arm: Postmenopausal women with a prior hysterectomy. 50 to 79 years at randomisation. Likely to reside in the area for 3 years. Exclusion criteria Medical conditions likely to be associated with a predicted survival of < 3 years, previous breast cancer, other cancer within the last 10 years (except for non-melanoma skin cancer), alcoholism, dementia, transportation problems.	Interventions Women with an intact uterus were randomly assigned to treatment with either 0.625mg conjugated equine oestrogens plus 2.5mg medroxyprogesterone acetate daily, or placebo. Women with a previous hysterectomy were randomly assigned to treatment with 0.625mg conjugated equine oestrogens daily, or placebo. Details Randomisation was was implemented at the WHI Clinical Coordinating Centre with a permuted block algorithm, stratified by clinical centre and age group. When the intervention phase ended, participants were continued to be monitored for trial endpoints as an observational cohort. Methods Clinical outcomes were collected through semi-annual mailed uestionnaires and annual clinic visits. Outcomes were verified by trained physician adjudicators at the local clinical centres by medical record review, followed by final adjudication at the WHI Coordinating Centre. All adjudicators were blinded to treatment assignment. Demographic characteristics and medical history were collected by self report using standardised questionnaires. Sample size Women with a uterus (oestrogen plus progestin arm)	Characteristics Oestrogen plus progestin arm HRT group Age, years† 63.2 (7.1) BMI, kg/m²‡ 27.5 (24.2 to 31.7) Current smokers 554 (6.5%) < 10 years since menopause 2780 (36.2%) Placebo group Age, years† 63.3 (7.1) BMI, kg/m²† 27.5 (24.3 to 31.7) Current smokers 490 (6.1%) < 10 years since menopause 2711 (36.1%) Oestrogen alone arm HRT group Age, years† 63.6 (7.3) BMI, kg/m²† 29.2 (25.7 to 33.7) Current smokers 669 (12.6%) < 10 years since menopause 827 (18.4%) Placebo group Age, years† 63.6 (7.3) BMI, kg/m²† 29.2 (25.7 to 33.5) Current smokers 709 (13.1%) < 10 years since menopause 817 (17.6%) † mean (standard deviation) ‡ median (interquartile range) Results Multiple publications have arisen from this trial and, for convenience, the relevant results from different publications are included below. Unless otherwise stated, VTE outcomes include both DVT and PE. Where different publications report	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Yes. There was adequate concealment of allocation. Yes. The groups were comparable at baseline. Yes. Bias: Low risk of bias Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Unclear. Individuals administering care were kept 'blind' to treatment allocation. Unclear. Bias: Unclear risk of bias Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in

Study details	Design	Comparison	Results	Other
Recruitment began in 1993. Trial suspended in July 2002 (oestrogen plus progesterone arm) and February 2004 (oestrogen only arm). Median intervention duration 5.2 years in combined therapy arm, 7.2 years for oestrogen only arm.		N = 16608 n = 8506 HRT n = 8102 placebo Women without a uterus (oestrogen alone arm) N = 10739 n = 5310 HRT n = 5429 placebo	different hazard ratios, the most up-to-date (recent) publication was used, representing the most complete follow up. The exception to this is where older publications report both DVT and PE outcomes, and newer publications only eported PE. In this instance the older data was used as it more accurately matches the review protocol (all VTE). Oestrogen plus progestin arm VTE during intervention phase in placebo group n/N: 102/8102 VTE during intervention phase in HRT group n/N: 209/8506 Relative risk for VTE in HRT group (95% CI): 1.95 (1.54 to 2.47)† Oestrogen alone arm VTE during intervention phase in placebo group n/N: 98/5429 VTE during intervention phase in HRT group n/N: 137/5310 Relative risk for VTE in HRT group (95% CI): 1.43 (1.11 to 1.85)† Both arms combined VTE during intervention phase in placebo group n/N: 200/13531 VTE during intervention phase in HRT group n/N: 346/13816 Relative risk for VTE in HRT group (95% CI): 1.69 (1.43 to 2.01)† Age of user Women aged 50 to 59 years at baseline, oestrogen plus progestin arm (Data from Cushman et al., 2004) VTE during intervention phase in placebo group n/N: 13/2683 VTE during intervention phase in HRT group n/N: 32/2837 Hazard ratio for VTE in HRT group (95% CI): 2.27 (1.19 to 4.33)‡	each group? not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Bias: Unclear risk of bias Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept ‘blind’ to participants’ exposure to the intervention. Unclear. Investigators were kept ‘blind’ to other important confounding and prognostic factors. Unclear. Bias: Low risk of bias

Study details	Design	Comparison	Results	Other
			Women aged 50 to 59 years at baseline, oestrogen alone arm (Data from Curb et al., 2006) VTE during intervention phase in placebo group n/N: 15/1674 VTE during intervention phase in HRT group n/N: 20/1639 Hazard ratio for VTE in HRT group (95% CI): 1.37 (0.70 to 2.68)‡ Women aged 60 to 69 years at baseline, oestrogen plus progestin arm Pulmonary embolism during intervention phase in placebo group n/N: 22/3655 Pulmonary embolism during intervention phase in HRT group n/N: 40/3854 Hazard ratio for pulmonary embolism in HRT group (95% CI): 1.69 (1.01 to 2.85)‡ Women aged 60 to 69 years at baseline, oestrogen alone arm (Data from Anderson et al., 2004) VTE during intervention phase in placebo group n/N: 39/2465 VTE during intervention phase in HRT group n/N: 49/2386 Hazard ratio for VTE in HRT group (95% CI): 1.31 (0.86 to 2.00)‡ Previous use of HRT, now discontinued - oestrogen alone arm (data from LaCroix et al., 2011) VTE during follow up period in placebo group n/N: 74/3867 VTE during follow up period in HRT group n/N: 52/3778 Hazard ratio for VTE in previous HRT group (95% CI): 0.72 (0.51 to 1.03)‡ Previous use of HRT, now discontinued - oestrogen plus	

Study details	Design	Comparison	Results	Other
			progestin arm (data from Heiss et al., 2008) VTE during follow up period in placebo group n/N: 45/7678 VTE during follow up period in HRT group n/N: 44/8052 Hazard ratio for VTE in previous HRT group (95% CI): 0.95 (0.63 to 1.44)‡ Time since menopause, in E+P arm (data reported by Canonico et al. 2014):, n/N, adjusted HR(95%CI): < 10 years: HRT users: 33/2758 Placebo users: 10/2694 HR: 3.4 (1.6-7.2) - Adjusted for age, BMI, race, history of events, smoking status, total energy expenditure, HRT use at baseline, and HRT use duration Time since menopause, in E-alone arm (data reported by Canonico et al. 2014): n/N, adjusted HR (95% CI): < 10 years: HRT users: 9/817 Placebo users: 8/802 HR: 1.1 (0.4-2.9) - Adjusted for age, BMI, race, history of events, smoking status, total energy expenditure, HRT use at baseline, and HRT use duration †Calculated by the NCC WCH technical team from data reported in the article ‡ Stratified by age, prior disease and randomisation in the WHI dietary intervention trial.	
Full citation Nachtigall,L.E., Nachtigall,R.H., Nachtigall,R.D., Beckman,E.M., Estrogen replacement therapy II: a prospective study in the relationship to carcinoma and cardiovascular and metabolic	Aim of the study To assess the long term effects of oestrogen replacement therapy on postmenopausal women. Inclusion criteria Postmenopausal women (LMP 2 or more years ago) hospitalised on a long term basis	Interventions The treatment group received conjugated equine oestrogens (Premarin) 2.5mg daily and medroxyprogesterone acetate (Provera) 10mg daily for 7 days in each month.	Characteristics HRT group Age, years (mean) 55.3 Time since LMP (years) 4.7 Ethnicity 70% white, 30% black Placebo group	Other information Limitations Very specific and unusual study population - women with long term chronic disease who are permanently hospitalised.

Study details	Design	Comparison	Results	Other
problems, Obstetrics and Gynecology, 54, 74-79, 1979 Ref Id 229959 Study type Randomised controlled double blind trial. Source of funding Not reported. Country/ies where the study was carried out USA Study dates 1965 to 1975. Trial duration 10 years.	at Goldwater Hospital in New York City. Elevated FSH level (>105.5mU) and total urinary oestrogen levels <10µg/dL. Exclusion criteria Previous use of HRT, acute heart disease, hypertension with blood pressure readings of 160/94, prior hysterectomy or any apparent malignancy.	The control group received a placebo matching the active medications in appearance. Details Occurence of adverse effects (including malignancy, hypertension, diabetes, cardiovascular disease, pneumonia, cirrhosis and pulmonary embolism) were recorded for the duration of the trial and compared between those taking HRT and those taking placebo. Methods 84 matched pairs of women were selected on the basis of age (within 2 years) and diagnosis. The research was given 84 matched pairs and randomly selected which member of each pair would be assigned to the treatment group and which to the placebo group. All patients were hospitalised for the duration of the study (10 years) due to the presence of other long term chronic diseases. Even when their diseases were not debilitating, the study patients had a more prolonged period of bed rest than a typical ambulatory patient. Sample size N = 168 n = 84 placebo group n = 84 HRT group	Age, years (mean) 54.9 Time since LMP (years) 4.5 Ethnicity 69% white, 31% black Results Occurence of pulmonary embolism in placebo group n/N: 1/84 Occurence of pulmonary embolism in HRT group n/N: 0/84 Relative risk of PE in HRT group (95% CI): 0.33 (0.01 to 8.07)	Randomisation process highly subject to bias. Study conducted in 1960's with much higher dose of oestrogen than would be typically used today. Unclear whether incidence of DVT was recorded but simply did not occur, or whether this was not recorded as an adverse event. Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Unclear - study nurse randomly selected which patient would be assigned to each group. Method not described. There was adequate concealment of allocation. Unclear. The groups were comparable at baseline. Yes. Bias: Unclear risk of bias Performance bias The comparison groups received the same care apart from the intervention(s) studied. Unclear. Participants receiving care were kept 'blind' to treatment allocation. Yes. Individuals administering care were kept 'blind' to treatment allocation. Unclear Bias: High risk of bias Attrition bias

Study details	Design	Comparison	Results	Other
				All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Follow-up was 100% The groups were comparable for treatment completion. Yes. For how many participants in each group were outcome data not available? None The groups were comparable with respect to the availability of outcome data. Yes. Bias: Low risk of bias Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. No. (the embolic phenomenon was a complication which was a cause of death) A valid and reliable method was used to determine the outcome. Unclear. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear (reported that an attempt was made to keep research physicians blinded to interventions) Investigators were kept 'blind' to other important confounding and

Study details	Design	Comparison	Results	Other
				prognostic factors. Unclear. Bias: Unclear risk of bias
Full citation Ohira,T., Folsom,A.R., Cushman,M., White,R.H., Hannan,P.J., Rosamond,W.D., Heckbert,S.R., Reproductive history, hormone replacement, and incidence of venous thromboembolism: the Longitudinal Investigation of Thromboembolism Etiology, British Journal of Haematology, 149, 606-612, 2010 Ref Id 301220 Study type Prospective cohort study. Source of funding Grants from the National Heart, Lung and Blood Institute. National Institute of Neurological Disorders and Stroke. Country/ies where the study was carried out USA Study dates Enrollment from 1987 to 1990. Follow up until December 31st 2001 or December 31st 2002.	Aim of the study To study the 12-year risk of VTE in relation to hormone replacement therpay use in postmenopausal women. The data were obtained from the combination of two prospective cohort studies: the Atherosclerosis Risk in Communities and the Cardiovascular Health Study. Inclusion criteria Postmenopausal white or black women aged over 45. Exclusion criteria Pre or perimenopausal women. Non-white or non-black ethnicity. Baseline history of VTE, cancer or warfarin use. Missing menopausal data.	Interventions Not applicable. Details Rate ratios of VTE were calculated with adjustment for age and other potential confounding factors using Cox proportional hazards model. Rates were compared between current users of HRT and those who were not currently using HRT. Methods Participants underwent baseline assessment of cardiovascular risk factors. Up to three follow up examinations were performed every three years for ARIC study participants, and up to 9 follow up examinations were performed annually for CHS participants. Subjects were followed to determine the incidence of VTE until December 31st 2002 for ARIC and December 31st 2001 for CHS. All participants were contacted annually by phone and asked about all hospitalizations in the past year. VTE events were validated by two physicians. Diagnosis of DVT or PE required positive imaging tests. -15-year follow-up Sample size N = 8236 n = 190 with VTE n = 8046 without VTE	Characteristics Only reported for cases of VTE compared to those without VTE, not for HRT users compared to non-users. Cases: Age, years (mean) 64.0 BMI, kg/m² (mean) 29.3 Race (% African American) 37% Never use of HRT 63.4% Former use of HRT 18.2% Current use of HRT 18.2% Controls: Age, years (mean) 61.0 BMI, kg/m² (mean) 27.6 Race (% African American) 29.1% Never use of HRT 63.3% Former use of HRT 19.2% Current use of HRT 17.5% Results Rate ratios (RR) are adjusted for age, race, BMI, diabetes mellitus and factor VIII at baseline, as well as other reproductive variables. They are expressed compared to the rate in never users of HRT. Current use of HRT RR (95% CI): 1.60 (1.06 to 2.36) Past use of HRT RR (95% CI): 1.07 (0.72 to 1.62)	Other information -Only clinically recognized VTE was ascertained in this study, which depended on participants' accurate reporting of hospitalization and on their physicians' diagnostic work-up of supspected VTE events. Limitations Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. Yes (population-based cohort study) Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes The groups were comparable at baseline, including all major confounding and prognostic factors. Unclear (Mostly comparable but the None VTE group were younger, had lower BMI and less African American women) Level of risk: Unclear Performance bias The comparison groups received the same care apart from the intervention(s) studied. N/A Participants receiving care were kept ‘blind’ to

Study details	Design	Comparison	Results	Other
				treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Level of risk: Unclear Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Not applicable. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Not applicable. Level of risk: Unclear Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors.

Study details	Design	Comparison	Results	Other
				Unclear. Level of risk: Unclear
Full citation Olie,V., Plu-Bureau, Conard,J., Horellou,M.H., Canonico,M., Scarabin,P.Y., Hormone therapy and recurrence of venous thromboembolism among postmenopausal women, Menopause, 18, 488-493, 2011 Ref Id 311435 Study type Retrospective cohort study. Source of funding Partially supported by a grant from PIerre Fabre Santé. Country/ies where the study was carried out France Study dates January 1st 2000 to December 31st 2008.	Aim of the study To evaluate the safety of transdermal oestrogens among postmenopausal women with a personal history of venous thromboembolism. Inclusion criteria Postmenopausal women aged 45 to 70 who attended the outpatient clinic of the Hotel Dieu hospital because of a first objectively confirmed episode of VTE (established with an imaging procedure). Exclusion criteria Superficial vein thrombosis, upper extremity VTE and central retinal vein thrombosis.	Interventions Not applicable. Details Cumulative incidence of recurrent VTE was estimated by the Kaplan Meier survival method, censoring at the time of thrombotic event recurrence or at the end of the study. Univariate and multivariate Cox proportional hazard models were used to estimate the risk of recurrent VTE associated with potential risk factors. Methods Women's characteristics were extracted from medical records using a standard questionnaire. Basline data included information on the first VTE event; medical history; reproductive factors; cardiovascular risk factors (e.g. height, weight, smoking status, diabetes, dyslipidaemia and hypertension) and the use of exogenous hormones. The presence of transient risk factors in the month preceding the first event was recorded. These factors included surgery, trauma, plaster, prolonged immobilization (> 10 days), oral contraceptive or HRT use, pregnany, venous sclerosis or air travel. In the absence of one of these conditions, VTE was considered idiopathic. The endpoint of the study was a documented recurrent VTE event. Recurrent events were adjudicated by a medical committee blinded to the use of HRT, using the same validation as for the initial event (diagnostic imaging was required). Follow up continued from the time of discontinuation of anti-coagulant	Characteristics Users of HRT: Age at baseline, years† 55.4 (5.5) BMI, kg/m²† 23.7 (4.1) Duration of follow up, months† 105 (104.7) Family history of VTE 50 (40.3%) Idiopathic first event 15 (11.7%) Thrombophilia 20 (15.4%) Non-users of HRT: Age at baseline, years† 58.3 (5.4) BMI, kg/m²† 25.2 (4.5) Duration of follow up, months† 75.2 (78.6) Family history of VTE 406 (48.2%) Idiopathic first event 212 (24.0%) Thrombophilia 246 (27.6%) † mean (standard deviation) Results Multivariate hazard ratios (HR) include age, overweight, obesity and characteristics of first event (idiopathic or secondary) and are compared to non-users of HRT. Route of administration Oral oestrogens HR (95% CI): 6.4 (1.5 to 27.3) Transdermal oestrogens HR (95% CI): 1.0 (0.4 to 2.4) HRT preparation Transdermal oestrogen alone HR (95% CI): 1.1 (0.2 to 8.1) Transdermal oestrogen and micronized progesterone HR (95% CI): 1.0 (0.3 to 3.2) Transdermal oestrogen and pregnane derivatives (no events therefore HR not calculable) Transdermal oestrogen and	Other information Limitations Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. No (participants were women with a confirmed first VTE) Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline, including all major confounding and prognostic factors. Unclear (mostly similar but different on characteristics of age (younger in HRT use group), duration of follow-up (longer for HRT use group etc) Level of risk: High Performance bias The comparison groups received the same care apart from the intervention(s) studied. Unclear. Participants receiving care were kept 'blind' to treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Level of risk: Unclear Attrition bias All groups were followed

Study details	Design	Comparison	Results	Other
		therapy from the first event to the time of recurrent VTE, or the date of the follow up questionnaire. Women were classified as HRT users if they had used HRT at any time during the 3 months before the date of recurrent VTE. All other women were classified as non-users (past- and never-users combined). -8-year follow-up Sample size N = 1023 n = 130 users of HRT n = 893 non-users of HRT	norpregnane derivatives HR (95% CI): 4.7 (1.1 to 20.0)	up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). No (about 2-yr longer follow-up in the HRT use group but reason not reported) How many participants did not complete treatment in each group? Not applicable. The groups were comparable for treatment completion. Yes. For how many participants in each group were outcome data not available? Not applicable. The groups were comparable with respect to the availability of outcome data. Yes. Level of risk: High Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. N/A Investigators were kept 'blind' to other important confounding and prognostic factors. N/A Level of risk: High
Full citation Su,I.H., Chen,Y.C., Hwang,W.T., Liu,Z., Su,T.P., Chen,T.J.,	Aim of the study To determine whether conjugated equine oestrogens with or	Interventions Not applicable. Details	Characteristics Oestrogen plus progestin HRT group	Other information -The study was a population-based study

Study details	Design	Comparison	Results	Other
Barnhart,K.T., Yang,Y.X., Risks and benefits of menopausal hormone therapy in postmenopausal Chinese women, Menopause, 19, 931-941, 2012 Ref Id 203512 Study type Retrospective cohort study. Source of funding ASRM/Ortho Research Grant in REproductive Medicine. Country/ies where the study was carried out Taiwan. Study dates Enrollment from June 1st 1997 to May 31st 2000. Follow up until 2007.	without medroxyprogesterone acetate increase the risks of cardiovascular disease and breast cancer in postmenopausal Chinese women. Inclusion criteria Women aged 50 to 80. Exclusion criteria Women using HRT preparations other than 0.625mg conjugated equine oestrogens (+/- medroxyprogesterone acetate). Medical condition associated with predicted survival < 3 years (AIDS, COPD, CHF, ESRD). Prior breast cancer. Other prior cancers within the last 10 years. Endometrial hyperplasia, alcoholism, drug dependency, dementia, mental illness. Acute MI, CVA or TIA within the past 6 months. Severe hypertension, chronic hepatitis or cirrhosis, previous PE or DVT.	Cox proportional hazard ratios were estimated for each primary outcome. Covariates that were clinically known confounders, or that changed the crude hazard ratio by more than 10% were included in the multivariable models. Methods Potential eligible participants who filed at least 2 monthly prescriptions for HRT within 3 consecutive months were categorized as exposured to HRT. This group subdivided into those who filled prescriptions for conjugated equine oestrogens (0.625mg daily) and medroxyprogesterone acetate (5mg daily), and those who only filled prescriptions for conjugated equine oestrogens (0.625mg daily). Unexposed participants were randomly selected from the remainder of the cohort. 2 age matched (within 5 years) unexposed participants were randomly selected for each exposed participant. Outcome data were collected from a National Insurance Registry data, as reported by ICD-9 codes. -Median follow-up was 110 months, Median duration of exposure in the E+P and E-only groups were 6.9 months and 9 months, respectively. Sample size N = 10715 n = 5920 exposed to HRT (n = 4712 oestrogen plus progestin, n = 1208 oestrogen only) n = 10125 not exposed to HRT (n = 8070 matched to oestrogen plus progestin group, n = 2055 matched to oestrogen only group)	Age, years† 58.2 (6.3) Current smokers 0 (0%) Obesity 2 (0.04%) Control group for oestrogen plus progestin (unexposed) Age, years† 58.9 (6.2) Current smokers 0 (0%) Obesity 2 (0.03%) Oestrogen alone HRT group Age, years† 59.2 (6.9) Current smokers 0 (0%) Obesity 1 (0.08%) Control group for oestrogen alone (unexposed) Age, years† 59.7 (6.7) Current smokers 0 (0%) Obesity 1 (0.01%) †mean (standard deviation) Results Hazard ratios (HR) are compared to non-exposed control group and are adjusted for age, statin use, hypercholesterolaemia, hypertension and use of diabetes medication. Risk of PE in combined HRT group (oestrogen plus progestin) HR (95% CI): 0.80 (0.35 to 1.85) Risk of DVT in combined HRT group (oestrogen plus progestin) HR (95% CI): 0.90 (0.51 to 1.60) Risk of PE in oestrogen alone HRT group HR (95% CI): 2.75 (0.45 to 16.8) Risk of DVT in oestrogen alone HRT group HR (95% CI): 3.63 (1.48 to 8.89)	carried out among Chinese women in Taiwan Limitations Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. Unclear Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes The groups were comparable at baseline, including all major confounding and prognostic factors. Yes. Level of risk: Unclear Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Level of risk: Unclear Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? 4% (follow-up was complete on 96% of

Study details	Design	Comparison	Results	Other
				participants) The groups were comparable for treatment completion. Not applicable. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Not applicable. Level of risk: Low Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Unclear. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear (data was extracted from health insurance datasets). Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear (data was extracted from health insurance datasets) Level of risk: Unclear
Full citation Vickers,M.R., MacLennan,A.H., Lawton,B., Ford,D., Martin,J., Meredith,S.K., DeStavola,B.L., Rose,S., Dowell,A., Wilkes,H.C., Darbyshire,J.H., Meade,T.W., WISDOM group., Main morbidities recorded in the women's international study of long duration oestrogen after menopause	Aim of the study To assess the balance of long term risks and benefits of hormone replacement therapy, with particular emphasis on cardiovascular disease and dementia. Inclusion criteria Postmenopausal women aged 50 to 69 years. Exclusion criteria History of breast cancer, any cancer in the	Interventions The combined therapy was 0.625mg conjugated equine oestrogens (CEE) plus 2.5mg medroxyprogesterone acetate (MPA) orally daily. Women with a uterus and within 3 years of their last period, those aged 50 to 53 and older women with unacceptable breakthrough	Characteristics HRT users: Age, years† 63.6 (4.7) BMI, kg/m²† 27.9 (4.9) Current smoker 256 (12%) Placebo users: Age, years† 63.3 (4.6) BMI, kg/m²† 28.0 (5.2) Current smoker 309 (14%)	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Yes. There was adequate concealment of allocation.

Study details	Design	Comparison	Results	Other
(WISDOM): a randomised controlled trial of hormone replacement therapy in postmenopausal women, BMJ, 335, 239-, 2007 Ref Id 230610 Study type Randomised controlled trial. Source of funding Wyeth Ayerst provided the active drugs and matched placebo but had no other involvement in the trial. UK Medical Research Council. British Heart Foundation. Department of Health for England. Scottish Office. Welsh Office. Department of Health and Social Services for Northern Ireland. Royal Australian and New Zealand College of Obstetricians and Gynaecologists. Australasian Menopause Society. National Health and Medical Research Council. National Heart Foundation of Australia. The Cancer Council of South Australia. The Cancer Society of New Zealand (Wellington Branch). NHS R&D Executive. Country/ies where the study was carried out UK, Australia and New Zealand Study dates Recruitment began in the UK in 1999, and in Australia and New Zealand in 2000. The trial was stopped in 2002 (whilst recruitment was still underway) following the publication of trial results for the combined oestrogen and progestagen arm of the WHI study. Median duration of treatment was	past 10 years (except basal and squamous cell skin cancer), endometriosis or endometrial hyperplasia, venous thromboembolism, gall bladder disease in womn who had not had a cholecystectomy, myocardial infarction, unstable angina, cerebrovascular accident, subarachnoid haemorrhage, transient ischaemic attack, or use of HRT within the past 6 months.	bleeding took 5.0mg MPA. Women with a uterus who experienced unacceptable spotting or bleeding with the combined therapy containing 5.0mg MPA were offered open label Premique cycle (0.625mg CEE orally daily plus MPA 10mg orally for the last 14 days of a 28 day cycle). Details Treatment was randomly allocated centrally with a computer based, stratified block randomisation system. Women with a uterus or subtotal hysterectomy were randomised to combined oestrogen plus progestogen, or to placebo, using a block size of 16. Women with no uterus were also included in the trial, but only for a comparison on oestrogen alone versus oestrogen plus progestagen therapy, therefore are not included for the purposes of this analysis. Hazard ratios were calculated under the Cox proprtional hazards model. Methods Women were to be seen at 4, 14, 27, 40 and 52 weeks after the start of treatment, and then at 6 months intervals. A final visit took place as soon as possible after the closure of the trial. At the start of treatment, and at all subsequent follow up visits, information was collected on all outcomes, adverse events and other medical history. A member of the study team (blinded to treatment allocation) obtained any data needed to confirm a clinical event from the general practice, hospital or coroner. Primary outcomes were major	† Mean (standard deviation) Results Risk of venous thromboembolism in users of HRT compared to placebo Hazard ratio (95% CI): 7.36 (2.20 to 24.60) Risk of fatal venous thromboembolism in users of HRT compared to placebo Relative risk (95% CI): 4.98 (0.24 to 103.76)	Yes. The groups were comparable at baseline. Yes. Bias: Low risk of bias Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Yes. Individuals administering care were kept 'blind' to treatment allocation. Yes. Bias: Low risk of bias Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 430 HRT arm, n = 203 placebo arm. The groups were comparable for treatment completion. Apparent increase in withdrawals in HRT arm - predominantly due to unacceptable vaginal bleeding. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Bias: High risk of bias

Study details	Design	Comparison	Results	Other
11.9 months (inter-quartile range 7.3 to 19.6 months).		cardiovascular disease, osteoporotic fractures and breast cancer. Secondary outcomes were breast cancer mortality, other cancers, death from all causes, venous thromboembolism, cerebrovascular disease and dementia. Participants were asked about symptoms and adverse events at each visit. Sample size N = 4385 n = 2196 HRT n = 2189 placebo		Detection bias The study had an appropriate length of follow up. No - trial was terminated prematurely and provided data for a median of 11.9 months follow up. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Unclear - not stated whether diagnostic imaging was required to define cases. Investigators were kept 'blind' to participants' exposure to the intervention. Yes. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear. Bias: High risk of bias
Full citation Whiteman,M.K., Cui,Y., Flaws,J.A., Espeland,M., Bush,T.L., Low fibrinogen level: A predisposing factor for venous thromboembolic events with hormone replacement therapy, American Journal of Hematology, 61, 271-273, 1999 Ref Id 230680 Study type Randomised controlled trial. Source of funding Research grants from the National Heart, Lung and Blood Institute; the National Institute of Child Health and Human Development; the National Institute of Arthritis and Musculoskeletal and Skin	Aim of the study To examine potential risk factors for VTE among women enrolled in the Postmenopausal Estrogen/Progestin Interventions (PEPI) trial. Inclusion criteria Surgically or naturally menopausal women (longer than 1 year, but less than 10 years since LMP) aged 45 to 64. Not taking oestrogens or progestins for at least 2 months prior to the first screening visit (> 4 months before randomization). If treated with thyroid hormone replacement, to have been on a stable dose for at least 3 months prior to initial screening. Exclusion criteria Extreme hyperlipidaemia, marked obesity, severe hypertension, recent myocardial infarction, congestive heart failure, stroke or	Interventions Participants were assigned to one of the following regimes in 28 day cycles: 1. Placebo 2. active treatment arms, which included four separate regimes: • conjugated equine estrogens (CEE) 0.625mg/day • CEE 0.625mg/day plus medroxyprogesterone acetate (MPA) 10mg/day for days 1 to 12 • CEE 0.625mg/day plus MPA 2.5mg/day • CEE 0.625mg/day plus micronized progesterone 200mg/day for day 1 to 12 For the purposes of this analysis data for the four active treatment	Characteristics Average age 56.1 years No significant differences in prior menopausal hormone use, smoking status, ethnicity or physical activity between the groups. Other characteristics reported separately for those taking HRT who suffered VTE and those who did not. In published analysis superficial phlebitis is regarded as VTE, whereas for the purposes of this analysis only DVT and PE were included. Therefore characteristics of women who developed DVT/PE are not identifiable. Results VTE in placebo group n/N: 0/174 VTE in HRT group n/N: 4/701	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Unclear. There was adequate concealment of allocation. Unclear. The groups were comparable at baseline. Yes. Bias: High risk of bias Performance bias The comparison groups received the same care

Study details	Design	Comparison	Results	Other
Diseases; the National Institute of Diabetes and Digestive and Kidney Diseases and the National Institute on Aging. Support was also provided by General Clinical Research Center Grants (University of California, Los Angeles; University of California, SanDiego and University of Iowa). Study medications were provided by Wyerth-Ayerst Laboratories, Philadelphia, Pa (conjugated equine estrogens), The Upjohn Company, Kalamazoo, Mich (medroxyprogesterone acetate) and Schering-Plough Research Institute, Kenilworth, NJ (micronized progesterone). Country/ies where the study was carried out USA Study dates Randomization occurred between December 1989 and February 1991. Trial duration was for three years.	TIA, anti-arrythmia medication use, diabetes mellitus requiring insulin, prior breast or endometrial cancer, melanoma, any non-basal cell skin cancer in the previous five years, an elevated thyroid stimulating hormone concentration, a history of trauma to the lower spine or hip fracture, chronic steroid use and severe menopausal symptoms.	arms were combined. Details After the first randomization visit, participants returned 3 times during the first year and biannually for the remaining 2 years. Symptoms, occurrence of vaginal bleeding, medications used, adherence to medications, adverse experiences (including fractures), blood pressure, weight and height were assessed at each visit. Methods No data are presented for women on individual HRT preparations, only for those taking and not taking HRT. Incidence of VTE in the two groups was compared. Sample size N = 875 n = 174 placebo group n = 701 active treatment group	Relative risk of VTE in HRT group (95% CI): 2.24 (0.12 to 41.48)	apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Yes. (no details reported) Individuals administering care were kept 'blind' to treatment allocation. Yes. (no details reported) Bias: Low risk of bias Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 11 placebo group, n = 28 HRT groups. The groups were comparable for treatment completion. Yes. For how many participants in each group were outcome data not available? n = 11 placebo group, n = 28 HRT groups. The groups were comparable with respect to the availability of outcome data. Yes. Bias: Low risk of bias Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to

Study details	Design	Comparison	Results	Other
				determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Yes Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear. Bias: Low risk of bias

H.8.2 Cardiovascular disease

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Full citation Cherry,N., McNamee,R., Heagerty,A., Kitchener,H., Hannaford,P., Long-term safety of unopposed estrogen used by women surviving myocardial infarction: 14-year follow-up of the ESPRIT randomised controlled trial, BJOG: An International Journal of Obstetrics and Gynaecology, 121, 700-705, 2014 Ref Id 321013 Country/ies where the study was carried out England and Wales Study type	Sample size N=1,017 Estrogen group: n=513 Placebo group: n=504 Characteristics Need check reference 1 Inclusion criteria All women aged 50-69 years admitted to coronary care units or general medical wards in participating hospitals in England and Wales between 1996 and 2000, provided that they: -met the diagnostic criteria for MI; were discharged alive from hospital within 31 days of admission. Exclusion criteria -Women who reported a history of cancer or use of HRT or vaginal bleeding in the previous 12 months; or active thrombophlebitis or a history of deep-vein thrombosis or pulmonary embolism, acute or chronic liver disease. -Rotor syndrome, Dubin-Johnson syndrome, or severe renal disease.	Interventions unopposed estrogen	Details Setting: Hospitals Methods: Randomisation: Randomisation was stratified bv hospital, where the trial statistician used a restricted randomsation scheme based on a block size of four to generate a list of treatment allocations Concealment of allocation: Consecutive study numbers were attached to the allocations. The lists were sent to Schering AC who prepared numbered packages that contained the corresponding treatments Blinding: The two treatments were of identical appearance and were supplied in identical packaging Outcome ascertainment: Cancer incidence, vital status and cause of death were determined from data routinely collected by the Office of National Statistics for England and Wales Statistical methods: Hazard ratio (HRs) comparing treatment arms were estimated using Cox regression. All HRs were adjusted	Results Risk of IHD death in relation to Estrogen, n/N (%), HR (95%CI) By age: 50-59 yr: Estrogen: 23/167 (13.8) Placebo: 14/134 (10.5) HR: 1.23 (0.63-2.41) -all models adjusted for age at risk	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-No, participants were originally recruited from an RCT A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders- Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Yes Level of risk-Unclear B. Performance bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Prospective cohort Aim of the study To compare health outcomes during 14-year observational follow-up in women initially randomised to unopposed estrogen or placebo. Study dates 1996-2002 (enrolment) to 2012 Source of funding UK National Health Services Research and Development Programme on Cardiovascular Disease and Stroke			for age at risk, using six 5-year age bands (50-55 to 75-80). Follow-up: mean follow-up 12.6 years (range: 10.9-14.5) for cancer and mean follow-up 14.1 years (range 12.4-16.0) for mortality.		(systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-No B.3 Individuals administering care were kept 'blind' to treatment allocation-No Level of risk:High C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Unclear C.2a How many participants did not complete treatment in each group?-N/a C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-Not reported C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					outcome data were not available)-N/A Level of risk: Unclear D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up- Yes D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-No D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-No Level of bias: Unclear Other information -During the extended follow-up of the original ESPRIT trial, researchers could not assess whether, over time, unopposed estrogen affects the risk of non-fatal re-infarction. Data were not available about use of HRT after the formal trial ended. Some women may have used these products subsequently, although the number is probably small due to the widespread publicity that occurred in the summer 2002 concerning the early stop of WHI.
Full citation Manson,J.A.E., Hsia,J., Johnson,K.C.,	Sample size N= 16,608 (Intervention (E+P) group: n=8506; conrol group: n= 8102)	Interventions estrogen plus progestin	Details Consent Informed written consent obtained from participants	Results Risk of CHD (including nonfatal myocardial infraction and death due	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Rossouw,J.E., Assaf,A.R., Lasser,N.L., Trevisan,M., Black,H.R., Heckbert,S.R., Detrano,R., Strickland,O.L., Wong,N.D., Crouse,J.R., Stein,E., Cushman,M., Estrogen plus progestin and the risk of coronary heart disease, New England Journal of Medicine, 349, 523-534, 2003 Ref Id 311345 Country/ies where the study was carried out US Study type RCT Aim of the study To present the final results of the WHI trial of the relation between the use of estrogen plus progestin and the risk of CHD; to provide an updated analysis of coronary end points reached through the termination of the trail on July 7, 2002 (previous analyses	(The sample analyzed here consists of the 16,608 women with an intact uterus at baseline who were enrolled in the double-blinded trial comparing esrogen plus progestin with placebo. The study regimen of combined estrogen and progestin was provided in one daily tablet containing 0.625 mg of oral conjugated equine estrogen and 2.5 mg of medroxyprogesterone acetate. The control group received matching placebo) Characteristics **Estrogen+progestin (n=8506)** / **Placebo (n=8102)** / **P value** Age at screening, mean (SD): 63.2 (7.1) / 63.3 (7.1) / 0.39 Age group at screening, y 50-59: 2839 (33.4) / 2683 (33.1) / 0.80 60-69: 3853 (45.3) / 3657 (45.1) 70-79: 1814 (21. / 1762 (21.		Setting Clinical trial, 40 clinical centre sites across the country Randomisation method The randomization procedure was developed at the WHI Clinical Coordinating Centre, using a randomized permuted block algorithm, stratified by clinical centre site and age group; Concealment of allocation All study medicate on bottles had a unique bottle number and bar code to allow for blinded dispensing Comparability of intervention groups at baseline The two groups were almost identical Blinding Considerable effort was made to maintain blinding of other participants and clinic staff. When required for safety or symptom management, an unblinding officer provided the clinic gynaecologist, who was not involved with study outcomes activities, with the treatment assignment. Statistical methods -sample size calculation (need durther check here from the design paper which is being ordered) -Primary analyses used time-to-event methods based on the intention-to-treat principle. Comparisons with regard to the primary outcome are presented as hazard ratios with 95% confidence intervals that were calculated from Cox proportional-hazards analyses, stratified according to age, presence or absence of CHD at baseline etc, and adjusted for the presence or absence of previous	to CHD) in relation to Estrogen + progestin, n (no. of cases of CHD, annualized percentage), adjusted hazard ratio (HR, 95%CI) By age: 50-59 yr: E+P: 37 (0.22) Placebo: 27 (0.17) HR: 1.27 (0.75-2.10) 60-69yr: E+P: 75 (0.35) Placebo: 68 (0.34) HR: 1.05 (0.75-1.35) -adjusted for the presence and absence of CHD at baseline; Confidence intervals here were reported by graph in the study and approximated by NCC-WCH based on it. By years since menopause (just for information giving in the evidence table): <10 yr: E+P: 31 (0.19) Placebo: 34 (0.22) HR: 0.89 (0.40-1.51) 10-19 yr: E+P: 63 (0.38) Placebo: 51 (0.32) HR: 1.22 (0.85-1.75) >=20 yr: E+P: 74 (0.75) Placebo: 44 (0.46) HR: 1.71 (1.25-2.6) -Adjusted for the presence or absence of CHD at baseline; Confidence intervals here were reported by	controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Unclear (with an average follow-up of 5.6 yrs, women taking HRT should have realized which group they were allocated to when HRT taking effect) B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Unclear C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes (48% in intervention arm versus 38% in the placebo arm) C3 - Were groups comparable for missing data - Yes Level of bias: High D Detection bias D1 - Was follow-up appropriate length - Unclear (the trial was stopped at an average follow-up of 5.6 years, which was earlier than planned) D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable

Study details	Participants				Interventions	Methods	Outcomes and Results	Comments
included end points reached through April 2002). Study dates Recruitment: 1993-1998 Ended in 2002 An average of 5.6 years of follow-up Source of funding NIH		3)	7)			CABG or PTCA. -Because CHD was the primary outcome of the hormone trial and was an important consideration for stopping the trial early, both nominal 95% intervals and 95% intervals adjusted for sequential monitoring are provided for the primary coronary end point. -Cox models for subgroup analyses were stratified according to age and the presence or absence of CHD at baseline. -Intention to treat analysis (ITT) -Analyses were performed according to ITT principle -Outcomes ascertainment: - CHD was defined as acute MI requiring overnight hospitalization, silent MI determined from serial electrcardiograms, or CHD deaths; -Stroke: At each semiannual contact, a standardized interview asked participants about symptoms, safety, and potential outcome events. When a potential outcome was identified, medical records and death certificates were obtained as necessary. Physician adjudicators at clinical sites reviewed the information to determine the cause of the event. Of locally adjudicated stroke, 94.5% were confirmed by the central adjudicators. Stroke data were centrally confirmed by neurologists. Local and central adjudicators were blinded to treatment assignment. Follow-up -an average of 5.2 yrs; follow-up for clinical events occured every 6 months, with annual in-clinic visits required. -Drop out-: 42% in CEE+MPA arm; 38% in the placebo arm; 10.7% cross-over from the placebo to treatment arm (drop-in)	graph in the study and approximated by NCC-WCH based on it. (All stroke and sroke stratified by age findings of WHI reported under Wassertheil-Smoller et al. 2003) Risk of all stroke (including ischemic and hemorrhagic stroke) in relation to Estrogen + progestin, n (%), adjusted hazard ratio (HR, 95%CI) All stroke (just for information in the evidence table): Estrogen+progestin group: 151 (0.31) Placebo group: 107 (0.24) HR (95%CI): 1.31 (1.02-1.68) By age: 50-59 yr: E+P: 24 (0.14) Placebo: 15 (0.10) HR: 1.46 (0.77-2.79) 60-69yr: E+P: 68 (0.32) Placebo: 47 (0.23) HR: 1.35 (0.93-1.96) 70-79 yr: E+P: 59 (0.61) Placebo: 45 (0.48) HR: 1.26 (0.86-1.86) -Adjusted for previous stroke and diabetes randomization treatment;	method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes (women aged 50-59) Intervention: yes Outcomes: yes Indirectness: Some Other information WHI trial is a trial involving predominantly healthy women with only 5% having a history of CVD. Their low-baseline risk is illustrated by the fact that even though the WHI cohort was much larger (N=16608) than other studies, only 335 CHDs and 258 strokes occured during the 5.6 year follow-up; -Because of the large number of subgroups considered (at least 36) in this study, the results should be interpreted with caution, since some significant findings (at least one or two, based on 0.05 nominal level of statistical significance) could have occured by chance alone. -The relatively high rate of discontinuation of HT in the trial, which tends to decrease the observed treatment effects and may lead to an underestimate of adverse CVD effects.
	Race/ethnicity							
	White	7140 (83.9)	6805 (84.0)	0.33				
	Black	549 (6.5)	575 (7.1)					
	Hispanic	472 (5.5)	416 (5.1)					
	American Indian	26 (0.3)	30 (0.4)					
	Asian/Pacific Islander	194 (2.3)	169 (2.1)					
	Unknown	125 (1.5)	107 (1.3)					
	Hormone use							
	Never	6280 (73.9)	6024 (74.4)	0.49				
	Past	1674 (19.7)	1588 (19.6)					
	Current	548 (6.4)	487 (6.0)					
	Duration of prior hormone							

Study details	Participants				Interventions	Methods	Outcomes and Results	Comments
	use, y							
	<5 yr	1538 (69.1)	1467 (70.6)	0.25			By duration of prior HRT use (for information giving in the evidence table): Never: E+P: 117 (0.33) Placebo: 80 (0.24) HR: 1.37 (1.03-1.82) <5 yr: E+P: 17 (0.19) Placebo: 17 (0.20) HR: 0.96 (0.49-1.88) 5-10 yr: E+P: 10 (0.41) Placebo: 7 (0.36) HR: 1.04 (0.40-2.73) >=10 yr: E+P: 7 (0.49) Placebo: 3 (0.22) HR: 2.17 (0.56-8.40)	
	5-10 yr	426 (19.1)	357 (17.2)					
	>= 10	262 (11.8)	253 (12.2)					
	BMI, mean (sd), kg/m2	28.5 (5.8)	28.5 (5.9)	0.66				
	<25	2579 (30.4)	2479 (30.8)	0.89				
	25-29	2992 (35.3)	2834 (35.2)					
	>=30	2899 (34.2)	2737 (34.0)					
	Systolic BP, mean (SD), mm Hg	127.6 (17.6)	127.8 (17.5)	0.51				
	Diastolic BP, mean (SD)	75.6 (9.1)	75.8 (9.1)	0.31				

Study details	Participants				Interventions	Methods	Outcomes and Results	Comments
	, mm Hg							
	Smoking							
	Never	4178 (49.6)	3999 (50.0)	0.85				
	Past	3362 (39.9)	3157 (39.5)					
	Current	880 (10.5)	838 (10.5)					
	Treated for diabetes	374 (4.4)	360 (4.4)	0.88				
	Treated for hypertension or BP >= 140/90 mm Hg	3039 (35.7)	2949 (36.4)	0.37				
	Elevated cholesterol levels requiring medication	944 (12.5)	962 (12.9)	0.50				

Study details	Participants				Interventions	Methods	Outcomes and Results	Comments
	Statin use at baseline	590 (6.9)	548 (6.8)					
	History of myocardial infraction	139 (1.6)	157 (1.9)	0.14				
	History of angina	238 (2.8)	234 (2.9)	0.73				
	History of CABG/PTCA	95 (1.1)	120 (1.5)	0.04				
	History of stroke	61 (0.7)	77 (1.0)	0.10				
	History of DVT or PE	79 (0.9)	62 (0.8)	0.25				
	Female relative had breast cancer	1286 (16.0)	1175 (15.3)	0.28				
	Frac	103	102	0.87				

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	ture at age >= 55 yr | 1 (13.5) | 9 (13.6) | (Extracted from: Hendrix et al. 2006 "Effects of conjugated equine estrogen on stroke in the WHI". Circulation, 113: 2425-2434" where updated data on an additional 19 stroke cases were included compared with the Anderson et al. 2004 publication) Inclusion criteria -Most women were recruited by population-b ased direct mailing campaigns to age-eligible women, in conjunction with media awareness progrems -women aged 50-79 at initial screening, post menopausal, likelihood of residence in the area for 3 years, and provision of written informed consent; -a 3-month washout period was required before baseline evaluation of women using postmenopausal hormones at initial screening; -women with an intact uterus at initial screening were eligible for the trial of combined postmenopausal hormones, while women with a prior hysterectomy were eligible for the trial of unopposed estrogen. This current report is limited to the 16608 women with an intact uterus at baseline who were enrolled in the trial component of estrogen plus progestin vs placebo. Exclusion criteria -Women who had medical conditions predictive of a survival time of less than 3 years;				

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	-Women were known to have conditions or characteristics inconsistent with study participation and adherence (alcoholism, drug dependency, mental illness, dementia); -Or if they were active participants in another RCT -Also, women were excluded from clinical trials for: reasons of competing risks (e.g., invasive cancer in the past 10 yrs; breast cancer at any time or suspicion of breast cancer at baseline screening; acute MI, stroke, or transient ischemic attack in the previous 6 months; reasons of safety (severe hypertension, or currently use of oral corticosteriods); and reasons relating to adherence or retention (unwillingness or inability to complete baseline study requirements). In addition, women were found to have femoral neck bone mineral density of more than 3 standard deviations below the corresponding age-specific mean were also excluded.				
Full citation Toh,S.D., Hernandez-Diaz,S., Logan,R., Rossouw,J.E., Hernan,M.A., Coronary heart disease in postmenopausal recipients of estrogen plus progestin therapy: Does the increased risk ever disappear? A randomized trial,	Sample size 16,608 (8506 in CEE/MPA group, and 8102 in placebo group) Characteristics As reported under Manson et al. 2003 Inclusion criteria As reported under Manson et al. 2003 Exclusion criteria As reported under Manson et al. 2003	Interventions CEE+MPA	Details Setting: As reported under Manson et al. 2003 Methods: As reported under Manson et al. 2003 Statistical methods: For the current re-analysis: -First, an intention-to-treat analysis was conducted to confirm that the authors' results were similar to those previously published by WHI investigators; -Second, the analyses were adjusted for adherence to assigned therapy to estimate the CHD risk for continuous hormone use versus no use. The adjustments used inverse probability weighting (i.e., more weight was given to observation from women with low	Results Risk of CHD in relation to continuous use of CEE+MPA by years since menopause and follow-up time: HR (95%CI): By age at baseline: 50-59 yrs: Overall follow-up (8-year cumulative use): 1.47 (0.57-3.77) <=2 years: 2.69 (1.46-6.36) >=2 years (6-year cumulative use): 1.22 (0.59-2.56)	Limitations As reported under Manson et al. 2003 Other information -This re-analysis found no suggestion of a reduced risk of CHD during the first 2 years of CEE+MPA therapy in subgroups of women defined by years since menopause and baseline age. A CVD protective effect of CEE+MPA among women within 10 years of menopause was only apparent after approximately 6 years of use; -Randomised trial and observational data from the

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Annals of Internal Medicine, 152, 211-217, 2010 Ref Id 311752 Country/ies where the study was carried out US Study type Re-analysis of WHI CEE+MPA trial data by adjusting for adherence using inverse probability weighting method. Aim of the study To estimate the effect of continuous estrogen-plus-progestin therapy on CHD risk over time and stratified by years since menopause, i.e., to estimate an adherence-adjusted effect. Study dates WHI: 1993-1998-2004 The current re-analysis: 2010 Source of funding Not reported			estimated probabilities than those with high probabilities to take her assigned threatment based on her measured prognostic factors). This approach allowed the authors to appropriately accommodate the variations in adherence over time and the effect of prior treatment use on subsequent adherence. -A two-stage modeling procedure was used to estimate a woman's probability of taking her assigned treatment. The models included SES, lifestyle, dietary, and medical factors; the number of years since randomisation; and the proportion of study pills taken during the previous year. Then the weights were stabilized. -Finally a weighted pooled logistic model was fitted to estimate the average hazard ratio of CHD for continuous use versus no use of hormone therapy. The effect of continuous use versus no use can be thought of as an adherence-adjusted effect: the effect the researchers would have observed had the women been fully adherent to their assigned therapy.	By years since menopause: of those less than 10 years since menopause: Overall follow-up (8-year cumulative use): 0.64 (0.21-1.99) <=2 years: 1.29 (0.52-3.18) >=2 years (6-year cumulative use): 0.63 (0.27-1.52)	WHI have been previously combined, but the WHI observational data contributed few events during the first 2 years after initiation of hormone therapy. -Refer to Manson et al. 2003 (the original publication for WHI CEE+MPA findings) for analyses results by intention-to-treat (ITT) principle: n/N, adjusted HR (95%CI), By age at baseline and follow-up time: 50-59 yrs: overall follow-up: CEE+MPA: 37/2839 Placebo: 27/2683 HR: 1.20 (0.79-2.15) <= 2 years: CEE+MPA: 16/2839 Placebo: 10/2683 HR: 1.60 (0.73-3.55) >=2 years: CEE+MPA: 21/2839 Placebo: 17/2683 HR: 1.14 (0.60-2.16) By years since menopause at baseline and follow-up time: of those less than 10 years since menopause: Overall follow-up: CEE+MPA: 31/2782 Placebo: 34/2712 HR: 0.89 (0.55-1.46) <= 2 years: CEE+MPA: 14/2782 Placebo: 12/2712 HR: 1.17 (0.54-2.52) >=2 years: CEE+MPA: 17/2782 Placebo: 22/2712 HR: 0.74 (0.39-1.40)
Full citation	Sample size	Interventions	Details	Results	Limitations

<table>
<tr><th>Study details</th><th>Participants</th><th>Interventions</th><th>Methods</th><th>Outcomes and Results</th><th>Comments</th></tr>
<tr>
<td>Anderson,G.L., Limacher,M., Assaf,A.R., Bassford,T., Beresford,S.A., Black,H., Bonds,D., Brunner,R., Brzyski,R., Caan,B., Chlebowski,R., Curb,D., Gass,M., Hays,J., Heiss,G., Hendrix,S., Howard,B.V., Hsia,J., Hubbell,A., Jackson,R., Johnson,K.C., Judd,H., Kotchen,J.M., Kuller,L., Lacroix,A.Z., Lane,D., Langer,R.D., Lasser,N., Lewis,C.E., Manson,J., Margolis,K., Ockene,J., O'Sullivan,M.J., Phillips,L., Prentice,R.L., Ritenbaugh,C., Robbins,J., Rossouw,J.E., Sarto,G., Stefanick,M.L., Van,Horn L., Wactawski-Wende,J., Wallace,R., Wassertheil-Smoller,S., Women's Health</td>
<td>N= 10,739 (CEE, n=5310; Placebo, n=5429)
Characteristics
<table>
<tr><th></th><th>CEE (n=5310)</th><th>Placebo (n=5429)</th></tr>
<tr><td>Age at screening, mean (SD)</td><td>63.6 (7.3)</td><td>63.3 (7.3)</td></tr>
<tr><td>Age group at screening, y</td><td></td><td>0.85</td></tr>
<tr><td>50-59</td><td>1637 (30.8)</td><td>1673 (30.8)</td></tr>
<tr><td>60-69</td><td>2387 (45.0)</td><td>2465 (45.4)</td></tr>
<tr><td>70-79</td><td>1286 (24.2)</td><td>1291 (23.8)</td></tr>
<tr><td>Race/ethnicity</td><td></td><td>0.81</td></tr>
<tr><td>White</td><td>4007 (75.5)</td><td>4075 (75.1)</td></tr>
<tr><td>Black</td><td>782 (14.7)</td><td>835 (15.4)</td></tr>
<tr><td>Hispanic</td><td>322 (6.1)</td><td>333 (6.1)</td></tr>
<tr><td>American Indian</td><td>41 (0.8)</td><td>34 (0.6)</td></tr>
<tr><td>Asian/Pacific Islander</td><td>86 (1.6)</td><td>78 (1.4)</td></tr>
<tr><td>Unknown</td><td>72 (1.4)</td><td>74 (1.4)</td></tr>
<tr><td>Smoking</td><td></td><td>0.33</td></tr>
<tr><td>Never</td><td>2723 (51.9)</td><td>2705 (50.4)</td></tr>
<tr><td>Past</td><td>1986 (37.8)</td><td>2089 (38.9)</td></tr>
<tr><td>Current</td><td>542 (10.3)</td><td>571 (10.6)</td></tr>
<tr><td>Hormone use</td><td></td><td></td></tr>
<tr><td>Never</td><td>2769</td><td>2770</td></tr>
</table>
</td>
<td>Conjugated equine estrogen (CEE)</td>
<td>Consent
Informed written consent obtained from participants

Setting
Clinical trial, 40 clinical cnetre sites across the country

Randomisation method
The randomization procedure was developed at the WHI Clinical Coordinating Centre, using a randomized permuted block algorithm, stratified by clinical centre site and age group;

Concealment of allocation
All study medication bottles had a unique bottle number and bar code to allow for blinded dispensing

Comparability of intervention groups at baseline
The two groups were almost identical

Blinding
Considerable effort was made to maintain blinding of other participants and clinic staff. When required for safety or symptom management, an unblinding officer provided the clinic gynecologist, who was not involved with study outcomes activities, with the treatment assignment.

Statistical methods
-sample size calculation: the trial design assumed 12,375 women would need to be randomised to achieve 81% power to detect a 21% reduction in CHD rates oever the projected 9-year average follow-up;
-Primary analyses used time-to-event methods based on the intention-to-treat principle. Comparisons of primary outcomes are presented as hazard ratios and 95% CI from Cox proportional</td>
<td>Risk of CHD (including nonfatal myocardial infraction and death due to CHD) in relation to Estrogen vs. placebo,
n (no. of cases of CHD, annualized percentage), adjusted hazard ratio (HR, 95%CI)

By age:

50-59 yr:
CEE: 16 (0.14)
Placebo: 29 (0.24)
HR: 0.56 (0.30-1.03)

60-69yr:
E+P: 87 (0.54)
Placebo: 98 (0.59)
HR: 0.92 (0.69-1.23)

-adjusted for previous history of coronary-artery bypass grafting or percutaneous transluminal coronary angioplasty

Risk of stroke in relation to Estrogen vs. placebo (the data for this outcome is from Hendrix et al. 2006 where an additional 19 cases were inclued compared with the 2004 report)

n (no. of cases of stroke, annualized percentage), adjusted hazard ratio (HR, 95%CI):

By age:

50-59 yr:</td>
<td>NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials
A Selection bias
A1 - Was there appropriate randomisation - Yes
A2 - Was there adequate concealment - Yes
A3 - Were groups comparable at baseline - Yes
Level of bias: Low

B Performance bias
B1 - Did groups get same level of care - Yes
B2 - Were participants blinded to treatment allocation- Unclear (with an average follow-up of 6.8 yrs, women taking HRT should have realized which group they were allocated to when HRT taking effect when vaginal bleeding occured)
B3 - Were individuals administering care blinded to treatment allocation- Yes
Level of bias: High

C Attrition bias
C1 - Was follow-up equal for both groups - Yes
C2 - Were groups comparable for dropout - Yes (overall about 54% dropped out)
C3 - Were groups comparable for missing data - Yes
Level of bias: High

D Detection bias
D1 - Was follow-up appropriate length - Unclear (the trial was stopped at an average follow-up of 6.8 years, which was earlier than planned)</td>
</tr>
</table>

Study details	Participants			Interventions	Methods	Outcomes and Results	Comments
Initiative Steering Committee., Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial, JAMA, 291, 1701-1712, 2004 Ref Id 228873 Country/ies where the study was carried out US Study type RCT Aim of the study To assess the effects on major disease incidence rates of the most commonly used postmenopausal hormone therapy in the US. Study dates 1993-1998 recruitment Ended in Feb, 2004, the study was stopped earlier than planned; An average of 6.8 yrs follow-up; This 2004 paper presents the results of the		(52.2)	(51.1)		hazard analyses, stratified by age, prior disease, and adjusted for previous history of coronary-artery bypass grafting or percutaneous transluminal coronary angioplasty. Cumulative hazard rates were estimated by the Kaplan-Meier method for each designated outcome; -Two forms of CIs were calculated, nominal and adjusted. This report primarily presents the nominal 95% CIs because they provide traditional estimates of variability and, as such, are comparable to most other reports of hormone therapy studies. To acknowledge multiple testing issues, adjusted CIs were calculated using group sequential methods. Unless other indicated, all CIs and P values are nominal. -Intention to treat analysis (ITT) -Analyses were performed according to ITT principle -Outcomes ascertainment: - CHD was defined as acute MI requiring overnight hospitalization, silent MI determined from serial electrocardiograms, or CHD deaths; -Stroke: At each semiannual contact, a standardized interview asked participants about symptoms, safety, and potential outcome events. When a potential outcome was identified, medical records and death certificates were obtained as necessary. Physician adjudicators at clinical sites reviewed the information to determine the cause of the event. Of locally adjudicated stroke, 94.5% were confirmed by the central adjudicators. Stroke data were centrally confirmed by neurologists. Local and central adjudicators were blinded to treatment assignment. Follow-up	CEE: 16 (0.13) Placebo: 15 (0.12) HR: 1.09 (0.54-2.21) 60-69yr: E+P: 68 (0.41) Placebo: 41 (0.24) HR: 1.72 (1.17-2.54) -adjusted for previous history of coronary-artery bypass grafting or percutaneous transluminal coronary angioplasty. Risk of global index in relation to Estrogen vs. placebo, n (no. of cases, annualized percentage), adjusted hazard ratio (HR, 95%CI): By age 50-59 yr: CEE: 104 (0.89) Placebo: 132 (1.11) HR: 0.80 (0.62-1.03) 60-69yr: E+P: 312 (1.95) Placebo: 327 (1.97) HR: 0.98 (0.84-1.15)	D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - No (During the follow-up, gynaecologists of those women who had an onset of vaginal bleeding were unblinded of patients' allocation status) D5 - Were investigators blinded to confounding factors - Unclear Level of bias: High Indirectness Does the study match the review protocol in terms of Population: yes (women aged 50-59) Intervention: yes Outcomes: yes Indirectness: Some Other information -High rates of discontinuation of study medications and higher than expected crossover from placebo to active hormone use
	Past	1871 (35.2)	1948 (35.9)				
	Current	669 (12.6)	708 (13.0)				
	Duration of prior hormone use, y						
	<5 yr	1352 (53.2)	1412 (53.1)				
	5-10 yr	469 (18.5)	515 (19.4)				
	>= 10	720 (28.3)	732 (27.5)				
	Hypertension	2386 (48.0)	2387 (47.4)				
	Systolic BP, mean (SD), mm Hg	130.4 (17.5)	130.2 (17.6)				
	Diastolic BP, mean (SD), mm Hg	76.6 (9.2)	76.5 (9.4)				
	Pulse pressure	53.8 (15.3)	53.7 (15.0)				
	Treated for diabetes	410 (7.7)	411 (7.6)				
	History of CVD	477 (9.1)	469 (8.7)				
	History of MI	165 (3.1)	172 (3.2)				
	History of stroke	76 (1.4)	92 (1.7)				
	BMI, mean (SD), kg/m2	30.1 (6.1)	30.1 (6.2)				
	Inclusion criteria -Most women were recruited by population-b ased direct mailing						

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
estrogen alone trial using available data through Feb 29,2004, prior to notifying participants of the decision on March 1, 2004. Source of funding NIH	campaigns to age-eligible women, in conjunction with media awareness progrems -women aged 50-79 at initial screening, post menopausal, likelihood of residence in the area for 3 years, and provision of written informed consent; -a 3-month washout period was required before baseline evaluation of women using postmenopausal hormones at initial screening; -women with an intact uterus at initial screening were eligible for the trial of combined postmenopausal hormones, while women with a prior hysterectomy were eligible for the trial of unopposed estrogen. Exclusion criteria -Women who had medical conditions predictive of a survival time of less than 3 years; -Women were known to have conditions or characteristics inconsistent with study participation and adherence (alcoholism, drug dependency, mental illness, dementia); -Or if they were active participants in another RCT -Also, women were excluded from clinical trials for: reasons of competing risks (e.g., invasive cancer in the past 10 yrs; breast cancer at any time or suspicion of breast cancer at baseline screening; acute MI, stroke, or transient ischemic attack in the previous 6 months; reasons of safety (severe hypertension, or currently use of oral corticosteriods); and reasons relating to adherence or retention (unwillingness or inability to complete baseline study requirements). In addition, women were found to have femoral		-an average of 6.8 yrs; follow-up for clinical events occured every 6 months, with annual in-clinic visits required. -Lost to follow-up: over the average of 6.8 yrs of follow-up, only 563 (5.2%) were considered lost to follow-up. -Drop-out: at the study termination, 53.8% of women had already stopped taking study medication. Dropout rates exceeded design projections, particularly early on, but did not differ significantly by randomisation assignment and were stable after year 1, even with the termination of the estrogen plus progestin. 5.7% of women in CEE group and 9.1% in the placebo group dropped in treatment by follow-up year 6. Reasons for initiating HRT outside the study were not captured.	-adjusted for previous history of coronary-artery bypass grafting or percutaneous transluminal coronary angioplasty.	

<table>
<tr><th>Study details</th><th>Participants</th><th>Interventions</th><th>Methods</th><th>Outcomes and Results</th><th>Comments</th></tr>
<tr><td></td><td>neck bone mineral density of more than 3 standard deviations below the corresponding age-specific mean were also excluded.</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Full citation
Lacroix,A.Z., Chlebowski,R.T., Manson,J.E., Aragaki,A.K., Johnson,K.C., Martin,L., Margolis,K.L., Stefanick,M.L., Brzyski,R., Curb,J.D., Howard,B.V., Lewis,C.E., Wactawski-Wende,J., Investigators,W.H.I., Health outcomes after stopping conjugated equine estrogens among postmenopausal women with prior hysterectomy: a randomized controlled trial, JAMA, 305, 1305-1314, 2011
Ref Id
229707
Country/ies where the study was carried out
US
Study type
Re-analysis of WHI CEE trial data after a mean of 10.7 years of follow-up through August 2009</td><td>Sample size
Original WHI CEE trial: N=10739; Post termination follow-up: N= 7645 [after the protocol-specified termination date of March 31,2005, subsequent participants follow-up required additional written consent, which was obtained from 77.9% of surviving participants in the CEE group (n=3778) and 78.4% of surviving participants in the placebo group (n=3867)]
Characteristics
<table><tr><th></th><th>CEE (n=3778)</th><th>Placebo (n=3867)</th><th>P value</th></tr><tr><td>Age group at screening, y</td><td></td><td></td><td></td></tr><tr><td>50-59</td><td>1223 (32.4)</td><td>1232 (31.9)</td><td>0.88</td></tr><tr><td>60-69</td><td>1740 (46.1)</td><td>1799 (46.5)</td><td></td></tr><tr><td>70-79</td><td>815 (21.6)</td><td>836 (21.6)</td><td></td></tr><tr><td>Race/ethnicity</td><td></td><td></td><td></td></tr><tr><td>White</td><td>2945 (78.0)</td><td>3001 (77.6)</td><td>0.27</td></tr><tr><td>Black</td><td>514 (13.6)</td><td>565 (14.6)</td><td></td></tr><tr><td>Hisp</td><td>189</td><td>181</td><td></td></tr></table></td><td>Interventions
CEE</td><td>Details
Setting:
As reported under Anderson et al. 2004
Methods:
As reported under Anderson et al. 2004
Statistical methods:
-Power calculation: with the actual randomised sample size, the power estimate was 72% for a 21% reduction in CHD
-The primary analyses included all randomised participants using time-to-event methods and were based on the intention-to-treat principle as described previously.
-The hazard ratios (HRs) were estimated using Cox proportional hazard models stratified by age, prior disease, and randomisation status in the WHI Dietary Modification Trial. Models were constructed for each clinical end point in which women contributed follow-up time until end of the interval, the date of their first relevant event, or the date of death or withdrawal from the study.
-To determine whether not providing consent to postintervention follow-up influenced risk estimates, inverse-probability weighting analyses were conducted. Adherence sensitivity analyses also were conducted by censoring follow-up at 6 months after participants became nonadherent.
Follow-up time:
-By the intervention phase ended after a mean 7.1 years in Feb, 2004, vital status was known for 95% of participants, of whome 5.4% died. By this time, 54% of participants had stopped taking their study medication. Median time receiving treatment was 5.9 yrs in the CEE group vs. 5.8 yrs in</td><td>Results
Risk of cardiovascular diseases in postmenopausal women with prior hysterectomy who stopped taking CEE after a median 5.9 years of use: n. (%) of events, HR (95% CI):
CHD:
By age of participants at WHI trial baseline (median 5.9 years after CEE termination and a total follow-up of 10.7 (mean) follow-up since the WHI trial's baseline):
50-59 yrs:
CEE: 33 (0.18)
Placebo: 56 (0.31)
HR: 0.59 (0.38-0.90)

60-69 yrs: (just for information giving in the evidence table)
CEE: 161 (0.65)
Placebo: 168 (0.65)
HR: 1.00 (0.80-1.24)

(P value for interaction across age groups: 0.06)

Total MI:
50-60 yrs:
CEE: 27 (0.15)
Placebo: 50 (0.27)
HR: 0.54 (0.34-0.86)

60-69 yrs: (just for information giving in the evidence table)
CEE: 126 (0.51)</td><td>Limitations
NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies
A. Selection bias (systematic differences between the comparison groups)
A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-Unclear

A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders- Yes
A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Yes
Level of risk-Unclear

B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation)
B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a
B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a
B.3 Individuals administering</td></tr>
</table>

Study details	Participants				Interventions	Methods	Outcomes and Results	Comments
(follow-up data analysis) Aim of the study To examine health outcomes associated with randomisation to treatment with conjugated equine estrogen (CEE) among women with prior hysterectomy after a mean of 10.7 years of follow-up through August 2009. Three objectives: 1) To assess the long-term effects of CEE intervention on health outcomes; 2) to determine whether effects of CEE on health outcomes differed between the intervention and postintervention periods; and 3) to determine if previously identified suggestions of age-specific differences in effects of CEE on health outcomes persisted after stopping the intervention. Study dates WHI: 1993-1998-	anic	(5.0)	(4.7)			the placebo group. The median adherent time receiving treatment (taking 80% of study pills) was 3.5 years in both groups (IQR: 1.5-6.5 yrs) -The current report reflects the mean (SD) postintervention follow-up duration of 47.2 (20.7) months through August 2009.	Placebo: 124 (0.48) HR: 1.05 (0.82-1.35) (P value for interaction across age groups: 0.07) Stroke: 50-59 yrs: CEE: 29 (0.16) Placebo: 28 (0.15) HR: 1.09 (0.65-1.83) 60-69 yrs: (just for information giving in the evidence table) CEE: 114 (0.46) Placebo: 94 (0.36) HR: 1.27 (0.97-1.67) (P value for interaction across age groups: 0.91) Global index: CEE: 184 (1.04) Placebo: 217 (1.22) HR: 0.85 (0.70-1.03) 60-69 yrs: (just for information giving in the evidence table) CEE: 544 (2.29) Placebo: 559 (2.29) HR: 1.00 (0.89-1.13) (P value for interaction across age groups: 0.09) -The results were similar when using inverse-probability weighting to account for censoring due to those not providing consent for postintervention follow-up. The results were	care were kept 'blind' to treatment allocation-N/a Level of risk:N/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes (another median 5.9 yrs after the termination of the WHI CEE trial which lasted a mean of 7.1 yrs) C.2a How many participants did not complete treatment in each group?-N/a C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-Not reported C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up- Yes
	American Indian	31 (0.8)	18 (0.5)					
	Asian/Pacific Islander	54 (1.4)	49 (1.3)					
	Unknown	45 (1.2)	53 (1.4)					
	Hormone Therapy Use							
	Never	1929 (51.1)	1916 (49.6)	0.43				
	Past	1304 (34.5)	1373 (35.5)					
	Current	544 (14.4)	575 (14.9)					
	Duration of hormone therapy use, y							
	<5	960 (51.9)	1036 (53.1)	0.52				
	5-10	348 (18.8)	377 (19.3)					
	>10	541 (29.3)	538 (27.6)					
	BMI							
	<25	785 (20.9	771 (20.1	0.21				

Study details	Participants				Interventions	Methods	Outcomes and Results	Comments
2004 The current re-analysis: 2011 Source of funding WHI: NIH The current re-analysis: not reported		)	)				also similar when women were censored 6 months after becoming nonadherent to study medication during the intervention period.	D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-No D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-No Level of bias: Unclear Indirectness Does the study match the review protocol in terms of; Population: Yes Outcome: Yes Indirectness: Some Other information -Statistically significant age interactions for CEE use suggested greater safety and possible benefit among women in their 50s and potential harm among older women, were observed for CHD, total MI, and the global index of chronic diseases.
	25-<30	1289 (34.3)	1391 (36.2)					
	>=30	1687 (44.9)	1683 (43.8)					
	Smoking status							
	Never	1988 (53.1)	1972 (51.5)	0.30				
	Past	1417 (37.9)	1489 (38.9)					
	Current	336 (9.0)	370 (9.7)					
	Medical history							
	Treated diabetes	243 (6.4)	250 (6.5)	0.95				
	Self-reported high blood pressure	1806 (51.1)	1844 (51.2)	0.92				
	High cholesterol	490 (14.3)	536 (15.5)	0.16				
	Angina	243 (6.5)	253 (6.6)	0.82				
	CABG or PTCA	69 (1.9)	70 (1.8)	0.96				
	Stroke	51 (1.3)	47 (1.2)	0.60				
	DVT	65	60	0.56				

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	or PE (1.7) (1.6) Inclusion criteria As reported under Anderson et al. 2004 Exclusion criteria As reported under Anderson et al. 2004				
Full citation Prentice,R.L., Manson,J.E., Langer,R.D., Anderson,G.L., Pettinger,M., Jackson,R.D., Johnson,K.C., Kuller,L.H., Lane,D.S., Wactawski-Wende,J., Brzyski,R., Allison,M., Ockene,J., Sarto,G., Rossouw,J.E., Benefits and risks of postmenopausal hormone therapy when it is initiated soon after menopause, American Journal of Epidemiology, 170, 12-23, 2009 Ref Id 230128 Country/ies where the study was carried out US Study type RCT Aim of the study To analyse the effects of CEE	Sample size -From CEE trial: 9129 (4493 in CEE arm and 4636 in placebo arm) women with a known age at first menopause and a known age at first use of HRT among prior hormone therapy users. From the observational study, a corresponding subcohort of 20,117 women who had undergone hysterectomy prior to enrollment was also included, including 10,582 women were using the same CEE regimen as the women in CEE trial or were not using any hormone therapy (9,535) at the time of WHI enrollment. -From CEE/MPA trial, 7,679 (90.3%) assigned to active CEE/MPA and 7,509 (92.7%) women assigned to placebo in the CEE/MPA trial and to a subcohort of 30,942 women with an intact uterus at observational study enrollment, which included 6,756 women who were using the same CEE/MPA regimen studied in the CEE/MPA trial and 24,186 women who were not using any HRT at the time of enrollment. In total: 9129+20117+7697+7509+30942=75,394 Characteristics Distribution of subjects from both the clinical trials and observational studies, by prior use of HRT and gap time from menopause to first use of HRT among HRT users, 1993-2004	Interventions HRT (CEE, CEE/MPA)	Details -As reported under Anderson et al. 2004 and Manson et al. 2003 with regard to the RCT components; -In the observational cohort, clinical outcomes were also reported semiannually. Medical record documentation of self-reported outcomes was obtained and diagnoses were confirmed at WHI clinical centres. Statistical methods: -"Time from WHI enrollment was the "basic time variable" in Cox regression analyses that stratified data on cohort (clinical trials vs. observational study) and baseline age. -Confounding in the observational study was addressed by including standard risk factors for each outcome in Cox regression models. The set of risk factors to include was the same as previous reports for CVD and breast cancer and otherwise based on the knowledge and experience of the investigator group, prior to data analysis. They included age, BMI, education, smoking, physical functioning construct, history of treated diabetes, family history of cancer, cholesterol etc. -"Prior hormone therapy" use in the clinical trials and in non-hormone-therapy group in the observational study was defined relative to th time of WHI enrollment. -Prior use for hormone therapy users in the observational study was defined relative to the beginning of the hormone	Results Risk of CVD in relation to use of CEE, HR (95%CI): By time from menopause to first use of HT: CHD: < 5 years: No prior HT: N/a Prior HT: 1.22 (0.89-1.87) >5 years (just for information giving in evidence table): No prior HT: 0.89 (0.67-1.20) Prior HT: 1.04 (0.58-1.86) P for gap time interaction: 0.40 Stroke: < 5 years: No prior HT: N/a Prior HT: 1.36 (0.98-1.90) >5 years (just for information giving in evidence table): No prior HT: 1.64 (1.12-2.41) Prior HT: 0.56 (0.20-1.28) for gap time interaction: 0.96 Global index: < 5 years:	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-Yes (observational study subjects were those who were unwilling to or unsuitable to participate in the clinical trials of WHI, although all participants across studies were selected from the same population) A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes (confounders in the observational study were controlled for in analyses, as reported by the authors) A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Unclear Level of risk-High B. Performance bias (systematic differences

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
and CEE/MPA (particularly longer-term effects), when initiated soon after menopause, on a range of clinical outcomes, including the global index. The analyses used both WHI clinical trial data and combined WHI clinical trial and observational study data. Study dates 1993-1998 to 2004 Source of funding NIH	**Gap time, years** Use of CEE Clinical trials No prior HT / Prior HT <5 yr / 5-14 yr / >=15 No. women (%): 198 (10%) / 618 (32%) / 1136 (84%) No. of cases CHD: 2 / 22 / 59 Stroke: 3 / 19 / 46 Global index: 15 / 68 / 202 Observational study No prior HT / Prior HT <5 yr / 5-14 yr / >=15 No. women (%): 6626 (76%) / 1454 (17%) / 597 (7%) No. of cases CHD: 104 / 28 / 15 Stroke: 119 / 39 / 13 Global index: 689 / 164 / 75 Gap time, years		therapy episode that was ongoing at enrollment. Going back in time, a change in hormone regimen or usage gap of 1 year or longer defined a new hormone therapy episode. -Nominal 95% CIs are presented for hazard ratio parameters; Follow-up -As reported under Anderson et al. 2004 and Manson et al. 2003 with regard to the RCT components; -For the observational study, the cohorts were followed through Dec 15, 2004 (CEE) AND Feb 28, 2003 (CEE+MPA), an average follow-up periods of 7.1 yrs and 5.5 yrs, respectively.	No prior HT: 0.90 (0.53-1.53) Prior HT: 1.22 (1.04-1.43) >5 years (just for information giving in evidence table): No prior HT: 0.98 (0.83-1.16) Prior HT: 0.71 (0.50-1.00) P for gap time interaction: 0.05 Risk of CVD in relation to use of CEE/MPA, HR (95%CI): By time from menopause to first use of HT: CHD: < 5 years: No prior HT: 0.99 (0.49-1.98) Prior HT: 1.57 (0.99-2.50) >5 years (just for information giving in evidence table): No prior HT: 1.19 (0.91-1.57) Prior HT: 1.45 (0.69-3.06) P for gap time interaction: 0.42 Stroke: < 5 years: No prior HT: 0.92 (0.38-2.24) Prior HT: 1.20 (0.71-2.03)	between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk: n/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-No, slight differences across trials and observationl study with regard to early-stopped times) C.2a How many participants did not complete treatment in each group?- High drop-out in the clinical trials as reported previously under Anderson et al. 2004 and Manson et al. 2003; for the observational cohort, drop-out rate was not reported in the current analysis) C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-Unclear (reasons not investigated) C.3a For how many

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Use of CEE/ MPA Clinical trials No prior HT / Prior HT <5 yr / 5-14 yr / >=15 No. women (%): 952 (17%) / 2338 (43%) / 2160 (40%) No. of cases CHD: 10 / 35 / 71 Stroke: 6 / 37 / 53 Global index: 54 / 205 / 281 Observational study No prior HT / Prior HT <5 yr / 5-14 yr / >=15 No. women (%): 4257 (75%) / 1115 (20%) / 338 (6%) No. of cases CHD: 30 / 13 / 7 Stroke: 27 / 7 / 3 88: 340 / 88 / 41 Inclusion criteria -To enhance comparablility with the clinical trial eligibility criteria, women from the observational subcohort were required to be without a personal history of breast cancer and to have had a			>5 years (just for information giving in evidence table): No prior HT: 1.31 (0.96-1.79) Prior HT: 1.10 (0.46-2.68) P for gap time interaction: 1.00 Global index: < 5 years: No prior HT: 1.13 (0.84-1.53) Prior HT: 1.11 (0.90-1.37) >5 years (just for information giving in evidence table): No prior HT: 1.12 (0.99-1.28) Prior HT: 1.09 (0.77-1.55) P for gap time interaction: 0.93 Risk of CVD in relation to use of CEE and CEE/MPA (among women who began HRT immediately following menopause), from combined analysis of clinical trial and observational study data, HR (95%CI): (subjects the following analyses were limited to those who adhered to their hormone therapy regime from both the clinical trials and observational studies, because of the high drop-out rates in trials	participants in each group were no outcome data available?- As reported in Anderson et al. 2004 and Manson et al. 2003 with regard to clinical trials; for the observational study, data not reported) C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-Yes Level of risk: High D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Unclear (all subcohorts were stopped early due to ethical reasons) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-Yes D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-Unclear (details about the observational study not reported) Level of bias: Unclear Indirectness Does the study match the review protocol in terms of; Population: Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	mammogram within 2 years prior to enrollment. -To have a known age at first use of HRT use. Exclusion criteria -As reported under Anderson et al. 2004 and Manson et al. 2003 as the same in/exclusion criteria were used for clinical trials and observtional study at baseline in WHI (besides that the observational cohort was comprised of clinical trial screenees who were either ineligible or unwilling to participate in the clinical trial). -			and the data from the observational study was combined) By year from HT initiation among women with no prior use of HT: CHD: <2 years: CEE: 1.12 (0.55-2.24) CEE/MPA: 1.42 (0.76-2.65) 2-4 years: CEE: 0.99 (0.49-2.00) CEE/MPA: 1.37 (0.71-2.67) >=5 years (just for information giving in the evidence table) CEE: 0.60 (0.35-1.04) CEE/MPA: 1.24 (0.61-2.50) Stroke: <2 years: CEE: 1.49 (0.68-3.28) CEE/MPA: 1.58 (0.69-3.66) 2-4 years: CEE: 2.45 (1.06-5.65) CEE/MPA: 2.17 (0.99-4.80) >=5 years (just for information giving in the evidence table) CEE: 2.46 (1.29-4.70) CEE/MPA: 3.48 (1.38-8.96) Global index: <2 years: CEE: 1.26 (0.86-1.83) CEE/MPA: 1.53 (1.14-2.05) 2-4 years: CEE: 1.23 (0.87-1.75) CEE/MPA: 1.56 (1.18-2.06)	Outcome: Yes Indirectness: Some Other information -According to this study, the effects of CEE and CEE/MPA did not depend significantly on gap time from menopause to first use of HRT for most clinical outcomes considered, either in further analyses of clinical trial data or in combined clinical trail and observational study data analyses. -The interpretation of these hazard ratios by years from HT initiation among women with or without prior use of HT should be interpreted with caution: there is multiple testing isue. One would expect approximately 3 of the 95% confidence intervals to exclude 1 by chance alone. Another limitation of the current analyses was that hazard ratio pertaining to 5 or more years from HRT initiation were derived mainly from the observational study.

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				>=5 years (just for information giving in the evidence table) CEE: 1.18 (0.89-1.69) CEE/MPA: 1.89 (1.42-2.49) By year from "current" HT episode among women with prior use of HT: CHD: <2 years: CEE: 1.26 (0.64-2.46) CEE/MPA: 2.70 (1.11-6.52) 2-4 years: CEE: 1.52 (0.81-2.86) CEE/MPA: 1.10 (0.46-2.63) >=5 years: CEE: 0.86 (0.48-1.52) CEE/MPA: 2.18 (0.77-6.19) Stroke: <2 years: CEE: 1.43 (0.61-3.39) CEE/MPA: 1.73 (0.53-5.59) 2-4 years: CEE:1.56 (0.81-3.03) CEE/MPA: 1.05 (0.45-2.45) >=5 years: CEE: 2.39 (1.25-4.56) CEE/MPA: 1.48 (0.51-4.29) Global index: <2 years: CEE: 1.29 (0.90-1.85) CEE/MPA: 1.28 (0.86-1.91) 2-4 years: CEE: 1.03 (0.76-1.39) CEE/MPA: 1.32 (0.94-	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				1.85) >=5 years: CEE: 1.53 (1.15-2.03) CEE/MPA: 1.43 (0.96-2.11)	
Full citation Rossouw,J.E., Prentice,R.L., Manson,J.E., Wu,L., Barad,D., Barnabei,V.M., Ko,M., Lacroix,A.Z., Margolis,K.L., Stefanick,M.L., Postmenopausal hormone therapy and risk of cardiovascular disease by age and years since menopause.[Erratum appears in JAMA. 2008 Mar 26;299(12):1426], JAMA, 297, 1465-1477, 2007 Ref Id 230240 Country/ies where the study was carried out US Study type RCT Aim of the study To explore whether the effects of homrone therapy on risk of CVD vary by age or years since menopause began. Study dates 1993-1998 to	Sample size N= 10739+16608 (10739 who had undergone a hysterectomy and were randomised to CEE or placebo trial; 16608 womeh who had not had a hysterectomy and were randomised to CEE+MPA or placebo trial) Characteristics Baseline characteristics of participants in the CEE trial by age group and years since menopause (n=10739) **No. (%) of participants** Randomisation assignment: CEE (n=5310); Placebo (n=5429); Age at randomisation Years since menopause <10 yr: 826 (15.6); 817 (15.0) 10-19 yr: 1436 (27.0); 1500 (27.6) >=20 yr: 2231 (42.0); 2319 (42.7) Age group, yr 50-59	Interventions HRT: CEE; and CEE+MPA	Details Details Consent As reported under Anderson et al. 2004 and Manson et al. 2003; Setting As reported under Anderson et al. 2004 and Manson et al. 2003; Randomisation method As reported under Anderson et al. 2004 and Manson et al. 2003; Concealment of allocation As reported under Anderson et al. 2004 and Manson et al. 2003; Comparability of intervention groups at baseline As reported under Anderson et al. 2004 and Manson et al. 2003; Blinding As reported under Anderson et al. 2004 and Manson et al. 2003; Statistical methods -The results of unadjusted models for all women are presented because "preliminary analyses showed no striking differences in HRs across categories of age or years of since menopause in women with and without prior CVD, or in unadjusted models or models adjusted for baseline risk factors". -The primary analyses of this study were based on the 2 trials combined.	Results Combined trials: Risk of cardiovascular and global index in relation to HRT by age at baseline: n/N, HR (95%CI): CHD: 50-59 yr: HRT: 59/4476 Placebo: 61/4356 HR: 0.93 (0.65-1.33) 60-69 yr: HRT: 174/6240 Placebo: 178/6122 HR: 0.98 (0.79-1.21) Stroke: 50-59 yr: HRT: 44/4476 Placebo: 37/4356 HR: 1.13 (0.73-1.76) 60-69 yr: HRT: 156/6240 Placebo: 102/6122 HR: 1.50 (1.17-1.92) Global index: 50-59 yr: HRT: 278/4476 Placebo: 278/4356 HR: 0.96 (0.81-1.14) 60-69 yr: HRT: 771/6240 Placebo: 661/6122 HR: 1.08 (0.97-1.20) CEE Trial Risk of cardiovascular and global index in relation to HRT by age at baseline: n/N, HR	Limitations As reported under Anderson et al. 2004 and Manson et al. 2003; Other information -This analysis of the WHI data provides some convergence with information from observational studies, which have focused on minaly on the effects of estrogen on women without clinical CVD. However, differences remain. -There is a divergency in regard to secondary prevention, with observational study but not trial data on women with existing disease suggesting CHD benefit for HRT users; -The low or absent excess risk of CHD in women with less than 10 years since menopause may be somewhat reassuring to women considering the use of HRT in the first five years after menopause.

Study details	Participants				Interventions	Methods	Outcomes and Results	Comments
2004 (combined data analyses for CEE and CEE+MPA trials of WHI) Source of funding NIH	yr					Separate tests for trend were performed to examine differences in hormone effects across 3 preselected, coded categories of age (50-59, 60-69, 70-79 years) or years since menopause (<10, 10-19, and >=20)using Cox regression model interaction terms. Interaction terms between age or years since menopause and active vs placebo groups tested whether there were differential effects of hormone therapy as a function of age or years since menopause. These models allow the data for the 2 trials to be combined because they do not make assumptions about baseline risk or the overall treatment effect of hormone therapy in each of the trials. -Outcomes ascertainment: -As reported under Anderson et al. 2004 and Manson et al. 2003; -Due to the compressed timeline for the initial publications, 13 additional adjudicated cases each of CHD and stroke from the CEE+MPA trial were available in this analysis; Follow-up -As reported under Anderson et al. 2004 and Manson et al. 2003;	(95%CI): CHD: 50-59 yr: CEE: 21/1637 Placebo: 34/1673 HR: 0.63 (0.36-1.09) 60-69 yr: CEE: 96/2387 Placebo: 106/2465 HR: 0.94 (0.71-1.24) Stroke: 50-59 yr: CEE: 18/1637 Placebo: 21/1673 HR: 0.89 (0.47-1.69) 60-69 yr: CEE: 84/2387 Placebo: 54/2465 HR: 1.62 (1.15-2.27) Global index: 50-59 yr: CEE: 114/1637 Placebo: 140/1673 HR: 0.82 (0.64-1.05) 60-69 yr: CEE: 333/2387 Placebo: 342/2465 HR: 1.01 (0.86-1.17) CEE+MPA trial CHD: 50-59 yr: CEE+MPA: 38/2839 Placebo: 27/2683 HR: 1.29 (0.79-2.12) 60-69 yr: CEE+MPA: 78/3853 Placebo: 72/3657 HR: 1.03 (0.74-1.43) Stroke: 50-59 yr: CEE+MPA: 26/2839 Placebo: 16/2683	
	60-69 yr							
	70-79 yr							
	Vasomotor symptoms							
	None	2962 (55.8)	3004 (55.3)					
	Mild	1377 (25.9)	1442 (26.6)					
	Moderate or severe	913 (12.6)	917 (16.9)					
	Prior use of hormone therapy							
	Never	2769 (52.1)	2770 (51.0)					
	Past	1871 (35.2)	1948 (35.9)					
	Current	669 (12.6)	708 (13.0)					
	Duration of prior hormone therapy use, yr							
	< 5 yr	1352 (25.5)	1412 (26.0)					
	5-9 yr	469 (8.8)	515 (9.5)					
	>=10 yr	720 (13.6)	732 (38.9)					
	Baseline characteristics of							

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	participants in the CEE+MPA trial by age group and years since menopause (n=16608) **No. (%) of participants** Randomisation assignment: CEE+MPA (n=8506); Placebo (n=8102); Age at randomisation Years since menopause <10 yr: 2783 (32.7); 2712 (33.5) 10-19 yr: 3947 (35.8); 2994 (37.0) >=20 yr: 1850 (21.7); 1803 (22.3) Age group, yr 50-59 yr 60-69 yr 70-			HR: 1.41 (0.75-2.65) 60-69 yr: CEE+MPA: 72/3853 Placebo: 48/3657 HR: 1.37 (0.95-1.97) Global index: 50-59 yr: CEE+MPA: 164/2839 Placebo: 138/2683 HR: 1.10 (0.87-1.38) 60-69 yr: CEE+MPA: 384/3853 Placebo: 319/3657 HR: 1.15 (0.99-1.34) Combined trials: Risk of cardiovascular and global index in relation to HRT by year since menopause at baseline: n/N, HR (95%CI): CHD: < 10 yr: HRT: 39/3608 Placebo: 51/3529 HR: 0.76 (0.50-1.16) 10-19yr: HRT: 113/4483 Placebo: 103/4494 HR: 1.10 (0.84-1.45) Stroke: < 10 yr: HRT: 41/3608 Placebo: 23/3529 HR: 1.77 (1.05-2.98) 10-19yr: HRT: 100/4483 Placebo: 79/4494 HR: 1.23 (0.92-1.66)	

Study details	Participants				Interventions	Methods	Outcomes and Results	Comments
	79 yr						Global index: < 10 yr: HRT: 222/3608 Placebo: 203/3529 HR: 1.05 (0.86-1.27) 10-19yr: HRT: 482/4483 Placebo: 440/4494 HR: 1.12 (0.98-1.27) CEE trial Risk of cardiovascular and global index in relation to HRT by year since menopause at baseline: n/N, HR (95%CI): CHD: <10yr: CEE: 8/826 Placebo: 16/817 HR: 0.48 (0.20-1.17) 10-19yr: CEE: 47/1436 Placebo: 50/1500 HR: 0.96 (0.64-1.44) Stroke: <10yr: CEE: 17/826 Placebo: 8/817 HR: 2.24 (0.92-5.44) 10-19yr: CEE: 43/1436 Placebo: 30/1500 HR: 1.47 (0.92-2.35) Global index: <10yr: CEE: 60/826 Placebo: 62/817 HR: 0.94 (0.65-1.36) 10-19yr: CEE: 179/1436 Placebo: 177/1500 HR: 1.05 (0.85-1.29)	
	Vasomotor symptoms							
	None	5162 (60.7)	4928 (60.8)					
	Mild	2190 (25.8)	2115 (26.1)					
	Moderate or severe	1072 (12.6)	974 (12.0)					
	Prior use of hormone therapy							
	Never	6277 (73.8)	6020 (74.3)					
	Past	1671 (19.6)	1588 (19.6)					
	Current	554 (6.5)	491 (6.1)					
	Duration of prior hormone therapy use, yr							
	< 5 yr	1539 (18.1)	1470 (18.1)					

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	5-9 yr: 427 (5.0) 356 (4.4) >=10 yr: 263 (3.1) 255 (3.1) Inclusion criteria As reported under Anderson et al. 2004 and Manson et al. 2003; Exclusion criteria As reported under Anderson et al. 2004 and Manson et al. 2003;			CEE+MPA trial Risk of cardiovascular and global index in relation to HRT by year since menopause at baseline: n/N, HR (95%CI): CHD: <10 yr: CEE+MPA: 31/2782 Placebo: 35/2712 HR: 0.88 (0.54-1.43) 10-19yr: CEE+MPA: 66/3047 Placebo: 53/2994 HR: 1.23 (0.85-1.77) Stroke: <10 yr: CEE+MPA: 24/2782 Placebo: 15/2712 HR: 1.59 (0.81-3.05) 10-19yr: CEE+MPA: 57/3047 Placebo: 49/2994 HR: 1.12 (0.76-1.64) Global index: <10 yr: CEE+MPA: 162/2782 Placebo: 141/2712 HR: 1.09 (0.87-1.37) 10-19yr: CEE+MPA: 303/3047 Placebo: 263/2994 HR: 1.17 (0.99-1.38) Combined trials: Risk of cardiovascular and global index in relation to HRT by vasomotor symptoms at baseline: n/N, HR (95%CI): CHD: Women with moderate	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				to severe vasomotor symptoms at baseline: 50-59 yr: HRT: 17/1097 Placebo: 19/1030 HR: 0.86 (0.44-1.65) 60-69 yr: HRT: 31/691 Placebo: 25/665 HR: 1.20 (0.70-2.04) Stroke: 50-59 yr: HRT: 14/1097 Placebo: 11/1030 HR: 1.09 (0.49-2.43) 60-69 yr: HRT: 16/691 Placebo: 20/665 HR: 0.75 (0.39-1.45) Global index: 50-59 yr: HRT: 69/1097 Placebo: 66/1030 HR: 0.98 (0.70-1.38) 60-69 yr: HRT: 88/691 Placebo: 85/665 HR: 1.02 (0.75-1.37) Women with moderate to severe vasomotor symptoms at baseline: Years since menopause: CHD: <10 yr: HRT: 13/833 Placebo: 17/757 HR: 0.84 (0.40-1.77) 10-19yr: HRT: 17/557 Placebo: 13/555 HR: 1.38 (0.63-3.00) Stroke:	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				<10 yr: HRT: 10/833 Placebo: 3/757 HR: 3.36 (0.92-12.24) 10-19yr: HRT: 13/557 Placebo: 11/555 HR: 1.02 (0.44-2.37) Global index: <10 yr: HRT: 55/833 Placebo: 47/757 HR: 1.15 (0.77-1.71) 10-19yr: HRT: 59/557 Placebo: 47/555 HR: 1.23 (0.82-1.84)	
Full citation Manson,J.E., Chlebowski,R.T., Stefanick,M.L., Aragaki,A.K., Rossouw,J.E., Prentice,R.L., Anderson,G., Howard,B.V., Thomson,C.A., LaCroix,A.Z., Wactawski-Wende,J., Jackson,R.D., Limacher,M., Margolis,K.L., Wassertheil-Smoller,S., Beresford,S.A., Cauley,J.A., Eaton,C.B., Gass,M., Hsia,J., Johnson,K.C., Kooperberg,C., Kuller,L.H., Lewis,C.E., Liu,S., Martin,L.W.,	Sample size N= 27,347 (16608 in CEE+MPA trial; and 10739 in CEE trial) The post intervention follow-up through September 30, 2010 is based on 81.1% surviving participants who provided additional written informed consent. Following stopping of the intervention, fewer than 4% women reported personal use of hormone therapy. Characteristics -As reported under Manson et al. 2003 for CEE+MPA trial and Anderson et al. 2004 for CEE trial Inclusion criteria -As reported under Manson et al. 2003 for CEE+MPA trial and Anderson et al. 2004 for CEE trial Exclusion criteria -As reported under Manson et al. 2003 for CEE+MPA trial and Anderson et al. 2004 for CEE trial	Interventions CEE+MPA and CEE alone	Details Setting: 40 clinical centres across the US Methods: -As reported under Manson et al. 2003 for CEE+MPA trial and Anderson et al. 2004 for CEE trial -CHD was defined as nonfatal myocardial infarction (MI) or coronary death; Results for total MI, which was a secondary end point, are reported separately. Statistical methods: -For each trial, intervention phase analyses included all randomised participants according to their randomisation assignment until last intervention contact, using time-to-event method based on the intention-to-treat principle. -Hazard ratios (HRs) were estimated using Cox proportional hazards models stratified by age, prior disease (if appropriate), and randomisation status in the WHI dietary modification trial. Comparisons during the postintervention phase include randomised participants in active follow-	Results Risk of CHD in relation to HRT for the overall combined phases of WHI trial- CEE+MPA trial (13.2 years follow-up): n. (annulized %) of events; HR (95%CI): by age: 50-59 yrs: CEE+MPA: 93 (0.26) Placebo: 69 (0.21) HR: 1.27 (0.93-1.74) 60-69 yrs: (just for information giving in the evidence table) CEE+MPA: 201 (0.44) Placebo: 199 (0.46) HR: 0.97 (0.79-1.18) Stroke: 50-59 yrs: CEE+MPA: 52 (0.15) Placebo: 35 (0.10) HR: 1.37 (0.89-2.11)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-No (only about 81% surviving participants of WHI trials consented to extension pahse participation) A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Ockene,J.K., O'Sullivan,M.J., Powell,L.H., Simon,M.S., Van,Horn L., Vitolins,M.Z., Wallace,R.B., Menopausal hormone therapy and health outcomes during the intervention and extended poststopping phases of the Women's Health Initiative randomized trials, JAMA, 310, 1353-1368, 2013 Ref Id 294268 Country/ies where the study was carried out US Study type Re-analyses of WHI clinical trials during the intervention and extended poststopping phases Aim of the study To report a comprehensive, integrated overview of findings from the 2 WHI hormone therapy trials with extended postintervention follow-up (median, 13			up and at risk for an initial diagnosis of the relevant outcome. -All statistical tests are 2-sided and nominal P values of 0.05 or less are regarded as significant. The p values do not adjust for multiple outcomes, sequential monitoring, or multiple subgroup comparisons due to the large number of tests conducted; therefore, the p values should be be interpreted cautiously. Inference on subgroup analyses rely primarily on tests for interaction, which are also subject to multiple testing limitations when a large number of tests are conducted. -Adherence sensitivity analyses, conducted by censoring follow-up 6 months after nonadherence, included time-varying weights (inversely proportional to the estimated probability of continued adherence) in proportional hazards models that adjusted for changes in the distribution of sample characteristics during follow-up. Follow-up: -CEE+MPA intervention: the cumulative results reported in the current re-analyses include a median postintervention follow-up of 8.2 years and a median cumulative follow-up of 13.2 years; -CEE intervention: the median postintervention follow-up was 6.6 years and the median cumulative follow-up was 13.0 years;	60-69 yrs: (just for information giving in the evidence table) CEE+MPA: 168 (0.36) Placebo: 138 (0.32) HR: 1.16 (0.92-1.45) Global index: 50-59 yrs: CEE+MPA: 431 (1.27) Placebo: 377 (1.17) HR: 1.08 (0.94-1.24) 60-69 yrs: (just for information giving in the evidence table) CEE+MPA: 999 (2.33) Placebo: 906 (2.21) HR: 1.05 (0.96-1.15) Total MI: 50-59 yrs: CEE+MPA: 75 (0.21) Placebo: 57 (0.17) HR: 1.25 (0.88-1.76) 60-69 yrs: (just for information giving in the evidence table) CEE+MPA: 165 (0.36) Placebo: 158 (0.36) HR: 0.99 (0.80-1.24) Risk of CHD in relation to HRT for the overall combined phases of WHI trial- CEE trial (13 years follow-up): n. (%) of events; HR (95%CI): CHD by age: 50-59 yrs: CEE: 42 (0.21) Placebo: 64 (0.32) HR: 0.65 (0.44-0.96)	factors-No Level of risk- High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk: N/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-Not reported C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
years of cumulative follow-up) and stratifcation by age and other important variables. Study dates For WHI clinical trials: 1993-1998-2002 (CEE trial), 204 (CEE+MPA trial) For the current re-analyses: 2013 Source of funding For WHI trials: NIH For the current re-analyses: not reported				60-69yrs: (just for information giving in the evidence table) CEE: 183 (0.67) Placebo: 188 (0.67) HR: 1.00 (0.82-1.23) Stroke 50-59 yrs: CEE: 33 (0.16) Placebo: 36 (0.18) HR: 0.96 (0.60-1.55) 60-69yrs: (just for information giving in the evidence table) CEE: 134 (0.49) Placebo: 114 (0.40) HR: 1.25 (0.97-1.60) Global index: by age: 50-59 yrs: CEE: 214 (1.10) Placebo: 264 (1.36) HR: 0.82 (0.68-0.98) 60-69yrs: (just for information giving in the evidence table) CEE: 637 (2.47) Placebo: 637 (2.40) HR: 1.03 (0.92-1.15) Total MI: by age: 50-59 yrs: CEE: 35 (0.17) Placebo: 58 (0.29) HR: 0.60 (0.39-0.91) 60-69yrs: (just for information giving in the evidence table) CEE: 140 (0.52) Placebo: 139 (0.49) HR: 1.03 (0.82-1.31)	data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Unclear D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-No D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-No Level of bias: High Indirectness Does the study match the review protocol in terms of; Population: Yes Outcome: Yes Indirectness: Some Other information -Event information collected poststopping represents unblinded reporting and nearly 20% of surviving participants did not consent to extended follow-up. Multiple outcomes and subgroups (some with lower power) were examined, potentially leading to both false-positive and false-

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					negative results.
Full citation Schierbeck,L.L., Rejnmark,L., Tofteng,C.L., Stilgren,L., Eiken,P., Mosekilde,L., Kober,L., Jensen,J.E., Effect of hormone replacement therapy on cardiovascular events in recently postmenopausal women: randomised trial, BMJ, 345, e6409-, 2012 Ref Id 230314 Country/ies where the study was carried out Denmark Study type open label, RCT Aim of the study To investigate the long term effect of hormone replacement therapy on cardiovascular outcomes in recently postmenopausal women. Study dates 1990-1993 to 2008 (Intervention was stopped after	Sample size N=1006 (502 allocated to HRT and 504 received no treatment) Characteristics **HRT group** / **Control group** Age (yrs): 50.0 (2.8) / 49.5 (2.7) BMI (kg/m2): 25.2 (4.50 / 25.3 (4.3) Total cholesterol concentration (mmol/L): 6.32 (0.98) / 6.28 (1.10) Systolic blood pressure (mm Hg): 130 (20) / 129 (18) Diastolic blood pressure (mm Hg): 81 (11) / 81 (11) Time since menopause (years): 0.61 (0.65) / 0.58 (0.63) No (%) of smokers: 255 (44.6) / 212 (42.3) Only age was significantly different between the two groups, p=0.007 Inclusion criteria -Healthy, recently postmenopausal white women aged 45-58, with last	Interventions HRT: (estrogen alone or combination therapy, namely triphasic estradiol and norethisterone acetate for women with an intact uterus; women who had undergone hysterectomy received estradiol)	Details Setting Denmark, multicentre trial Methods: -Open label trial -HRT exposure: -All participants enrolled underwent a physical examinaton and biochemical screening at baseline. They were subsequently seen after 6 months, one year, and two, three, five, and 10 years. The study drug were posted to the women randomised to HRT and they were offered an annual visit. -Outcomes ascertainment: -The study was planned for 20 years but stopped at 10 years. After that participants in the randomized HRT arm were followed up for another 6 years in national registers, which provided data on all hospital contacts or death (no participants were lost to follow up in these 6 yrs, with only 2 women emigrated. In the randomised treatment, at 5 yrs, 75% of the women adhered to the randomisation arm to which they were allocated for 80% or more of the time). -Evaluations of endpoints in the 10 year randomised trial were carried out using a PROBE (prospectively, randomised, open with blinded endpoint evaluation) design; -The extra 6 year follow-up data was retrieved on all participants from the Danish civil registration system and the national hospital discharge register. Statistical methods: -All analyses were done on the intention to treat population; -The analyses were carried out, with August 1,2002 as the stopping date, about 10 years after randomisation (when the randomised treatment was stopped). Secondary analyses with an	Results Results at the 10-year randomised treatment follow-up: Risk of mortality, heart failure, or myocardial infraction (composite): adjusted hazard ratio (95%CI) 0.48 (0.26-0.87) by age: age >=50 (50-58) yr: 0.63 (0.29-1.36) age < 50 (45-49) yr: 0.35 (0.13-0.89) Risk of stroke: adjusted hazard ratio (95%CI): among women aged 45-58 years: 0.77 (0.35-1.70) Risk of breast cancer: adjusted hazard ratio (95%CI): 0.58 (0.27-1.27) By age: age >=50: 0.98 (0.33-2.92) age < 50: 0.34 (0.11-1.08) -adjusted for age Results at the 16-year total follow-up: (the use of HRT during this non-randomised follow-up time was uncertain) Risk of mortality, heart failure, or myocardial infraction (composite): adjusted hazard ratio (95%CI) 0.61 (0.39-0.94) By age:	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-Yes A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Yes (mostly besides age) Level of risk-Low B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-Unclear B.2 Participants receiving care were kept 'blind' to treatment allocation-No (open-label trial) B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk: High

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
about 11 yrs owing to adverse reports from other trials, but participants were followed to death, CVD, and cancer for up to 16 yrs) Source of funding Novo Nordisk, Novartis, and Leo Pharma Denmark provided the study drug free of charge	menstrual bleeding 3-24 months before study entry or perimenopausal symptoms (including irregular menstruations) in combination with recorded postmenopausal serum follicle simulating hormone values. -Women who had had hysterectomy if they were aged 45-52 and had records showing an increase in serium follicle simulating hormone levels. Exclusion criteria -A history of bone disease (including non-traumatic vertebral fractures on radiography), uncontrolled chronic disease, previous or current cancer or thromboembolic disease, current or past treatment with glucocorticoids for more than 6 months, current or previous use of hormone replacement therapy within the past 3 months, and alcohol or drug dependency.		additional 6 years of non-randomised follow-up were also conducted. -Chi-square test for dichotomous variables and continous variables with students t test; -Hazard ratios (95% CI) were determined using Cox proportional hazards regression analyses, adjusting for age.	age>= 50 (50-58) years:: 0.68 (0.38-1.21) age< 50 (45-49) years: 0.55 (0.29-1.05) Risk of stroke: adjusted hazard ratio (95%CI): Among women aged 45-58 years: 0.89 (0.48-1.65) Risk of breast cancer: adjusted hazard ratio (95%CI): 0.90 (0.52-1.57) By age: age >=50: 1.58 (0.73-3.44) age < 50: 0.50 (0.22-1.14) -adjusted for age	C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-None C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-Yes C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					'blind' to participants' exposure to the intervention-No D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-No Level of bias: Unclear Indirectness Does the study match the review protocol in terms of; Population: Yes Outcome: Yes Indirectness: Some Other information Breat cancer data available -Using a population based approach, recruiting participants by direct mail to a random sample of Danish women in the perimenopausal to postmenopausal age range, the study participants were as representative as possible for a randomised trial. -The additional 6 years of follow-up after discontinuation of the randomised treatment was difficult to interpret; it was uncertain whether women continued treatment after information of the results of the WHI in 2002.
Full citation Stampfer,M.J., Willett,W.C., Colditz,G.A., Rosner,B., Speizer,F.E., Hennekens,C.H., A prospective study of postmenopausal estrogen therapy and coronary	Sample size N=121,964 Characteristics **Variable** / **Estrogen use**: Never / Ever / Current Percentage	Interventions Conjugated estrogen (the 1976 questionnaire did not include the type of dose of hormone. On the 1978 questionnarie, about 74% of the users reported using conjugated estrogens (premarin in most cases), nearly all of which were unopposed progestins)	Details Setting: Survey study among female registered nurses in the US Methods: -In 1976, questionnaires covering questions on a variety of health conditions, including prior CHD, menopause, parental history of myocardial infraction, height and weight, current and past smoking, and use of postmenopausal hormones were sent	Results Non fatal myocardial infraction: -65 cases of nonfatal myocardial and 25 confirmed coronary deaths during 105,786 person-years of follow-up among those without a prior coronary disease. Total coronary disease	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant

Study details	Participants				Interventions	Methods	Outcomes and Results	Comments
heart disease, New England Journal of Medicine, 313, 1044-1049, 1985 Ref Id 202650 Country/ies where the study was carried out US Study type Prospective follow-up study Aim of the study To examine the effect of hormones on the risk of nonfatal myocardial infraction and fatal coronary disease in a large prospective cohort of postmenopausal women. Study dates 1976-1980 Source of funding NIH		of subjects				out; -In 1978 and 1980, follow-up quesstionnaries that updated the information on most of these variables and inquired about the development of new illnesses, including myocardial infraction. -Measurement of HRT exposure: In 1976 the subjects were asked whether they had used postmenopausal hormones after menopause, if so, how long. -Current HRT users: women were considered current users if the duration of use was equal (within 12 months) to the interval between menopause and the time the questionnaire was completed; -Past HRT users: women whose duration of use was less than interval between menopause and the return of the questionnaire (by more than 12 months) were considered past users. -Information on hormone use was updated in 1978 with explicit questions about current use and the duration of use between 1976 and 1978. -Measurement of CHD outcome: -nonfatal myocardial infraction and fatal coronary heart disease. Nurses reporting nonfatal myocardial infarction on the 1978 and 1980 questionnaires were asked to grant permission for a review of their medical records and was verified in the medical record. -Myocardial infarctions that required hospitalisation and were corroborated by additional confirmatory information but for which the records could not be obtained were designated as probable. -a death was considered to be due to coronary disease if a fatal myocardial infarction was confirmed by hospital records or autopsy. Coronary death also included cases in which coronary disease was listed as underlying cause, without another plausible cause, on the	(including non fatal myocardial infarction plus fatal coronary disease) in relation to HRT use: adjusted relative risk* (RR, 95%CI) By user type: Non users: 1.00 (reference group) Current users: 0.30 (0.14-0.64) Past users: 0.59 (0.33-1.66) * -adjusted for risk factors listed in the baseline characteristics table Nonfatal infraction only: adjusted relative risk* in relation to HRT use: (RR, 95%CI): by user type: Non users: 1.00 (reference group) Current users: 0.34 (0.14-0.82) Past users: 0.65 (0.33-1.28) * -adjusted for risk factors listed in the baseline characteristics table Risk of total CHD in relation to ever and current HRT users compared with nonusers: n(caess)/person years; adjusted RR* (95%CI): be user type and age: 30-34 yrs: Never: 0/228.3; 1.00 (Reference group)	allocation to treatment groups is not expected to affect the outcome(s) under study)-No (participants were registered nurses) A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-No, more leaner women in estrogen use group Level of risk- High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk: N/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants
	Maternal history of myocardial infraction (MI)	11.3	1.4	10.9				
	Paternal history of MI	23.0	24.4	24.6				
	Smoking status							
	Never	41.2	39.1	40.8				
	Former	20.2	23.6	24.2				
	Current	38.2	36.9	34.5				
	Hypertension	17.8	18.6	18.1				
	High serum cholesterol	4.9	6.6	6.2				
	Diabetes	2.9	2.4	2.1				
	Bilateral oophorectom	12.4	53.6	60.3				

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	y Quetelet's index (kg/m2) <+21.2 19.8 23.0 24.0 21.3-24.6 37.5 42.2 43.3 24.6 41.6 33.9 31.8 Inclusion criteria -Female, married, registered nurses aged 30-55 who were living in 1 of 11 large US states. Exclusion criteria -Since women with a diagnosis of coronary disease may alter their pattern of hormone use and are also at increased risk for progression of the disease, their inclusion could have distorted the results. Therefore, nurses who reported either myocardial infraction or angina on the 1976 questionnaire were excluded. Similarly, women with such reports on the 1978 questionnaire were excluded from follow-up after 1978, so that the base population for each period was always free of reported coronary disease at the start of the period.		death certificate. Statistical methods: -age-specific rates of HRT and non-HRT users were individually calculated, and aged-adjusted relative risks were calculated over five-year age strata. -to adjust for multiple potential risk factors simultaneously, proportional-hazards models were developed for total coronary disease (including nonfatal myocardial infraction and fatal heart disease) and for nonfatal infraction alone. Proportional-hazards models were not used for fatal coronary disease alone because of the relatively small number of cases.	Ever: 0/789.5; RR: n/a Current: 0/644.4; RR: n/a 35-39 yrs: Never: 0/663.1; RR: 1.00 (reference group) Ever: 0/2170; RR: n/a Current: 0/1593.9; RR: n/a 40-44 yrs: Never: 1/2073.3; RR: 1.00 (reference group) Ever: 2/5401.9; RR: 0.8 (0.1-4.6) Current: 1/3833.0; RR: 0.6 (0.2-2.4) 45-49 yrs: Never: 11/9106.9; RR: 1.00 (reference group) Ever: 3/11,064.3; RR: 0.2 (0.1-0.7) Current: 2/6,890.1; RR: 0.2 (0.1-0.9) 50-55 yrs: Never: 40/34197.6; RR: 1.00 (reference group) Ever: 323/30,045.8; RR: 0.6 (0.4-1.1) Current: 8/15,239.2; RR: 0.4 (0.2-0.9) 56-59 yrs: Never: 8/5238.7; RR: 1.00 (reference group) Ever: 2/4837.2; RR: 0.3 (0.1-1.1) Current: 0/1721.4; RR: 0 Overall age-adjusted RR: Never: 60/51,477.5; RR: 1.00 (reference group) Ever: 30/54,308.7; RR:	did not complete treatment in each group?-N/a C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/a C.3a For how many participants in each group were no outcome data available?-N/a C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)- Yes Level of risk: N/a D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Unclear (just 4-yrs follow-up data in this study) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/a D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/a Level of bias: Unclear Indirectness

<table>
<tr><th>Study details</th><th>Participants</th><th>Interventions</th><th>Methods</th><th>Outcomes and Results</th><th>Comments</th></tr>
<tr><td></td><td></td><td></td><td></td><td>0.5 (0.3-0.8)
Current: 11/29,922.0; RR: 0.3 (0.2-0.6)

*-other risk factors adjusted for or not not clearly reported in the study.</td><td>Does the study match the review protocol in terms of:
Population: Some (only registered nurses)

Outcome: Yes
Indirectness: Some</td></tr>
<tr><td>Full citation
Grodstein,F., Stampfer,M.J., Manson,J.E., Colditz,G.A., Willett,W.C., Rosner,B., Speizer,F.E., Hennekens,C.H., Postmenopausal estrogen and progestin use and the risk of cardiovascular disease.[Erratum appears in N Engl J Med 1996 Oct 31;335(18):1406], New England Journal of Medicine, 335, 453-461, 1996
Ref Id
229374
Country/ies where the study was carried out
US
Study type
Propective follow-up study (The Nurses' Health Study)
Aim of the study
To examine the relation betwee ncardiovascular disease and</td><td>Sample size
N=59,337 (in 1976, a total of 21,726 postmenopausal women were included in the analysis, and 37,611 women were added during follow-up as they became postmenopausal; 662,891 person-years of follow-up were accrued from 1976 to 1992.
Characteristics
<table><tr><th>Characteristics</th><th>Hormone use</th><th></th><th></th></tr><tr><td></td><td>Never users (n=27,034)</td><td>Past users (n=12,503)</td><td>Current users</td></tr><tr><td></td><td></td><td></td><td>Estrogen alone (n=7776)</td></tr><tr><td>Parental MI before age 60 (%)</td><td>29.6</td><td>26.7</td><td>21.8</td></tr><tr><td>Hypertension</td><td>32.9</td><td>35.9</td><td>35.6</td></tr></table></td><td>Interventions
Combined hormone therapy (estrogen + progestin)</td><td>Details
Setting:
As reported under Stampfer et al. 1985
Methods:
As reported under Stampfer et al. 1985
Statistical methods;
As reported under Stampfer et al. 1985 -for the current analyses, proportional-hazards models were used to calculate relative risks, with adjustments for age, age at menopause, BMI, smoking, hypertension, diabetes, elevated cholesterol levels, myocardial infraction in a parent before the age of 60 years, prior use of oral contraceptives, type of menopause, and two-year interval

Follow-up:
16 years with 662,891 person-years of follow-up (information was missing for 3.2% of the follow-up time)</td><td>Results
Risk of coronary heart disease (nonfatal myocardial infarction and death due to coronary diseaes) among current users compared with non-users: n (no. of cases)/person years; adjusted RR* (95% CI): (based on data from 1978-1992)
By HRT preparation:
Never users: 431/304,744; RR: 1.00 (reference group)
Current estrogen users: 47/82,626; RR:0.60 (0.43-0.83)
Current estrogen with progestin users: 8/27,161; RR: 0.39 (0.19-0.78)

*-- RR adjusted for age, age at menopause, BMI, smoking, hypertension, diabetes, elevated cholesterol levels, myocardial infraction in a parent before the age of 60 years, prior use of oral contraceptives, type of menopause, and two-year interval

Risk of stroke among current users compared</td><td>Limitations
As reported under Stampfer et al. 1985; up to 1992 information was missing for 3.2% of the follow-up time.
Other information</td></tr>
</table>

Study details	Participants				Interventions	Methods	Outcomes and Results	Comments
postmenopausal hormone therapy (combined therapy: esterogen plus progestin) during up to 16 years of follow-up in 59,337 women from the Nurses' Health Study, who were 30 to 55 years of age at base line. Study dates 1976-1992 (Information on hormone use was ascertained with biennial questionnaries. From 1976-1992, 770 cases of MI or death from coronary disease in this group and 572 storkes were documented. Source of funding NIH	(%)						with non-users: n (no. of cases)/person years; adjusted RR (95% CI): By HRT preparation: Never users: 270/304,744; RR: 1.00 (reference group) Current estrogen users: 74/82,626; RR: 1.27 (0.95-1.69) Current estrogen with progestin users: 17/27,161; RR: 1.09 (0.66-1.80) --*-- RR adjusted for age, age at menopause, BMI, smoking, hypertension, diabetes, elevated cholesterol levels, myocardial infraction in a parent before the age of 60 years, prior use of oral contraceptives, type of menopause, and two-year interval Risk of coronary heart disease (nonfatal myocardial infarction and death due to coronary diseaes) among current users compared with non-users: n (no. of cases)/person years; adjusted RR* (95% CI): (based on data from 1976-1992) By user type: Never users: 452/324,748; RR: 1.00 (reference group) Current users:	
	Diabetes (%)	5.8	5.6	3.8				
	High serum cholesterol	35.6	41.9	43.9				
	Moderate smoker	9.4	8.9	5.5				
	Bilateral oophorectomy (%)	4.2	27.6	47.9				
	Past use of oral contraceptives (%)	30.6	37.9	42.0				
	Mean age (yr)	60.1	61.6	58.5				
	Mean age at menopause (yr)	50.9	46.3	44.7				
	Mean BMI	26.3	25.9	25.1				
	Mean alcohol consumption (g/day)	4.7	5.5	6.4				

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Mean consumption of saturated fat (g/day) 31.2 34.4 41.9 Inclusion criteria As reported under Stampfer et al. 1985 Exclusion criteria As reported under Stampfer et al. 1985			98/166,371; RR: 0.60 (0.47-0.76) past users: 195/150,238; RR: 0.85 (0.71-1.01) *-- RR adjusted for age, age at menopause, BMI, smoking, hypertension, diabetes, elevated cholesterol levels, myocardial infraction in a parent before the age of 60 years, prior use of oral contraceptives, type of menopause, and two-year interval Risk of stroke among current users compared with non-users: n (no. of cases)/person years; adjusted RR* (95% CI): (based on data from 1976-1992) By user type: Never users: 279/324,748; RR: 1.00 (reference group) Current users: 121/166,371; RR: 1.03 (0.82-1.31) past users: 152/150,238; RR: 0.99 (0.80-1.22) *-- RR adjusted for age, age at menopause, BMI, smoking, hypertension, diabetes, elevated cholesterol levels, myocardial infraction in a parent before the age of 60 years, prior use of oral contraceptives, type of menopause, and two-year interval	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				Risk of ischemic stroke among current users compared with non-users: n (no. of cases)/person years; adjusted RR* (95% CI): (based on data from 1976-1992) By user type: Never users: 133/324,748; RR: 1.00 (reference group) Current users: 73/166,371; RR: 1.40 (1.02-1.92) past users: 75/150,238; RR: 1.01 (0.74-1.36) *-- RR adjusted for age, age at menopause, BMI, smoking, hypertension, diabetes, elevated cholesterol levels, myocardial infraction in a parent before the age of 60 years, prior use of oral contraceptives, type of menopause, and two-year interval Risk of subarachnoid stroke among current users compared with non-users: n (no. of cases)/person years; adjusted RR* (95% CI): (based on data from 1976-1992) By user type: Never users: 79/324,748; RR: 1.00 (reference group) Current users: 33/166,371; RR: 0.90 (0.57-1.41)	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				past users: 32/150,238; RR: 0.81 (0.52-1.25) *-- RR adjusted for age, age at menopause, BMI, smoking, hypertension, diabetes, elevated cholesterol levels, myocardial infraction in a parent before the age of 60 years, prior use of oral contraceptives, type of menopause, and two-year interval Risk of coronary heart disease (nonfatal myocardial infarction and death due to coronary diseaes) among current users compared with non-users: n (no. of cases)/person years; adjusted RR* (95% CI): By user type: By age: (exact follow-up time not reported for this outcome) <50 yr: Never users: 22/29,881; RR: 1.00 (reference group) Current users: 4/35,379; RR: 0.18 (0.05-1060) 50-59 yr: Never users: 272/213,636; RR: 1.0 (Reference group) Current users: 61/92,922; RR: 0.71 (0.52-0.96)	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				60-71yr: (just for information giving in evidence table) Never users: 158/81,231; RR: 1.0 (Reference group) Current users: 33/38,070; RR: 0.66 (0.44-1.01) *-- RR adjusted for age, age at menopause, BMI, smoking, hypertension, diabetes, elevated cholesterol levels, myocardial infraction in a parent before the age of 60 years, prior use of oral contraceptives, type of menopause, and two-year interval Risk of Cardiovascular death in relation to HRT use, n (no. of cases), adjusted RR (95%CI): (based on 1976 to 1994 data) By user type: Death due to coronary heart desease: Never users: 289; RR: 1.00 (Reference group) Current users: 43; RR: 0.47 (0.32-0.69) Past users: 129; RR: 0.99 (0.75-1.30) Death due to stroke: Never users: 91; RR: 1.00 (Reference group) Current users: 28; RR: 0.68 (0.39-1.16) Past users: 48; RR: 1.07 (0.68-1.69)	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Full citation Grodstein,F., Manson,J.E., Colditz,G.A., Willett,W.C., Speizer,F.E., Stampfer,M.J., A prospective, observational study of postmenopausal hormone therapy and primary prevention of cardiovascular disease, Annals of Internal Medicine, 133, 933-941, 2000 Ref Id 229378 Country/ies where the study was carried out US Study type Prospective follow-up (The Nurses' Health Study; 20-yr follow-up report) Aim of the study To investigate duration, dose, and type of postmenopausal homrone therapy and primary prevention of cardiovascular disease. Study dates 1976-1996 (20-yr follow-up) Source of funding NIH	Sample size N= 70, 533 Characteristics Age in years: 30-55 (other characteristics not reported in this publication) Inclusion criteria -Female nurses aged 30-55 yrs of age Exclusion criteria -Women who reported stroke, , myocardial infarction, angina, coronary revascularization, or cancer on the 1976 questionnaire were excluded	Interventions HRT- analyses were limited to users of oral conjugated estrogen with or without oral medroxyprogesterone acetate (the most common hormone regimens)	Details Setting: questionnaire survey among registered nurses in 1976, and biennial follow-up Methods: Ascertainment of HRT: -Self-reported use and duration of HRT after menopause; beginning in 1978, information on type of HRT was collected; all information was updated biennially; Ascertainment of CVDs: -self-reported first occurrence of CVDs between the return of 1976 questionnaire and 1996. Permission to review of medical records of the reported cases was obtained throughout the study; Statistical analysis: -for a total of 70533 participants, 808, 825 per-years of follow-up were accrued from 1976-1996; -Analyses of type of HRT were limited to users of oral conjugated estrogen with or without oral medroxyprogesterone acetate (the most common hormone regimens) -Pooled logistic regression across the ten 2-yr time periods to adjust simultaneously for potential confounding factors; Simulation studies have established the asymptotic equivalence of pooled logistic regression to Cox regression with time-dependent covariates. The necessary conditions for this equivalence include relatively short time intervals and small probability of the outcome during each interval, both of which were satisfied. Follow-up: 20-yr	Results Major coronary heart disease: n/person-years, adjusted RR (95%CI), by HRT use type and duration of current users: Never users: 662/358,125; RR:1.0 (reference) Past users: 337/185,497; RR: 0.82 (0.72-0.94) Current users: 259/265,203; RR: 0.61 (0.52-0.71) <1yr: 9/20,091; RR: 0.40 (0.21-0.77) 1-1.9 yr: 9/19,155; RR: 0.41 (0.21-0.80) 2-4.9 yr: 60/78,928; RR: 0.53 (0.41-0.70) 5-9.9 yr: 74/77,435; RR: 0.58 (0.45-0.74) >=10 yr: 107/69,594; RR: 0.74 (0.59-0.91) -Confounders adjusted for: age, BMI, history of diaberes, hypertension, high cholesterol level, age at menopause, smoking, and parental history of premature heart disease; -Duration of use was underestimated by an average of 1 yr, since duration during each 2-yr follow-up period was established at the start of each period; All stroke: n/person-years, adjusted RR (95%CI), by HRT use type and duration of current	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-No A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Unclear Level of risk-High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk: N/a C. Attrition bias (systematic

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				users: Never: 312/358,125; RR: 1 (reference group) Past: 217/185,497 RR: 1.02 (0.85-1.24) Current: 238/265,203; RR: 1.13 (0.94-1.35) <1 yr: 13/20,091; RR: 1.32 (0.76-2.32) 1-1.9 yr: 10/19,155; RR: 1.04 (0.55-1.97) 2-4.9 yr: 61/78,928; RR: 1.14 (0.86-1.52) 5-9.9 yr: 63/77,435; RR: 1.05 (0.79-1.38) >=10 yr: 91/65,594; RR: 1.17 (0.91-1.49) Ischemic stroke: n/person-years, adjusted RR (95%CI), by HRT use type and duration of current users: Never: 170/358,125; RR: 1 (reference group) Past: 120/185,497; RR: 1.01 (0.79-1.30) Current: 142/265,203; RR: 1.26 (1.00-1.61) <1yr: 6/20,091; RR: 1.07 (0.44-2.61) 1-1.9yr: 6/19,155; RR: 1.32 (0.58-3.00) 2-4.9yr: 36/78,928; RR: 1.31 (0.90-1.92) 5-9.9yr: 42/77,435; RR: 1.36 (0.96-1.92) >=10yr: 52/69,594; RR: 1.17 (0.84-1.63) Hemorrhagic stroke: n/person-years, adjusted RR (95%CI), by HRT use type and duration of current users:	differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-Not reported C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-Not reported C.3a For how many participants in each group were no outcome data available?- not reported (for the whole cohort about 10% dopped out) C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)- yes Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up- Yes (20 yrs) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				Never: 79/358,125; RR: 1 (reference group) Past users: 45/185,497; RR: 0.95 (0.65-1.40) Current: 50/265,203; RR: 0.93 (0.64-1.34) < 1 yr: 5/20,091; RR: 1.56 (0.63-3.90) 1-1.9 yr: 2/19,155; RR: 0.63 (0.15-2.59) 2-4.9yr: 14/78,928; RR: 0.95 (0.54-1.67) 5-9.9yr: 12/77,435; RR: 0.74 (0.40-1.36) >=10 yr: 17/65,594; RR: 1.03 (0.59-1.78) -Confounders adjusted for: age, BMI, history of diaberes, hypertension, high cholesterol level, age at menopause, smoking, and parental history of premature heart disease; -Duration of use was underestimated by an average of 1 yr, since duration during each 2-yr follow-up period was established at the start of each period	D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/a D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/a Level of bias:Low Indirectness Does the study match the review protocol in terms of: Population: No (only registered nurses were included) Outcome: Yes Indirectness: Some Other information The NIH was not a general population study
Full citation Grodstein,F., Manson,J.E., Stampfer,M.J., Hormone therapy and coronary heart disease: the role of time since menopause and age at hormone initiation, Journal of Women's Health, 15, 35-44, 2006	Sample size N=121,700 (1976-2000 follow-up data for the current analyses) Characteristics As reported under Stampfer et al. 1985 Inclusion criteria As reported under Stampfer et al. 1985 Exclusion criteria As reported under Stampfer et al. 1985	Interventions HRT	Details Setting: -As reported under Stampfer et al. 1985 Methods: -As reported under Stampfer et al. 1985 Statistical methods: -As reported under Stampfer et al. 1985 -Confounding factors adjusted for: age, BMI, smoking, history of hypertension, elevated cholesterol, parental MI before age 60. For certain analyses, husband's education was also adjusted for as an additional measure of socioeconomic status. Follow-up:	Results Risk of coronary heart disease among current HRT users compared to never users, n/person-years, adjusted RR (95%CI): --Analyses excluding women with prevalent heart disease (1976-2000 data): Never users: 795/429,032; RR: 1.00 (reference group) Current estrogen alone	Limitations As reported under Stampfer et al. 1985 Other information The inability to assess acute effects of hormone use is a limitation of the current study. The issue of incomplete capture of early clinical events in observational studies has been suggested as a possible explanation for the apparent discrepancey between observational and the WHI. The NHS do not have

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Ref Id 229382 Country/ies where the study was carried out US Study type Prospective follow-up Aim of the study To explore the relation of heart disease to type of hormones used and dose of estrogen, in addition to the possible influences of women's CHD risk factor profile, the timing of their HT initiation, and incomplete capture of early clinical events. Study dates 1976-2000 (24-year follow-up analyses) Source of funding NIH			Cohort follow-up was >90%	users: 225/206,383; RR: 0.65 (CI not reported) Current estrogen plus progestin: 112/118,735; RR: 0.64 (CI not reported) -Adjusted for age, BMI, hypercholesterolemia, hypertension, parental history of premature heart disease, diabetes, smoking (1980-2000 data) Never users: 795/429,032; RR: 1.00 (reference group) Current estrogen alone users: 225/206,383; RR:0.71 (0.61-0.83) Current estrogen plus progestin: 112/118,735; RR: 0.68 (0.55-0.83) -Adjusted for age, BMI, hypercholesterolemia, hypertension, parental history of premature heart disease, diabetes, smoking, and husband's education, physical activity, vitamin E and multivitamin supplementation, aspirin use. --Analyses similar with WHI inclusion criterion-including women with and without prevalent heart disease: (herein, about 6% of women with prevalent coronary disease in NHS were included as WHI included about 4%-6% of women with preexisting CHD	sufficient data to indentify women who had begun HT shortly before their coronary event (follow-up every two years), and in the primary analysis, these subjects would be generally categorized among those who had never taken HRT.

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				conditions) (1976-2000 data): Never users: 922/449,599; RR: 1.00 (reference group) Current estrogen alone users: 274/220,368; RR: 0.66 (CI not reported) Current estrogen plus progestin: 131/124,391; RR: 0.64 (CI not reported) -Adjusted for age, BMI, hypercholesterolemia, hypertension, parental history of premature heart disease, diabetes, smoking (1980-2000 data) Never users: 922/449,599; RR: 1.00 (reference group) Current estrogen alone users: 274/220,368; RR:0.72 (0.62-0.82) Current estrogen plus progestin: 131/124,391; RR: 0.69 (0.57-0.83) -Adjusted for age, BMI, hypercholesterolemia, hypertension, parental history of premature heart disease, diabetes, smoking, and husband's education, physical activity, vitamin E and multivitamin supplementation, aspirin use. Risk of coronary heart disease in relation to current HRT use and timing of hormone therapy initiation with respect to onset of	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				menopause, n (no. of cases)/person-years; adjusted RR (95% CI): --Analyses excluding women with prevalent heart disease , near menopause (within 4 years of menopause), 1976-2000 data: Never users: 666/329,604; RR: 1.00 (reference group) Initiated estrogen alone: 116/133,194; RR: 0.48 (CI not reported) Initiated estrogen + progestin: 78/91,985; RR: 0.45 (CI not reported) 1980-2000 data: Never users: 666/329,604; RR: 1.00 (reference group) Initiated estrogen alone: 116/133,194; RR: 0.66 (0.54-0.80) Initiated estrogen + progestin: 78/91,985; RR: 0.72 (0.56-0.92) --Analyses excluding women with prevalent heart disease , HRT initiated 10 + years after menopause, 1976-2000 data: Never users: 400/152,205; RR: 1.00 (reference group) Initiated estrogen alone: 59/34,000; RR: 0.68 (CI not reported) Initiated estrogen + progestin: 23/11,945; RR: 0.70 (CI not reported) --Adjusted for age, BMI,	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				hypercholesterolemia, hypertension, parental history of premature heart disease, diabetes, smoking --Analyses excluding women with prevalent heart disease , HRT initiated 10+ years after menopause, 1980-2000 data: Never users: 400/152,205; RR: 1.00 (reference group) Initiated estrogen alone: 59/34,000; RR: 0.76 (0.57-1.00) Initiated estrogen + progestin: 23/11,945; RR: 0.80 (0.53-1.23) --Adjusted for age, BMI, hypercholesterolemia, hypertension, parental history of premature heart disease, diabetes, smoking, and husband's education, physical activity, vitamin E and multivitamin supplementation, aspirin use. --Analyses similar with WHI inclusion criterion-including women with and without prevalent heart disease: (herein, about 6% of women with prevalent coronary disease in NHS were included as WHI included about 4%-6% of women with preexisting CHD conditions) near menopause (within	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				4 years of menopause), 1976-2000 data: Never users: 773/346,219; RR: 1.00 (Refernce group) initiated estrogen alone: 130/140,515; RR: 0.46 (CI not reported) Initiated estrogen + progestin: 89/95,847; RR: 0.45 (CI not reported) ---Adjusted for age, BMI, hypercholesterolemia, hypertension, parental history of premature heart disease, diabetes, smoking 1980-2000 data: Never users: 773/346,219; RR: 1.00 (Refernce group) initiated estrogen alone: 130/140,515; RR: 0.62 (0.52-0.76) Initiated estrogen + progestin: 89/95,847; RR: 0.71 (0.56-0.89) --Adjusted for age, BMI, hypercholesterolemia, hypertension, parental history of premature heart disease, diabetes, smoking, and husband's education, physical activity, vitamin E and multivitamin supplementation, aspirin use. HRT initiated 10+ years after menopause, 1976-2000 data: Never users:	

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				481/164,537; RR: 1.00 (Reference group) Initiated estrogen alone: 84/37,978; RR: 0.78 (CI not reported) Initiated estrogen + progestin: 31/13,133; RR: 0.78 (CI not reported) ----Adjusted for age, BMI, hypercholesterolemia, hypertension, parental history of premature heart disease, diabetes, smoking --1980-2000 data: Never: 481/164,537; RR: 1.00 (Reference group) Initiated estrogen alone: 84/37,978; RR: 0.87 (0.69-1.10) Initiated estrogen + progestin: 31/13,133; RR: 0.90 (0.62-1.29) --Adjusted for age, BMI, hypercholesterolemia, hypertension, parental history of premature heart disease, diabetes, smoking, and husband's education, physical activity, vitamin E and multivitamin supplementation, aspirin use.	
Full citation Grodstein,F., Manson,J.E., Stampfer,M.J., Rexrode,K., Postmenopausal hormone therapy and stroke: role of time since menopause and	Sample size N= 121 700 Characteristics Not reported in this publication Inclusion criteria -Women aged 30-55 yrs, who returned a mailed questionnaire including detailed information on menopause and postmenopausal hormone use as well as on	Interventions Estrogen, estrogen and progestin	Details Setting: questionnaire survey among registred nurses in 1976, and biennial follow-up Methods: Ascertainment of HRT: -Self-reported use and duration of HRT after menopause; beginning in 1978, information on type of HRT was collected; all information was updated	Results Risk of total stroke: n/person-years; adjusted RR (95% CI): by user type: Never users: 360/485,987; 1.00 (reference group) Current users of estrogen alone:	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
age at initiation of hormone therapy, Archives of Internal Medicine, 168, 861-866, 2008 Ref Id 301080 Country/ies where the study was carried out US Study type Prospective follow-up (The Nurses' Health Study Cohort) Aim of the study To evaluate stroke risk associated with hormone therapy (HT) in younger women, in recently menopausal women, and in older women. To explore the effects of initiating HT at varying intervals since menopause and at different ages. Study dates 1976-2004 (28 yrs) Source of funding NIH	diagnoses of CVD and CVD risk factors. Exclusion criteria -Women who reported stroke as well as myocardial infarction, angina, CVD, or cancer on the 1976 questionnaire;		biennially; Ascertainment of stroke cases: -The first occurrences of nonfatal and fatal stroke between the return of the 1976 questionnaire and June 2004 were identified. Medical records for the nonfatal stroke cases were reviewed. Deaths were ascertained by reports from relatives or postal authorities and a search of the National Death Index. Only fatal stroke cases documented by medical records were included for analysis. Statistical analysis: -Analyses were based on incidence rates using person-years of follow-up as the denominator; -Mantel-Haenszel rate ratios with 95% confidence interval for age-adjusted RRs; -Cox proportional hazards models were used to calculate adjusted RRs controlling for age, BMI, height, smoking, history of hypertension, diabetes, and elevated cholesterol level, husband's education, and parental MI before the age of 60 yrs.	276/256,437; 1.39 (1.18-1.63) Current users of estrogen and progestin: 138/153,192; 1.27 (1.04-1.56) Risk of ischemic stroke: n/person-years; adjusted RR (95% CI): by user type: Never users: 235/485,987; 1.00 (reference group) Current users of estrogen alone: 183/256,437; 1.43 (1.17-1.74) Current users of estrogen and progestin: 103/153,192; 1.53 (1.21-1.95) Risk of hemorrhagic stroke: n/person-years; adjusted RR (95% CI): by user type: Never users: 85/485,987; 1.00 (reference group) Current users of estrogen alone: 61/256,437; 1.37 (0.98-1.91) Current users of estrogen and progestin: 103/153,192; 0.87 (0.55-1.39) Risk of fatal stroke: n/person-years; adjusted RR (95% CI): by user type: Never users: 50/485,987; 1.00 (reference group)	confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-No (participants were registered nurses) A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Unclear Level of risk-High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk: N/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				Current users of estrogen alone: 33/256,437; 1.22 (0.78-1.90) Current users of estrogen and progestin: 15/153,192; 1.03 (0.57-1.86) Risk of nonfatal stroke: n/person-years; adjusted RR (95% CI): by user type: Never users: 310/485,987; 1.00 (reference group) Current users of estrogen alone: 243/256,437; 1.41 (1.19-1.68) Current users of estrogen and progestin: 123/153,192; 1.31 (1.05-1.62) (Adjusted for age, BMI, height, smoking, history of hypertension, diabetes, and elevated cholesterol level, husband's education, and parental MI before the age of 60 yrs) Risk of total stroke: n/person-years; adjusted RR (95% CI): by timing of HT initiation with respect to onset of menopause: HT initiation near menopause (defined as 4-yr in the study) Never users: 312/370,831; 1.00 (reference group) Estrogen alone: 146/163,092; 1.29	C.2a How many participants did not complete treatment in each group?-10% (90% follow-up was achived by the study) C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-Not reported C.3a For how many participants in each group were no outcome data available?- Unclear (not reported) C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)- yes Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up- Yes (24 yrs) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-Yes D.5 Investigators were kept 'blind' to other important confounding and prognostic

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				(1.06-1.58) Estrogen and progestin: 93/119,912; 1.22 (0.95-1.55) Risk of total stroke: n/person-years; adjusted RR (95% CI): HT iniation >=10 yr after menopause Never users: 240/193,066; 1.00 (reference group) Estrogen alone: 133/87,038; 1.31 (1.06-1.63) Estrogen and progestin: 53/35,909; 1.18 (0.87-1.60) Risk of total stroke: n/person-years; adjusted RR (95% CI): By HT initiation age: HT initiation at age 50-59 yr: Never: 108/239,967; 1.00 (reference group) Estrogen alone: 31/49,590; 1.58 (1.06-2.37) Estrogen and progestin: 25/51,904; 1.34 (0.84-2.13) HT initiation at age >=60 yr: Never: 242/202,856; 1.00 (reference group) Estrogen alone: 41/18,513; 1.82 (1.30-2.54) Estrogen and progestin: 37/17,588; 1.72 (1.21-2.44) (Adjusted for age, BMI, height, smoking, history of hypertension,	factors-Unclear Level of bias:Low Indirectness Does the study match the review protocol in terms of: Population: No (only registered nurses were included) Outcome: Yes Indirectness: Some Other information -The NHS study was carried out among registered nurses; -Compared with the previous NHS publication with follow-up through 1996, the present data represent substaintially greater power to detect effects, with a 36% increase in person-years among women who had never used HT and 54% increase among women who were currently taking HT; -The NHS' results on the relation of HT to stroke were entirely consistent with those from the WHI trials;

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				diabetes, and elevated cholesterol level, husband's education, and parental MI before the age of 60 yrs)	
Full citation Corrao,G., Zambon,A., Nicotra,F., Fornari,C., La,Vecchia C., Mezzanzanica,M., Nappi,R.E., Merlino,L., Cesana,G., Persistence with oral and transdermal hormone replacement therapy and hospitalisation for cardiovascular outcomes, Maturitas, 57, 315-324, 2007 Ref Id 301026 Country/ies where the study was carried out Italy Study type Prospective cohort study Aim of the study To compare the effects of transdermal and oral routes of HRT administration, and to investigate the role of income as a potential	Sample size - 88,050 women for whom at least one drug used for HRT dispensed during the study period - 11,175 women excluded because they had already experienced at least one prescription of HRT and/or had been hospitalised for cardiovascular or neoplastic disease and/or accumulated less than 6 months of follow-up - Remaining cohort: 76,875 Characteristics AT COHORT ENTRY Age in years, mean (SD) ≤ 6 months persistence with HRT: 56.1 (5.3) 7-12 months persistence with HRT: 56.0 (5.1) 13-24 months persistence with HRT: 54.5 (4.8) 25-36 months persistence with HRT: 53.4 (4.4) >36 months persistence with HRT: 52.4 (3.9) Total: 54.7 (5.0) Taxable income in 1000 Euros, median (interquartile range) ≤ 6 months persistence with HRT: 11.4 (3.9 to 21.0) 7-12 months persistence with HRT: 12.2 (4.3 to 22.0) 13-24 months persistence with HRT: 13.7 (4.9 to 24.0) 25-36 months persistence with HRT: 14.0 (2.3 to 25.0) >36 months persistence with HRT: 14.3 (3.5 to 24.3) Total: 12.7 (3.9 to 22.8)	Interventions HRT use	Details Setting Data obtained from the Health Services databases of Lombardia HRT exposure assessment Drug types, dosages and number of canisters dispensed at each cohort member during follow-up were retrieved from the Regional outpatient prescription drug database and used to construct the cumulative measure of HRT exposure. The conjugated-estrogen dose equivalent was calculated for each dispensed canister and the resultant defined daily dose units, established as the typical adult's daily maintenance dose was calculated for each prescribed drug. For overalapping prescriptions, the individual was assumed to have refilled early and completed the first prescription before starting the second. An indicator of cumulative persistence with HRT during follow up was constructed by summing the number of days with medication available and categorized according to progressively increasing exposure duration (≤6, 7-12, 13-24, 25-36 and >36 months) Outcome assessment The Regional hospital discharge database was used to identify cohort members who during follow-up experienced at least one hospitalisation for any disease of the circulatory system (ICD9: 390-459) and among those for ischaemic heart disease (410-414) and cerebrovascular disease (430-438), recorded as main cause of hospitalisation. The earliest date of	Results Hazard ratios* (95%CI) of cumulative persistence with every form and with different routes (transdermal vs oral) of HRT administration on the risk of hospitalisation for disease of ischaemic heart disease, and of cerebrovascular disease Ischaemic heart disease Every route of administration: ≤6 months persistence with HRT - 1.00 (reference), 7-12 months persistence with HRT - 1.00 (0.80 to 1.26), 13-24 months persistence with HRT: 0.85 (0.65 to 1.11), 25 to 36 months persistence with HRT - 0.83 (0.58 to 1.20), >36 months - 0.61 (0.37 to 0.99) Transdermal administration: ≤6 months persistence with HRT - 1.00 (reference), 7-12 months persistence with HRT - 1.03 (0.82 to 1.30), 13-24 months persistence with HRT: 0.79 (0.59 to 1.05), 25 to 36 months persistence with HRT - 0.83 (0.56 to 1.24), >36 months - 0.59 (0.33 to 1.05) Oral administration: ≤6 months persistence with	Limitations Based on NICE guidelines manual 2012: Cohort studies checklist A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-No (all participants of this study were HRT users at baseline) A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-No (women of longer HRT use duration had higher income at baseline) Level of risk-High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
confounder of the HRT effect on the risk of the considered outcomes Study dates 1998 to 2000 (all women received at least one HRT prescription during this period) Source of funding Supports for the study comes from grants of the Italian Minister for University and Research	Route of HRT administration Transdermal, % ≤ 6 months persistence with HRT: 83.9 7-12 months persistence with HRT: 91.9 13-24 months persistence with HRT: 91.6 25-36 months persistence with HRT: 91.9 >36 months persistence with HRT: 92.3 Total: 89.1 Oral, % ≤ 6 months persistence with HRT: 16.1 7-12 months persistence with HRT: 8.1 13-24 months persistence with HRT: 8.4 25-36 months persistence with HRT: 8.1 >36 months persistence with HRT: 7.7 Total: 10.9 DURING FOLLOW-UP Route of HRT administration Only transdermal, % ≤ 6 months persistence with HRT: 69.6 7-12 months persistence with HRT: 68.5 13-24 months persistence with HRT: 54.6 25-36 months persistence with HRT: 49.9 >36 months persistence with HRT: 38.2 Total: 57.7 Only oral, % ≤ 6 months persistence with HRT: 14.7 7-12 months persistence with HRT:		hospitalisation was considered as that of outcome onset. Statistical methods Follow-up 1998-2000 to 2003; each women accumulated person-years of follow up from the date of the first recorded prescription of a drug for HRT to the earliest of the dates of: hospitalisation for CVD or cancer, death for any cause, emigration or 31 December 2003.	HRT - 1.00 (reference), 7-12 months persistence with HRT - 1.08 (0.75 to 1.55), 13-24 months persistence with HRT: 0.60 (0.31 to 1.14), 25 to 36 months persistence with HRT - 1.02 (0.38 to 2.75), >36 months - 1.80 (0.66 to 4.88) Cerebrovascular disease Every route of administration: ≤6 months persistence with HRT - 1.00 (reference), 7-12 months persistence with HRT - 0.82 (0.61 to 1.10), 13-24 months persistence with HRT: 0.74 (0.53 to 1.06), 25 to 36 months persistence with HRT - 0.57 (0.34 to 0.94), >36 months - 0.53 (0.30 to 0.94) Transdermal administration: ≤6 months persistence with HRT - 1.00 (reference), 7-12 months persistence with HRT - 0.73 (0.53 to 0.99), 13-24 months persistence with HRT: 0.81 (0.58 to 1.15), 25 to 36 months persistence with HRT - 0.50 (0.29 to 0.87), >36 months - 0.39 (0.18 to 0.82) Oral administration: ≤6 months persistence with HRT - 1.00 (reference), 7-12 months persistence with HRT - 1.21 (0.78 to 1.90), 13-24 months persistence with HRT: 1.26 (0.69 to 2.31), 25 to	B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk: N/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	4.9 13-24 months persistence with HRT: 5.2 25-36 months persistence with HRT: 4.7 >36 months persistence with HRT: 5.1 Total: 8.4 Either transdermal and oral, % ≤ 6 months persistence with HRT: 15.7 7-12 months persistence with HRT: 26.6 13-24 months persistence with HRT: 40.2 25-36 months persistence with HRT: 45.4 >36 months persistence with HRT: 56.7 Total: 33.9 Inclusion criteria - All women aged 45 to 65 years who received at least one HRT prescription anytime during 1998 to 2000 identified from the outpatient prescription drug database (these drugs included all those that have been used to treat symptoms of menopause with different hormone regimen (estrogens or estradiol alone or conjugated with progestin) and mode of administration (ovules, gels, patches and pills) Exclusion criteria - Women younger than 45 years or older than 65 years at the date of their first recorded prescription - Those at whom at least one prescription of HRT was dispensed in the period ranging from 1 January 1997 through the date of entry into the cohort - Those who previously experienced at least one hospitalisation for CVD or cancer - Those reporting CVD as			36 months persistence with HRT - 0.73 (0.18 to 2.93), >36 months - 0.54 (0.08 to 3.86) *Adjusted for age at entry (continuous), exposures to cardiac drugs, antihypertensives, lipid modifying agents, drugs used in diabetes, raloxifene, and other sex hormones during follow-up	up-Unclear (1998-2000 to 2003) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-No D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-No Level of bias: High Indirectness Does the study match the review protocol in terms of; Population: Yes Outcome: Yes Indirectness: Some Other information This study reported findings on "circulatory system disease" but the results were not included here, because circulatory disease included hypertension and hypercholesterol which were not of interest to the review.

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	'secondary diagnosis' or as 'other relevant condition' in presence of another primary diagnosis during follow-up - Those who did not reach at least 6 months of follow up				
Full citation Alexander,K.P., Newby,L.K., Hellkamp,A.S., Harrington,R.A., Peterson,E.D., Kopecky,S., Langer,A., O'Gara,P., O'Connor,C.M., Daly,R.N., Califf,R.M., Khan,S., Fuster,V., Initiation of hormone replacement therapy after acute myocardial infarction is associated with more cardiac events during follow-up, Journal of the American College of Cardiology, 38, 1-7, 2001 Ref Id 228857 Country/ies where the study was carried out US Study type Prospective study Aim of the study To explore the association	Sample size N=1,857 Participants were postmenopausal women who were originally subjects enroled in a RCT [Coumadin Aspirin Reinfarction Study (CARS) Investigators] Characteristics Demographics: Age in years, mean (sd): Never users: 67 (60,73) Prior/current users: 59 (52,66) New users: 58 (51, 65) Race (%white): Never users: 82 Prior/current users: 91 New users: 86 Education (% college): Never users: 22 Prior/current users: 43 New users: 32 CVD risk factors (%): Current smoker: Never users: 24 Prior/current users: 31 New users: 39 Diabetes: Never users: 30 Prior/current users:20 New users:24 Hypertension Never users:60 Prior/current users:58 New users:51 Cardiac history prior to index MI (%): Prior MI: Never users:18	Interventions HRT	Details Setting: follow-up secondary analysis of data collected in a prior RCT, among women who have had an acute MI Methods: -participants consisted 1,857 postmenopausal women enrolled in CARS HRT exposure assessment: -Prior/current users: those who reported use of HRT at the time of randomization or within the prior two years -New users: those who did not use HRT prior to randomization but reported use during follow-up -Never users: those had not recorded use Outcome assessment: -Composite of CVD death, reinfarction and unstable angina requiring hospitalisation; -Individual components of the triple end point and on subsequent use of revascularization were further looked at; Statistical methods: -Cox proportional hazards survival models for death, MI were developed which included the foregoing 11 predictors as well as randomized treatment and HRT -Counfounder adjusted for included age, previous angina, congestive heart failure, current smoker, hypertension, prior MI, PVD, prior stroke or TIA, race, weight, and randomised treatment. Follow-up:	Results Cardiac events, adjusted HR (95%CI): Composite of death/MI(myocardial infarction)/UA(unstable angina): Prior/current users (duration > 2 yrs) vs. never users: 0.94 (0.75-1.18) New users (duration < 2 yrs) vs. never users: 1.44 (1.05-1.99) Death: Prior/current users vs. never users (duration > 2 yrs): 0.36 (0.17-0.77) New users (duration < 2 yrs) vs. never users: n/a MI: Prior/current users vs. never users (duration > 2 yrs):0.88 (0.58-1.33) New users (duration < 2 yrs) vs. never users: n/a -adjusted for included age, previous angina, congestive heart failure, current smoker, hypertension, prior MI, PVD, prior stroke or TIA, race, weight, and randomised treatment	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-No (subjects were participants enrolled in a RCT, not representative) A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-No Level of risk- High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
between the initiation of hormone replacement therapy (HRT) and early cardiac events (<1 year) in women with a recent myocardial infarction (MI). Study dates Not reported Source of funding Not reported	Prior/current users:14 New users:16 Prior stroke or TIA: Never users:4 Prior/current users:5 New users:2 Congestive heart failure: Never users:17 Prior/current users:14 New users:10 Angina: Never users:33 Prior/current users:34 New users:2 Inclusion criteria -Women were either postmenopausal or surgically sterilized -women who were >=50 years, or who used HRT Exclusion criteria Not reported		2-year		B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk: N/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					up-No (2-year) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Unclear D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/a D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/a Level of bias: High Indirectness Does the study match the review protocol in terms of: population: No Outcome: yes Indirectness: yes Other information -Note that non-users in this study were older than prior and new users (those who initiated HRT use after enrolment of the RCT) -During the follow-up period of the study, there were few MIs and no deaths among the new users of HRT. Therefore, the ability to detect clear associations between HRT use and end points of death and MI was diminished.
Full citation Lokkegaard,E., Andreasen,A.H., Jacobsen,R.K., Nielsen,L.H., Agger,C., Lidegaard,O., Hormone therapy and risk of myocardial	Sample size N= 698,098 Characteristics	Interventions HRT	Details Setting: the Danish Sex Hormone Register Study, which is based on five national registers Methods: -Ascertainment of HRT use: exposure to HRT was recorded from the National Register of Meidicinal Product Statistics, which has collected data on redeemed	Results Risk of myocardial infraction in relation to HRT use: rate [n (MI cases)/n (women-years)], adjusted RR (95%CI): by HRT user categories and age group: Never users:	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential

Study details	Participants				Interventions	Methods	Outcomes and Results	Comments
infarction: a national register study, European Heart Journal, 29, 2660-2668, 2008 Ref Id 311315 Country/ies where the study was carried out Denmark Study type Prospective follow-up study Aim of the study To assess the risk of myocardial infarction as a result of hormone therapy, with focus on the influence of age, duration of HT, various regimens and routes, progestagen type, and oestrogen dose. Study dates 1995-2001 Source of funding Copenhagen County University Hospital		**Year of birth**	**MI rate, %, (n/women-years)**	**Current HRT users (%)**		prescriptions by Danish citizens since Jan 1994, and is considered complete as of Jan 1995. HT exposure was considered a time-varying covariate in the statistical model. -Ascertainment of myocardial infarction: The first event of MI was recorded in either the NPR or cause of death registry receiving information from death certificates; Statistical methods: -Data was analysed by Poisson regression analysis on a data set consisting of risk time (women-years) and number of MI events for each combination of exposure axis, age band, and included confounders. Rate ratio estimates and 95% confidence intervals were calculated for each model. -Confounders adjusted for included age, calendar year, education, employment status, habitation, medication for hypertension, heart conditions, hyperlipidamia, or diabetes; Follow-up: 6 years	51-54 years: 0.61 (374/610,880); RR: 1.00 (reference group) 55-59 years: 1.16 (660/569,331); RR: 1.00 (reference group) 60-64 years: 2.17 (1110/510,776); RR: 1.00 (reference group) 65-69 years: 3.27 (1598/488,409); RR: 1.00 (reference group) Previous users: 51-54 years: 0.57 (38/66,689); RR: 0.84 (0.60-1.18) 55-59 years: 1.08 (76/70,228); RR: 0.94 (0.74-1.19) 60-64 years: 1.53 (67/43,800); RR: 0.74 (0.57-0.94) 65-69 years: 2.34 (64/27,338); RR: 0.77 (0.60-0.99) Current users: 51-54 years: 0.81 (143/177,340); RR: 1.24 (1.02-1.51) 55-59 years: 1.08 (207/192,103); RR: 0.96 (0.82-1.12) 60-64 years: 2.28 (274/120,274); RR: 1.11 (0.97-1.27) 65-69 years: 2.80 (211/75,473); RR: 0.92 (0.80-1.06) By duration and age group: < 1 year duration: 51-54 years: 0.77 (42/54,291); RR: 1.18 (0.86-1.63) 55-59 years: 1.01 (42/41,516); RR: 0.84	confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-Yes A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Unclear (information on important confounder such as BMI, smoking, alcohol consumption, physicial activity not available) Level of risk- Unclear B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk:N/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to
	Age	1925 - 1929	3.4 (856/250,838)	n/a				
		1930 - 1934	2.8 (1740/610,737	13.9				
		1935 - 1939	1.7 (1221/728,707)	19.3				
		1940 - 1944	0.9 (847/919,428)	23.2				
		1945 - 1949	0.6 (283/477,359)	20.3				
	Education	Elementary school	2.2 (3454/1,570,921)	17.4				
		Occupational practice	1.2 (1071/901,304)	21.4				
		Further education	0.7 (319/458,301)	23.6				
		Unknown	1.8 (103/56,542)	16.7				

Study details	Participants				Interventions	Methods	Outcomes and Results	Comments
	Medication	Lipid lowering	5.6 (227/40,178)	16.8			(0.61-1.15) 60-64 years: 2.96 (69/23,297); RR: 1.33 (1.04-1.70) 65-69 years: 3.18 (50/15,717); RR: 0.85 (0.72-1.27) 1-4 years duration: 51-54 years: 0.77 (78/101,337); RR: 1.20 (0.94-1.53) 55-59 years: 1.06 (115/108,221); RR: 0.96 (0.79-1.17) 60-64 years: 2.29 (148/54,511); RR: 1.13 (0.95-1.35) 65-69 years: 2.74 (111/40,547); RR: 0.91 (0.75-1.11) >4 years duration: 51-54 years: 1.06 (23/21,672); RR: 1.59 (1.04-2.44) 55-59 years: 1.18 (50/42,366); RR: 1.07 (0.80/1.44) 60-64 years: 1.76 (57/32,439); RR: 0.89 (0.68-1.16) 65-69 years: 2.60 (50/19,209); RR: 0.89 (0.67-1.19) - adjusted for included age, calendar year, education, employment status, habitation, medication for hypertension, heart conditions, hyperlipidamia, or diabetes;	allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: N/A D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up- Yes (6-year) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-No D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-No Level of bias: High
		Antiarrhythmic	12.6 (458/36,231)	20.3				
		Anti-hypertensive	3.9 (2911/751,268)	23.0				
		Anti-diabetic	7.4 (481/64,761)	11.4				
	Inclusion criteria -In the Civil Registration System (CRS) that registers all Danish inhabitants' age and address, a national cohort of all Danish women aged at least 51 years by Jan 1995 or reaching 51 years during the period from Jan 1995 to Dec 2001 were identified. Exclusion criteria -Women recorded in the National Register of Patients (NRP) with cardiovascular diseases or hormone-related cancers prior to entrance were excluded; -Additionally, women were excluded upon emigration or death from reasons other than MI, or at turning 70 years of age;							

<table>
<tr><th>Study details</th><th>Participants</th><th>Interventions</th><th>Methods</th><th>Outcomes and Results</th><th>Comments</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Indirectness
Does the study match the review protocol in terms of;
Population: Yes

Outcome: Yes
Indirectness: Some
Other information
-Information on HT exposure is based on whether prescription are redeemed. Older women who used HT in their 50s was likely to be misclassified as having never used HT instead of previous users because of truncation of the database.
(detailed definition previous and never HRT users were not reported)</td></tr>
<tr><td>Full citation
Sourander,L., Rajala,T., Raiha,I., Makinen,J., Erkkola,R., Helenius,H., Cardiovascular and cancer morbidity and mortality and sudden cardiac death in postmenopausal women on oestrogen replacement therapy (ERT).[Erratum appears in Lancet 1999 Jan 23;353(9149):330], Lancet, 352, 1965-1969, 1998
Ref Id
230428</td><td>Sample size
N= 7,944

Characteristics
<table><tr><th></th><th>Never users</th><th>Former users</th><th>Current users</th></tr><tr><td>Total number</td><td>5572</td><td>757</td><td>988</td></tr><tr><td>Age in years, mean (sd)</td><td>60.9 (2.5)</td><td>61.0 (2.6)</td><td>59.9 (2.5)</td></tr><tr><td>BMI, mean (sd)</td><td>26.7 (4.3)</td><td>26.1 (4.3)</td><td>25.5 (3.5)</td></tr><tr><td>Social</td><td></td><td></td><td></td></tr></table></td><td>Interventions
HRT (oestrogen)</td><td>Details
Setting:
Questionnaire survey among women attending a mammography screening
Methods:
HRT exposure measurement:
-a validated questionnaire was filled in by participants with the help of a trained nurses who confirmed and checked answers. The questionnaire contained inquires about former and present use of hormone therapy.
-HRT users were classified into 3 groups according to their estrogen use: never users, former users, and current users;
-The mammography and interview were repeated with 2-yr intervals three times during follow-up. These data were linked with those derived from the national registers.
-The mean duration of current ERT before baseline was 8.2 (sd 5.4) years.
Outcomes (CVDs, CVD related death) ascertainment:
-The National death register was used</td><td>Results
Cardiovascular morbidity, adjusteds hazards ratio (HR, 95%CI):
by HRT user category:
Never users: 1
Former users: 1.11 (0.89-1.39)
Current users: 1.07 (0.86-1.32)

Cardiovascular mortality, adjusteds hazards ratio (HR, 95%CI):
by HRT user category:
Never users: 1
Former users: 0.75 (0.41-1.37)
Current users: 0.21 (0.08-0.59)

Coronary artery disease (CAD) morbidity, adjusted hazards ratio (HR, 95%CI):</td><td>Limitations
NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies
A. Selection bias (systematic differences between the comparison groups)
A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-No (participants were women attending a mammography screening program)

A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes
A.3 The groups were comparable at baseline,</td></tr>
</table>

Study details	Participants				Interventions	Methods	Outcomes and Results	Comments
Country/ies where the study was carried out Finland Study type Prospective follow-up study Aim of the study To analyse the relation between postmenopausal oestrogen replacement therapy (ERT), cardiovascular disease, and cancer. Study dates 1987-1988 to 1995 Source of funding Not reported	class, n (%)					to collect mortality data -The National Agency for Welfare and Health register was used to obtain morbidity information on hospital discharges Statistical methods: -One-way ANOVA for differences in mean values between groups; -Cox's proportional-hazards model adjusting for social class, smoking, age, BMI, diabetes, hypertension, CVA, and cardiac failure. Follow-up: 7-yr	by HRT user category: Never users: 1 Former users: 1.23 (0.88-1.71) Current users: 1.05 (0.76-1.46) Coronary artery disease (CAD) mortality, adjusted hazards ratio (HR, 95%CI): by HRT user category: Never users: 1 Former users: 0.64 (0.27-1.47) Current users: 0.19 (0.05-0.77) Stroke morbidity, adjusted hazards ratio (HR, 95%CI): by HRT user category: Never users: 1 Former users: 1.08 (0.55-2.10) Current users: 0.86 (0.42-1.75) Stroke mortality, adjusted hazards ratio (HR, 95%CI): by HRT user category: Never users: 1 Former users: 1.05 (0.41-2.68) Current users: 0.16 (0.02-1.18) Breast cancer morbidity, adjusted hazards ratio (HR, 95%CI): by HRT user category: Never users: 1 Former users: 0.94 (0.47-1.90) Current users: 0.57 (0.27-1.20)	including all major confounding and prognostic factors-No Level of risk-High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk:n/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes (8 yrs) C.2a How many participants did not complete treatment in each group?-N/A C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were
	Highest	340 (6.1 %)	72 (9.5 %)	147 (14.9 %)				
	Upper middle	934 (16.8 %)	176 (23.2 %)	246 (24.9 %)				
	Lower middle	2575 (46.2 %)	283 (37.4 %)	360 (36.4 %)				
	Lowest	1477 (26.5 %)	198 (26.2 %)	214 (21.7 %)				
	Not recorded	246 (4.4 %)	28 (3.7 %)	21 (2.1 %)				
	Clinical							
	Diabetes	134 (2.4 %)	12 (1.6 %)	8 (0.81 %)				
	Smoking	96 (1.7 %)	19 (2.5 %)	16 (1.6 %)				
	Hypertension	1196 (21.5 %)	150 (19.8 %)	151 (15.3 %)				
	CAD	192 (3.5 %)	25 (3.3 %)	27 (2.7 %)				
	Cardiac failure	135 (2.4 %)	12 (1.6 %)	16 (1.6 %)				
	Inclusion criteria -All women born between 1923 and 1930 living in Turku Exclusion criteria -Those started ERT during follow-up (n=627) and those who had missing data on occupation,							

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	smoking, weight, or height were excluded from multivariate survival analyses;			Breast cancer mortality, adjusted hazards ratio (HR, 95%CI): by HRT user category: Never users: 1 Former users: 1.27 (0.38-4.29) Current users: 5.06 (2.47-10.4)	comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: N/a D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes (8 yrs) D.2 The study used a precise definition of outcome-Yes (from national registers) D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-Unclear (not reported) D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-Unclear (not reported) Level of bias: moderate Indirectness Does the study match the review protocol in terms of: Population: Yes Outcome: Yes Indirectness: Some Other information -Self-selected group of women taking HRT who may have healthier lifestyles with fewer risk factors. In the present study, HRT use was more prevalent in the higher social classes.

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Full citation Lafferty,F.W., Fiske,M.E., Postmenopausal estrogen replacement: a long-term cohort study, American Journal of Medicine, 97, 66-77, 1994 Ref Id 229713 Country/ies where the study was carried out US Study type Prospective study Aim of the study To assess the long-term effects of estrogen replacement therapy in 157 post-menopausal women, a prospective, non-randomised, cohort study was conducted from 1964 to 1989. Study dates 1964-1989 (25 yrs) Source of funding University Hospitals, Cleveland, Ohio	Sample size N=157 Characteristics **Non-Estrogen users mean (SD)** / **Estrogen users Mean (SD)** No. of patients: 76 / 81 Age at entry in yrs: 54.7 (3.8) / 52.6 (4.8) Age at menopause: 49.6 (4.1) / 47.8 (4.4) Years menopause to entry: 5.1 (5.3) / 4.7 (4.6) Duration of follow-up: 12.7 (5.1) / 11.5 (5.1) BMI (kg/m2): 24.4 (3.4) / 22.3 (3.2) Hypertension (BP>150/90) in percentages: 23 (30) / 12 (15) Alcohol use (%): 12 (16) / 18 (22) Smoker (%): 20 (26) / 17 (21) Prior hysterectomy: 11 (14) / 35 (43)	Interventions ERT (conjugated equine estrogens, 0.625mg)	Details Setting: Department of medicine, university of Cleveland Methods: HRT exposure: -ERT was offered to all women seen at the private practice, 76 denied. CVD ascertainment: -subjects were followed up prospectively with annual or bienial physical examinations; Cardiovascular disease was detected by the clinic who served as the primary physician of all subjects. Abnormal findings from electrocardigrams were reviewed by a cardiologist unaware of a subject's status Statistcal methods: -Comparisons of demographic variables and serum lipids were analysed using a Student's t-test, chi-square statistics or Mann-Whitney test depending on the distribution of the sample data; -The effect of estrogen on major CVD outcomes controlling for potential confounders was evaluated by using a Cox proportional hazards model. Follow-up: 14 yrs	Results Risk of CVD events associated with ERT, n/1000 patient-years, adjusted RR (95%CI): Myocardial infarction: Non ERT users: 5/1000 ERT users: 1.08/1000 Non ERT users vs. ERT users: 0.34 (0.09-1.34) Cerebrovascular accident: Non ERT users: 4.15/1000 ERT users: 0/1000 Non ERT users vs. ERT users: n/a (p=0.025) -Adjusted for age only;	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-No (ERT was offered to 157 women but 76 declined to use) A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes (though only age adjusted in analyses) A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Unclear Level of risk-High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to

Study details	Participants			Interventions	Methods	Outcomes and Results	Comments
	(%)						treatment allocation-N/a Level of risk: N/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-Not reported (but the study reported that 95% follow-up was achieved) C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/a C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes (14 yrs) D.2 The study used a precise definition of outcome-Yes
	Activity (previous decade)						
	Secondary	22 (37)	24 (40)				
	Moderate/vigorous	38 (63)	36 (60)				
	Education level (median)	13.7 (2.5)	12.8 (2.0)				
	Inclusion criteria -women aged 43-60 years seen at the private practice of Department of medicine, university of Cleveland were offered ERT -healthy, ambulatory, White women with no abonrmality by physical examinaton Exclusion criteria -Past or present history of major diseases including cancer, severe hypertension or CVD, osteroporosis, diabetes, alcoholism, and miscellaneous diseases						

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-Yes D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-Unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of: Population: Yes Outcome: Yes Indirectness: Some (mainly middle-class women with health insurance were included in the study) Other information -The patients population from which the subjects were selected draws predominantly from middle-class neighborhoods in suburban Cleveland. The majority of patients carried some form of health insurance. This limits the ability to generalise the results of the study.
Full citation Hernandez,Avila M., Walker,A.M., Jick,H., Use of replacement estrogens and the risk of myocardial infarction, Epidemiology, 1, 128-133, 1990 Ref Id 229459 Country/ies	Sample size N= 310,000 Characteristics Age in years: 50-64 Ethnicity (%): White: 90% Education: 12 yrs of education: 66% High school: 92% Unemployment (%): 4% Inclusion criteria Not reported	Interventions HRT (conjugated estrogens)	Details Setting: Retrospective chart review Methods: Ascertainment of HRT: -all prescriptions for conjugated estrogens were identified Ascertainment of MI: -cases were women aged 54-60 yrs with a primary diagnosis of myocardial infarction (MI) Statistical methods: Poisson regression models for the cohort analysis and conditional logistic	Results Hospitalisation for MI in relation to duration of estrogens use in women aged 50-64; n/person years; adjusted RR (95%CI) By duration of current use: Non-users: 108/110,971; 1 year duration: 1/1,383; RR: 0.8 (0.1-6.1) 2 years: 1/1,833; RR: 0.6 (0.1-4.1)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is,

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
where the study was carried out US Study type Retrospective cohort study Aim of the study To explore further the relation between estrogen and coronary heart disease and to elucidate the reasons for conflict in previous findings, data from women aged 50-64 years at the Group Cooperative of Puget Sound in Seattle, Washington were examined. Study dates 1978-1984 (6-yr follow-up) Source of funding Not reported	Exclusion criteria Not reported		regression for the case-control analysis; Follow-up: 6-yr	3 years: 0/1,930; RR: - 4 years: 0/1,339; RR:- 5 + years: 4/5,033; RR: 0.9 (0.3-2.6) Unknown: 6/5,995; RR: 0.9 (0.4-2.2) > 1 year: - ; RR: 0.7 (0.3-1.3) -Confounders adjusted for: age in 5-yr intervals and for period in 2-yr intervals	the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-Yes A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes (only age and period effects adjusted for in analyses) A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Unclear Level of risk-High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk:N/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Unclear D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes (6-yr) D.2 The study used a precise definition of outcome-Yes (hospitalisation records)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					D.3 A valid and reliable method was used to determine the outcome-Unclear D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/a D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/a Level of bias:Low Indirectness Does the study match the review protocol in terms of; Population: Unclear Outcome: Yes Indirectness: Some Other information -The authors did not have access to data on major predictors of MI such as smoking, blood lipid levels etc. -The present study was restricted to women who survived MI long enough to be hospitalised
Full citation Su,I.H., Chen,Y.C., Hwang,W.T., Liu,Z., Su,T.P., Chen,T.J., Barnhart,K.T., Yang,Y.X., Risks and benefits of menopausal hormone therapy in postmenopausal	Sample size - 16,045 subjects were in the final dataset - 4,712 subjects were exposed to E + P MHT - 1,208 subjects were exposed to E-only MHT - For E + P MHT exposed participants, there were 8070 E + P MHT unexposed controls - For E only MHT exposed participants, there were 2055 E only unexposed controls	Interventions - HT exposure: E + P HT, E-only HT - No HT exposure: E + P unexposed, E-only unexposed	Details Exposure status - Potential eligible subjects who filled at least 2 monthly prescriptions within 3 continuous months during the enrollment interval were categorized as exposed to MHT - For each MHT exposed participant, the first date when the MHT prescription was filled was deemed her study enrollment date - Two MHT exposure groups were selected based on prescription data	Results Comparison of outcomes between E-only MHT and unexposed participants aged ≤ 55 years at study entry Acute MI E-only MHT: 0 (0) E-only unexposed: 2 (0.04)	Limitations Based on NICE guidelines manual 2012: Cohort studies checklist Other information Based on NICE guidelines manual 2012: Cohort studies checklist A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Chinese women, Menopause, 19, 931-941, 2012 Ref Id 203512 Country/ies where the study was carried out USA Study type Retrospective cohort study Aim of the study To assess risks and benefits of conjugated equine estrogens (CEE) and medroxyprogesterone acetate (MPA) in postmenopausal Chinese women Study dates Enrollment interval June 1 1997 to May 31 2000 Source of funding ASRM/Ortho Research Grant in Reproductive Medicine	*During the study, 551 (3.4%) were lost to follow up Characteristics Age at study entry in years, mean (SD) E + P MHT: 58.2 (6.3) E + P unexposed: 58.9 (6.2) E-only MHT: 59.2 (6.9) E-only unexposed: 59.7 (6.7) Smoking, n (%) E + P MHT: 0 (0) E + P unexposed: 0 (0) E-only MHT: 0 (0) E-only unexposed: 0 (0) Obesity, n (%) E + P MHT: 2 (0.04) E + P unexposed: 2 (0.03) E-only MHT: 1 (0.08) E-only unexposed: 1 (0.01) Hypertension, n (%) E + P MHT: 503 (10.6) E + P unexposed: 529 (6.6) E-only MHT: 157 (13.0) E-only unexposed: 143 (7.0) Hypercholestrolemia, n (%) E + P MHT: 194 (4.1) E + P unexposed: 126 (1.6) E-only MHT: 52 (4.3) E-only unexposed: 41 (2.0) Treated for diabetes, n (%) E + P MHT: 373 (7.9) E + P unexposed: 662 (8.2) E-only MHT: 137 (11.3) E-only unexposed: 178 (8.7) Inclusion criteria - Age 50 to 79 - Assumed menopausal - Controls age matched 1:2 Exclusion criteria - Medical condition associated with predicted survival <3 years		- Those who filled prescriptions for daily CEE (0.625mg daily) and MPA (5mg daily) were considered exposed to E + progestin; subjects who filled prescriptions for only CEE (0.625mg daily) and no P were considered exposed to E-only MHT. - Unexposed subjects were randomly selected from the remainder of the cohort - Matched by date of birth within 5 years, two age-matched unexposed subjects were randomly selected for each exposed subjects and designated the same enrollment date Outcomes - CHD deaths were defined as death occurring within 28 days of hospitalisation when MI diagnosis was given - The global index was a composite outcome summarizing the earliest occurrence of breast cancer, stroke, PE, endometrial cancer, colorectal cancer, hip fracture or death Follow-up - Follow-up period of each subject was determined from the subject's enrollment date to the date of the respective outcome diagnosis, death, loss of NHI coverage or December 31, 2007, whichever was earliest Statistical analysis - Cox proportional hazard ratios were estimated for each primary outcome	Adjusted* HR (95%CI): N/A CHD death E-only MHT: 0 (0) E-only unexposed: 0 (0) Adjusted* HR (95% CI): N/A Stroke E-only MHT: 17 (0.41) E-only unexposed: 18 (0.37) Adjusted* HR (95%CI): 0.99 (0.50-1.95) Global index E-only MHT: 53 (1.3) E-only unexposed: 53 (1.1) Adjusted* HR (95%CI): 1.12 (0.77-1.66) *Adjusted for age, statin use, aspirin use, hypercholesterolemia, diabetes medication use and hypertension	to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-Yes A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-No Level of risk-High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk:N/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	- Previous breast cancer - Other previous cancers within 10 years - Endometrial hyperplasia - Alcoholism, drug dependency - Dementia, mental illness - Acute MI, CVA, TIA within 6 months - Severe hypertension - Chronic hepatitis or cirrhosis - Previous PE or DVT				allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Unclear D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Unclear D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/a D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/a

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Level of bias:Low Indirectness Does the study match the review protocol in terms of; Population: the present study was carried out among Chinese women Outcome: Yes Indirectness: Some
Full citation Gast,G.C., Pop,V.J., Samsioe,G.N., Grobbee,D.E., Nilsson,P.M., Keyzer,J.J., Wijnands-van Gent,C.J., van der Schouw,Y.T., Hormone therapy and coronary heart disease risk by vasomotor menopausal symptoms, Maturitas, 70, 373-378, 2011 Ref Id 226543 Country/ies where the study was carried out Sweden or Holland? check Study type Prospective study Aim of the study To examine whether the association	Sample size N= 8,865 (women aged between 46-64) Characteristics **Never HRT users (n=4794)** / **Ever HRT users (n=4071)** Follow-up time in mths, means (sd): 129.7 (25.4) / 116.0 (22.9) Age in years , mean (sd): 52.8 (4.1) / 55.0 (3.7) BMI (kg/m2), mean, sd: 25.6 (4.4) / 25.2 (3.9) CHD, n (%): 142 (3.0) / 110 (2.7) Hot flushes , yes, n (%): 2140 (44.6) / 2333 (57.3) Intens: 391 / 375	Interventions HRT	Details Setting: Questionnaire survey and linkage to official registries Methods: -HRT use: self-reported HT classified as never or ever -CHD: morbidity data was from the Hospital Discharge Registries Statistical methods: -Cox regression model controlling for age, education level, smoking, physical activity, hypertension, hypercholesterolemia, menopausal status, and oral contraceptive use Follow-up: about 10-yr (whenevery multiple CHD events occured, the first clinical diagnosis was taken as endpoint)	Results Coronary heart disease (CHD), adjusted HR (95% CI) According to presence of vasomotor symptoms Presence of flushing: Absent: 1.11 (0.73, 1.69) Present: 1.18 (0.78-1.79) p interaction: 0.66 HRT use among women with presence of (night) sweat Absent: 1.35 (0.91, 2.01) Present: 0.89 (0.57, 1.38) p interaction: 0.15 HRT use among women with intense VMS Absent: 1.26 (0.92, 1.72) Present: 0.51 (0.21, 1.23) p interaction: 0.02	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-Yes A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-No Level of risk-Unclear B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation)

Study details	Participants			Interventions	Methods	Outcomes and Results	Comments
between HRT use and coronary haret disease (CHD) risk differred between women with and without vasomotor symptoms (VMS). Study dates 1994-1995; 1995-2000; Source of funding Board of the UMCU, Utrecht	e VMS, n (%)	(8.2)	(9.2)				B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk: n/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Low D. Detection bias (bias in how
	Hypertension, n (%)	2648 (51.5)	1959 (48.1)				
	Hysterectomy, n (%)	581 (12.2)	743 (18.3)				
	Education completed n (%)						
	Low	766 (16.4)	619 (15.5)				
	Medium	2971 (63.5)	2180 (54.5)				
	High	943 (20.2)	1205 (30.1)				
	Smoking status n (%)						
	Never	2152 (45.3)	2288 (56.5)				
	Past	1411 (29.7)	828 (20.4)				
	Current	1184 (24.9)	935 (23.1)				
	Physically active, n (%)	2031 (43.2)	1714 (42.6)				
	Menopausal status (%)						
	Perimenopausal	1751 (36.5)	1999 (49.1)				
	Postmenopausal	3043 (63.5)	2072 (50.9)				

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Inclusion criteria Not reported Exclusion criteria -Premenopausal women -women who did not consent to linkage with vital status registries; could not be traced in these registries, had unknown date of inclusion or deaht or did not provide information on VMS or HT use -prevalent cases of CHD, stroke, or cancer				outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes (about 10 yrs) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Unclear D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/a D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/a Level of bias: low Indirectness Does the study match the review protocol in terms of; Population: Yes Outcome: Yes Indirectness: Some
Full citation Li,C., Engstrom,G., Hedblad,B., Berglund,G., Janzon,L., Risk of stroke and hormone replacement therapy. A prospective cohort study, Maturitas, 54, 11-18, 2006 Ref Id 311292 Country/ies where the study was carried out Sweden	Sample size N=16,906 Characteristics Sociodemographic characteristics Age in years, mean (sd): Non users: 58 (8) HRT uses: 56 (6) Married (%): Non users: 64.9 HRT uses: 63.7 College/univesity education (%): Non users: 22.5 HRT uses: 29.0 Non-manual occupation (%): Non users: 27.6 HRT uses: 35.1 Life style factors Current smokers (%): Non users: 23.4	Interventions HRT use	Details Setting Malmo Diet and Cancer study -HRT exposure assessment: women who reported they have taken systemic hormone therapy regularly were considered as HRT users (information on past use of HRT was not available in the questionnaire -Outcome assessment: the records of patients with stroke were retrieved by the data linkage to the "Stroke Register in Malmo" and National Hospital Discharge Register Statistical methods: -Cox-regression analysis was applied to assess the relative risk of stroke in relation to HRT use controlled for age and other covariates	Results Ischemic stroke, adjusted HR (95% CI) BY age: < 60 years: 1.01 (0.60-1.70) > 60 years: 1.24 (0.76-2.00) (RRs were adjusted for age, smoking, alcohol consumption, BP, BMI, diabetes, use of BP lowering agents, lipid-lowering agents or and aspirin)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-Yes A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Study type Prospective study Aim of the study To examine the risk of first-ever stroke in relation to use of hormone replacement therapy (HRT) among middle-aged and older Swedish women. Study dates 1991-1996 (baseline examination) to 2004 (mean follow-up time 10.5 yrs) Source of funding Swedish council for Working life and Research	HRT uses: 26.1 Alcohol intake in mean g/day (sd): Non users: 0.77 (0.5) HRT uses: 0.91 (0.4) Low physical activity (%): Non users: 24.8 HRT uses: 23.1 Clinical characteristics: Diabetes (%): Non users: 2.6 HRT uses: 1.1 Hypertension (%): Non users: 56.2 HRT uses: 46.8 History of myocardial infarction (%): Non users: 0.6 HRT uses: 0.3 BMI, mean (sd): Non users: 25.6 (4.3) HRT uses: 24.7 (3.6) Gynecological characteristics: age of menopause in years, mean (sd): Non users: 49.0 (4.8) HRT uses: 48.5 (5.1) postmenopausal (%): Non users: 67.0 HRT uses: 65.0 Prior oral contraceptive (%): Non users: 46.8 HRT uses: 65.3 Oopherectomy (%): Non users: 1.4 HRT uses: 2.3 Inclusion criteria -Women born between 1923-1950 and living in Malmo city Exclusion criteria -Participants with incomplete response to the questions of medication -a history of stroke before baselin examination		-RRs were adjusted for age, smoking, alcohol consumption, BP, BMI, diabetes, use of BP lowering agents, lipid-lowering agents or and aspirin Follow-up time: an average of 10.5 years		A.3 The groups were comparable at baseline, including all major confounding and prognostic factors- No Level of risk-Moderate B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk: N/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Unclear D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Unclear D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/a D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/a Level of bias: Moderate Indirectness Does the study match the review protocol in terms of: Population: Yes Outcome: Yes Indirectness: Some
Full citation Folsom,A.R., Mink,P.J., Sellers,T.A., Hong,C.P., Zheng,W., Potter,J.D., Hormonal	Sample size N=41,837 Analyses were restricted to 41,070 postmenopausal women with hormone replacement therapy data Characteristics HRT status:	Interventions HRT	Details Setting: questionnaire survey among women with a valid Iowa driving license Methods: Ascertainment of HRT use: -a mailed questionnarie provided	Results Risk of CHD in relation to HRT, adjusted RR* (95%CI): By duration: current HRT users >5 yrs: 0.77 (0.61-0.96) current HRT users >5	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
replacement therapy and morbidity and mortality in a prospective study of postmenopausal women, American Journal of Public Health, 85, 1128-1132, 1995 Ref Id 229297 Country/ies where the study was carried out US Study type Prospective follow-up study Aim of the study To assess the association of hormonal replacement therapy with mortality and incidence of multiple diseases in over 40,000 postmenopausal women followed for 6 years as part of the Iowa Women's Health Study. Study dates 1985-1991 (6-year follow-up) Source of funding The National Cancer Institute	Never users: n= 25,275 Former users: n= 11,439 Current users: n=4356 Age 55-59 yr, (%): Never users: 36 Former users: 29 Current users: 46 Current smoker, (%): Never users: 9 Former users: 10 Current users: 8 Alcohol drinker, (%): Never users: 42 Former users: 44 Current users: 51 Currently married, (%): Never users: 75 Former users: 77 Current users: 82 BMI>28kg/m2 (%): Never users: 37 Former users: 35 Current users: 27 Waist/hip ratio > 0.80 (%): Never users: 66 Former users: 65 Current users: 54 High physical activity (%): Never users: 25 Former users: 24 Current users: 28 Hypertension (%): Never users: 36 Former users: 40 Current users: 37 Diabetes (%): Never users: 7 Former users: 6		information on currrent and HRT use; -during the three follow-up questionnaires in 1987,89,92, information on current HRT was also updated. Ascertainment of outcomes: -disease end points between 1986 and 1991 were ascertained (details not reported); -Deaths were identified through the Health Registry and the National Death Index Statistical methods: -Person-years of follow-up were calculated; age-adjusted and multivariate-adjusted relative risks and 95% confidence intervals were determined by proportional hazards regression modelling. -Associations between HRT and end poins were based on baseline HRT use category only. Follow-up: 6 years (response rates in three follow-up questionnaires in 1987,89,92 were 91%,90%, and 83%, respectively)	yrs (excluding women with cancer and heart disease at baseline): 0.90 (0.47-1.72) -*analyses adjusted for age, marital status, physical activity level, alcohol use, smoking, BMI, waist/hip ratio, hypertension, and diabetes Risk of stroke in relation to HRT, adjusted RR* (95%CI): By duration: current HRT users >5 yrs: 1.05 (0.41-2.64)	to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-Unclear (only women with a valid driving license were included) A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Unclear (detailed statistics not reported) Level of risk-High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk: N/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Current users: 4 Inclusion criteria Not reported Exclusion criteria Depending on the end point, the following additional exclusions were made: -breast cancer at baseline (3780) and 348 with prior partial or total mastectomy -endometrial cancer at baseline -any cancer, colon cancer, and other cancer -fracture (7205 with previous fracture at baseline)				up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes (6-year) C.2a How many participants did not complete treatment in each group?-N/A (for the whole cohort the response rates were 91%,90%, and 83% during three follow-ups) C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Unclear D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes D.2 The study used a precise definition of outcome-No (ascertainment of CHD and stroke cases not clearly reported) D.3 A valid and reliable method was used to determine the outcome-Unclear

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					D.4 Investigators were kept 'blind' to participants' exposure to the intervention-No D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-No Level of bias: High Indirectness Does the study match the review protocol in terms of; Population: Yes Outcome: Yes Indirectness: Some
Full citation Shlipak,M.G., Angeja,B.G., Go,A.S., Frederick,P.D., Canto,J.G., Grady,D., Hormone therapy and in-hospital survival after myocardial infarction in postmenopausal women, Circulation, 104, 2300-2304, 2001 Ref Id 230366 Country/ies where the study was carried out US Study type Retrospective cohort study Aim of the study To test the hypothesis that use of HRT before	Sample size N=114,724 (women with documented MI) Characteristics **HRT** Characteristics: Users (n=7353), %; Non-users (n=107,370), % Age, mean: 71; 77 Age, y 55-64: 32; 14 65-74: 36; 27 75-84: 26; 36 >84: 7; 23 Race White: 91; 85 Black: 4; 8 Other: 5; 7 Diabetes: 25; 35 Hypertension: 65; 66 Hypercholesterole: 40; 26	Interventions HRT use	Details Setting: 1674 hospitals chart reviews using data from the national registry Methods: -Ascertainment of HRT: HRT was defined as the NRMI-3 as the use of estrogen, progestin, or estrogen/progestin for reasons other than contraception. -Ascertainment of MI: diagnosis of MI required a principal discharge diagnosis of MI, presentation of or autopsy evidence; Statistical methods: -t-test for the comparison of continuous variables and the Chi-square test for categorical variables; -to determine association of HRT use with MI complications, multivariate logistic regression was used adjusting for differences in baseline characteristics, severity of presentation, and treatments received in hospital;	Results Risk of in-hospital mortality after MI in relation to HRT use, n/N, adjusted OR (95%CI): By age: 55-64 yrs: Non HRT users: 9/15,835; HRT users: 3/2332 OR: 0.54 (0.41-0.71) -adjusted for age, race, diabetes, hypertension, smoking, hypercholesterolemia, prior MI, prior stroke, prior agina, prior heart failure, presence of chest pain, time to presentation to hospital, BP, heart rate, admission diagnosis etc.	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-No (retrospective study) A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-No (HRT users in this study were younger, more likely to be Level of risk-High

Study details	Participants			Interventions	Methods	Outcomes and Results	Comments
hospitalisation would be associated with decreased in-hospital mortality among postmenopausal women with acute MI. Study dates 1998-2000 Source of funding Health Services Research and Development Division of the Veterans Administration, US	mia						
	Current smoker	21	14				
	Angina	14	15				
	Heart failure	14	25				
	Prior event						
	MI	19	24				
	Stroke	9	14				
	PTCA	10	8				
	CABG	10	10				
	Family history of coronary artery disease	30	20				
	First BP (mm Hg)						
	Systolic	146	144				
	Diastolic	79	78				
	Anterior myocardial infarction (MI)	26	24				
	Admission diagnosis of MI	41	36				
	Inclusion criteria Women enrolled in the National						B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk: N/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Registry of Myocardial Infarction-3, aged >=55 yrs and with documented MI. Exclusion criteria Patients who were transferred to another hospital because of the lack of information				differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: N/a D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Unclear (only in-hospital mortality was assessed) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-No D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-No Level of bias: N/a Indirectness Does the study match the review protocol in terms of; Population: Yes Outcome: Yes Indirectness: Some
Full citation Hedblad,B., Merlo,J., Manjer,J., Engstrom,G., Berglund,G., Janzon,L., Incidence of cardiovascular disease, cancer and death in	Sample size N=5,721 (a total of 5,862 peri- or post-menopausal women were identified, analyses were based on 5,721 women without a history of breast or endommetrial cancer at baseline) Characteristics	Interventions HRT	Details Setting: Screening programme conducted between 1983 and 1992 and followed up until 1995; Methods: Ascertainment of HRT use: -a self-administered questionnaire was used to assess use of HRT and other lifestyle factors; Ascertaiment of endpoints:	Results Risk of myocardial or CHD deaths: n/N, adjusted RR (95%CI): Non users: 92/4,759 HRT users: 5/962 RR: 0.37 (0.15-0.90), P=0.029 -adjusted for age, BMI, hypertension, diabetes,	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is,

Study details	Participants			Interventions	Methods	Outcomes and Results	Comments
postmenopausal women affirming use of hormone replacement therapy, Scandinavian Journal of Public Health, 30, 12-19, 2002 Ref Id 229444 Country/ies where the study was carried out Sweden Study type Prospective follow-up study Aim of the study To evaluate the incidence of myocardial infarction, cancer and death in relation to use of hormone replacement therapy (HRT). Study dates 1983-1992 Source of funding The City of Malmo, the Swedish Medical Research Council, and the Swedish Heart and Lung Foundation and government	**Characteristics**	**Non-users (n=4,759)**	**Users (n=962)**		-information on morbidity and mortality following the health examination was obtained by record linkage with the national inpatient register, the Swedish Causes of Death Register, the Swedish Cancer Registry and the Malmo Heart Infarction register. Underlying causes of death or treatment diagnosis was coded in accordance with the 9th ICD system. Statistical methods: -The Kaplan=Meier method, with the generalized Wilcoxon rank sum test, was used for computation of all-cause mortality rate, incidence of cardiac events and cancer; -Cox's proportional hazards model was used to estimate the influence of HRT on incidence of cardiac events and death; adjustment was made for BMI, hypertension, diabetes, smoking, hyperlipidaemia, age at menopause, history of myocardial infraction or stroke, marital status and social class; Follow-up time: 9.21 years (median), ranged from 0.03 to 12.58 years	hyperlipidemia, smoking habits, use of HRT, age at menopause, history of MI or stroke, marital status, and social class.	the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-No A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-No (HRT users were younger, better educated, had lower BMI at baseline) Level of risk-High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk: N/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants
	Age in years, mean (sd)	54.1 (3.0)	53.8 (3.1)				
	Menopausal status						
	Perimenopausal	9.1	28.0				
	Postmenopausal	90.9	72.0				
	Marital status						
	Living alone	34.9	37.2				
	Cohabiting	65.1	62.8				
	Missing values	0.1	0				
	Social class						
	Others	7.4	4.6				
	Manual workers	74.5	70.7				
	Non-manual workers	18.1	24.7				
	Missing values	1.2	0.6				
	Education						
	Primary educati	61.8	54.6				

Study details	Participants			Interventions	Methods	Outcomes and Results	Comments
	on						did not complete treatment in each group?-N/A C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes (median 9.2 years) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-No D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-No Level of bias: High Indirectness Does the study match the
	Some secondary education	23.6	25.2				
	Complete secondary education	11.7	17.0				
	Missing values	2.9	3.2				
	BMI (kg/m2)						
	< 26	64.2	74.7				
	26-30	22.6	18.3				
	>30	13.1	7.0				
	Blood pressure						
	Diastolic blood pressure (mm Hg)	82.7 (9.0)	81.2 (8.7)				
	Systolic blood pressure (mm Hg)	127.8 (17.2)	125.8 (16.1)				
	Smoking habits						
	Never smoked	47.5	45.8				
	Former smokers	19.5	21.4				
	Current	33.0	32.7				

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	smokers History of cardiovascular disease: 1.5 / 1.5 Missing values: 0.1 / 0 History of myocardial infarction: 0.9 / 0.9 History of stroke: 0.7 / 0.6 Inclusion criteria Women born between 1928 and 1942 attending a screening program for early detection of high-risk individuals for CVD Exclusion criteria Women with a history of breast cancer or endometrial cancer were excluded, while those with other forms of cancer were included.				review protocol in terms of; Population: Yes Outcome: Yes Indirectness: Some Other information -Absence of information on type, dose, and duration of HRT use is a limitation in this study. Further, change of exposure is also an inherent methodological problem in long-term cohort studies, such as smoking habit change, change in exposure to HRT, e.g., discontinuation of treatment or dose or change of dose and type, could have been confounders.
Full citation Ettinger,B., Friedman,G.D., Bush,T., Quesenberry,C. P.,Jr., Reduced mortality associated with long-term postmenopausal estrogen therapy, Obstetrics and Gynecology, 87, 6-12, 1996 Ref Id	Sample size N=454 (232 women who began using estrogen within 3 years of menopause and used it for at least 5 years; 222 aged-mathced postmenopausal nonusers) Characteristics **Estrogen users** / **Non users** / **p** Abnormal electrocardiogra: 7.8% / 13.5% / <0.05	Interventions Estrogen	Details Setting: Pharmacy records review, Kaiser Permanente Medical Centre, US Methods: -Ascertainment of HRT exposure: The review was carried out by a medical record analyst who determined the eligibility of each subject without knowledge of the outcome measurements or the hypotheses to be tested. 1110 women born during 1900-1915 who had filled at least two prescriptions for an oral estrogen preparation were identified. Included were those who met the inclusion	Results Risk of CHD-specific mortality in relation to HRT use (among women who began using estrogen within 3 years of menopause, and taken for at least 5 years), n/N, adjusted RR (95%CI): CHD (ICD9 410-444, specific conditions included please see information): Non users: 24/222; RR: 1.00 (Reference group)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)- Unclear

Study details	Participants				Interventions	Methods	Outcomes and Results	Comments
229267 Country/ies where the study was carried out US Study type Restropective follow-up study Aim of the study To compare all-cause and specific-cause mortality rates in women who had or had not used long-term postmenopausal estrogen replacement therapy (ERT). Study dates 1980: pharmacy records between 1969 and 1973 were reviewed; in 1993, updated medical charts were reviewed. Source of funding National Cancer Institute and the Northern California Kaiser Foundation Hospitals	m (ECG)					criteria (n=232); -Non HRT users were women matched for age and length of membership in the health plan who were found from the same computer pharmacy records to have filled prescription for medication other than oral estrogen. They also satisfied all inclusion and exclusion criteria, except that none used estrogen for as long as 1 year. -Ascertainment of outcomes: -Deaths related to reasons documented in the computer pharamacy records were validated by review of the decedent's medical record and hospital discharge data. All death determination were made without knowledge of subjects' estrogen-use status; Statistical methods: -Student t test and chi-square test were used to assess the significance of differences between estrogen users and nonusers; -Cox proportional hazards models were used to estimate relative risks and associated 95% confidence interval for death from any cause and for each of four cause categories including coronary heart disease, other caridovascular disease. Confounders adjusted for included age, BMI, smoking, alcohol consumption, hypertension, abnormal ECG, and total serum cholesterol level above 260 mg/dL; Follow-up: Follow-up was ended at death or the end of 1992, whichever came first; -women using estrogen were followed up to a mean of 26.8 (6.9) years after menopause, and, on average, had taken estrogen for about two-thirds of this time; -non users were followed-up to a mean of 27.9 (6.2) years after menopause	Esterogen users: 10/232; RR: 0.40 (0.16-1.02) -Adjusted for age, BMI, current smoking, alcohol intake, hypertension, total serum cholesterol level >=260 mg/dL, and abnormal electrocardiogram CVD (ICD9 420-444, specific conditions included please see information): Non users: 25/222; RR: 1.00 (Reference group) Estrogen users: 10/232; RR: 0.27 (0.10-0.71) -Adjusted for age, BMI, current smoking, alcohol intake, hypertension, total serum cholesterol level >=260 mg/dL, and abnormal electrocardiogram	A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Yes (besides nonusers drank more and had higher serum cholesterol) Level of risk-Unclear B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk: N/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A C.2b The groups were comparable for treatment
	Diabetes	2.3%	1.5%	0.79				
	Hypertension, treated	36.2%	41.0%	0.30				
	Diastolic BP>90 mm Hg	26.3%	29.8%	0.43				
	Systolic BP > 160 mm Hg	16.0%	19.2%	0.39				
	Cholesterol > 260 mg/dL	37.3%	44.5%	0.16				
	Smoking							
	Current	32.0%	36.0%	0.43				
	Ever	57.5%	48.0%	0.07				
	Alcohol use, drinks/day							
	None, < 1	36.4%	43.3%	0.04				
	<=2	57.4%	47.4%					
	>2	6.2%	9.3%					
	Obesity (BMI	19.6%	25.4%	0.16				

Study details	Participants				Interventions	Methods	Outcomes and Results	Comments
	> 27)					and, although 13.8% began using estrogen, non took it for as long as 1 year.		completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-Yes D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-Yes Level of bias: Low Indirectness Does the study match the review protocol in terms of; Population: some (black women were excluded; and participants were limited to those who were members of
	Surgical menopause	23.1 %	836 %	<0.001				
	BP, mm HG							
	Systolic	133.8 (23.0)	138.6 (21.6)	0.05				
	Diastolic	80.6 (13.6)	82.9 (12.6)	0.10				
	Serum cholesterol (mg/dL)	247.0 (44.6)	257.6 (45.6)	0.02				
	Inclusion criteria -Two groups were included; one included women who had used postmenopausal estrogen for at least 5 years and the other was of age-matched women who had not used estrogen as long as 1 year; -Included in the estrogen group were those subjects who satisfied two criteria: date of menopause documented by either bilateral oophorectomy or spontaneous cessation of meses, and ERT at a dosage equivalent to at least 0.3 mg of conjugated estrogens begun within 3 years of menopause and taken for at least 5 years; Exclusion criteria -Because the original purpose was to study osteoporotic fractures, subjects who used thyroid preparations in dosages exceeding 2 grains daily or who used							

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	anticonvulsants or glucocorticoids or had chronic alcoholism, chronic renal or hepatic disease, hypoparathyroidism, insulin-requiring diabetes, hyperthyroidism, or other conditions known to adversly affect skeletal integrity. -Black women were excluded because they were not considered prone to osteoporotic fractures. -Also women, before the index pharmacy visit, had suffered either myocardial infarction or stroke or who had been diagnosed with any cancer except squamous cell or basal cell skin neoplasm.				large health maintenance organization) Outcome: Yes Indirectness: Some Other information -No information on dosage or dosage change was available over the follow-up years; -specific conditions of outcomes assessed: CHD 410-414: 410 Acute myocardial infarction 411 Other acute and subacute forms of ischemic heart disease 412 Old myocardial infarction 413 Angina pectoris 414 Other forms of chronic ischemic heart disease CVD 420-444: 420 Acute pericarditis 421 Acute and subacute endocarditis 422 Acute myocarditis 423 Other diseases of pericardium 424 Other diseases of endocardium 425 Cardiomyopathy 426 Conduction disorders 427 Cardiac dysrhythmias 428 Heart failure 429 Ill-defined descriptions and complications of heart disease Subarachnoid hemorrhage 431 Intracerebral hemorrhage 432 other and unspecified intracranial hemorrhage 433 Occlusion and stenosis of precerebral arteries 434 Occlusion of cerebral

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					arteries 435 Transient cerebral ischemia 436 Acute, but ill-defined, cerebrovascular disease 437 Other and ill-defined cerebrovascular disease 438 Late effects of cerebrovascular disease etc.
Full citation Graff-Iversen,S., Hammar,N., Thelle,D.S., Tonstad,S., Hormone therapy and mortality during a 14-year follow-up of 14 324 Norwegian women, Journal of Internal Medicine, 256, 437-445, 2004 Ref Id 311098 Country/ies where the study was carried out Norway Study type Prospective study Aim of the study To compare total, cardivascular disease (CVD) and CHD mortality associted with the use of any HT and HT combined with norethisterone or levonorgestrel during 14-yr of	Sample size N= 14,324 (aged 35-62 yrs) Characteristics Age in years, mean: Non users: 51.2 HT users: 48.8 History of MI in percentages: Non users: 0.6 HT users: 0.1 History of angina pectoris in percentages: Non users: 0.7 HT users: 3.1 Use of blood pressure lowering medication in percentages: Non users: 15.5 HT users: 7.8 All causes death, n/N: Any HT type: 41/702 Oestradiol with norethisterone or levonorgestrel: 17/363 Non users: 1141/13,622 CVD death, n/N: Any HT type: 7/702 Oestradiol with norethisterone or levonorgestrel: 4/363 Non users: 324/13,622 CHD death, n/N: Any HT type: 6/702 Oestradiol with norethisterone or levonorgestrel: 4/363 Non users: 169/13,622	Interventions Any HRT, and oestradiol with norethisterone or levonorgestrel	Details Setting: Health screening for CVD risk factors; questionnaires survey in three Norwegian counties Methods: Ascertainment of HRT use: -During health examination following the screening a nurse encouraged attendees to complete the questionnaire with questions on HT use. Ascertainment of death causes: -Information on all deaths in the cohort during follow-up was obtained from the Causes of Death Registry Statistical methods: -The RR of death during 14-year follow-up was analysed for users of HT compared with non users, by means of proportional hazard regression; -Analyses were also performed separately for subgroups according to baseline self-reported CVD status Follow-up: 14-yr	Results Relative mortality risks by use of HT regimens of oestradiol with norethisterone or levonorgestrel: adjusted RR (95%CI): Among all women including both of those with and without CVD health problems at entry (n=13,985): CVD any cause of death: HT use versus non HT use: 0.96 (0.43-2.17) -Adjusted for age and CVD health CVD main cause of death: HT use versus non HT use: 0.94(0.35-2.54) CHD any cause of death HT use versus non HT use: 1.87 (0.76-4.60) -Adjusted for age and CVD health CHD main cause of death HT use versus non HT use: 1.85 (0.68-5.06) Among women without CVD health problems at entry (n=11,350): CVD any cause of death:	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-Unclear A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes (though only age was adjusted in analyses) A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-No (HRT users were "healthier" compared with non-users) Level of risk-High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
follow-up, taking life-style, social factors and baseline cardiovascular health into account. Study dates 1985-1988 to 2002 (14-yr follow-up) Source of funding Not reported	Death due to stroke: Any HT type: 0/702 Oestradiol with norethisterone or levonorgestrel: 0/363 Non users: 87/13,622 -The HT users had higher level of education and personal income, less likely to live in the northernmost county and had less often domestic work as their main occupation; -Mean level of TC, triglycerides, BMI and blood pressure were lower amongst HT users than non-users, whilest mean body height and HDL cholesterol level was higher. Inclusion criteria -women aged between 40-62 Exclusion criteria Not reported			HT use versus non HT use: 0.44 (0.11-1.78) -Adjusted for age CVD main cause of death: HT use versus non HT use: n/a CHD any cause of death HT use versus non HT use: 0.61 (0.08-4.39) -Adjusted for age CHD main cause of death HT use versus non HT use: n/a Among women with CVD health problems at entry (n=2,635): CVD any cause of death: HT use versus non HT use: 2.61 (0.95-7.13) -Adjusted for age and CVD health CVD main cause of death: HT use versus non HT use: 3.40 (1.23-9.37) CHD any cause of death HT use versus non HT use: 4.77 (1.70-13.3) -Adjusted for age and CVD health CHD main cause of death HT use versus non HT use: 5.94 (2.10-16.9) Relative mortality risks by use of any use of HRT: adjusted RR (95%CI): Among all women including both of	received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk: N/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: N/a D. Detection bias (bias in how outcomes are ascertained,

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				those with and without CVD health problems at entry (n=14,324): CVD any cause of death: HT use versus non HT use: 0.69 (0.35-1.33) -Adjusted for age and CVD health CVD main cause of death: HT use versus non HT use: 0.77(0.36-1.64) CHD any cause of death HT use versus non HT use: 1.40 (0.68-2.86) -Adjusted for age and CVD health CHD main cause of death HT use versus non HT use: 1.30 (0.50-2.97) Among women without CVD health problems at entry (n=11,658): CVD any cause of death: HT use versus non HT use: 0.43 (0.16-1.16) -Adjusted for age CVD main cause of death: HT use versus non HT use: 0.32(0.08-1.31) CHD any cause of death HT use versus non HT use: 0.86 (0.27-2.74) -Adjusted for age CHD main cause of death HT use versus non HT use: 0.69 (0.17-2.85) Among women with CVD health problems at	diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes (14-yr) D.2 The study used a precise definition of outcome-Yes (from Causes of Death Registry) D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/a D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/a Level of bias: Low Indirectness Does the study match the review protocol in terms of; Population: Yes Outcome: Yes Indirectness: Some Other information -HT exposure information was taken only once at the entry of the study, there was no information regarding exposure HT during the follow-up. -At baseline HT users were of better health status comapred with non-users.

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				entry (n=2,666): CVD any cause of death: HT use versus non HT use: 1.43 (0.59-3.51) -Adjusted for age and CVD health CVD main cause of death: HT use versus non HT use: 1.96 (0.75-4.38) CHD any cause of death HT use versus non HT use: 2.66 (1.07-6.64) -Adjusted for age and CVD health CHD main cause of death HT use versus non HT use: 2.70 (0.97-7.52)	
Full citation Pentti,K., Honkanen,R., Tuppurainen,M.T., Sandini,L., Kroger,H., Saarikoski,S., Hormone replacement therapy and mortality in 52- to 70-year-old women: the Kuopio Osteoporosis Risk Factor and Prevention Study, European Journal of Endocrinology, 154, 101-107, 2006 Ref Id 230079 Country/ies where the study was carried out	Sample size N=11,667 Characteristics Age in years, mean (sd) No use: 57.5 (3.0) HRT use <= 5 yrs: 56.8 (2.9) HRT use > 5 yrs: 57.6 (2.7) Total: 57.3 (2.9) BMI (kg/m2), mean (sd) No use: 22.2 (3.9) HRT use <= 5 yrs: 21.8 (3.5) HRT use > 5 yrs: 21.1 (3.0) Total: 21.9 (3.6) Parity, mean (sd) No use: 2.5 (1.7) HRT use <= 5 yrs: 2.5 (1.5) HRT use > 5 yrs: 2.2 (1.4) Total: 2.4 (1.6) Time (years) since menopausal (for postmenopausal), mean (sd): No use: 8.1 (4.4) HRT use <= 5 yrs: 6.4 (4.0) HRT use > 5 yrs: 9.3 (3.8) Total: 7.7 (4.3)	Interventions HRT	Details Setting population-based study with data obtained from national registry and surveys HRT exposure assessment: - In 1989, the lifetime use of HRT in years and the indication for HRT was recorded - in 1994, HRT form and duration of use in months were asked for separately for each year from June 1989 to 1994 -HRT use was classified as: no use; 0.05-5 yrs of HRT; and > 5 yrs of HRT use Outcome ascertainment: -Mortality data were obatined from the National Cause of Death Register Statistical methods: The chi-square test and one-way ANOVA were used to compare differences among groups; -Cox's proportional-hazards models were used to study the association of HRT use with mortality from different causes after adjustment for 6-11 covariates.	Results In all women (N=11,667) during the 7-yr follow-up CHD death, n/N, RR (95% CI), P value No HRT use: 33/5519; 1.0 (reference group) HRT use <= 5 yrs: 11/3945; 0.79 (0.36-1.73) p=0.557 HRT use > 5 yrs: 10/2203; 2.16 (0.93-4.98) p=0.072 Death from any cause, n/N: RR (95% CI), P value: No HRT use: 203/5519; 1.0 (reference group) HRT use <= 5 yrs:	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-Yes A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-No A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Yes Level of risk-Low B. Performance bias (systematic differences

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Finland Study type Prospective study Aim of the study To analyse prospectively the association between hormone replacement therapy (HRT) and mortality in women before old age. Study dates 1994-2001 (7-year follow-up) Source of funding Grant from Kuopio University, National Statistics Finland and Academy of Finland	No. of chronic health disorders none (%): No use: 27.9 HRT use <= 5 yrs: 26.1 HRT use > 5 yrs: 26.0 Total: 26.9 one (%) No use: 31.1 HRT use <= 5 yrs: 29.8 HRT use > 5 yrs: 27.5 Total: 30.0 2-3 (%) No use: 30.9 HRT use <= 5 yrs: 33.0 HRT use > 5 yrs: 35.3 Total: 32.4 >=4 (%) No use: 10.1 HRT use <= 5 yrs: 11.2 HRT use > 5 yrs: 11.2 Total: 10.7 Hysterectomy (%): No use: 15.0 HRT use <= 5 yrs: 22.2 HRT use > 5 yrs: 34.2 Total: 21.1 Bilateral oophrorectomy (%): No use: 3.9 HRT use <= 5 yrs: 9.7 HRT use > 5 yrs: 19.5 Total: 8.8 Diabetes (%) No use: 3.6 HRT use <= 5 yrs: 1.8 HRT use > 5 yrs: 1.1 Total: 2.5 Smoking history (%): No use: 18.6		-Covariates adjusted for were: age, parity, BMI, hysterectomy, bilateral oophorectomy, number of chronic health disorders and time since menopause (in postmenopausal group); further, hypertension, daibetes and smoking history were fitted into the multivariate model to study the association of HRT use with the risk of CHD death. Follow-up time: 7 years	95/3945; 1.05 (0.80-1.36) p=0.748 HRT use > 5 yrs: 63/2203; 1.06 (0.78-1.46) p=0.704 In postmenopausal women (N=9,111) during the 7-yr follow-up CHD death, n/N, RR (95% CI), P value No HRT use: 29/4233; 1.0 (reference group) HRT use <= 5 yrs: 8/3276; 0.84 (0.32-2.17) p=0.710 HRT use > 5 yrs: 9/1845; 1.97 (0.80-4.86) p=0.142 Death from any cause, n/N: RR (95% CI), P value: No HRT use: 156/4233; 1.0 (reference group) HRT use <= 5 yrs: 78/3276; 1.07 (0.79-1.46) p=0.661 HRT use > 5 yrs: 56/1845; 0.99 (0.71-1.39) p=0.971	between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/A B.2 Participants receiving care were kept 'blind' to treatment allocation-N/A B.3 Individuals administering care were kept 'blind' to treatment allocation-N/A Level of risk: Unclear C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	HRT use <= 5 yrs: 20.2 HRT use > 5 yrs: 17.9 Total: 19 Inclusion criteria -Women resident in Kuopio Province and born in 1932-1941 (aged 47-57 yrs in 1989) Exclusion criteria -Women whose menopause could not be defined due to hysterectomy; -women whose time since menopause could not defined due to imcomplete data;				available)-N/A Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up- Unclear D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome- Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/A D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/A Level of bias: Moderate Indirectness Does the study match the review protocol in terms of: Population: Yes Outcome: Yes Indirectness: Some Other information -The study did not distinguish between unopposed estrogen and combined therapy.
Full citation Stram,D.O., Liu,Y., Henderson,K.D., Sullivan-Halley,J., Luo,J., Saxena,T., Reynolds,P., Chang,E.T., Neuhausen,S.L., Horn-Ross,P.L., Bernstein,L., Ursin,G., Age-	Sample size N=71,237 Characteristics **36-59 yrs n=30080**: BMI: <18: 337 (1.1) **60-64 yrs n=10816**: BMI: <18: 120 (1.1)	Interventions HRT use	Details Setting: Questionnaire survey Methods: HRT exposure assessment: -on the baseline questionnaire, participants' current, past, or never use of menopausal estrogen and progestin, information on Premarin dose, ages at and years of use were collected; -A later follow-up questionnaire updated information about current use of HT	Results Ischemic heart diease (IHD) death, adjusted HR (95%CI): By age at questionnaire and HRT use type: 36-59: Former HRT: 4/23189 person years Never use: 23/48219 person years HR: 0.37 (0.13-1.06)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups

Study details	Participants			Interventions	Methods	Outcomes and Results	Comments
specific effects of hormone therapy use on overall mortality and ischemic heart disease mortality among women in the California Teachers Study, Menopause, 18, 253-261, 2011 Ref Id 230473 Country/ies where the study was carried out US Study type Prospective study Aim of the study To examine whether age modified the association between HT and the relative risk of overall mortality and ischemic heart diease (IHD) death in the large, prospective California Teachers Study (CTS) cohort. Study dates 1995-1996 through 2004 (5 to 7-year follow-up) Source of funding National Insitute of Health	18-22.5	9844 (32.7)	2925 (27.0)		begining in May 2000 Outcome assessment: -Death were identified by annual linkage with California mortality files and the Social Security Administration death file. Cause of death was obtained from the California mortality files. Statistical methods: Cox regression models controlling for the following confounders: BMI, smoking status, alcohol consumption, physical activity, total caloric intake, and cholesterol during the year before baseline, Self-reported history of diabetes, high blood pressure, MI or heart disease, cancer and stroke. Follow-up: 5-7 year follow-up	Current HRT: 26/178190 person years Never use: 23/48219 person years HR: 0.38 (0.22-0.67) 60-64: Former HRT: 6/13042 person years Never use: 19/20983 person years HR: 0.52 (0.21-1.27) Current HRT: 24/55742 person years Never use: 19/20983 person years HR: 0.53 (0.30-0.93) By age at which HRT was started: <45 years: 1:00 (reference group) 45-54 years of age: 1.05 (0.87-1.27) 55-64 years of age: 0.91 (0.72-1.15) >=65 years of age: 0.99 (0.75-1.31) By years from menopause to hormone therapy: 0: 1.00 (reference group) 1-5: 1.06 (0.85-1.32) 5-10: 1.11 (0.85-1.46) > 10: 0.99 (0.76-1.30)	is not expected to affect the outcome(s) under study)-No (participants were teachers) A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-No Level of risk-High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk: Unclear C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-Not reported C.2b The groups were comparable for treatment
	22.5-25	6771 (22.5)	2473 (22.9)				
	>30	4769 (15.9)	1730 (16.0)				
	Unknown	784 (2.6)	458 (4.2)				
	Smoking:						
	Never	17893 (59.5)	5963 (55.1)				
	Former	10214 (4.0)	4109 (38.0)				
	Current	1973 (6.6)	744 (6.7)				
	Alcohol:						
	Never	4745 (15.8)	1839 (17.0)				
	Former	4250 (14.1)	1361 (12.6)				
	Current	20163 (66.9)	7229 (66.8)				
	HRT use:						
	Never	5525 (18.4)	2429 (22.5)				
	Former	2658 (8.8)	1510 (14.0)				
	Current	20111 (66.9)	6351 (58.7)				
	other	1786 (5.9)	526 (4.9)				
	Death:						
	No	29227 (97.2)	10196 (94.3)				
	Yes	853 (2.8)	620 (5.7)				

Study details	Participants			Interventions	Methods	Outcomes and Results	Comments
	IHD death:						completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-Not reported C.3a For how many participants in each group were no outcome data available?-Not reported C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-Not reported Level of risk: Unclear D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up- Yes D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/a D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/a Level of bias: Low Indirectness Does the study match the review protocol in terms of: Population: Unclear (teachers only) Outcome: Yes
	No	30017 (99.8)	10756 (99.5)				
	Yes	55 (0.2)	54 (0.5)				
	Prior heart attack:						
	No	29839 (99.2)	10632 (98.3)				
	Yes	156 (0.5)	147 (0.4)				
	Prior stroke:						
	No	29752 (98.9)	10643 (98.4)				
	Yes	243 (10.8)	136 (1.3)				
	Prior diabetes:						
	No	29243 (89.4)	9318 (86.2)				
	Yes	3197 (10.6)	1498 (13.9)				
	Inclusion criteria Current and retired female public school teachers and administrators who participated in the CTS Exclusion criteria Women who were: -premenopausal or of unknown menopausal status -who reported a hysterctomy with at least part an ovary left intact and who were less than 56 yrs at baseline -with incomplete information on ever use of HT -older than 94 at baseline -with missing data on smoking						

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	status -younger than 36 yrs				Indirectness: Some Other information -The study may be subject to the "health woman effect"
Full citation Brownley,K.A., Hinderliter,A.L., West,S.G., Grewen,K.M., Steege,J.F., Girdler,S.S., Light,K.C., Cardiovascular effects of 6 months of hormone replacement therapy versus placebo: differences associated with years since menopause, American Journal of Obstetrics and Gynecology, 190, 1052-1058, 2004 Ref Id 310824 Country/ies where the study was carried out US Study type Randomised, double blind placebo-controlled trial Aim of the study To assess the cardiovascular and neuroendocrine effects of HRT versus placebo	Sample size N=84 Characteristics Age Women HRT/ < 5 Y (N=19): 50.6 ± 0.9 Placebo: 53.2 ± 1.2 Ethnicity HRT/ < 5 Y (N=19): Black: 5 White: 14 Placebo (n = 23): Black: 7 white: 16 Inclusion criteria - 9 months or more post menses cessation - pretreatment follicle stimulating levels exceeding 30 IU/mL and mean estradiol level was 19.1 ± 26.7 pg/mL - Satisfactory adherence to 7 months of testing (including 1 month run-in phase) determined by monthly pill counts and plasma estradiol change - Peri-menopausal symptom free at entry Exclusion criteria - History of stage 2 or stage 3 hypertension, MI, CHD or other serious CVH, gall blader disease, liver disorder, thrombophlebitis, thromboembolism or any other cancer or other serious physical or mental illness - Current use of cardiovascular medications - Women with endometrial hyperplasia on biopsy, a first degree relative having breast cancer, and without a negative	Interventions HRT - Oral CEE - E + EP, Premarin daily + Cycrin +	Details Setting: Not reported Sample size calculation: Not reported Randomisation: Method of randomisation unclear. Women with hysterectomy randomly assigned to receive CEE or placebo for 3 months. Women with intact uterus randomly assigned to receive ESTROGEN + PROGESTORONE Allocation concealment and blinding Unclear. "All participants and research staff were blinded to treatment conditions" Statistical methods A series of 3 mixed-model repeated measures ANCOVA Follow-up: 6 months	Results HRT/< 5 y (N=19) SBP (mmHg): 124.0 ± 3.5 - Significant reduction at follow-up compared to placebo (p<0.0007) DBP (mmHg): 80.8 ± 1.7 - Significant reduction at follow-up compared with placebo (p < 0.0001) Placebo (N= 23) SBP (mmHg): 118.9 ± 2.4 DBP (mmHg): 77.7 ± 1.3 *no significant association observed when compared to placebo (p > 0.15)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Unclear A3 - Were groups comparable at baseline - Yes Level of bias: Unclear B Performance bias B1 - Did groups get same level of care - Unclear B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Unclear Level of bias: Unclear C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Yes Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - No D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
in postmenopausal women grouped according to time since menopause. Study dates Not reported. Source of funding NIH grants HL50778 GCRC RR00046 Unrestricted funds from Wyeth-Ayerst	mammogram within past 12 months.				D5 - Were investigators blinded to confounding factors - Unclear Level of bias: Unclear Other information
Full citation The Writing Group for the PEPI Trial, Effects of estrogen or estrogen/progestin regimens on heart disease risk factors in postmenopausal women. The Postmenopausal Estrogen/Progestin Interventions (PEPI) Trial. The Writing Group for the PEPI Trial.[Erratum appears in JAMA 1995 Dec 6;274(21):1676], JAMA, 273, 199-208, 1995 Ref Id 228823 Country/ies where the study was carried out US Study type	Sample size N= 845 CEE, 0.625 mg/d: N = 175 CEE, 0.625 mg/d, + MPA, 10 mg/d for first 12 days: N = 174 CEE, 0.625 mg/d, + MPA, 2.5 mg/d: N = 174 CEE, 0.625 mg/d, + MP, 200 mg/d for first 12 days: N = 178 Placebo: N = 174 Characteristics Age 45 - 64, average: 56.1 years Race: White: 89% Hispanic: 5% African American: 4% Asian: 2% Native American: 0.5% Smoking: Never smoked: 49% Smoked/previous smoker: not reported Hysterectomy Approximately 32% had hysterectomy at average age of 41.8 years.	Interventions HRT (orally): CEE, 0.625 mg/d: CEE, 0.625 mg/d, + MPA, 10 mg/d for first 12 days CEE, 0.625 mg/d, + MPA, 2.5 mg/d CEE, 0.625 mg/d, + MP, 200 mg/d for first 12 days	Details Setting: 7 clinical centres in US: George Washington University, The John Hopkins University, Stanford University, The University of California (LA), The University of California (San Diego), University of Iowa, The University of Texas Health Science Centre, San Antonio Sample size calculation: Designed to provide statistical power exceeding 80%, with overall type I error controlled to be 0.05. Randomisation method: Treatment assignment determined by a computer program that verified all eligibility criteria prior to randomisation. A blocked randomisation scheme was used to assign eligible women in equal numbers to one of five treatment groups (placebo + 4 HRTs), stratified by clinical centre and hysterectomy status. It was expected that women with hysterectomy would differ with regards to bleeding and subsequent unblinding, equal proportions of hysterectomized women were targeted into each PEPI clinic. Allocation concealment and blinding: All pills and capsules were provided in blister packs designed to be opened	Results Results of ANOVA across treatment groups No significant differences in systolic BP or diastolic BP found in groups. Baseline Systolic BP values (mmHg): Placebo: 115 ± 1.1 CEE only: 114.6 ± 1.1 CEE+MPA (cyc*): 114.8 ± 1.0 CEE+MPA (con**): 115.4 ± 1.0 CEE+MP (cyc): 114.2 ± 1.0 Baseline Diastolic BP Values: Placebo: 72.6 ± 0.6 CEE only: 71.8 ± 0.6 CEE+MPA (cyc*): 72.2 ± 0.6 CEE+MPA (con**): 72.1 ± 0.6 CEE+MP (cyc): 71.1 ± 0.6 Unadjusted mean	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes A3 - Were groups comparable at baseline - Yes Level of bias: low B Performance bias B1 - Did groups get same level of care - Unclear B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Unclear C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Multicenter, randomised, double-blind, placebo-controlled trial (RCT) Aim of the study To assess pairwise differences between placebo, unopposed estrogen and each of three estrogen/prgestin regimens on selected heart disease risk factors in healthy postmenopausal women. Study dates December 1989 - February 1991 Source of funding National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health (NIH). Four other NIH institutes: NIA, NIDDK, NIAMS, NICHD provided technical and financial support for the study.	Other: More than half had previous used noncontraceptive estrogen. Inclusion criteria - Aged 45 - 64 years - With or without a uterus - Naturally or surgically menopausal. If natural menopausal: at least 1 year to 10 years past their last menstrual cycle. If surgically: at least 2 months after hysterectomy and with a follicle stimulating hormone level greater than or equal to 40 IU/L. - Normal baseline results of mammography and endometrial biopsy required. Exclusion criteria - Women with severe menopausal symptoms (to minimise potential for unblinding) - Women who had estrogens or progestins within 3 months. - Women treated with thyroid hormone who had not been taking a stable dose for at least 3 months and who did not have a normal thyroid stimulating hormone level. - Serious illness (MI within 6 months, congestive heart failure, stroke, transient ischemic attack) or contraindications to estrogen, including prior breast/endometrial cancer. - Inability to adhere to placebos for 28 days after the third screening visit. Laboratory exclusions included BP ≥ 160 mm/Hg systolic or 95 mmHg diastolic.		once a day. Active drugs and placebo prepared in identical forms. Statistical methods: Intention to treat. General mixed linear models fitted using restricted maximum likelihood and evaluated using F tests, t-tests used to assess pairwise treatment differences. For BP, treatment effects were assessed by rates of change based on linear models. Follow-up: 3 years	changes (95% CI) Systolic BP (mmHg): Placebo: 1.2 [-0.1, 2.6] CEE only: 0.5 [-0.7 , 1.8] CEE+MPA (cyc*): 0.7 [-0.6, 2.1] CEE+MPA (con**): 1.8 [0.6, 3.0] CEE+MP (cyc): 0.1 [-1.0 , 1.1] Diastolic BP (mmHg): Placebo: 0.0 [-0.9, 0.9] CEE only: -0.7 [-1.5, 0.1] CEE+MPA(cyc): -1.0 [-1.8, -0.1] CEE+MPA(con): 0.2 [-0.5, 0.9] CEE+MP(cyc): -0.6 [-1.3, 0.0] *= cyclic administration (days 1 - 12 of each month) **= administered daily for 1 month	for missing data - Yes Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Yes D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - Unclear Level of bias: low Other information
Full citation Weiner,M.G., Barnhart,K., Xie,D., Tannen,R.L., Hormone	Sample size N= 26,536 (aged 50-79) Characteristics	Interventions HRT (Conjugated estrogens 0.625 mg/d PO, Norgestrel 150 µg PO)	Details Setting: The UK General Practice Research Database (GRPD) study Methods: -HRT exposure: all women aged 50-79	Results Adjusted HRs (95%CI) By age < 55 yr old (n=50756): MI: 0.90 (0.69-1.17)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the

Study details	Participants				Interventions	Methods	Outcomes and Results	Comments
therapy and coronary heart disease in young women, Menopause, 15, 86-93, 2008 Ref Id 230653 Country/ies where the study was carried out UK Study type Prospective study Aim of the study Given the similarity between the UK General Practice Research Database (GPRD) study of older women and the WHI RCT, the GPRD methodology was used to study a cohort of younger women. Study dates 1990-April 1999 Source of funding Not reported		**Women > 55 yr old**		**Women <55 yr old**		and treated with any estrogen-containing preparation during the recruitment interval were identified -Potential unexposed women were age matched to this exposed group using a computer-generated random-number selection program Statistical analysis: -Cox proportional hazard analysis with multiple imputations for missing data on BP, BMI, and smoking and use of the same confounders; -In addition, a propensity score analysis, in which virtually all baseline data were considered potential confounders, was used to determine an overall adjusted HR by combining the HRs of the five quintiles. Follow-up: 9-yr	Stroke: 1.46 (1.11-1.92) Breast cancer: 1.46 (1.24-1.69) Death: 0.79 (0.67-0.93) Among women with no previous HT use (n=41701): MI: 0.86 (0.62-1.20) Stroke: 1.51 (1.09-2.09) Breast cancer: 1.43 (1.20-1.71) Death: 0.84 (0.69-1.02)	comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-No A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-No Level of risk-Low B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk: N/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length
		HRT use	Non-HRT use					
	Age in years	59.2	59.8					
	BMI, mean kg/m2	25.1	26.4					
	BMI >30, %	11.4	19.8					
	Hypertension, %	13.5	15.5					
	Smoker							
	Past, %	34.5	34.4					
	Current, %	20.3	24.1					
	Diabetes, %	1.5	2.7					
	High chol, %	6.9	4.6					
	Previous MI, %	0.26	0.85					
	Previous CVA, %	0.26	0.67					
	HT use							
	Past, %	14.4	1.8					

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Current, % 39.6 0.1 Inclusion criteria Exposure: -Conjugated estrogens 0.625 mg/d PO -Norgestrel 150 µg PO Exclusion criteria -Hysterectomy -Acute MI, CVA, or TIA within 6 mo of entry (H/O: history of): -H/O breast or endometrial cancer -H/O maglignant melanoma -H/O other maligancies in the past 10 yr -Abnormal Pap smear, pelvic examination -Endometrial hyperplasia -H/O nontraumatic pulmonary embolus or DVT -Severe hypertension -Chronic hepatitis or cirrhosis -Corticosteroid, tamoxifen, or anticoagulant treatment at entry -Medical condition with predicted survival < 3 yrs -Condition inconsistent with study adherence Those taking other HT preparations other than the two above				of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes D.2 The study used a precise definition of outcome-No D.3 A valid and reliable method was used to determine the outcome-No (how outcome was ascertained was not clearly reported) D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/a D.5 Investigators were kept 'blind' to other important confounding and prognostic

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					factors-N/a Level of bias: Unclear Indirectness Does the study match the review protocol in terms of: Population: Yes Outcome: Yes Indirectness: Some Other information -The amount of missing data on potential confoudners was much greater in the unexposed than exposed group, and the risk profile for cardiovascular disease was higher in the unexposed group. -USE of HT before the start of the study was substantially greater in the exposed than unexposed gorup; however, the subset without any HT exsposure in the year before study start exhibited findings similar to those of the overall cohort, suggesting that previous HT use did not greatly influence the results.

H.8.3 Development of type 2 diabetes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Full citation Manson,J.E., Rimm,E.B., Colditz,G.A., Willett,W.C., Nathan,D.M., Arky,R.A., Rosner,B., Hennekens,C.H., Speizer,F.E.,	Sample size 21,028 participants who were postmenopausal and free from diagnosed diabetes mellitus, CHD, stroke and cancer in 1976, as well as who subsequently became postmenopausal during the follow-up period. Characteristics Hormone use, n	Interventions HRT use -broken down into: Never, past, current use	Details Consent Not applicable Setting Survey carried out through mailed questionnaires Methods -Mailed questionnaire survey among	Results non-insulin-dependent diabetes (NIDDM), RR (95% CI) BY HRT use category: Never: 1.0 (reference group) past: 1.07 (0.93-1.23) Current: 0.80 (0.67-0.96) Analysis restricted to women	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Stampfer,M.J., A prospective study of postmenopausal estrogen therapy and subsequent incidence of non-insulin-dependent diabetes mellitus, Annals of Epidemiology, 2, 665-673, 1992 Ref Id 229840 Country/ies where the study was carried out US Study type Prospective study Aim of the study To examine prospectively the association between postmenopausal estrogen therapy and subsequent incidence of clinical NIDDM among postmenopausal women followed up for up to 12 years in the Nurses' Health Study. Study dates 1976 to 1988 Source of funding Research grant from the NIH, US.	Never: 9761 past: 3953 Current: 7314 Total: 21,028 Age in years, mean (SD) Never: 50.9 (3.5) past: 50.4 (4.3) Current: 48.6 (5.2) BMI, mean (SD) Never: 24.6 (4.4) past: 24.3 (4.2) Current: 23.7 (3.7) Family history of diabetes in percentages, % Never: 16.1 past: 17.8 Current: 17.4 Inclusion criteria Not reported Exclusion criteria -Women reporting a diagnosis of diabetes before 1976 -Women with insulin-dependent (type 1) diabetes, defined as confirmed diabetes and 1) continuous insulin therapy begun within 1 year of diabetes diagnosis, plus 2) ketonuria (more than trace) on at least two occasions or hospitalization for ketoacidosis. -women classified as having gestational diabetes only		registered nurses in the US (the Nurse's Health Study cohort was established in 1976 when 121,700 female registered nurse, aged 30 to 55 years and residing in one of 11 US states, responded to mailed questionnaries regarding their medical history, exogenous hormone use, and life-style). -Baseline questionnaries mailed in 1976 elicited information about a previous diagnosis of DM and other major illnesses, as well as age, height, weight, menopausal status, and use of postmenopausal hormones -In 1976, women were asked whether they had used hormone supplements following menopause and, if so, the duration of use. Biennial follow-up questionnaires from 1978 to 1988 updated information on hormone use -Women reporting DM, CHD, stroke, or cancer on previous questionnaires were excluded from subsequent follow-up -Incidence of diabetes was confirmed if at least one of the following was reported: one or more classic symptoms (thirst, polyuria, weight loss, hunger, etc) plus fasting plasma glucose level of at least 140 mg/dL or random plasma glucose level of at least 200 mg/dL; or 2) at least two elevated plasma glucose levels on different occasions (fasting >= 140mg/dL and/or random >= 200 mg/dL and/or glucose level >= 200 mg/dL at >= 2 hrs on oral glucose tolerance testing) in the absence of symptoms; or 3) treatment with hypoglcemic medication (insulin or oral hypoglycemic agent). Statistic methods -Incidence rates for NIDDM during the 12 years of follow-up were computed according to postmenopausal hormone use at baseline in 1976 and updated by questionnaire every 2 years -Rate ratios (RR) were computed as the rate of occurence of NIDDM in a specific	with natural menopause , RR (95%CI) Never: 1.0 (reference group) past: 1.08 (0.88-1.33) Current: 0.69 (0.48-0.99) By duration of current and past HRT use NIDDM, RR (95% CI), current use in years 0 yr: 1.0 (reference group) <1 yr: 0.84 (0.50-1.40) 1-3 yrs: 0.47 (0.31-0.69) 4-6 yrs: 0.89 (0.64-1.24) 7+ yrs: 1.08 (0.84-1.38) NIDDM, RR (95% CI), past use in years 0 yr: 1.0 (reference group) <1 yr: 0.86 (0.67-1.12) 1-3 yrs: 1.05 (0.85-1.29) 4-6 yrs: 1.29 (0.97-1.71) 7+ yrs: 1.13 (0.84-1.52) By type of postmenopausal hormone, RR (95% CI) Never use: 1.0 (reference group) Premarin only (conjugated estrogens): 0.86 (0.69-1.08) Other (combination conjugated estrogens and progesterone, progesteron alone, and miscellaneous categories of postmenopausal hormones): 0.65 (0.42-0.99) Unknow: 0.90 (0.37-2.16) (Follow-up from 1978-1988 when information on type of Hormon was available) By dose of paremarin (conjugated estrogens), RR (95% CI) Never use: 1.0 (Reference group) ≤ 0.3mg daily: 0.90 (0.52-1.58) 0.625 mg daily: 0.56 (0.38-	allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-No A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Unclear (only age, BMI, family history of DM were reported) Level of risk-High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-Not reported B.2 Participants receiving care were kept 'blind' to treatment allocation-N/A B.3 Individuals administering care were kept 'blind' to treatment allocation-N/A Level of risk: Moderate C. Attrition bias (systematic differences between the comparison groups with respect to loss of

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
			category of HRT use, divided by the incidence rate in never users of postmenopausal hormones (confounders controlled for were age and BMI, 12 yrs follow-up time) -proportional hazards models were used to evaluate the effects of postmenpausal estrogen therapy, age, BMI, family history of diabetes, past oral contraceptive hormone use, smoking, hypertension, high serum cholesterol level, parental history of myocardial infarction at age 60 years or younger, and time period in relation to the risk of diabetes Follow-up 12 yrs	0.83) 1.25mg daily: 1.16 (0.82-1.64) >1.25mg daily: 0.35 (0.05-2.37) (Follow-up from 1980-1988 when information on dose of Hormon was available)	participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?- About 7.2% were lost to follow up C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-Yes C.3a For how many participants in each group were no outcome data available?- not reported in each group, follow-up rate of the whole cohort was high (92.8%) and comparable across categories of hormone use; C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)- Yes Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up- Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					D.2 The study used a precise definition of outcome- Yes D.3 A valid and reliable method was used to determine the outcome- Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/A D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/A Level of bias: Low
Full citation de Lauzon-Guillain,B., Fournier,A., Fabre,A., Simon,N., Mesrine,S., Boutron-Ruault,M.C., Balkau,B., Clavel-Chapelon,F., Menopausal hormone therapy and new-onset diabetes in the French Etude Epidemiologique de Femmes de la Mutuelle Generale de l'Education Nationale (E3N) cohort, Diabetologia, 52, 2092-2100, 2009 Ref Id 203247 Country/ies where the study was carried out France Study type Cohort study	Sample size 63,624 (64% of the original 98,998 subjects enrolled in 1990) Characteristics Participants, n By MHT use -Non-user: 18,230 -User: 45,394 By route of oestrogen administration -Oral: 11,263 -Transdermal/cutaneous: 25740 -Other/unknow: 8,391 By type of MHT -Oestrogen alone: 4,656 -Oestrogen + progestagen: 30,905 -Other/unknown: 9,833 Age in years at start of follow-up, mean (SD) By MHT use -Non-user: 57.0 (5.5) -User: 54.8 (4.7) By route of oestrogen administration -Oral: 53.6 (4.1) -Transdermal/cutaneous: 54.5 (4.3) -Other/unknow: 57.1 (5.4) By type of MHT -Oestrogen alone: 54.8 (5.1) -Oestrogen + progestagen: 54 (4.1) -Other/unknown: 56.9 (5.4)	Interventions MHT use, stratified by -duration of use -MHT user type (current, past, unknown) -route of oestrogen administration	Details Consent All women signed an informed consent Setting survey by follow-up questionnaires Methods -In 1990 and at follow-up (1992,1993,1995,1997,2000,2002 and 2005), women completed self-administered questionnaires -cases of diabetes were identified through self-reporting or drug-reimbursement record linkage, and further validated Statistical methods -the association between MHT use and new-onset diabetes was investigated by Cox regression analysis (HR, 95% CI) -confounders adjusted for: age, age at menarche (<13 yrs, ≥13yrs), parity (nullparous/parous), breastfeeding, age at menopause, type of menopause, family history of diabetes, physical activity in 1993, alcohol intake, total energy intake exclusive of alcohol, education level, baseline cholesterol level, hypertension, smoking, and baseline BMI, and BMI as a time-dependent variable	Results New onset diabetes, n/N, adjusted HR (95%CI): According to MHT use: MHT non-users (Reference group): 518/18,230; 1 MHT users: 702/45,394; 0.75 (95%CI: 0.66-0.85) According to duration of MHT use 0-2 yrs: 144/7,300; 0.75 (95%CI: 0.61-0.91) 2-5 yrs: 202/11,868; 0.84 (95%CI: 0.70-1.00) >5 yrs: 294/23,460; 0.70 (955CI: 0.59-0.82) Unknown duration: 62/2,766; 0.75 (95%CI: 0.57-1.00) p value for homogeneity in duration of use: 0.32 According to MHT user type Current use: 422/7,657; 0.78 (95%CI: 0.65-0.89) past use (> 1 yr before): 244/35,384; 0.90 (95%CI: 0.76-1.07) Unknow recency: 36/2,353; 0.99 (95%CI: 0.70-1.39)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)- No A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-No Level of risk-Moderate

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Aim of the study To evaluate the influence of menopausal hormone therapies (MHTs), and their type and route of administration, on the risk of new-onset diabetes in a cohort of postmenopausal French women. Study dates 1990-2005 Source of funding MGEN; European Community; French League against Cancer (LNCC);	Age in years at menopause, mean (SD) By MHT use --Non-user: 50.7 (3.9) -User: 50.1 (3.7) By route of oestrogen administration -Oral: 50.2 (3.6) -Transdermal/cutaneous: 50.2 (3.5) -Other/unknow: 49.7(4.4) By type of MHT -Oestrogen alone: 49.4 (4.4) -Oestrogen + progestagen: 50.3 (3.3) -Other/unknown: 49.8 (4.4) Parent with diabetes, n(%) By MHT use --Non-user: 5,341 (29.3%) -User: 10,597 (23.3%) By route of oestrogen administration -Oral: 2,537 (22.5%) -Transdermal/cutaneous: 5,964 (23.2%) -Other/unknow: 2,096 (25%) By type of MHT -Oestrogen alone: 1,144 (24.6) -Oestrogen + progestagen: 7,073 (22.9%) -Other/unknown: 2,380 (24.2%) Smoker, n(%) By MHT use --Non-user: 5,282 (29%) -User: 14,536 (32%) By route of oestrogen administration -Oral: 3,778 (33.5%) -Transdermal/cutaneous: 8,120 (31.5%) -Other/unknow: 2,638 (31.4%) By type of MHT -Oestrogen alone: 1,469 (31.6%) -Oestrogen + progestagen: 9,964 (32.2%) -Other/unknown: 3,103 (31.6%) BMI (Kg/m2), mean (SD) By MHT use		Follow-up 14 yrs	p value in homogeneity in recency: 0.09 According to route of oestrogen administration oral: 121/11,263; 0.61 (95%CI: 0.50-0.76) cutaneous: 425/25,740; 0.78 (95%CI: 0.67-0.90) other route: 49/2,533; 0.76 (95%CI: 0.56-1.04) unknown route: 103/5,858; 0.73 (95%CI: 0.59-0.92) p value for homogeneity in oral and cutaneous routes: 0.031	B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/A B.2 Participants receiving care were kept 'blind' to treatment allocation-N/A B.3 Individuals administering care were kept 'blind' to treatment allocation-N/A Level of risk: Moderate C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-No C.2a How many participants did not complete treatment in each group?- About 36% were excluded or lost during follow up C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-Not clear (loss to follow-up across groups not reported) C.3a For how many

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	--Non-user: 23.8 (3.8) -User: 22.9 (3.1) By route of oestrogen administration -Oral: 22.7 (3.0) -Transdermal/cutaneous: 23.0 (3.1) -Other/unknow: 23.1 (3.1) By type of MHT -Oestrogen alone: 23.4 (3.4) -Oestrogen + progestagen: 22.8 (3.0) -Other/unknown: 23.1 (3.1) Alcohol intake (g/day), mean (SD) By MHT use --Non-user: 10.5 (14.1) -User: 11.5 (14.1) By route of oestrogen administration -Oral: 11.9 (14.5) -Transdermal/cutaneous: 11.4 (13.9) -Other/unknow: 11.2 (14) By type of MHT -Oestrogen alone: 10.9 (13.5) -Oestrogen + progestagen: 11.6 (14.2) -Other/unknown: 11.3 (14.1) Inclusion criteria The prospective cohort included 98,995 women living in France, aged 40-65 ys in 1990, who were covered by the national insurance plan for teachers and co-workers. Exclusion criteria women -who did not repsond to a dietary history questionnaire -had miscoding of dietary questionnaire -did not agree to be followed -reported unreasonable energy				participants in each group were no outcome data available?- not reported C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)- Not clear Level of risk: High D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up- Yes D.2 The study used a precise definition of outcome- Yes D.3 A valid and reliable method was used to determine the outcome- Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/A D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/A Level of bias: Low

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	intake -reported no health status information -with non-validated diabetes status -who have been diagnosed diabetes before the dietary questionnaire or first report of menopause -with no follow-up -with missing data on MHT use				
Full citation Bonds,D.E., Lasser,N., Qi,L., Brzyski,R., Caan,B., Heiss,G., Limacher,M.C., Liu,J.H., Mason,E., Oberman,A., O'Sullivan,M.J., Phillips,L.S., Prineas,R.J., Tinker,L., The effect of conjugated equine oestrogen on diabetes incidence: The Women's Health Initiative randomised trial, Diabetologia, 49, 459-468, 2006 Ref Id 203608 Country/ies where the study was carried out US Study type double masked RCT Aim of the study To determine the effect of conjugated equine oestrogen (CEO) alone on the incidence of diabetes mellitus in postmenopausal women, results of the WHI oestrogen-	Sample size N=9,712 (reported no diagnosis of diabetes at baseline) (CEO group, n= 4,806 Placebo group, n= 4,906) Characteristics Age group in at screen (yrs), n (%), p value: -CEO (N=4,806) 50-59: 1,504 (31.3) 60-69: 2,138 (44.5) 70-79: 1,164 (24.2) -Placebo (N=4,906) 50-59: 1,542 (31.4) 60-69: 2,203 (44.9) 70-79: 1,161 (23.7) P=0.81 Hormone use, n (%), p value: -CEO (N= 4,806) Never: 2,459 (51.2) Past user: 1,716 (35.7) Current user: 630 (13.1) -Placebo (N=4,906) Never: 2,477 (50.5) Past user: 1,759 (35.9) Current user: 667 (13.6) p= 0.40 Duration of prior hormone use in years, n (%), p value: -CEO (N=4,806) < 5: 1,241 (52.9) 5-10: 435 (18.5) > 10: 670 (28.6) -Placebo (N= 4,906) < 5: 1,278 (52.7) 5-10: 1,759 (35.9)	Interventions CEO versus placebo	Details Consent Informed consent was obtained from participants Setting 40 clinical centres throughout the US Randomisation method A randomised permuted block algorithm, stratified by clinical centre site and age, was developed at the WHI Clinical Coordinating Centre and implemented locally through a distributed study database. Concealment of allocation -details not reported in this study Comparability of intervention groups at baseline The two groups were comparable in terms of age, weight, and comorbidity at baseline, there were no significantly differences between them Blinding -Participants, clinical staff, investigators and outcomes adjudicators were blinded to treatment assignment. -Neither the clinic gynaecologist nor any of the staff or investigators involved with the clinical care of the participants was involved with study outcomes assessment Statistical methods -Baseline variables were compared with either X2 or Fisher's exact tests for categorical variables or two-sample t tests	Results Self-reported diabetes incidence, n/N, HR (95%CI): CEO: 397/4,787 (1.16%); Placebo: 455/4,887 (1.30%); CEO vs Placebo: 0.88 (0.77-1.01) (after 7.1 yrs follow-up) By age group (age at screening), n (%), HR (95%CI): 50-59: CEO: 131 (1.17%); Placebo: 159 (1.39%); CEO vs placebo: 0.83 (0.66-1.05) 60-69: CEO: 181 (1.20%); Placebo: 198 (1.28%); CEO vs placebo: 0.94 (0.77-1.15) 70-79: CEO: 85 (1.06%); Placebo: 98 (1.22%); CEO vs placebo: 0.85 (0.64-1.14) (age subgroup models were only stratified by randomisation status in the low-fat-diet trial which participants of this trial also took part in)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes (WHI trial, details not reported in this study) A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Unclear (participants were blinded at baseline allocation, but during the trial some participants should be able to realise which group they had been assigned to when the HRT took effects on their menopausal symptoms) B3 - Were individuals administering care blinded to treatment allocation-Yes Level of bias: Unclear C Attrition bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
alone trial were analysed. Study dates (7.1 yrs follow-up) Source of funding The National Heart, Lung and Blood Institute, US Department of Health and Human Services	> 10: 667 (13.6) p=0.83 BMI (kg/m2),n (%), p value -CEO, (N=4,806) <25: 1,073 (22.4) 25-30: 1,677 (35.1) >30: 2,032 (42.5) -Placebo (N=4,906) <25: 1,046 (21.5) 25-30: 1,749 (35.9) >30: 2,079 (42.7) p=0.47 Smoking, n(%), p value: -CEO (N=4,806) Never: 2,480 (52.1) Past: 1,776 (37.3) Current: 500 (10.5) -Placebo (N=4,906) Never: 2,430 (50.1) Past: 1,891 (39.0) Current: 528 (10.9) p=0.14 Alcohol use > 1 drink/week, n/N (%), p value: CEO: 1,437/4,806 (30.0) Placebo: 1,514/4,906 (31.1) p=0.27 Lipid-lowering medication use, n (%), p value: CEO: 393 (8.2) Placebo: 403 (8.2) p=0.95 Aspirin use, n (%), p value: CEO: 914 (19.0) Placebo: 943 (19.2) p=0.80 History of myocardial infarction, n (%), p value: CEO: 132 (2.7) Placebo: 132 (2.7) p=0.87		for continous variables; -The incidence of diabetes was assessed using a Cox proportional hazards model, stratified by age -Intention to treat analysis Not reported Follow-up -7.1 years		C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Yes Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Unclear D2 - Were outcomes defined precisely - Unclear D3 - Was a valid and reliable method used to assess outcome - No D4 - Were investigators blinded to intervention - Yes D5 - Were investigators blinded to confounding factors - No (not all possible for this outcome, e.g., BMI could be a confounder) Level of bias: Unclear Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information -There was no confirmation of the self-reported diabetes diagnosis with medical records, nor was it possible to determine the incidence of undiagnosed diabetes.

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	History of angina, n (%), p value: CEO: 241 (5.0) Placebo: 234 (4.8) p=0.58 History of stroke, n (%), p value: CEO: 61 (1.3) Placebo: 71 (1.4) p=0.45 History of DVT or PE, n (%), p value: CEO: 79 (1.6) Placebo: 77 (1.6) p=0.77 Inclusion criteria -women of 50-79 yrs of age; had undergone hysterectomy Exclusion criteria -women with a history of previous breast cancer, any cancer within the previous 10 yrs except non-melanoma skin cancer, current use of corticosteroids, anticoagulants, tamoxifen or other selective oestrogen receptor modifiers (SERMs), and triglyeerides > 4.56 mmol/l. A history of venous thromboembolism was added as an exclusion criterion in 1997. -women who were unwilling to discontinue the use of HRT were also excluded, and a 3-month washout period was required for women who were current hormone users at the initial screening visit. -self-reported diabetes at baseline				
Full citation Zhang,Y., Howard,B.V., Cowan,L.D., Yeh,J., Schaefer,C.F., Wild,R.A., Wang,W., Lee,E.T., The effect of estrogen use on levels of glucose	Sample size n=857 (the current study was based on women who were both nondiabetic and postmenopausal at the baseline examination and who completed a second examination an average 4 yr later) -there were 2,703 women at baseline, among them, 2,109 were	Interventions HRT	Details Consent: Not reported Setting: Survey carried out among vlunteers from 13 Indian tribes/communities Methods:	Results By HRT user category (Past and never users vs current users of estrogen): Adjusted Odds Ratio (95%CI) for fasting glucose >=7.0mmol/l (126 mg/dl) Past and never users: 1.0 (reference group)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
and insulin and the risk of type 2 diabetes in american Indian postmenopausal women : the strong heart study, Diabetes Care, 25, 500-504, 2002 Ref Id 301383 Country/ies where the study was carried out US Study type Longitudinal study Aim of the study To examine the association between estrogen use and levels of insulin and glucose as well as well the effect of estrogen use on the risk of type 2 diabetes. Study dates 1989-1992 (Baseline examination) to 1993-1995 (the second examination) Source of funding The National Heart, Lung, and Blood Institute	postmenopausal). Characteristics No detailed data reported; The study reported that - "compared with never users (of HRT), past and current users were more educated; had a higher hysterectomy rate; had lower American Indian heritage, gravity, and parity; were more active; and had a lower WHR"; "compared with past users and never users, current users wer younger, with a lower BMI" Inclusion criteria -Postmenopausal women who did not have a history of diabetes, did not take diabetic medication, and had a fasting plasma glucose level <7.0 mmol/l (126 mg/dl) and a 2-h post challenge glucose level < 11.1 mmol/l (200 mg/dl) at the baseline examination were eligible for the present analysis; Exclusion criteria -Women who had inconsistent information on estrogen use at the baseline and the 2nd examination.		-Three definitions of diabetes have been used in the analysis: one is based on a fasting plasma glucose >=7.0mmol/l or 2-h glucose level >=11.1 mmol/l; one is based on fasting glucose >=11.1 mmol/l. The third one is based on elevated 2-h postchallenge glucose level (>=11.1 mmol/l; 75-g oral glucose tolerance test) -The cohort for analysis was divided into three groups: never users (n=604), past users (n=119), and current users (n=134) of estrogen, based on women's use at the bsaeline examination. Never users had never used estrogen; Past users had used estrogen but were not taking estrogen at baseline; Current users were using estrogen at the time of the baseline examination. (Estrogen use was ascertained by interview and was confirmed by examination of pills and prescription broughts brought to the visit) Statistic methods: -Logistic regression was used to assess the independent contributions of estrogen use and duration of estrogen use to the incidence of type 2 diabetes, adjusted for covariates which remained in the final selected logisc model after step-wise selections. -Covairates included in the model included BMI, waist-to-hip ratio, American Indian Heritage, SHS centre, education etc. Follow-up: 4 yrs	Current users: 0.48 (0.20-1.14) Covariates adjusted for in the model: BMI, waist to hip ratio, American Indian heritage Adjusted odds ratio (95%CI) for fasting glucose >=7.0mmol/l or 2-h glucose >=11.1mmol/l Past and never users: 1.0 (reference group) Current users: 1.11 (0.62-1.97) Covariates adjusted for in the model: BIM, American Indian Heritage, SHS centre Adjusted odds ratio (95%CI) for 2-h glucose >=11.1 mmol/l (200mg/dl): Past and never users: 1.0 (reference group) Current users: 1.58 (0.81-3.1) Covariates adjusted for the model: BMI, education (yrs), family history, hysterectomy status By duration of estrogen use (n=134; duration as a continouse variable) Adjusted Odds Ratio (95%CI): duration of estrogen use and the risk of of fasting glucose >=7.0mmol/l (126 mg/dl): 1.01 (0.9-1.12) Covariates: none Adjusted Odds Ratio (95%CI): duration of estrogen use and the risk of fasting glucose >=7.0mmol/l (126 mg/dl) or 2-h glucose >=11.1 mmol/l: 1.10 (1.01-1.18) Covariates: BMI, hysterectomy status (yes or no) The risk of T2DM increased by 10% f or each year of current estrogen use;	A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-Unclear A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-No Level of risk-Low B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/A B.2 Participants receiving care were kept 'blind' to treatment allocation-N/A B.3 Individuals administering care were kept 'blind' to treatment allocation-N/A Level of risk: n/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				Adjusted Odds Ratio (95%CI): duration of estrogen use and the risk of 2-h glucose >=11.1 mmol/l: 1.10 (1.01-1.19) Covariates: BMI, hysterectomy status (yes or no) The risk of T2DM increased by 10% f or each year of current estrogen use;	followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?- n/a C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-n/a C.3a For how many participants in each group were no outcome data available?- n/a C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)- N/a Level of risk: low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up- Unclear (4 yrs) D.2 The study used a precise definition of outcome- Yes D.3 A valid and reliable method was used to determine the outcome- Yes D.4 Investigators were kept

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					'blind' to participants' exposure to the intervention-N/A D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/A Level of bias: moderate Other information -Participants were volunteers from American Indian Tribes -Estrogen use was ascertained by interview and was confirmed by examination of pills and prescriptions brought to the visit, while whether women using estrogen were also taking a progestogen agent was not ascertained at the baseline.

H.8.4 Type 2 diabetes management – control of blood sugar

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Full citation Kernohan,A.F., Sattar,N., Hilditch,T., Cleland,S.J., Small,M., Lumsden,M.A., Connell,J.M., Petrie,J.R., Effects of low-dose continuous combined hormone replacement therapy on glucose homeostasis and markers of cardiovascular risk in women with type 2 diabetes, Clinical Endocrinology, 66, 27-34, 2007 Ref Id 202962 Country/ies where the study was carried out	Sample size N=30 randomised (n=15 in HRT group, n=15 in placebo group) N=28 analysed (n=14 in HRT group, n=14 in placebo group Characteristics HRT/placebo Mean age, year (SD) 62.2 (5.8)/62.1 (3.8) Years since menopause, mean year (SD) 13.0 (1.4)/14.0 (4.7) Weight, mean kg (SD) 82.0 (16.4)/80.5 (20.3) BMI, mean kg/m2 (SD) 34.0 (6.3)/33.0 (8.9) Hypertension, %	Interventions Oral 17β oestradiol (1mg) and norethisterone (0.5mg) Matching placebo tablet	Details Setting Diabetes centres of North Glasgow University Hospitals NHS trust Randomisation method Participants were randomly assigned to HRT or placebo in blocks of six, stratified for presence or absence of hypertension, method not clearly reported Statistical methods Baseline and after treatment data were reported as means and SDs, or median and	Results HbA1c Reported as mean percentage (SD) HRT/placebo Baseline: 7.4 (1.1)/ 7.6 (0.9) 3 months treatment (final): 7.4 (1.3)/ 8.1 (1.1) P= 0.11 Fasting glucose Reported as mean mmol (SD) HRT/placebo Baseline: 8.1	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes, reported, but method of randomisation not reported A2 - Was there adequate concealment - Unclear, methods of concealment not reported A3 - Were groups comparable at baseline - Yes Level of bias: Moderate

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
UK Study type Randomised, double-blind placebo controlled trial Aim of the study To assess the effects on glucose homeostasis and cardiovascular risk factors of continuous oral 17b oestradiol (1mg) and norethisterone (0.5mg) in postmenopausal women with type 2 diabetes Study dates Not reported Source of funding British Heart Foundation	78.6/78.6 Mean number of antihypertensive drugs 1.6/1.9 Inclusion criteria Postmenopausal women, >1 year from last menstrual period Age <70 years and had type 2 diabetes according to national guidelines Women on stable oral anti-diabetic therapy and/or diet for at least 3 months prior to entry and regular medication was not changed during the study Exclusion criteria Poor glycaemic control, (glycated haemoglobin (HbA1c) >10%), severe hypertriglyceridaemia (>70 mmol/l), serum creatinine >120µmol/l, blood pressure >160/110 mmHg, HRT use within 2 years, insulin therapy, or other standard contraindication to HRT		interquartile range for parameters not exhibiting normal distribution Results after treatment expressed as mean (or median) and as percentage change from baseline. Between group differences assessed by two-sample t test or Mann-Whitney U test P value of <0.05 was considered significant Pearson's correlation coefficients (r) were calculated using Minitab A priori power calculation based on previous studies in subjects with type 2 diabetes estimated that a sample size of n=15 in each group would give 80% power to detect a 10-15% change in EGP, fasting plasma glucose, HbA1c and total cholesterol (α=0.05, two-sided)	(1.9)/8.5 (2.1) 3 months treatment (final): 7.2 (1.9)/ 8.9 (1.6) P=0.02	B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Yes B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Low C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Unclear, not reported Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Yes D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Unclear, not reported D5 - Were investigators blinded to confounding factors - Unclear, not reported Level of bias: Moderate Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no
Full citation Darko,D.A., Dornhorst,A.,	Sample size N=41 recruited, N=33 completed	Interventions Three cycles were taken	Details Randomisation method	Results HbA1c	Limitations NICE guidelines manual 2012:

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Kennedy,G., Mandeno,R.C., Seed,M., Glycaemic control and plasma lipoproteins in menopausal women with Type 2 diabetes treated with oral and transdermal combined hormone replacement therapy, Diabetes Research and Clinical Practice, 54, 157-164, 2001 Ref Id 203073 Country/ies where the study was carried out UK Study type Randomised open parallel study Aim of the study To compare the effect of a fixed combination of an oestrogen (17b-oestradiol) with cyclical progestogen (norethisterone) on glycaemic control, plasma lipoproteins and haemostatic factors in women with type 2 diabetes Study dates Not reported Source of funding Coronary Thrombosis Trust at Charing Cross Hospital	study Characteristics HRT (oral)/HRT (transdermal)/control BMI, mean kg/m2 (SD) 28.2 (6.8)/33.5 (8.0)/33.5 (9.1) Fasting plasma glucose, mean mmol (SD) 8.2 (1.6)/11.2 (5.5)/8.7 (3.9) HbA1c, mean percentage (SD) 7.4 (1.4)/7.8 (1.7)/7.4 (1.2) Inclusion criteria Postmenopausal women (cessation of menses for >1 year in the presence of climacteric symptoms, or biochemically, follicular stimulating hormone >25IU with serum oestradiol <100pmol-1) with type 2 diabetes (diagnosed after age of 40 years and treated with either diet alone or diet and oral hypoglycaemic agents) recruited from outpatient clinics from hospital or from local GPs Exclusion criteria Women taking insulin or lipid lowering therapy within the last 6 months or HRT within the last 3 months Women consuming >20 units of alcohol a week or had significant medical co-morbidity	continuously for 12 weeks Oral preparation: 28 day cycle of 17β oestradiol 2mg for 16 days followed by norethisterone 1 mg for 12 days Transdermal preparation: patch releasing 17β oestradiol 50µg per 24 hours transdermally for 14 days followed by a second patch releasing both 17β oestradiol 50µg and norethisterone 170µg per 24 hours for 14 days Control group: no treatment	At visit one, participants were randomised and allocated to one of the three study groups, and biochemical, demographic and clinical data was recorded At visit two (at 12 weeks), all measurements were repeated Samples were obtained at start of HRT use and also at the second visit for future analysis Statistical methods All values were expressed as mean (SD) ANOVA was used to analyse paired data and P value of <0.05 as significant	Reported as mean percentage (SD) Oral HRT/transdermal HRT/control At 12 weeks: 6.8 (1.2)/ 7.8 (1.8)/ 7.4 (1.6) Control P value at baseline and 12 weeks: not significant Oral HRT P value at baseline and 12 weeks: <0.005 Transdermal HRT P value at baseline and 12 weeks: not significant Fasting plasma glucose Reported as mean mmol/l (SD) Oral HRT/transdermal HRT/control 8.4 (2.4)/ 10.7 (3.0)/ 9.2 (4.2) P value for all treatment groups at baseline and 12 weeks: not significant	Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes, randomisation by drawing of lots into one of three treatment groups A2 - Was there adequate concealment - No. The study was an open parallel study A3 - Were groups comparable at baseline - Unclear, not reported Level of bias: High B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- No. The study was an open trial B3 - Were individuals administering care blinded to treatment allocation- No, the study was an open trial Level of bias: High C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Unclear, not reported Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Yes D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Unclear, not reported D5 - Were investigators blinded to confounding factors - Unclear, not reported Level of bias: High Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no
Full citation Ferrara,A., Karter,A.J., Ackerson,L.M., Liu,J.Y., Selby,J.V., Northern California Kaiser Permanente Diabetes Registry., Hormone replacement therapy is associated with better glycemic control in women with type 2 diabetes: The Northern California Kaiser Permanente Diabetes Registry, Diabetes Care, 24, 1144-1150, 2001 Ref Id 323433 Country/ies where the study was carried out USA Study type Cross sectional study of cohort from the Kaiser Permanente Diabetes Registry Aim of the study To examine whether HbA1c levels varied by current HRT among women with type 2 diabetes Study dates Diabetes registry was started in	Sample size N=15,435 women with T2DM Characteristics Characteristics during 2 year study period HRT/no HRT Mean age, years (SD) 61.2 (7.6)/65.9 (8.8) BMI, mean kg/m2 (SD) 30.7 (6.5)/30.4 (6.8) HbA1c, mean %, SD 8.1 (1.7)/8.4 (2.0) Ethinicity, % Non-Hispanic: 60.9/53.2 African-American: 9.4/15.0 Hispanic: 12.9/12.3 Asian/Pacific Islanders: 9.4/11.5 Other/unknown: 7.4/8.0 Therapy, % Diet: 13.9/12.2 OHA: 51.5/53.4 Insulin: 34.6/34.4 Diabetes duration, % <5 years: 38.0/36.2 5-9 years: 23.9/21.6 ≥10 years: 38.1/42.2 SMBG practice, % Never: 19.9/26.4 <1/week: 18.2/17.1	Interventions Current HRT (oestrogen and/or progestin) No current HRT	Details Setting Kaiser Permanente Medical Care Programme of Northern California, group practice pre-paid health plan Statistical methods Two sample t test was used to compare current HRT and no current HRT use for continuous variables and X2 for categorical variables HbA1c and BMI means were age-adjusted (ANOVA) Generalised estimating equation model was constructed to assess association between HRT and HbA1c level (after taking into account clustering of patients characteristics treated by the same physician and adjusting for age, ethnicity, education, BMI, hypoglycaemic therapy, diabetes duration, SMBG,	Results Age adjusted mean (SE) HbA1c (%) during 2 year study HRT/no HRT 7.9 (0.03)/8.5 (0.02) P=0.0001 Regression coefficient for HRT in predicting HbA1c: HRT use/HbA1c: β coefficient= -0.475 (SE 0.04), P=0.0001	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies 1 Objectives 1.1 Are the objectives of the study clearly stated? Yes 2 Design 2.1 Is the research design clearly specified and appropriate for the research aims? Yes 2.2 Were the subjects recruited in an acceptable way? Yes 2.3 Was the sample representative of a defined population? Yes Risk of bias: Low 3 Measurement and observation 3.1 Is it clear what was measured, how it was measured and what the outcomes were? Yes 3.2 Are the measurements valid? Partly. Duration of HRT use prior to study was not reported. 3.3 Was the setting for data

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
1993, patients included in study from 1995 to 1997 Source of funding American Heart Association and SmithKline Beecham Pharmaceuticals	≥1/week: 61.8/56.5 Smoking,% Current: 9.7/8.9 Former: 36.0/31.6 Never: 54.3/59.5 Exercise, % 52.4/46.9 Inclusion criteria Women aged ≥50 years age who were members of the diabetes registry, Women who filled an HRT prescription, women who were continuously enrolled in the health plan (without gaps), confirmed type 2 diabetes, HbA1c measured at least once Exclusion criteria Women not continuously enrolled in the health plan, women who stated that they did not have diabetes on the survey, women with type 1 diabetes or unclassified for type of diabetes		and exercise Confounders were included in the GEE models if their inclusion resulted in appreciable changes in the HRT coefficient or if the variable was shown by previous scientific publications to be associated with both outcome and exposure All P values were for two-tailed tests with statistical significance defined as P≤0.05		collection justified? Yes 3.4 Were all important outcomes/results considered? Partly. Only HbA1c was considered, not blood glucose levels. Risk of bias: Low 4 Analysis 4.1 Are tables/graphs adequately labelled and understandable? Yes 4.2 Are the authors' choice and use of statistical methods appropriate, if employed? Yes, they want to see the correlation of HbA1c in women currently taking HRT 4.3 Is there an in-depth description of the analysis process? Yes 4.4 Are sufficient data presented to support the findings? Partly. This is a cross-sectional study, but the HbA1c results are reported at an unknown time point during the 2 year study Risk of bias: Low 5 Discussion 5.1 Are the results discussed in relation to existing knowledge on the subject and study objectives? Yes, other studies are also discussed 5.2 Can the results be generalised? Yes Risk of bias: Low Indirectness Does the study match the review protocol in terms of; Population:Yes Outcome: Yes Indirectness: None
Full citation McKenzie,J., Jaap,A.J.,	Sample size n=50	Interventions Active medication (1 mg	Details Setting	Results Glycaemic control	Limitations NICE guidelines manual 2012:

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Gallacher,S., Kelly,A., Crawford,L., Greer,I.A., Rumley,A., Petrie,J.R., Lowe,G.D., Paterson,K., Sattar,N., Metabolic, inflammatory and haemostatic effects of a low-dose continuous combined HRT in women with type 2 diabetes: potentially safer with respect to vascular risk?, Clinical Endocrinology, 59, 682-689, 2003 Ref Id 203263 Country/ies where the study was carried out Scotland, UK Study type Double-blind, randomized placebo-controlled trial. Aim of the study To assess the metabolic effects of a continuous combined HRT containing 1 mg oestradiol and 0·5 mg norethisterone or matching placebo Study dates Study only stated women with type 2 diabetes aged under 70 years of age were recruited between December 1998 to September 2000 Source of funding Not reported	Active n=25 randomized/22 completed trial/19 demonstrated compliance Placebo n=25 randomized/23 completed trial Characteristics Active/placebo Mean age, year (SD): 60.7 (5.5)/61.3 (4.8) BMI (kg/m2) (SD): 30.5 (6.5)/29.8(5.61) Waist circumference,cm (SD): 93.9 (11.3)/93.7 (13.6) Years postmenopausal (SD): 14.6 (8.5)/14.2(6.3) Inclusion criteria -women with type 2 diabetes aged under 70 years of age -clinically and biochemically postmenopausal, i.e. at least 1 year since last menses and a FSH concentration of greater than 20 IU/l. Menopause could be either natural or surgically induced Exclusion criteria -poor glycaemic control -severe hypertriglyceridaemia (> 10 mmol/ l) -moderate to severe hypertension (systolic > 160 mmHg, diastolic > 110 mmHg) -renal impairement (serum creatinine greater than twice the upper limit of normal range) -liver disease (serum transaminases and bilirubin greater than twice the upper limit of normal range) -established cardiovascular, cerebrovascular, or peripheral vascular disease -subjects with either a personal history of – or first-degree relative with – breast cancer	oestradiol plus 0·5 mg norethisterone) or identical placebo daily for 6 months	General diabetic clinics in Glasgow Hospitals Randomisation method In blocks of four using computer-generated number Statistical methods Mean differences in changes from baseline between the two treatment groups were compared using the unpaired t-test; 95% confidence interval for change in active group data relative to change in control group data are presented. Adjustment for baseline concentrations was made by linear regression. Baseline data are presented as mean and SD or median and interquartile range (IQR) for parameters exhibiting skewed distribution.	-HbA1c (%) Reported as mean (SD) Active/Placebo Baseline: 10.2 (1.8) / 10.2 (1.3) Mean change: -0.37/0.22 Mean difference for change active relative to change placebo (95%CI) / p: -0.59 (-1.45 to 0.27)/ 0.17 -Blood glucose Reported as Glycaemia glucose (mmol/l), mean (SD) Active/Placebo Baseline: 12.4 (4.2) / 11.3 (3.2) Mean change: -1.74/0.42 Mean difference for change active relative to change placebo (95%CI) / p: -2.16 (-4.06 to -0.28)/ 0.026 Health related quality of life Not reported Mortality Not reported Adverse events (complications resulting from diabetes) Not reported	Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Unclear, methods of concealment not reported A3 - Were groups comparable at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Unclear, methods of blinding not reported B3 - Were individuals administering care blinded to treatment allocation- Unclear, methods of blinding not reported Level of bias: High C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Yes C3 - Were groups comparable for missing data - Unclear, not reported Level of bias: Low D Detection bias D1 - Was follow-up appropriate length - Yes D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Unclear, not reported

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					D4 - Were investigators blinded to intervention - Unclear, not reported D5 - Were investigators blinded to confounding factors - Unclear, not reported Level of bias: High Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no Other information Study does not report the sample size analysed for each treatment outcome.
Full citation Perera,M., Sattar,N., Petrie,J.R., Hillier,C., Small,M., Connell,J.M.C., Lowe,G.D.O., Lumsden,M.A., The effects of transdermal estradiol in combination with oral norethisterone on lipoproteins, coagulation, and endothelial markers in postmenopausal women with type 2 diabetes: A randomized, placebo-controlled study, Journal of Clinical Endocrinology and Metabolism, 86, 1140-1143, 2001 Ref Id 311478 Country/ies where the study was carried out Scotland, UK Study type Randomised placebo-controlled trial Aim of the study To assess the effect of transdermal oestradiol (80-µg patches) in combination with	Sample size Continuous combined HRT [transdermal oestradiol (80-µg patches) in combination with oral norethisterone (1 mg daily; n = 22] or identical placebos (n = 21) Characteristics HRT/Placebo Mean age, year (SD): 61.2 (3.7)/62.8(4.9) Duration of diabetes, median year (ranges): 2 (1-20)/4 (1-14) Mean BMI (kg/m2), (SD): 31 (7.8)/31.6(4.3) Inclusion criteria Not reported Exclusion criteria Not reported	Interventions Continuous transdermal oestradiol (80-µg patches) in combination with oral norethisterone (1 mg daily) or identical placebos for 6 months	Details Setting Diabetes Centers in Glasgow Randomisation method Not reported Statistical methods The adequacy of the randomization process was checked by comparing the baseline values in the two groups (unpaired t test or Mann-Whitney U test as appropriate). Differences in changes from baseline between the two treatment groups were compared using t tests if the changes were normally distributed. Baseline values in parameters of interest and in age, smoking status, and diabetes duration were adjusted for using linear regression. Correlation	Results Glycaemic control -HbA1c (%): Reported as mean (SD) HRT/placebo Baseline: 6.6(1.3)/6.4(1.3) 6 months (final): 6.6(1.2)/6.8(1.6) p value change (differences in changes from baseline between groups): 0.35 -Blood glucose: Reported as mean fasting blood glucose (mmol/L) (SD) HRT/placebo Baseline: 8.1 (1.7)/8.5(2.7) 6 months (final): 8.6(2.5)/8.6(2.6) p value change	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Unclear, not reported A2 - Was there adequate concealment - Unclear, not reported A3 - Were groups comparable at baseline - Yes Level of bias: High B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Unclear, not reported B3 - Were individuals administering care blinded to treatment allocation- Unclear, not reported Level of bias: High

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
continuous oral norethisterone (1 mg daily) on conventional anthropometric parameters, lipoprotein concentrations, coagulation (fibrinogen, factor VII, and fibrin D dimers), and endothelial factors [tissue plasminogen activator (t-PA), and von Willebrand factor (vWF)] in postmenopausal women with type 2 diabetes. Study dates Not reported Source of funding Not reported			analysis was performed using the Spearman rank correlation. Data are presented as the mean and SD for normally distributed data and as the median and range for data with a nonparametric distribution.	(differences in changes from baseline between groups): 0.57 Health related quality of life Not reported Mortality Not reported Adverse effects (complications resulting from diabetes) Not reported	C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - Unclear, not reported C3 - Were groups comparable for missing data - Unclear, not reported Level of bias: High D Detection bias D1 - Was follow-up appropriate length - Yes D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Unclear, not reported D4 - Were investigators blinded to intervention - Unclear, not reported D5 - Were investigators blinded to confounding factors - Unclear, not reported Level of bias: High Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no
Full citation Sutherland, W. H., Manning, P. J., de Jong, S. A., Allum, A. R., Jones, S. D., Williams, S. M., Hormone-replacement therapy increases serum paraoxonase arylesterase activity in diabetic postmenopausal women, Metabolism: Clinical & ExperimentalMetabolism, 50, 319-24	Sample size N=47 HRT group=28 Placebo group=19 Characteristics Age (years, mean, SD): 64±8 BMI (kg/mg2, mean, SD): 32.3±5.7 HbA1c (%, mean, SD): 7.5±1.9	Interventions HRT: conjugated equine oestrogen (Premarin 0.625mg) and medroxyprogesterone acetate (Provera 2.5 mg) combined in a single capsule Placebo (single capsule identical to HRT)	Details Treatment: Written informed consent obtained from participants HRT was titrated upward over a 4-week period to minimise acute side effects. At end of 4 weeks women were taking either HRT or placebo treatment (1 capsule/daily)Patients	Results Glycaemic control -HbA1c (%) Reported as mean (SD) HRT/Placebo Baseline: 7.3 (1.6) / 7.8 (2.3) 6 months: 7.9 (1.6) / 8.5 (2.1)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - Yes A2 - Was there adequate concealment - Yes A3 - Were groups comparable

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Ref Id 325988 Country/ies where the study was carried out New Zealand Study type Randomised placebo-controlled, cross-over study Aim of the study To test the effect of HRT on plasma concentrations of lipids, lipoproteins, and apolipoproteins in postmenopausal diabetic women Study dates Recruitment of participants ended in 1996 Source of funding Health Research Council of New Zealand	Fasting glucose (mmol, mean, SD): 10.2±3.9 Inclusion criteria Postmenopausal women with type 2 diabetes (postmenopausal defined as absence of menstrual periods for more than 2 years Cardiovascular disease was present in 14% of the diabetic women Exclusion criteria Poorly controlled diabetes (glycosylated [HbA1c] >10%) Concomitant significant medical disorder Contraindications to HRT (history of breast or endometrial cancer) Undiagnosed vaginal bleeding Uncontrolled hypertension Severe liver dysfunction or they met the current national criteria for lipid-lowering therapy with statins		were seen at 3 month intervals to check for adverse effects (reaction to medication, suffered serious concurrent illness contraindicating HRT or receiving lipid-lowering therapy), compliance (capsule counting: defined as tablet count >80%), record body weight, measure blood lipids Laboratory methods: Plasma gluocose was measured enzymatically by automated methods using a commercial kit HbA1c was measured using a commercial kit Statistics: Values expressed as means±SD Multivariate linear regression analysis with final (6 month) and baseline values to test for differences between HRT and placebo treatment Paired t test was used to estimate treatment effect if significant difference was observed between HRT and placebo treatments Two-tailed tests of significance were used, and a P value of <0.05 was considered statistically significant	-Blood glucose Reported as glucose (mmol/l), mean (SD) HRT/Placebo Baseline: 9.97 (3.30) / 10.66 (4.69) 6 months: 8.37 (2.1) / 10.38 (4.1)	at baseline - Yes Level of bias: Low B Performance bias B1 - Did groups get same level of care - Yes B2 - Were participants blinded to treatment allocation- Unclear, methods of blinding not reported B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: Moderate C Attrition bias C1 - Was follow-up equal for both groups - Yes C2 - Were groups comparable for dropout - No. 13 participants (40%) in the placebo group dropped out compared with 1 in the HRT group C3 - Were groups comparable for missing data - Unclear, not reported Level of bias: High D Detection bias D1 - Was follow-up appropriate length - Yes D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - Unclear, not reported D5 - Were investigators blinded to confounding factors - Unclear, not reported Level of bias: High Indirectness

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: no indirectness

H.8.5 Breast cancer

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Full citation Jernstrom,H., Bendahl,P.O., Lidfeldt,J., Nerbrand,C., Agardh,C.D., Samsioe,G., A prospective study of different types of hormone replacement therapy use and the risk of subsequent breast cancer: The women's health in the Lund area (WHILA) study (Sweden), Cancer Causes and Control, 14, 673-680, 2003 Ref Id 300068 Country/ies where the study was carried out Sweden Study type Prospective Cohort Study Aim of the study To establish whether breast cancer risk depends on the type of HRT formula. Study dates 1995-2000 Source of funding Skane County Council Foundation for Research and Development	Sample size 6,586 participants Characteristics Women aged 50-64 years Mean (SD) age at study entry, years Cases: 56.5 (2.9) Controls: 56.4 (3.0) Mean (SD) age at menarche, years Cases: 13.4 (1.4) Controls: 13.4 (1.4) Body weight (SD), kg Cases: 68.2 (11.5) Controls: 66.9 (9.0) Inclusion criteria Women with no reported history of breast cancer Exclusion criteria Women with previous breast cancer	Interventions Continuous combined estrogen plus progestin (CCEP, 0.625 mg of conjugated equine estrogens and 2.5 mg of medroxyprogesterone acetate) Other HRT formulas	Details All women born between December, 2, 1935 and December 1, 1945 were invited for health assessment. Women matched to the South Swedish tumor registry to obtain data on newly diagnosed breast cancers	Results 101 breast cancer cases disgnosed Median follow-up: 4.1 years Hazard Ratios for Breast Cancer With Use of Different Types of HRT CCEP exclusively: 3.3 (1.9-5.6) CCEP and other HRT: 2.8 (1.4-5.5) Other HRT only: 1.5 (0.84-2.50) Adjusted for baseline age	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: No A3. The groups were comparable at baseline, including all major confounding and prognostic factors: Yes Level of risk: High risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? N/A C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: High risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important confounding and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Overall risk of bias: High

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Other information
Full citation Beral,V., Million Women,Study Collaborators, Breast cancer and hormone-replacement therapy in the Million Women Study.[Erratum appears in Lancet. 2003 Oct 4;362(9390):1160], Lancet, 362, 419-427, 2003 Ref Id 300217 Country/ies where the study was carried out UK Study type Prospective Cohort Study Aim of the study To investigate the effects of specific types of HRT on incident and fatal breast cancer. Study dates 1996-2001 Source of funding Cancer Research UK NHS Breast Screening Programme Medical Research Council	Sample size 1,084,110 women Characteristics Average age at recruitment: 55.9 years Inclusion criteria 1. Women aged 50-64 years Exclusion criteria Women with cancer registered before recruitment, except if they had a previous non-melanoma skin cancer	Interventions Estrogen Estrogen-Progestagen Tibolone	Details Women recruited from a screening programme Women classified according to their repported use of HRT, menopausal status, and other relevant factors Endpoints included incident invasive breast cancer and deaths due to breast cancer	Results Average follow-up for cancer incidence: 2.6 years Average follow-up for cancer mortality: 4.1 years Incident breast cancer: 9,364 Breast cancer deaths: 637 Relative Risk of Incident Breast Cancer in Relation to Recency of Use of HRT Never use: ref Current users: 1.66 (1.60-1.72) Past users: 1.01 (0.95-1.08) Last use < 5 years previously: 1.04 (0.95-1.12) Last use 5-9 years previously: 1.01 (0.88-1.16) Last use ≥ 10 years previously: 0.90 (0.72-1.12) Relative Risk of Incident Breast Cancer in Relation to Type of HRT Never use: ref Estrogen: 1.30 (1.22-1.38) Estrogen-Progestagen: 2.00 (1.91-2.09) Tibolone: 1.45 (1.25-1.67) Relative Risk of Incident Breast Cancer in Relation to Duration and Type of HRT Estrogen < 1 year: 0.81 (0.55-1.20) 1-4 years: 1.25 (1.10-1.41) 5-9 years: 1.32 (1.20-1.46) ≥ 10 years: 1.37 (1.22-1.54) Estrogen+Progestin < 1 year: 1.45 (1.19-1.78) 1-4 years: 1.74 (1.60-1.89) 5-9 years: 2.17 (2.03-2.33) ≥ 10 years: 2.31 (2.08-2.56)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: Yes Level of risk: Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				Relative Risk of Fatal Breast Cancer in Relation to Use of HRT at Baseline Never use: ref Current users: 1.22 (1.05-1.41) Past users: 1.05 (0.85-1.29) Confounders adjusted for: Age Time since menopause Parity and age at first birth Family history of breast cancer BMI Region Deprivation Index	care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? Not reported C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: High risk of

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important confounding and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Overall risk of bias: High Other information
Full citation Fournier,A., Berrino,F., Riboli,E., Avenel,V., Clavel-Chapelon,F., Breast cancer risk in relation to different types of hormone replacement therapy in the E3N-EPIC cohort, International	Sample size 54,548 participants Characteristics Women born between 1925 and 1950 Mean age at inclusion: 52.8 years	Interventions HRT: Estrogens Progestogens	Details Women were part of a health insurance scheme HRT categorised according to type and route of administration Follow-up started either at	Results Mean duration of follow-up: 5.8 years 948 primary cancers diagnosed Relative Risk of Breast Cancer	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Journal of Cancer, 114, 448-454, 2005 Ref Id 300256 Country/ies where the study was carried out France Study type Prospective Cohort Study Aim of the study Effects of different types of HRT and routes of administration on breast cancer risk Study dates 1990-1992 Source of funding French League Against Cancer The European Community 3M Company etc	Mean duration of HRT use: 2.8 years Inclusion criteria Postmenopausal women Exclusion criteria Women who only replied the baseline questionnaires Women who had reported a cancer other than a basal cell carcinoma before the start of followup In situ cancer during followup Women who had reported using HRT before the year preceeding the start of follow-up		the date of return of the baseline questionnaire for women already postmenopausal at that time, or at date of menopause as reported in the follow-up questionnaire	for Ever Users Never users: ref Ever uses: 1.2 (1.1-1.4) Relative Risk of Breast Cancer by Type of HRT Never users: ref Estrogens alone: 1.1 (0.8-1.6) Estrogens + Progestogens: 1.3 (1.1-1.5) Relative Risk of Breast Cancer by Duration of HRT Use Never users: ref < 2 years: 1.2 (1.0-1.5) 2-4 years: 1.2 (1.0-1.5) ≥ 4 years: 1.2 (0.9-1.6) Fully adjusted analyses.	between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: Yes Level of risk: Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? NR C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: High risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important confounding and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Overall risk of bias: High Other information
Full citation Sourander,L., Rajala,T., Raiha,I., Makinen,J., Erkkola,R., Helenius,H., Cardiovascular and cancer morbidity and mortality and sudden cardiac death in postmenopausal women on oestrogen replacement therapy (ERT).[Erratum appears in Lancet 1999 Jan 23;353(9149):330], Lancet, 352, 1965-1969, 1998 Ref Id 230428 Country/ies where the study was carried out	Sample size 7944 postmenopausal women Characteristics Significant differences between never users and current users of ERT in age, social class, BMI, hypertension, and diabetes Mean age at baseline, years Never users: 60.9 Former users: 61.0 Current users: 59.9 Mean BMI at baseline,	Interventions ERT	Details Women born between 1923-1930 were asked to participate in a free mammography screening for breast cancer Validated questionnaire filled in by participants with the help of trained nurses Participants divided into three groups by their estrogen use: never users, former users, and current users Data linked to Finnish Cancer Registry	Results Current users of ERT: 988 Former usrs of ERT: 757 Cases of breast cancer: 97 Relative Risk of Breast Cancer According to Use of ERT Never users: ref Past users: 0.94 (0.47-1.90) Current users: 0.57 (0.27-1.20) Ever users: 0.74 (0.45-1.24) Multivariate adjusted.	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Finland Study type Prospective Cohort Study Aim of the study To analyse the relation between estrogen replacement therapy (ERT) and breast cancer Study dates 1987-1995 Source of funding Samfundet Folkhalsan	kg/m^2 Never users: 26.7 Former users: 26.1 Current users: 25.5 Inclusion criteria Postmenopausal women Exclusion criteria NR		Participants were followed up from 1987 to 1995. Multivariate analyses used Cox proportional hazards model		expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: No Level of risk: High risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? NR C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: High risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants'

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important confounding and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Overall risk of bias: High Other information Estimates for Ever users calculated by fixed effects analysis of current and past users
Full citation Schuurman,A.G., van den Brandt,P.A., Goldbohm,R.A., Exogenous hormone use and the risk of postmenopausal breast cancer: results from The Netherlands Cohort Study, Cancer Causes and Control, 6, 416-424, 1995 Ref Id 300595 Country/ies where the study was carried out Netherlands Study type Prospective Cohort Study (Case-cohort) Aim of the study Association between use of exogenous hormones (oral contraceptives or HRT) in relation to postmenopausal	Sample size 62,573 women Characteristics Women aged 55-69 years Inclusion criteria Cohort members who completed a mailed self-administered questionnaire Exclusion criteria Incident breast cancer cases with in situ carcinoma Women who reported as history of cancer at baseline, other than skin cancer	Interventions HRT	Details Case-cohort approach used Follow-up status of sub-cohort was 100% Follow-up of cancer incidence was at least 95%	Results 3.3 years of follow-up 553 breast cancer cases Mean duration of HRT use was 3.6 years in subcohort 3.4 years in cases Relative Risk of Breast Cancer by HRT in Women Aged < 50 Years Never use: ref Ever use: 1.4 (0.8-2.4) Confounders adjusted for: Age Benign breast disease Mother with breast cancer Sisters with breast cancer Parity Age at first birth Age at menarche Age at menopause	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
breast cancer incidence Study dates 1986 Source of funding Dutch Cancer Society				Induced menopause Education Current cigarette smoking BMI Alcohol use Energy consumption Use of oral contraceptives	comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: Yes Level of risk: Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? See details

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					section C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important confounding

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Overall risk of bias: Low
Full citation Folsom,A.R., Mink,P.J., Sellers,T.A., Hong,C.P., Zheng,W., Potter,J.D., Hormonal replacement therapy and morbidity and mortality in a prospective study of postmenopausal women, American Journal of Public Health, 85, 1128-1132, 1995 Ref Id 229297 Country/ies where the study was carried out USA Study type Prospective Cohort Study Aim of the study The association of HRT with mortality and incidence of multiple diseases including breast cancer. Study dates 1986-1991 Source of funding National Cancer Insitute	Sample size 41,070 postmenopausal women Characteristics Age 55-59 years Never users of HRT: 36% Former users of HRT: 29% Current users of HRT: 46% Current smokers Never users of HRT: 9% Former users of HRT: 10% Current users of HRT: 8% Body mass index > 28 kg/m² Never users of HRT: 37% Former users of HRT: 35% Current users of HRT: 27% Inclusion criteria Women aged 55 through 69 years who had a valid Iowa drivers' license in 1985. Postmenopausal women with HRT data Exclusion criteria Women with baseline cancer	Interventions HRT	Details Cancer incidence detected through the State Health Registry of Iowa HRT categorized as current use, former use, and never use Relative risks determined by Cox proportional hazards regression	Results Follow-up: 6 years Incident Breast Cancer: 468 Relative Risk of Breast Cancer Incidence by HRT Never use: ref Ever use: 1.24 (0.99-1.56) Relative Risk of Breast Cancer Incidence by Duration of HRT Never use: ref ≤ 5 years: 1.45 (1.03-2.06) > 5 years: 1.21 (0.92-1.60) Multivariate adjusted.	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: Yes Level of risk: Low risk of bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? N/A C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: High risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important confounding and prognostic factors: N/A Level of bias: High risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Outcomes: Yes Indirectness: No serious Overall risk of bias: High
Full citation Lando,J.F., Heck,K.E., Brett,K.M., Hormone replacement therapy and breast cancer risk in a nationally representative cohort, American Journal of Preventive Medicine, 17, 176-180, 1999 Ref Id 300686 Country/ies where the study was carried out USA Study type Prospective Cohort Study Aim of the study Assess the association of postmenopausal HRT with risk of breast cancer. Study dates 1971-1974 Source of funding National Center for Health Statistics National Institute of Aging National Cancer Institute	Sample size 5,761 Characteristics Mean age at study entry: 55.5 years Never used HRT: 3564 Ever used HRT: 2197 Family history of breast cancer: 9.4% Inclusion criteria 1. Women older than 55 years 2. Menopause status based on report that menstrual periods had stopped entirely Exclusion criteria Breast cancer diagnosed prior to baseline	Interventions Postmenopausal HRT	Details 1. Multi-stage stratified probability sample of the non-institutionalized population of the US 2. Age at menopause defined either as the age at which menstruation naturally ceased entirely, the age at bilateral oophorectomy, or the assigned age of 49 for women who had a hysterectomy without bilateral oophorectomy.	Results Mean follow-up: 12.7 years Incident cases of breast cancer: 219 Relative Risk of Cancer by HRT Use Never use: reference Ever use: 0.80 (0.60-1.10) Relative Risk of Cancer by Duration of HRT Use Never use: reference < 3 years: 0.9 (0.6-1.4) 3-9 years: 0.5 (0.3-0.9) ≥ 10 years: 0.8 (0.5-1.3) Covariates adjusted for: Age Race Education Body mass index Age at first child Age at menopause Type of menopause Family history of breast cancer	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: Yes Level of risk: Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? 4.4% lost to follow-up C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): Yes C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					outcome data were not available): N/A Level of risk: High risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important confounding and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Overall risk of bias: High
Full citation Bakken,K., Alsaker,E., Eggen,A.E., Lund,E., Hormone replacement therapy and incidence of hormone-dependent cancers in the	Sample size 35,456 postmenopausal women 31,451 included in analyses Characteristics Women aged 45-64 years	Interventions HRT Estrogen Estrogen+Progestagen Estriol	Details 2 subsamples of the general population provided information on reproductive, lifestyle, and use of HRT and were	Results 624 incident breast cancer cases Relative Risk of Breast Cancer by Recency of HRT Use	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Norwegian Women and Cancer study, International Journal of Cancer, 112, 130-134, 2004 Ref Id 300704 Country/ies where the study was carried out Norway Study type Prospective Cohort Study Aim of the study Relation between use of HRT and risk of hormone-dependent cancers Study dates 1996-1998 Source of funding Community Pharmacy Foundation	Mean age: 53 years Mean BMI: 25 kg/m² Ever use of HRT was reported by 43.5% Majority of women use oral HRT preparations Inclusion criteria Postmenopausal women Age range 45-64 years Exclusion criteria NR		followed up for cancer incidence Follow-up information was based on linkage to the Cancer Registry of Norway Cox proportional hazards used for analyses	Never user: ref Ever user: 1.9 (1.5-2.5) Past user: 1.0 (0.6-1.6) Relative Risk of Breast Cancer by Duration of HRT Use Never user: ref 0-1 year: 1.4 (1.0-2.1) 2-4 years: 2.4 (1.6-2.9) 5-9 years: 2.2 (1.5-3.1) 10+ years: 2.2 (1.4-3.6) Relative Risk of Breast Cancer by Type of HRT Estrogen: 1.8 (1.1-2.9) Estrogen+Progestin: 2.5 (1.9-3.2) Relative Risk of Breast Cancer by Duration of HRT Use Estrogen < 5 years: 2.5 (1.4-4.5) ≥ 5 years: 1.0 (0.4-2.5) Estrogen+Progestin < 5 years: 2.3 (1.7-3.2) ≥ 5 years: 2.8 (2.0-4.0) Multivariate-adjusted	(systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: Yes Level of risk: Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? NR C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: High risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important confounding and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Overall risk of bias: High
Full citation Tjonneland,A., Christensen,J., Thomsen,B.L., Olsen,A., Overvad,K., Ewertz,M., Mellemkjaer,L., Hormone replacement therapy in relation to breast carcinoma incidence rate ratios: a prospective Danish cohort study, Cancer, 100, 2328-2337, 2004 Ref Id 300709 Country/ies where the study was carried out Denmark Study type Prospective Cohort Study	Sample size 23,618 postmenopausal women Characteristics Age at entry, years Never used: 57.2 Tried HRT: 57.5 Previously used: 59.0 Currentl use: 56.3 Median BMI, kg/m² Never used: 25.1 Tried HRT: 25.6 Previously used: 25.5 Currentl use: 24.4 Inclusion criteria Women aged 50-64 years	Interventions Unopposed estrogen Sequential estrogen plus progestin Continuous estrogen plus progestin	Details Participants completed a detailed, 192-item food frequency questionnaire Records were linked to Danish Cancer Registry Each cohort member was followed for breast cancer detection from the date of study entry	Results Breast cancer cases: 423 Median follow-up: 4.8 years Breast Cancer Incidence Rate Ratios Associated With HRT Use Never use: 1.00 Past use: 1.35 (0.90-2.02) Current use: 2.22 (1.80-2.75) Confounders adjusted for: Duration of schooling BMI Parity Number of births Age at birth of first child	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Aim of the study Relation between HRT and breast cancer in postmenopausal women Study dates 1993-1997 Source of funding Danish Cancer Society and the Europe Against Cancer Program	Exclusion criteria 1. Malignancy 2. Participants who did not respond to significant portions of lifestyle questionnaire 3. Premenopausal women 4. Women who reported a lifetime history of no menstruation 5. Women for whom data on duration of HRT use or time since cessation were unavailable			History of benign breast tumour surgery Alcohol consumption	outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: Yes Level of risk: Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					follow-up): Yes C2a. How many participants did not complete treatment in each group? N/A C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					intervention: N/A D5. Investigators were kept 'blind' to other important confounding and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Overall risk of bias: Low
Full citation Ewertz,M., Mellemkjaer,L., Poulsen,A.H., Friis,S., Sorensen,H.T., Pedersen,L., McLaughlin,J.K., Olsen,J.H., Hormone use for menopausal symptoms and risk of breast cancer. A Danish cohort study, British Journal of Cancer, 92, 1293-1297, 2005 Ref Id 300739 Country/ies where the study was carried out Denmark Study type Prospective Cohort Study Aim of the study Risk of developing breast cancer in relation to HRT Study dates 1989-2002 Source of funding NR	Sample size 78,380 women Characteristics Women aged 40-67 years Inclusion criteria Women aged 40-66 years at any time during study period and resident in study area Women who had received at least two prescriptions for systemic HRT Exclusion criteria Women who had a cancer diagnosis before 1989 of before age 40 years Women who received prescriptions for sex hormones other than those used in HRT including androgens, durung 1989-2002, and women who had used systemic HRT before the age of 40 years	Interventions HRT	Details Women were linked to the Danish Cancer Registry Prescription of nonsystemic HRT was not judged as HRT exposure Followup for breast cancer started on 1 January 1989 or at 40 years	Results 1462 cases of breast cancer Mean follow-up of 10 years Relative Risk of Incident Breast Cancer for HRT in Women Aged < 65 Years Never use: ref Ever use: 1.33 (1.19-1.49) Confounders adjusted for: Calendar period Number of children Age at first child	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					confounding and prognostic factors: Yes Level of risk: Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? NR C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: High risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important confounding and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Overall risk of bias: High Other information Relative risks for breast cancer in those aged < 65 years was calculated by meta-analysing provided estimates for different age-groups
Full citation Hedblad,B., Merlo,J., Manjer,J., Engstrom,G., Berglund,G., Janzon,L., Incidence of cardiovascular disease, cancer and death in postmenopausal women affirming use of hormone replacement therapy, Scandinavian Journal of Public Health, 30, 12-19, 2002 Ref Id 229444 Country/ies where the study was carried out Sweden Study type Prospective Cohort Study Aim of the study Incidence of breast cancer in relation to use of HRT Study dates 1974-1992 Source of funding Government grants	Sample size 5,862 per- or postmenopausal women Characteristics Women usng HRT had longer general education and a greater proportion of them had non-manual jobs. were leaner and the percentage with diabetes, hypertension, or hyperlipidemia was smaller Inclusion criteria Peri- or postmenopausal women Exclusion criteria NR	Interventions HRT	Details Self-administered questionnaire to assess smoking habits, medical history, parity, menopause, and use of HRT Incidence of cancer based on data linkage to National Cancer Registry and the National Cause of Death Registry Cox proportional hazards model used to estimate the influence of HRT on incidence of cancer	Results 9 years of follow-up 136 incident breast cancer cases Relative Risk of Breast Cancer in Relation to HRT Never use: ref Ever use: 1.52 (1.01-2.28) Multivariate adjusted.	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: No Level of risk: High risk of bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? NR C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: High risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important confounding and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Outcomes: Yes Indirectness: No serious Overall risk of bias: High
Full citation Manjer,J., Malina,J., Berglund,G., Bondeson,L., Garne,J.P., Janzon,L., Increased incidence of small and well-differentiated breast tumours in post-menopausal women following hormone-replacement therapy, International Journal of Cancer, 92, 919-922, 2001 Ref Id 267698 Country/ies where the study was carried out Sweden Study type Prospective Cohort Aim of the study Assess whether HRT is associated with an increase risk of breast cancer Study dates 1974-1992 Source of funding NR	Sample size 5,865 postmenopausal women Characteristics Age at baseline, years HRT users: 53.8 Non-users: 54.1 BMI at baseline, kg/m² HRT users: 24.3 Non-users: 25.2 Inclusion criteria Postmenopausal women Exclusion criteria Women diagnosed with invasive breast cancer at baseline	Interventions HRT	Details Cohort of postmenopausal women followed for an average of 9.8 years for invasive breast cancer Data linked to Swedish Cancer Registry Cox proportional hazards used to estimate relative risk of breast cancer	Results Number of breast cancer cases HRT users: 106 Non-users: 35 Relative Risk of Breast Cancer in Relation to HRT Exposure 1.66 (1.12-2.45) Multivariate-adjusted	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: Yes Level of risk: Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? NR C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					available): N/A Level of risk: High risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important confounding and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Overall risk of bias: High
Full citation Stahlberg,C., Pedersen,A.T., Lynge,E., Andersen,Z.J., Keiding,N., Hundrup,Y.A., Obel,E.B., Ottesen,B., Increased risk of breast cancer following different regimens of hormone	Sample size 10,874 women Characteristics Women above the age of 44 years 25.1% were current users of HRT	Interventions HRT Estrogen Estrogen+Progesterone	Details Women identified through membership of the Danish Nurses Organization Breast cancer cases were identified by linkage to the Danish Cancer Registry	Results Mean duration of HRT use: 7.2 years 244 breast cancer cases during followup. Mean duration of follow-up: 6.34 years	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
replacement therapy frequently used in Europe, International Journal of Cancer, 109, 721-727, 2004 Ref Id 300784 Country/ies where the study was carried out Denmark Study type Prospective Cohort Study Aim of the study To investigate whether different treatment regimens influence risk of breast cancer differently. Study dates 1993-1999 Source of funding Danish Cancer Society	14.5% were past users 60.4% had never used HRT at baseline Inclusion criteria Danish postmenopausal nurses above the age of 44 years Exclusion criteria Breast cancer cases at baseline Other invasive cancers except for nonmelanoma skin cancer Women with missing information Premenopausal women Women with a surgical menopause Hysterectomized women		Women were considered postmenopausal if the menstrual bleeding had ceased, or they were bleeding while currently taking HRT	Relative Risk of Breast Cancer for HRT Never use: ref Past use: 1.16 (0.76-1.77) Current use: 2.42 (1.81-3.26) Current ≤ 1 year: 2.28 (1.26-3.15) Current 2-4 years: 1.84 (1.07-3.15) Current 5-9 years: 2.58 (1.64-4.05) Current 10-14 years: 3.08 (1.87-5.06) Current 15+ years: 2.56 (1.49-4.39) Relative Risk of Breast Cancer by Type of HRT Never use: ref Estrogen: 1.95 (1.15-3.32) Estrogen+Progesterone: 3.02 (1.80-5.05) Multivariate adjusted.	between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: Yes Level of risk: Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? N/A C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: High risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important confounding and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Overall risk of bias: High
Full citation Bakken,K., Fournier,A., Lund,E., Waaseth,M., Dumeaux,V., Clavel-Chapelon,F., Fabre,A., Hemon,B., Rinaldi,S., Chajes,V., Slimani,N., Allen,N.E., Reeves,G.K., Bingham,S., Khaw,K.T., Olsen,A., Tjonneland,A., Rodriguez,L., Sanchez,M.J., Etxezarreta,P.A., Ardanaz,E., Tormo,M.J., Peeters,P.H., Van,GilsC, Steffen,A., Schulz,M., Chang-Claude,J., Kaaks,R., Tumino,R., Gallo,V., Norat,T., Riboli,E., Panico,S., Masala,G., Gonzalez,C.A., Berrino,F., Menopausal hormone therapy	Sample size N=133,744 Characteristics Mean age at recruitment (y, SD): 58.1 Type of menopause (%): Artificial=6.7 Natural=93.3 BMI (kg/m2)(%): <18.5=1.7 18.5-25=51.2 25-30=32.9 Inclusion criteria Postmenopausal women at baseline Postmenopausal women who had undergone a bilateral ovariectomy or if	Interventions Oestrogen Oestrogen+progestin Tibolone Other/unknown	Details Study population: Multicentre study, 23 contributing centres in 10 European cities, participants mainly recruited from the general population with exception to Norway, Utrecht, France and Naples which included women only. Turin, Ragusa, and Spain=mostly from blood donors France=teachers Oxford=high proportion of health-conscious individuals	Results Breast cancer risk and type of HRT used at baseline (cases, RR and 95%CI): Current use of oestrogen only Reference=HRT never use Denmark: 68, RR 1.56 (1.17-2.09) France: 80, RR 1.32 (1.04-1.67) Germany: 50, RR 2.07 (1.42-3.00) Italy: 12, RR 1.09 (0.61-1.97) Norway: 17, RR 1.61 (0.90-2.88) Spain: 6, RR 1.25 (0.52-3.00) The Netherlands: 24, 1.48 (0.96-2.27)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study) -

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
and breast cancer risk: Impact of different treatments. The European Prospective Investigation into Cancer and Nutrition, International Journal of Cancer, 128, 144-156, 2011 Ref Id 300918 Country/ies where the study was carried out Denmark, France, Germany, Great Britain, Greece, Italy, Norway, Spain, Sweden, The Netherlands Study type Prospective cohort study Aim of the study To investigate the association of menopausal hormone therapy and the risk of breast cancer according to different hormones, regimens and routes of administration using data from the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort Study dates Recruitment =1992-1999 Follow-up started in mid-1990s to 2009 Source of funding Not reported	menseshad stopped since 12 months or more (unless due to hysterectomy) Women who were still menstruating and using exogenous hormones, women for whom menopause had been obscured by hysterectomy, and women with no information on number of menses over 12 months were considered menopausal if they were 55 years or older Exclusion criteria Women with prevalent cancer at any site at baseline Women with missing non-dietary questionnaire data Women from the Swedish and Greek cohorts excluded due to lack of data on hormone use Women from the Dutch centre excluded due to missing information on some reproductive adjustment variables Women who never menstruated Women with no information on hormone use (ever or current)		Utrecht and Florence= women attending mammographic screening programmes Study was based on 344,581 women Cancers identified by self-reports and registration Menopause status defined according to information on ovariectomy, hysterectomy, menstruation status, and exogenous hormone use Final analytical cohort =133,744 women from 8/10 participating countries Identification of breast cancer cases and follow-up: Population cancer registries (Denmark, Italy, the Netherlands, Norway, Spain, and United Kingdom) or active follow-up (France, Germany, health insurance records, cancer and pathology registries, contacts with next of kin) Mortality data=mortality registries at regional and national level Women followed-up from study start to first cancer diagnosis (except nonmelanoma skin cancer), death and emigration or until end of follow-up (2002 to 2005, depending on country) Identification of menopausal HT use: Country-specific questionnaire, ever and current use of HT, brand name, age at start and total duration of use,	UK: 49, RR 1.11 (0.80-1.54) Current use of oestrogen+progestin Reference =HRT never use Denmark: 207, RR 2.71 (2.23-3.28) France: 635, RR 1.48 (1.31-1.67) Germany: 110, RR 2.20 (1.60-3.01) Italy: 17, RR 1.60 (0.96-2.66) Norway: 90, RR 1.65 (1.10-2.46) Spain: 4, RR 0.51 (0.18-1.41) The Netherlands: 13, RR 1.58 (0.89-2.80) UK: 143, RR 1.88 (1.50-2.37) Breast cancer risk and total duration of HRT use for current users at baseline (cases, RR and 95%CI) in United Kingdom: Current use of oestrogen only Reference=HRT never use <1 yr use: 2, RR 0.36 (0.09-1.48) 1-3 yrs use: 6, RR 0.67 (0.30-1.53) 3-5 yrs use: 16, RR 1.81 (1.07-3.06) 5-10 yrs use: 15, RR 1.25 (0.73-2.13) >10 yrs use: 5, RR 0.80 (0.33-1.95) Current use of oestrogen+progestin Reference=HRT never use <1 yr use: 16, RR 1.23 (0.73-2.09) 1-3 yrs use: 45, RR 1.88 (1.33-2.66) 3-5 yrs use: 28, RR 1.60 (1.06-2.04) 5-10 yrs use:39, RR 2.46 (1.74-3.48) >10 yrs use: 6, RR 1.58 (0.70-3.58)	No A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders - Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors - yes Moderate risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - N/A B2. Participants receiving care were kept 'blind' to treatment allocation - N/A B3. Individuals administering care were kept 'blind' to treatment allocation - N/A Unclear/unknown risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - yes C2a. How many

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
			administration and regimen. For past HT users, time since last use not available Progestins grouped=Micronised progesterone, progesterone derived progestins and testosterone-derived progestins For combination HT, Oestrogen+progestin was sequential (oestrogen with added progestin 10-14 d a month) or fixed continuous (oestrogen+progestin daily) Statistical analysis: Risk ratios and 95%CI for breast cancer estimated using Cox proportional hazards models, adjusting for age, type of menopause, BMI, ever use of oral contraceptives, number of full term pregnancies, age at first full-term pregnancy, age at menarche, and alcohol consumption Sensitivity analysis to investigate duration of HT use or age at menopause were confounders in comparison of two regimens regarding breast cancer risk	Breast cancer risk in current users, type of HRT, and regimen (cases, RR and 95%CI) in United Kingdom: Type of oestrogen only Reference=HRT never use Oestradiol compounds: 20/22,303, RR 1.08 (0.67-1.74), P=0.48 CEE: 25/22,303, RR 1.16 (0.76-1.78), P=0.09 Progestin component in sequential regimen Reference=HRT never use Testosterone derivatives: 126/22,303, RR 1.08 (1.48-2.38), P=0.15 Regimen of HRT Sequential HRT: 131/22,303, RR 1.91 (1.51-2.42), P=0.09 Fixed continuous HRT: 11/22,303, RR 1.78 (0.97-3.29), P=0.07 Adjusted for age, type of menopause, BMI, number of full term pregnancies, age at full term pregnancy, age at menarche, alcohol consumption	participants did not complete treatment in each group? - N/A C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - N/A C3a. For how many participants in each group were no outcome data available?- Swedish, Dutch and Greek centres were excluded due to lack of data and missing data C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					intervention - N/A D.5 Investigators were kept 'blind' to other important confounding and prognostic factors - N/A Low risk of bias.
Full citation Manson,J.E., Chlebowski,R.T., Stefanick,M.L., Aragaki,A.K., Rossouw,J.E., Prentice,R.L., Anderson,G., Howard,B.V., Thomson,C.A., Lacroix,A.Z., Wactawski-Wende,J., Jackson,R.D., Limacher,M., Margolis,K.L., Wassertheil-Smoller,S., Beresford,S.A., Cauley,J.A., Eaton,C.B., Gass,M., Hsia,J., Johnson,K.C., Kooperberg,C., Kuller,L.H., Lewis,C.E., Liu,S., Martin,L.W., Ockene,J.K., O'Sullivan,M.J., Powell,L.H., Simon,M.S., Van,HornL, Vitolins,M.Z., Wallace,R.B., Menopausal hormone therapy and health outcomes during the intervention and extended poststopping phases of the women's health initiative randomized trials, JAMA - Journal of the American Medical Association, 310, 1353-1368, 2013 Ref Id 300923 Country/ies where the study was carried out USA Study type Randomized Controlled Trial (Estrogen+Progestin vs. placebo component) Aim of the study Menopausal hormone therapy and risks and benefits for chronic disease prevention Study dates	Sample size 16,608 with uterus randomized to Conjugated Equine Estrogens plus medroxyprogesterone acetate (CEE+MPA) or placebo Characteristics Age (SD) at screening, years CEE+MPA: 63.2 (7.1) Placebo: 63.3 (7.1) Baseline characteristics were well balanced according to demographic and disease risk factors. Inclusion criteria Data extracted in a previous publication. Exclusion criteria Data extracted in a previous publication.	Interventions CEE+MPA Placebo	Details Intervention phase of the CEE+MPA trial ended after a median of 5.6 years due to increased breast cancer risk and an unfavourable risk-to-benefit ratio with CEE+MPA. After the intervention phase, the follow-up phase continued among surviving participants who provided additional written consent.	Results Median follow-up of 5.6 years for intervention phase Median follow-up of 8.2 years for postintervention follow-up phase Hazard Ratio for Breast Cancer Comparing CEE+MPA Versus Placebo Among 50-59 Year Group in Intervention Phase 1.21 (0.81-1.80) Hazard Ratio for Breast Cancer Comparing CEE+MPA Versus Placebo Among 50-59 Year Group in Intervention Phase + Postintervention Follow-up Phase (Combined) 1.34 (1.03-1.75)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Risk of bias: Low B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
1993-1998 Source of funding National Heart, Lung, and Blood Institute National Institutes of Health US Department of Health and Human Services					groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Risk of bias: Low C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - Trial was terminated. C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - No C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - No Risk of bias: High D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Risk of bias: Low Overall Risk of Bias: High Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious
Full citation Colditz,G.A., Stampfer,M.J., Willett,W.C., Hunter,D.J.,	Sample size 23,965 women were followed-up	Interventions Conjugated Estrogen	Details Endpoint for primary analyses was incident	Results 1,050 incident cases of breast cancer	Limitations NICE guidelines manual 2012: Appendix D:

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Manson,J.E., Hennekens,C.H., Rosner,B.A., Speizer,F.E., Type of postmenopausal hormone use and risk of breast cancer: 12-year follow-up from the Nurses' Health Study, Cancer Causes and Control, 3, 433-439, 1992 Ref Id 301487 Country/ies where the study was carried out USA Study type Prospective Cohort Study Aim of the study Use of HRT in relation to breast cancer incidence. Study dates 1976-1988 Source of funding National Cancer Institute NIH Department of Health and Human Services	Characteristics Women aged 30-55 years 33% were current users of HRT 18% were past users Inclusion criteria Female registered nurses Postmenopausal women Exclusion criteria All women who reported breast or other cancer on 1976 questionnaire. Carcinomas in situ		breast cancer Women were followed for 12 years.	Relative Risks of Breast Cancer by Duration of Use of ERT Never use: ref < 2 years: 1.07 (0.77-1.49) 2 to < 5 years: 1.32 (1.02-1.70) 5 years to < 10 years: 1.60 (1.25-2.06) 6 years plus: 1.50 (1.12-2.01) Relative Risks of Breast Cancer by Past Duration of Use of ERT Never use: ref < 2 years: 0.92 (0.74-1.14) 2 to < 5 years: 0.87 (0.67-1.14) 5 years to < 10 years: 1.09 (0.80-1.48) 6 years plus: 1.18 (0.83-1.67) Relative Risks of Breast Cancer by Type of ERT Never use: ref Conjugated Estrogen: 1.42 (1.19-1.70) Estrogen-Progestin: 1.54 (0.99-2.39) Progestin: 2.52 (0.66-9.63) Confounders adjusted for: Age at menopause Type of menopause Time period Age at first birth Age at menarche History of benign breast disease Family history of breast cancer BMI	Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: Yes Level of risk: Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					allocation: No Level of risk: High risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? Follow-up was 85% amd 98% complete for nonfatal and fatal breast cancer respectively. C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: Low risk of bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important confounding and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Overall risk of bias: Low
Full citation Grodstein,F., Stampfer,M.J., Colditz,G.A., Willett,W.C., Manson,J.E., Joffe,M., Rosner,B., Fuchs,C., Hankinson,S.E., Hunter,D.J., Hennekens,C.H., Speizer,F.E., Postmenopausal hormone therapy and mortality, New England Journal of Medicine,	Sample size 23,965 women were followed-up Characteristics Women aged 30-55 years Among cases 15.8% were current users of HRT 27.8% were past users 56.4% never users	Interventions HRT	Details Endpoint for primary analyses was breast cancer mortality Women were followed for an average of 14 years Conditional logistic regression used to estimate relative risks	Results 425 breast cancer mortality cases Relative Risks of Breast Cancer among HRT users Never use: ref Current use: 0.76 (0.56-1.02) Past use: 0.83 (0.63-1.09)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
336, 1769-1775, 1997 Ref Id 229375 Country/ies where the study was carried out USA Study type Prospective Cohort Study Aim of the study Use of HRT in relation to breast cancer mortality Study dates 1976-1994 Source of funding National Cancer Institute NIH Department of Health and Human Services	Among controls 24.5% were current users of HRT 24.9% were past users 50.6% never users Inclusion criteria Female registered nurses Postmenopausal women Exclusion criteria All women who reported breast or other cancer on 1976 questionnaire. Carcinomas in situ			Multivariate-adjusted	allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: Yes Level of risk: Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias C. Attrition bias (systematic differences between the comparison

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? NR C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important confounding and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Overall risk of bias: Low
Full citation Lund,E., Bakken,K., Dumeaux,V., Andersen,V., Kumle,M., Hormone replacement therapy and breast cancer in former users of oral contraceptives--The Norwegian Women and Cancer study, International Journal of Cancer, 121, 645-648, 2007 Ref Id 314666 Country/ies where the study was carried out Norway Study type Cohort study (NOWAC study) Aim of the study To investigate the risk of breast cancer in HRT users Study dates	Sample size N=35453 Characteristics Never oral contraceptive group: Age at baseline (y) Never HRT (n=11305):58.8 Current HRT (n=5838):56.7 Former HRT (n=1604):59.0 BMI (kg/m2): Never HRT:25.3 Current HRT:24.7 Former HRT:25.7 Ever oral contraceptive group: Age at baseline (yrs): Never HRT (n=5167):54.0 Current HRT (n=5170):54.2 Former HRT (n=1034):55.3 BMI (kg/m2):	Interventions Oestrogen only Combined oestrogen+progestin	Details Cohort consisted of 2 parts: 1. 11777 women completed postal questionnaire in 1991/1992, and 1998 2. 23676 women completed postal questionnaire in 1996/1997 Menopause (at start of follow-up) was defined as irregular periods or stopped, or whether women did not know Postmenopause defined as hysterectomised women and when reached age of 53 years. Age 45-52 yrs was defined as unknown menopausal status	Results Mean follow-up=7.0 yrs Risk of breast cancer and HRT (all types)use: Never OC/never HRT: RR 1.00 (reference) Never OC/current HRT: RR 1.53 (1.18-1.98) Never OC/former HRT: RR0.87 (0.53-1.44) Ever OC/never HRT: RR 1.06 (0.77-1.45) Ever OC/current HRT: RR 2.30 (1.77-2.99) Ever OC/former HRT: RR 0.85 (0.44-1.62) Risk of breast cancer and oestrogen use: Never OC/Never HRT: 1.00 (reference) Never OC/Current oestrogen	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
1996-2004 Source of funding Norwegian research council	Never HRT:24.9 Current HRT:24.3 Former HRT:25.2 Inclusion criteria Postmenopausal women Born between 1927-1957 Exclusion criteria Not reported		Duration of use was recorded HRT use was divided into three groups: Current, former, or never HRT groups were treated all together, then divided into two groups: oestrogen users only, or combined users BMI was based on last questionnaire for entire cohort Statistical analysis: Cox proportional hazard model ws used and adjusted for age, BMI, family history of breast cancer, mammography, menarche, parity and age at first delivery	only:RR 0.88 (0.49-1.58) Never OC/former oestrogen only:RR 2.38 (1.16-4.85) Ever OC/never HRT oestrogen only:RR 1.10 (0.82-1.49) Ever OC/current HRT oestrogen only:RR 2.63 (1.65-4.20) Ever OC/former HRT oestrogen only:RR 0.79 (0.11-5.68) Risk of breast cancer and oestrogen+progestin use: Never OC/never HRT: 1.00 (reference) Never OC/current HRT oestrogen+Progestin: RR 1.95 (1.49-2.56) Never OC/former HRT oestrogen+progestin: RR 0.54 (0.22-1.33) Ever OC/never HRT oestrogen+Progestin: RR 1.15 (0.85-1.55) Ever OC/current HRT oestrogen+progestin: RR 2.55 (1.94-3.35) Ever OC/former HRT oestrogen+progestin: RR 0.85 (0.35-2.07)	analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: Yes Level of risk: Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: N/A B3. Individuals administering care were kept 'blind' to treatment allocation: N/A Level of risk: Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					each group? No loss to follow-up C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					important confounding and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Overall risk of bias: Low
Full citation Mills,P.K., Beeson,W.L., Phillips,R.L., Fraser,G.E., Prospective study of exogenous hormone use and breast cancer in Seventh-day Adventists, Cancer, 64, 591-597, 1989 Ref Id 314783 Country/ies where the study was carried out California, USA Study type Prospective cohort study Aim of the study To analyse the risk of breast cancer in a large cohort of Seventh-day adentist women who completed a lifestyle questionnaire in 1976 to obtain information on history of use of exogenous hormones (either OC or HRT) and who were subsequently followed for breast (and other) cancer incidence until the end of 1982 Study dates 1974-1976 Follow-up= 6 years Source of funding National cancer institute, USA	Sample size N=60,000 identified through census questionnaire (response rate=75%) (N=20,341 HRT group; N=20,341 oral contraceptive (OC) group) Characteristics Age (mean,y): 55.4 Race: Non-Hispanic white Distribution of exoqenous hormones in cohort in 1976: HRT group (n=20,341): Premenopausal=8873 (43.7%) Postmenopausal ever used HRT=7580 (66%) Postmenopausal never used HRT=3888 (33.9%) Duration of use among ever users: <1 y=1645 (21.7%) 1-5 y=2556 (33.7%) 6-10 y=1434 (18.9%) 10+y=1945 (25.7%) Inclusion criteria Women aged 25 years and over	Interventions HRT or OC	Details Population selection: 60,000 women were identified from census questionnaire in 1974. Eligible women were mailed a second questionnaire on lifestyle to ascertain exogenous hormone use. 35,000 respondents annually monitored for any hospitalisation in previous 12 months. Any reported hospitalisation was recoorded and medical records reviewed with permission for evidence of cancer diagnosis. 99% of the cohort completed follow-up. Outcomes: All newly diagnosed breast cancer (ICDO:174) occuring in the cohort between return of lifestyle questionnaire (1976) to end of follow-up (1982)	Results During follow-up: 215 primary breast cancers detected (primarily infiltrating ductal carcinomas) Mean age of cases=62.4 yrs Mean age at diagnosis=65.8 yrs (primarily postmenopausal women) 171 (80%) cases in 1976 were menopausal Relative risk (RR) of breast cancer and HRT use (age-adjusted): Never= 1.00 (52 cases) Ever= 1.67 (1.17 to 2.39) (101 cases) Past use only=1.44 (0.95 to 2.17) (44 cases) Current use only=2.53 (1.62 to 3.98) (52 cases) Overall X2=18.47, P=0.0001 Relative risk (RR) of breast cancer and HRT duration (age-adjusted): Never=1.00 (52 cases) <1 yr=2.28 (1.38 to 3.97) (24 cases) 1-5 yrs=1.56 (0.95 to 2.56) (27 cases)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-No A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Unclear (only use of

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Exclusion criteria Not reported		Statistical analysis: Person years at risk from 1976 to end of year, at follow-up, or at time of death. Age-adjusted univariate analyses conducted to obtain relative risk estimates (Mantel-Haenszel procedure). 3 or more categories of exposure examined to detect dose-response gradients between exposure and outcome. Cox-proportional hazards regression models (multivariate) constructed to evaluate age-adjusted relative risk. All multivariate adjusted relative risks accompanied by 95% CI, all P vaues 2-sided.	6-10 yrs=2.75 (1.64 to 4.64) (26 cases) 10+yrs=1.53 (0.92 to 2.54) (24 cases) Overall X2=18.18, P=0.001 Trend P=0.01 Relative risk (RR) of breast cancer, HRT use and menopause type (age-adjusted): Never use: Natural menopause=1.00 Hysterectomy=1.00 Ever use: Natural menopause=1.74 (1.10 to 2.74) Hyterectomy=1.30 (0.78 to 2.18) Past use only: Natural menopause=1.43 (0.85 to 2.44) Hysterectomy=1.00 (0.55 to 1.85) Current use only: Natural menopause=2.71 (1.48 to 4.96) Hysterectomy=1.55 (0.84 to 2.84) Overall X2=11,73, P=0.02, trend P=0.07 Relative risk (RR) of breast cancer, duration of HRT and menopause type (age-adjusted): Never: Natural menopause=1.00 Hysterectomy=1.00 <1yr: Natural menopause=2.47 (1.32 to 4.62) Hysterectomy=1.52 (0.72 to 3.21) 1-5 yrs: Natural menopause=1.29 (0.65 to 2.55) Hysterectomy=0.98 (0.48 to 1.99)	exogenous hormone use at end of screening was reported) Level of risk-High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-The cohort was selected for a particular group of Seventh day adventists takeing either OC or HRT-yes B.2 Participants receiving care were kept 'blind' to treatment allocation-N/A B.3 Individuals administering care were kept 'blind' to treatment allocation-N/A Level of risk: Moderate C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?- Participant numbers at follow-up not reported C.2b The groups were

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				6-10 yrs: Natural menopause=2.66 (1.34 to 5.28) Hysterectomy=1.67 (0.81 to 3.42) 10+yrs: Natural menopause=1.49 (0.68 to 3.28) Hysterectomy=1.15 (0.60 to 2.21) Overall X2=11,73, P=0.02, trend P=0.52 Relative risk (RR) of breast cancer within strata of age at menopause, menopause status, and use of hormones (age-adjusted): <50 years age at menopause: Hysterectomy+no hormone use=1.00 (18 cases) Hysterectomy+hormone use=1.24 (0.70 to 2.20) (46 cases) No hysterectomy+no hormone use=0.63 (0.33 to 1.21) (19 cases) No hysterectomy+hormone use=1.14 (0.59 to 2.19) (21 cases) >50 years at menopause: Hysterectomy+no hormone use=1.23 (0.36 to 4.24) (3 cases) Hysterectomy+hormone use=1.76 (0.85 to 3.61) No hysterectomy+no hormone use=0.91 (0.44 to 1.85) No hysterectomy+hormone use+1.56 (0.82 to 2.96) Cox proportional hazard (HR) regression analysis* of HRT and breast cancer: Total group: Never=1.00 Ever=1.39 (1.00 to 1.94) Current only=1.69 (1.12-2.55) (95%CI does not include 1.0)	comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-Yes C.3a For how many participants in each group were no outcome data available?- not reported in each group, follow-up rate for non-hispanic white group reported (75%) C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)- Yes Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up- Yes (6 yrs) D.2 The study used a precise definition of outcome- Yes (newly detected BC) D.3 A valid and reliable method was used to determine the outcome- Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/A D.5 Investigators were kept 'blind' to other

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				Natural menopause: Never=1.00 Ever=1.44 (0.91 to 2.29) Current only=2.07 (1.14 to 3.78) (95%CI does not include 1.0) Hysterectomy: Never=1.00 Ever=1.05 (0.64 to 1.75) Current only=1.18 (0.66 to 2.14) Menopause <44 yr: Never=1.00 Ever=1.05 (0.57 to 1.94) Current only=1.42 (0.69 to 2.92) Menopause>44 yr: Never=1.00 Ever=1.56 (1.04 to 2.34) Current only=1.79 (1.08 to 2.96) Maternal breast cancer-yes: Never=1.00 Ever=0.83 (0.25 to 2.77) Current=1.34 (0.28 to 6.53) Maternal breast cancer-no: Never=1.00 Ever=1.45 (1.03 to 2.05) Current=1.71 (1.12 to 2.63) Menarche >14 yrs: Never=1.00 Ever=1.70 (0.95 to 3.06) Current=2.44 (1.16 to 5.14) Menarche <14 yrs: Never=1.00 Ever=1.26 (0.85 to 1.87) Current=1.49 (0.91 to 2.43) Age at first birth <24 yrs: Never=1.00 Ever=1.58 (0.95 to 2.62) Current=2.43 (1.29 to 4.55) (CI does not include 1.0) Age at first birth >24 yrs: Never=1.00 Ever=1.14 (0.67 to 1.94) Current=1.26 (0.64 to 2.48)	important confounding and prognostic factors-N/A Level of bias: Low

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				*All adjusted for ages at menarche, first birth, and menopause, educational attainment, Quetelet's index, maternal breast cancer and benign breast cancer.	
Full citation Saxena,T., Lee,E., Henderson,K.D., Clarke,C.A., West,D., Marshall,S.F., Deapen,D., Bernstein,L., Ursin,G., Menopausal hormone therapy and subsequent risk of specific invasive breast cancer subtypes in the California Teachers Study, Cancer Epidemiology, Biomarkers and Prevention, 19, 2366-2378, 2010 Ref Id 315161 Country/ies where the study was carried out Norway Study type Prospective cohort study Aim of the study To investigate hormone therapy use and breast cancer risk in the California Teachers Study cohort Study dates Study start in 1995 to first diagonsis of breast cancer through to 31 December 2006 Source of funding National cancer institute California breast cancer research fund California department of health services	Sample size Cohort N=133, 479 Analysed for breast cancer risk or death N=56,867 Characteristics Invasive breast cancer cases (n): Total: 2,857 HT never users: 493 ET users only: 764 EPT only users: 1153 Mixed HT/unknown: 447 Age at baseline (mean, SD): Total (n): 60,492 HT never users: 63.3 (9.3) ET users only: 63.7 (9.7) EPT only users: 56.7 (7.2) Mixed HT/unknown: 61.2 (9.1) Race: Non-hispanic white: Total (n): 50,681; HT never users: 10,498; ET users only: 14,730; EPT users only: 17,880; mixed HT/unknown: 7,573 Black: Total (n):1628; HT never users:583; ET users only:567; EPT users only:305; mixed/unknown:173 Hispanic: Total (n):1410; HT never users:363; ET users only: 386; EPT users only:465; mixed/unknown: 196 Asian/pacific islander: Total (n):1719; HT never users: 504; ET users only: 397; EPT users only:611;	Interventions HT never use ET (oestrogen use only) PT (progestin use only) EPT (combined oestrogen and progestin use only)	Details The California Teachers Study cohort was assessed for confirmed invasive breast cancer at mean follow-up of 9.8 years HT use was ascertained from detailed questionnaire about type of HT, duration, current or past use Statistical analysis involved using multivariate Cox proportional hazards regression models to estimate association of HT and risk of breast cancer	Results Overall risk of breast cancer and HT use (RR 95%CI): HT never users: 1.00 (reference) HT users: RR 1.40 (1.26-1.55) (adjusted for age, race, family history of breast cancer, BMI, smoking, alcohol consumption, mammographic screening, parity and age at full-term pregnancy, age at menopause, age at menarche, and history of breast biopsy) Risk of breast cancer and type of HT use (RR 95%CI): HT never users: 1.00 (reference) ET only: RR 1.21 (1.07-1.36) EPT only: RR 1.59 (1.42-1.78) PT only: RR 1.22 (0.85-1.75) Mixed ET+EPT: RR 1.42 (1.23-1.63) Mixed PT+EPT: RR 1.59 (1.14-2.22) Mixed PT+ET: 0.59 (0.28-1.24) (adjusted for age, race, family history of breast cancer, BMI, smoking, alcohol consumption, mammographic screening, parity and age at full-term pregnancy, age at menopause, age at menarche, and history of breast biopsy) Risk of breast cancer and duration of HT use (RR 95%CI): Duration ≤5 yrs: HT never users: 1.00 ET only: RR 0.99 (0.88-1.12) EPT only: RR 1.26 (1.14-1.39)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: Yes Level of risk: Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	mixed/unknown: 207 Other/mixed/unknown: Total (n):1429; HT never users: 383; ET users only: 449; EPT users only: 402; mixed/unknown:195 BMI (Kg/m2): <25.0: Total (n):30,474; HT never users: 5871; ET users only: 8277; EPT users only:11,680; mixed HT/unknown:4664 25.0-29.9: Total (n):15,440; HT never users:3373; ET users only:4790; EPT users only:5070; mixed HT/unknown:2207 ≥30.0: Total (n):8154; HT never users:2221; ET users only:2450; EPT users only:2367; mixed HT/unknown: 1116 Menopausal age (y): <35: Total (n):969; HT never users:109; ET users only:494; EPT users only:137; mixed HT/unknown: 229 35-39: Total (n):1751; HT never users:213; ET users only:856; EPT users only:308; mixed HT/unknown:374 40-43: Total (n):3458; HT never users:670; ET users only:1370; EPT users only: 798; mixed HT/unknown:620 44-46: Total (n):5417; HT never users:1202; ET users			Duration 6-14 yrs: HT never users: 1.00 ET only: RR 1.03 (0.90-1.17) EPT only: RR 1.57 (1.40-1.76) Duration 15+yrs: HT never users: 1.00 ET only: RR 1.19 (1.03-1.37) EPT only: RR 1.83 (1.48-2.26) Duration of current use: HT never users: 1.00 Current ET (≤5 yrs): RR 1.23 (1.02-1.49) Current ET (6-14 yrs): RR 1.28 (1.08-1.51) Current ET (15+yrs): RR 1.35 (1.15-1.58) Current EPT (≤5 yrs): RR 1.61 (1.41-1.83) Current EPT (6-14 yrs): RR 1.78 (1.55-2.03) Current EPT (15+ yrs): RR 1.94 (1.53-2.44) Duration of past use: HT never users: 1.00 Past ET or EPT: 1.04 (0.90-1.20) Effects and duration of HT through 2002: HT never users: 1.00 Current ET (≤5 yrs): RR 1.34 (1.06-1.70) Current ET (6-14 yrs): RR 1.52 (1.24-1.85) Current ET 15+ yrs): RR 1.44 (1.19-1.75) Current EPT (≤5 yrs): RR 1.81 (1.53-2.12) Current EPT (6-14 yrs): RR 2.18 (1.86-2.56) Current EPT (15+ yrs): RR 2.25 (1.71-2.96) Duration of past use (through 2002): HT never users: 1.00 Past ET or EPT: RR 1.09 (0.91-1.30) Stratified by age and adjusted	care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: Unclear B3. Individuals administering care were kept 'blind' to treatment allocation: Unclear Level of risk: Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? No loss to follow-up C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	only:1913; EPT users only:1495; mixed HT/unknown:807 47-49: Total (n):8462; HT never users:2252; ET users only:1990: EPT users only:3095; mixed HT/unknown:1125 50-52: Total (n):11628; HT never users:3509; ET users only:2053; EPT users only:4650; mixed HT/unknown:1416 53-55: Total (n):7537; HT never users:2336; ET users only:1133, EPT users only:3075; mixed HT/unknown:993 Hyserectomy: No: Total (n):36,474; HT never users:10,472; ET users only:3386; EPT users only:18,243; mixed HT/unknown:4373 Yes: Total (n):19,343; HT never users:1638; ET users only:12,797; EPT users only:1072; mixed HT/unknown:3827 Inclusion criteria Perimenopausal women Postmenopausal women Age <35 to 55 years Exclusion criteria Not California residents at time of completing baseline questionnaire Previous/unknown history of breast cancer Older than 80 yrs of age at baseline			for categories of race, family history of breast cancer, BMI, smoking, alcohol consumption, mammographic screening, parity and age at full term pregnancy, age at menarche, and history of breast biopsy	or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important confounding and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Overall risk of bias: Low

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Premenopausal Unknown menopausal status Unknown history of ever using HT				
Full citation Schairer,C., Lubin,J., Troisi,R., Sturgeon,S., Brinton,L., Hoover,R., Menopausal estrogen and estrogen-progestin replacement therapy and breast cancer risk.[Erratum appears in JAMA 2000 Nov 22-29;284(20):2597], JAMA, 283, 485-491, 2000 Ref Id 268450 Country/ies where the study was carried out USA Study type Prospective cohort study Aim of the study To examine the relationship between menopausal estrogen and estrogen-progestin replacement therapy and risk of breast cancer Study dates 1980-1995 Source of funding American Cancer Society US National Cancer Institute	Sample size 46,355 postmenopausal women Characteristics Average age at start of follow-up: 58 years Race (%) White: 89 Blacks: 5 Asian-Americans: 5 Menopause type (%) Natural No hormone use: 61 Estrogen only: 32 Estrogen-progestin: 6 Hysterectomy No hormone use: 31 Estrogen only: 58 Estrogen-progestin: 6 Bilateral oophectomy No hormone use: 20 Estrogen only: 73 Estrogen-progestin: 7 First-degree familyhistory of breast cancer (%) No No hormone use: 46 Estrogen only: 47 Estrogen-progestin: 6 Yes No hormone use: 47 Estrogen only: 46 Estrogen-progestin: 6 Inclusion criteria Women who did not have a menstrual period for at least	Interventions Estrogen Estrogen and Progestins	Details Subjects were participants in a breast cancer screening program. Follow-up study carried out in three phases. Breast cancer risk factors collected at baseline interview.	Results Mean duration of follow-up: 10.2 years 2,082 cases ascertained at follow-up Relative Risk of Incident Breast Cancer Associated With Type of HRT Never use: reference Estrogens only: 1.1 (1.0-1.3) Estrogens+progestins: 1.3 (1.0-1.6) Progestin: 0.9 (0.5-1.6) Relative Risk of Incident Breast Cancer According to Time Since Last Use Estrogen 1-2 years: 1.4 (1.1-1.8) > 2-4 years: 1.2 (0.9-1.6) > 4-6 years: 0.9 (0.6-1.3) > 6 years: 1.1 (0.9-1.2) Estrogen+Progestin 1-2 years: 1.2 (0.6-2.4) > 2-4 years: 1.2 (0.5-2.5) > 4-6 years: 0.6 (0.2-2.6) > 6 years: 0.6 (0.3-1.6) Relative Risk of Incident Breast Cancer According to Duration of Use Estrogen Never use: reference < 8 years: 1.00 (0.83-1.21) 8- <16 years: 1.30 (1.06-1.60) ≥ 16 years: 1.23 (0.97-1.56) Estrogen+Progestin Never use: reference < 2 years: 1.13 (0.75-1.69)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: Yes Level of risk: Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	3 months prior to an inteview for one of the following reasons: natural menopause; bilateral oopherectomy with or without hysterectomy; or a hysterectomy with at least one ovary retained. Exclusion criteria 1. Women with uncertain ages at menopause or types of menopause 2. Reported bilateral prophylactic mastectomies or a diagnosis of breast cancer before the start of follow-up 3. Cases of breast cancer diagnosed between the end of the screening program and start of follow-up study 4. Premenopausal cases of breast cancer			2- <4 years: 1.27 (0.82-1.97) ≥ 4 years: 1.75 (1.24-2.47) Adjusted for age, age at menopause, education, mammographic screening, and BMI	intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? 0.5% lost to follow-up C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					of those for whom outcome data were not available): N/A Level of risk: Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important confounding and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Overall: Low risk of bias
Full citation Stahlberg,C., Lynge,E., Andersen,Z.J., Keiding,N., Ottesen,B., Rank,F., Hundrup,Y.A., Obel,E.B.,	Sample size N=19898 included N=10874 analysed Characteristics Not reported	Interventions HRT use No HRT use	Details Population: Postmenopausal women were identified from the Danish Nurse cohort and	Results Risk of breast cancer and HRT use: Never use (n):110/6566 breast cancer cases; HR=1.00	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Pedersen,A.T., Breast cancer incidence, case-fatality and breast cancer mortality in Danish women using hormone replacement therapy - A prospective observational study, International Journal of Epidemiology, 34, 931-935, 2005 Ref Id 304857 Country/ies where the study was carried out Denmark Study type Prospective cohort study Aim of the study To investigate the effect of HRT on risk of breast cancer and breast cancer mortality in natural post-menopausal women Study dates 1993-2004 Source of funding Danish cancer society	Inclusion criteria Natural posmenopausal women Age >44 yrs at start of study Invasive breast cancer cases Complete HRT use information Exclusion criteria Non-melanoma skin cancer Missing information on HRT use Surgical menopause Hysterectomised women Premenopause women		information ascertained by questionnaire. Breast cancer cases were identified by linkage through the unique personal identification number to the Danish nationwide registries Follow-up started in 1993 until 1999 (6 yrs), and for mortality ended in 2004 (11 yrs) Prognostic characteristics obtained from Danish breast cancer cooperative group, mortality data obtained from Danish civil registration. Cause of death obtained from the National causes of death register Statistical analysis: Conditional Cox proportional hazards model was used for time to cancer prognosis and time to death outcomes. HRT exposure was estimated using HR and 95%CI and adjusted for age, smoking, alcohol use, BMI and physical activity	(reference) Past use (n):31/1582 breast cancer cases; HR=1.16 (0.76-1.77) Current use (n):103/2726 breast cancer cases; HR=2.42 (1.81-3.26) Adjusted for smoking, alcohol, BMI, and physical activity Breast cancer mortality and HRT use: Never use (n):37; HR=1.00 (reference) Past use (n):12; HR=1.31 (0.68-2.52) Current use (n):22; HR=1.97 (1.14-3.42) Adjusted for age	A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: Unclear, not reported Level of risk: Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: Unclear, not reported B3. Individuals administering care were kept 'blind' to treatment

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					allocation: Unclear, not reported Level of risk: Unclear risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? No loss to follow-up C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: Low risk of bias D. Detection bias (bias in

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important confounding and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: Some indirectness, the cohort was not representative of the general population as they were all nurses Overall risk of bias: Low
Full citation Vickers,M.R., MacLennan,A.H., Lawton,B., Ford,D., Martin,J., Meredith,S.K., DeStavola,B.L., Rose,S., Dowell,A., Wilkes,H.C., Darbyshire,J.H., Meade,T.W., WISDOM group., Main morbidities recorded in the	Sample size Combined therapy versus placebo Combined therapy: 2,196 Placebo: 2,189 Combined therapy versus oestrogen therapy	Interventions Conjugated equine ostrogens 0.625 mg orally daily versus placebo Conjugated equine ostrogens plus medroxyprogesterone acetate 2.5/5.0 mg orally daily versus placebo	Details 1. Treatment was by random allocation with a computer based, stratified block randomisation program. 2. Stratification based on hysterctomy status and	Results Trial closed prematurely during recruitment after a median follow-up of 11.9 months after publication of early results of the WHI study. OR for Incident Breast Cancer	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
women's international study of long duration oestrogen after menopause (WISDOM): a randomised controlled trial of hormone replacement therapy in postmenopausal women, BMJ, 335, 239-, 2007 Ref Id 230610 Country/ies where the study was carried out UK, Australia, and New Zealand Study type Multi-centre RCT Aim of the study To assess the long term risks and benefits of HRT Study dates 1999-2000 Source of funding UK Medical Research Council British Heart Foundation Department of Health for England Scottish Office Welsh Office etc.	Combined therapy: 815 Oestrogen therapy: 826 Characteristics Combined therapy versus placebo Mean (SD) age at randomisation, yrs Combined therapy: 63.3 (4.7) Placebo: 63.3 (4.6) Mean (SD) body mass index Combined therapy: 27.9 (4.9) Placebo: 28.0 (5.2) Mean (SD) SBP Combined therapy: 136 (21) Placebo: 137 (22) Combined therapy versus oestrogen therapy Mean (SD) age at randomisation, yrs Combined therapy: 61.7 (5.1) Oestrogen: 61.9 (5.1) Mean (SD) body mass index Combined therapy: 28.0 (4.7) Oestrogen: 27.9 (5.0) Mean (SD) SBP Combined therapy: 137 (21) Placebo: 135 (20) Inclusion criteria 1. Postmenopausal women (no menstrual period in the past 12 months or had undergone hysterectomy) 2. Women aged 50-69 years Exclusion criteria 1. History of breast cancer 2. Any other cancer in the past 10 years except basal		intended use of HRT. 3. Women with a uterus or subtotal hysterctome were randmoised to combined oestrogen plus progestogen or to placebo 4. Women with no uterus and unwilling to take a placebo were randmised to either oestrogen only or combined oestrogen and progestogen therapy. 5. Planned treatment duration was 10 years (range 9-12)	Combined therapy versus placebo 12 incident breast cancer cases 0.71 (0.18-2.61) Combined therapy versus oestrogen alone 5 incident breast cancer cases 1.52 (0.17-18.24)	between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Risk of bias: Low B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - As far as possible B3. Individuals administering care were kept 'blind' to treatment allocation - As far as possible Risk of bias: Low

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	and squamous cell skin cancer 3. Endometriosis or endometrial hyperplasia 4. Venous thromboembolism 5. Gall bladder disease in women who had not had a cholecystectomy 6. Myocardial infarction 7. Unstable angina 8. Cerebrovascular accident 9. Subarachnoid haemorrhage 10. Transient ischaemic attack 11. Use of HRT within the past 6 months				C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? -Trial was terminated prematurely C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - No C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - No Risk of bias: High D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					D1. The study had an appropriate length of follow-up - No D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - As far as possible D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Risk of bias: High Overall Risk of Bias: High Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information Odds ratios were calculated from raw figures using STATA.
Full citation Willis,D.B., Calle,E.E., Miracle-McMahill,H.L., Heath,C.W.,Jr., Estrogen replacement therapy and risk of fatal breast cancer in a prospective cohort of postmenopausal women in the United States, Cancer Causes and Control, 7, 449-457, 1996 Ref Id	Sample size N=422,373 Characteristics Age, yrs Breast cancer cases: 61.4 Other women: 59.2 Ever use of ERT, % Breast cancer cases: 39.8 Other women: 44.7	Interventions Estrogen replacement therapy	Details Women who were cancer free at study entry and supplied information on estrogen use were followed up for cancer deaths. Endpoints ascertained through National Death Index and death certificates.	Results Average follow-up: 9 years Breast cancer deaths: 1,469 Relative risk of breast cancer mortality by categories of estrogen use Use of estrogen Never: reference Ever: 0.84 (0.75-0.94)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
315522 Country/ies where the study was carried out USA Study type Prospective cohort study Aim of the study To examine the relationship between fatal breast cancer and use of estrogen replacement therapy (ERT) in a cohort of postmenopausal women Study dates 1982 Source of funding Not reported	Inclusion criteria Postmenopausal women Exclusion criteria 1. Women with incomplete race informaton 2. Women with prevalent cancer (except non-melanoma skin cancer) at study entry 3. Unknown menopausal status at study entry 4. No data on estrogen use 5. Women who could not be classified as a baseline/former use/duration of use			Recency of use Never: reference Baseline: 0.90 (0.75-1.09) Former: 0.78 (0.68-0.89) Years of use Never: reference ≤ 1: 0.85 (0.71-1.02) 2-5: 0.78 (0.65-0.93) 6-10: 0.78 (0.62-0.98) 11+: 0.93 (0.75-1.15) Age at first use Never: reference < 40: 0.65 (0.51-0.85) 40-49: 0.84 (0.73-0.97) 50+: 0.89 (0.76-1.05) Years since stopping estrogen use Never: reference 0-5: 0.82 (0.64-1.05) 6-10: 0.70 (0.56-0.89) 10+: 0.84 (0.70-1.01) Covariates adjusted for Age at interview, race, menopausal status, smoking status, age at menarche and menopause, body mass index, alcohol consumption, age at 1st livebirth, first-degree family history of breast cancer, history of breast cysts, DES use, and use of oral contraceptives	allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): Yes A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: Yes Level of risk: Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: Yes B2. Participants receiving care were kept 'blind' to treatment allocation: N/A B3. Individuals administering care were kept 'blind' to treatment allocation: N/A Level of risk: Low risk of bias C. Attrition bias (systematic differences between the comparison

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? See results section C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: Unclear risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					outcome: Yes D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: Unclear D5. Investigators were kept 'blind' to other important confounding and prognostic factors: Unclear Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious
Full citation Schierbeck,L.L., Rejnmark,L., Tofteng,C.L., Stilgren,L., Eiken,P., Mosekilde,L., Kober,L., Jensen,J.E.B., Effect of hormone replacement therapy on cardiovascular events in recently postmenopausal women: Randomised trial, BMJ (Online), 345, -, 2012 Ref Id 288651 Country/ies where the study was carried out Denmark Study type Open-label Randomised Controlled Trial Aim of the study To investigate long-term effect of HRT on cardiovascular outcomes in recently	Sample size 1006 women HRT group: 502 Control: 504 Characteristics Healthy women aged 45-58 years Mean age: 49.7 years Mean BMI: 25.2 kg/m² Mean time since menopause: 0.59 years Inclusion criteria 1. Healthy recently postmenopausal white women aged 45-58 years 2. Last menstrual bleeding 3-24 months before study entry or perimenopausal symptoms in combination with recorded serum FSH values (> 2 standard deviations over the	Interventions Women with an intact uterus 2 mg synthetic 17-ß-estradiol for 12 days 2 mg 17-ß-estradiol plus 1 mg norethisterone acetate for 10 days 1 mg 17-ß-estradiol for 6 days Women who had undergone hysterctomy 2 mg synthetic 17-ß-estradiol a day	Details Women enrolled in a prospective followed cohort Randomly allocated (open label) to receive HRT or no treatment Participants recruited by direct mailing to a randomised sample Participants stratified according to centre and randomised to treatment in blocks of 10 using sealed envelopes Planned duration of study was 20 years Intervention was stopped at about 11 years owing to adverse reports from other trials After termination of randomisation, women	Results Mean duration for randomised treatment: 10.1 years Mean duration after termination of randomisation: 15.8 years Hazard Ratios for Breast Cancer Associated With HRT During Randomisation Phase Age ≥ 50 years: 0.98 (0.33-2.92) Age < 50 years: 0.34 (0.11-1.08)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators,

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
postmenopausal women Study dates 1990-1993 Source of funding University of Aarhus Elise Jensen's Foundation Novo Nordic Novartis LEO Pharma	premenopausal mean) 3. Women who had had a hysterectomy aged 45-52 years and had records showing an increase in serum FSH levels Exclusion criteria 1. History of bone disease 2. Uncontrolled chronic disease 3. Previous or current cancer or thromboembolic disease 4. Current or past treatment with glucocorticoids for more than 6 months 5. Current or previous use of HRT within the past three months 6. Alcohol or drug dependency		were followed for an additional 5.7 years		clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Risk of bias: Low B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - No B3. Individuals administering care were kept 'blind' to treatment allocation - No Risk of bias: High C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - None C2b. The groups were

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? None C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - N/A Risk of bias: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - No D5. Investigators were kept 'blind' to other important confounding and prognostic factors - No Risk of bias: High

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Overall Risk of Bias: High Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious
Full citation Anderson,G.L., Limacher,M., Effects of Conjugated Equine Estrogen in Postmenopausal Women with Hysterectomy: The Women's Health Initiative Randomized Controlled Trial, Journal of the American Medical Association, 291, 1701-1712, 2004 Ref Id 295534 Country/ies where the study was carried out 40 centres in the USA Study type Randomised Controlled Trial (Estrogen alone component of the WHI) Aim of the study To assess the effects of HRT on major disease incidence rates Study dates 1993-1998 Source of funding The National Heart, Lung, and Blood Institute	Sample size 10,739 Conjugated Equine Estrogen (CEE) arm: 5,310 Placebo: 5,429 Characteristics Study participants were healthy and at average risk of CHD and breast cancer. Intervention groups were balanced at baseline on key demographic and disease risk factor characteristics Inclusion criteria 1. Women 50-79 years old at baseline 2. Had undergone hysterectomy 3. Were likely to reside in area of recruitmenty for 3 years Exclusion criteria 1. Any medical condition likely to be associated with a predicted survival < 3 years) 2. Safety (prior breast cancer, other prior cancer within the last 10 years except nonmelanoma skin cancer 3. Adherence and retention concerns	Interventions 0.625 mg/day of CEE Matching placebo	Details Participants recruited by population-based direct mailing campaigns to age-eligible women 3-month washout period was required of women using postmenpausal hormones at initial screening Eligible women randomly assigned to HRT or matching placebo in equal proportions Study participants contacted via telephone 6 weeks after randomization to assess symptoms and reinforce adherence	Results Average follow-up: 6.8 years 563 (5.2%) participants withdrew, lost to follow-up. Were comparable between treatment groups Hazard Ratio of Breast Cancer for CEE Compared to Placebo in 50-59 Year Group 0.72 (0.43-1.21)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Risk of bias: Low B. Performance bias (systematic differences between groups in the

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Risk of bias: Low C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - See results section C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Risk of bias: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Risk of bias: Low Overall Risk of Bias: Low Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Indirectness: No serious
Full citation Cherry,N., McNamee,R., Heagerty,A., Kitchener,H., Hannaford,P., Long-term safety of unopposed estrogen used by women surviving myocardial infarction: 14-year follow-up of the ESPRIT randomised controlled trial, BJOG: An International Journal of Obstetrics and Gynaecology, 121, 700-705, 2014 Ref Id 321013 Country/ies where the study was carried out UK Study type Randomised Controlled Trial Aim of the study To compare health outcomes during 14-year observational follow-up in postmenopausal women initially randomised to unopposed estrogen or placebo Study dates 1996-2000 Source of funding UK National Health Services Research and Development Programme on Cardiovascular Disease and Stroke	Sample size 1017 women Estradiol group: 513 Placebo: 504 Characteristics Women aged 50-69 years who had survived a first myocardial infarction Inclusion criteria Exclusion criteria Women who reported a history of cancer or use of HRT in the previous 12 months	Interventions 2 mg Estradiol valerate Placebo	Details Women recruited at time of hospitalisation for MI Women randomised to recieve treatment or placebo for 2 years Cancer incidence and mortality collected from Office of National Statistics for England and Wales	Results Breast cancer deaths Estradiol group: 1 Placebo group: 4 Breast cancer incidence Estradiol group: 7 Placebo group: 15 Hazard Ratio for Breast Cancer Incidence for Treatment Group Compared to Placebo (Age 50-59 year old group) 0.33 (0.06-1.68)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Risk of bias: Low B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Risk of bias: Low C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - NR C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Risk of bias: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Risk of bias: Low Overall Risk of Bias: Low Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious
Full citation Fournier,A., Berrino,F., Clavel-Chapelon,F., Unequal risks for breast cancer associated with different hormone replacement therapies: Results from the E3N cohort study, Breast Cancer Research and Treatment, 107, 103-111, 2008 Ref Id	Sample size 80,377 postmenopausal women Characteristics Women aged 40-65 years 70% of women had used HRT, for a mean duration of 7 years Mean age at start of treatment: 52.4 years	Interventions HRT	Details Women who agreed to participate filled a first questionnaire and an informed consent form Breast cancer patients were identified from self-reports, health insurance register, or information on deaths	Results 2,354 invasive breast cancer cases Relative Risks of Breast Cancer by Type of HRT and Duration of Exposure Estrogen < 2 years: 1.26 (0.83-1.89) 2-4 years: 1.13 (0.70-1.81)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
321031 Country/ies where the study was carried out French Study type Prospective Cohort Study Aim of the study Assess and compare the association between different HRTs and breast cancer risk Study dates 1990-2002 Source of funding European Community French League against Cancer etc.	Inclusion criteria 1. Postmenopausal women 2. Were considered postmenopausal if they had had 12 consecutive months without menstrual periods, had undergone bilateral oophorectomy, had ever used HRT, or self-reported that they were postmenopausal. Exclusion criteria 1. Women who reported a cancer other than a basal cell carcinoma before the start of followup 2. Women for whom no age at first HRT use was available		Women for whom age at menopause could not be determined were considered menopausal at age 47 if menopause was artificial, and at age 51 otherwise	4-6 years: 1.50 (0.88-2.56) 6+ years: 1.31 (0.76-2.28) Estrogen+Progesterone < 2 years: 0.71 (0.44-1.14) 2-4 years: 0.95 (0.67-1.36) 4-6 years: 1.26 (0.87-1.82) 6+ years: 1.22 (0.89-1.67) Relative Risks of Breast Cancer by Type of HRT and Recency of Use Estrogen Last use 0-2 years previously: 1.22 (0.90-1.65) Last use 2-5 years previously: 2.10 (1.04-4.21) Last use ≥ 5 years previously: 1.17 (0.69-1.99) Estrogen + Progesterone Last use 0-2 years previously: 1.03 (0.84-1.26) Last use 2-5 years previously: 1.93 (0.99-3.72) Confounders adjusted for: Time since menopause Age at menarche Parity and age at fiurst full-term pregnancy Breast feeding Age at menopause Type of menopause Personal history of benign breast disease Family history of breast cancer in first-degree relatives Family history of breast cancer in other relatives Physical activity Previous mammography	allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study): N/A A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders: Yes A3. The groups were comparable at baseline, including all major confounding and prognostic factors: Yes Level of risk: Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied: N/A B2. Participants receiving care were kept 'blind' to treatment allocation: No B3. Individuals administering care were kept 'blind' to treatment allocation: No Level of risk: High risk of bias C. Attrition bias (systematic differences between the comparison

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					groups with respect to loss of participants) C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up): Yes C2a. How many participants did not complete treatment in each group? NR C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment): N/A C3a. For how many participants in each group were no outcome data available? N/A C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available): N/A Level of risk: High risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up: Yes D2. The study used a precise definition of outcome: Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					D3. A valid and reliable method was used to determine the outcome: Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention: N/A D5. Investigators were kept 'blind' to other important confounding and prognostic factors: N/A Level of bias: Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Overall risk of bias: High

H.8.6 Osteoporosis

Study details	Study design	Comparison	Results	Other
Full citation Aitken,J.M., Hall,P.E., Rao,L.G., Hart,D.M., Lindsay,R., Hypercortisolaemia and lack of skeletal response to oestrogen in postmenopa	Aim of the study To assess the value of oestrogen mestranol in the prevention of bone mineral loss with age after oophorectomy. Inclusion criteria Healthy women who had	Details Oral 20 µg oestrogen mestranol Placebo tablets Methods Women were given either oestrogen replacement therapy or placebo and were instructed to take two daily. Samples of venous blood and urine were obtained from participants at the start of the treatment and at yearly intervals. An X-ray of the right hand was taken for densitometric and morphological measurements at the start of treatment alone, and photon absorptiometric measurement was made at midpoint of the third metacarpal at the start of treatment and at yearly	Characteristics Age (years, mean, SE): Two months post oophorectomy: Placebo: 44.1 (2.3) ; oestrogen: 45.0 (0.7) Three years post oophorectomy: Placebo: 49.1 (0.5); oestrogen: 49.1 (0.6) Six years post oophorectomy: Placebo: 51.6 (0.4); oestrogen: 50.4 (1.0) Whole bone density (percentile, mean, SE): Two months post oophorectomy: placebo:47.4 (6.3); oestrogen:52.8 (9.1)	Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept ‘blind’ to treatment allocation. No. Individuals administering care were kept ‘blind’ to treatment allocation. No. High risk of bias

Study details	Study design	Comparison	Results	Other
usal women, Clinical Endocrinology, 3, 167-174, 1974 Ref Id 295514 Study type Double blind controlled trial Source of funding Scottish Hospitals Endowments Research Trust National Fund for Research into Crippling Diseases Country/ies where the study was carried out UK Study dates Not reported	undergone hysterectomy and bilateral oophorectomy for non-malignant disease two months, three years, or six years previously. Exclusion criteria History of hepatitis or either deep venous thrombosis or pulmonary embolism, or both, or specific diseases known to be associated with bone mineral loss. Women who had taken hormone therapy between oophorectomy and the time of review were also excluded.	intervals. Biochemical measurements including serum and urine were made by standard procedures. Calcium was estimated by atomic aborption spectrophotometry. Creatinine, phosphorus, serum aspartate, alanine transaminases, blood sugar were estimated as well as lactic dehydrogenase. Urinary calcium and phosphorus excretion was calculated, as well as the whole bone density at the metacarpal midpoint, and were converted to percentile values. The metacarpal mineral content was measured by photon absorptiometry, and was standardised to allow for participants of different size by dividing the ash per unit length by the metacarpal length to give the standardised metacarpal ash. Statistical method used was Students t test. Sample size N=114	Three years post oophorectomy: placebo: 39.0 (4.1); oestrogen:36.9 (3.5) Six years post oophorectomy: placebo: 37.4 (9.1); oestrogen: 30.1 (6.4) Standardised metacarpal ash (mg ash/mm/cm, mean,SE): Two months post oophorectomy: placebo:7.23 (0.24); oestrogen: 7.44 (0.33) Three years post oophorectomy: placebo:6.79 (0.15); oestrogen: 6.76 (0.10) Six years post oophorectomy: placebo:6.64 (0.25); oestrogen: 6.77 (0.15) Results Any non-vertebral fracture (oestrogen versus placebo): Oestrogen: 0/68 Placebo: 2/66	Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 15 placebo group, n = 16 HRT group. The groups were comparable for treatment completion. Yes. For how many participants in each group were outcome data not available? n = 15 placebo group, n = 16 HRT group. The groups were comparable with respect to the availability of outcome data. Yes. Moderate risk of bias Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.

Study details	Study design	Comparison	Results	Other
				Moderate risk of bias
Full citation Lacroix,A.Z., Chlebowski, R.T., Manson,J.E., Aragaki,A.K., Johnson,K.C., Martin,L., Margolis,K.L., Stefanick,M.L., Brzyski,R., Curb,J.D., Howard,B.V., Lewis,C.E., Wactawski-Wende,J., Investigators, W.H.I., Health outcomes after stopping conjugated equine estrogens among postmenopausal women with prior hysterectomy: a randomized controlled trial, JAMA, 305, 1305-1314, 2011 Ref Id 229707 Study type Randomised controlled trial followed by post	Aim of the study To examine health outcomes associated with randomisation to treatment with conjugated equine oestrogen (CEE) among women with prior hysterectomy after a mean of 10.7 years of follow-up through August 2009. Inclusion criteria Postmenopausal women aged 50-79 years, with prior hysterectomy, were not taking hormone therapy, and had an anticipated 3 year survival. Exclusion criteria Women with prior breast cancer or other cancer within 10 years (except non-melanoma skin cancer), or prior venous thromboembolism (if screened after 1997).	Details CEE (0.625mg/d) Placebo Methods Intervention phase (Cauley et al.,2003) Post intervention phase (current study focus on 47.2 months follow-up duration through 2009): Participants were instructed to discontinue taking study pills. Subsequent participant follow-up consent was obtained from 77.9% of surviving participants in the CEE group and 78.4% in the placebo group. Outcomes were identified from annual questionnaires and verified by medical review. Annual mammograms were encouraged and tracked by annual review. During the post intervention phase 3.6% to 4.7% women from CEE group and 2.7% to 3.0% women from the placebo group reported oestrogen alone use (any route of administration) on annual questionnaires. Statistical analysis Primary analysis included all randomised participants using time to event methods and were based on ITT method. Baseline characteristics of women who gave additional consent were compared with X2 and t tests. Annualised rates of clinical events were estimated for intervention period, Sample size Post intervention analysis (n): CEE: 3778 Placebo: 3867	Characteristics Age at screening (mean years (SD)): 50-59: CEE:1223/3778; placebo:1232/3867 60-69: CEE:1740/3778; placebo:1799/3867 70-79: CEE:815/3778; placebo: 836/3867 Hormone therapy use (n): Never: CEE:1929/3778; placebo:1916/3867 Past: CEE:1304/3778; placebo: 1373/3867 Current: CEE:544/3778; placebo:575/3867 Duration of hormone therapy use (y, n): <5 years: CEE:960/3778; placebo:1036/3867 5-10 years: CEE:348/3778; placebo:377/3867 >10 years: CEE:541/3778; placebo:538/3867 BMI (n): <25: CEE:785/3778; placebo:771/3867 25-<30: CEE: 1289/3778; placebo:1391/3867 ≥30:CEE: 1687/3778; placebo: 1683/3867 Hysterectomy age group (y, n): <40: CEE: 1495/3778; placebo: 1501/3867 40-49: CEE: 1643/3778; placebo: 1662/3867 50-54: CEE: 345/3778; placebo: 412/3867 ≥55: CEE:275/3778; placebo: 271/3867 Fracture and age ≥55 years (n): CEE:455/3778; placebo:447/3867 Results Hip fracture Intervention: CEE: 48/3778; placebo:74/3867; HR: 0.64 (95%CI 0.46-0.96) Post intervention: CEE: 66/3778; placebo:53/3867; HR: 1.27 (95%CI 0.88-1.82) Overall: CEE: 114/3778; placebo:127/3867; HR: 0.92 (95%CI 0.71-1.18) Cumulative annualised incidence rates for hip fracture (age, n): 50-59: CEE:8/3778; placebo:5/3867; HR: 1.55 (95%CI 0.51-4.75) 60-69: CEE:38/3778; placebo:45/3867; HR: 0.87 (95%CI 0.57-1.35) 70-79: CEE:68/3778; placebo:77/3867; HR: 0.97 (95%CI 0.65-1.25)	Other information Limitations Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. Yes. Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline, including all major confounding and prognostic factors. Yes. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Yes.

Study details	Study design	Comparison	Results	Other
intervention observational study Source of funding Wyeth Ayerst (dontated study drugs) National Heart, Lung, and Blood Institute NIH US Department of Health and Human Services Country/ies where the study was carried out USA (multicentre) Study dates Recruitment of participants:1993-1998 Intervention phase end: 2004 Post intervention phase started: 2004-2009				For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. No. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Manson,J.E., Chlebowski, R.T., Stefanick,M. L., Aragaki,A.K., Rossouw,J.E	Aim of the study To report a comprehensive, integrated overview of findings from the two WHI trials with extended	Details CEE+MPA (combined equine oestrogen plus medroxyprogesterone acetate) versus placebo CEE (combined equine oestrogen) alone versus placebo Methods Fracture was defined as which was a secondary end point, are reported separately. For each trial, intervention phase analyses included all	Characteristics Age at screening (mean, SD, y): CEE: 63.6 (7.3); placebo: 63.6 (7.3) CEE+MPA: 63.2 (7.1); placebo: 63.3 (7.1) Years since menopause (y, n): CEE versus placebo: <10 years: 827/5310; 817/5429 10-<20 years: 1438/5310; 1500/5429	Other information Limitations Study quality NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias

Study details	Study design	Comparison	Results	Other
., Prentice,R.L., Anderson,G., Howard,B.V., Thomson,C.A., LaCroix,A.Z., Wactawski-Wende,J., Jackson,R.D., Limacher,M., Margolis,K.L., Wassertheil-Smoller,S., Beresford,S.A., Cauley,J.A., Eaton,C.B., Gass,M., Hsia,J., Johnson,K.C., Kooperberg,C., Kuller,L.H., Lewis,C.E., Liu,S., Martin,L.W., Ockene,J.K., O'Sullivan,M.J., Powell,L.H., Simon,M.S., Van,Horn L., Vitolins,M.Z., Wallace,R.B., Menopausal hormone therapy and health outcomes during the intervention	post-intervention follow-up. Inclusion criteria Post menopausal women aged 50 to 79 years, with uterus (CEE+MPA trial). Post menopausal women aged 50 to 79, with prior hysterectomy (CEE trial). Exclusion criteria Not reported in paper, reported in previous WHI studies.	randomised participants according to their randomisation assignment until last intervention contact, using time-to-event method based on the intention-to-treat principle. -Hazard ratios (HRs) were estimated using Cox proportional hazards models stratified by age, prior disease (if appropriate), and randomisation status in the WHI dietary modification trial. Comparisons during the postintervention phase include randomised participants in active follow-up and at risk for an initial diagnosis of the relevant outcome. -All statistical tests are 2-sided and nominal P values of 0.05 or less are regarded as significant. The p values do not adjust for multiple outcomes, sequential monitoring, or multiple subgroup comparisons due to the large number of tests conducted; therefore, the p values should be interpreted cautiously. Inference on subgroup analyses rely primarily on tests for interaction, which are also subject to multiple testing limitations when a large number of tests are conducted. -Adherence sensitivity analyses, conducted by censoring follow-up 6 months after non adherence, included time-varying weights (inversely proportional to the estimated probability of continued adherence) in proportional hazards models that adjusted for changes in the distribution of sample characteristics during follow-up. CEE+MPA intervention: the cumulative results reported in the current re-analyses include a median post intervention follow-up of 8.2 years and a median cumulative follow-up of 13.2 years; -CEE intervention: the median post intervention follow-up was 6.6 years and the median cumulative follow-up was 13.0 years; Sample size N= 27,347 (16608 in CEE+MPA trial; and 10739 in CEE trial) The post intervention follow-up through September 30, 2010 is based on 81.1% surviving participants who provided additional written informed consent. Following stopping of the intervention, fewer than 4% women reported personal use of hormone therapy.	≥20 years: 2230/5310; 2319/5429 CEE+MPA versus placebo: <10 years: 2780/8506; 2771/8102 10-<20 years: 3044/8506; 2992/8102 ≥20 years: 1850/8506; 1805/8102 Hormone use (n): CEE versus placebo Never use: 2760/5310; 2769/5429 Past use: 1871/5310; 1947/5429 Current use: 669/5310; 709/5429 CEE+MPA versus placebo: Never use: 6277/8506; 6022/8102 Past use: 1671/8506; 1587/8102 Current use: 554/8506; 490/8102 BMI (kg/m2, median (IQR)): CEE versus placebo: 29.2 (25.7-33.7); 29.2 (25.7-33.5) CEE+MPA versus placebo: 29.2 (25.7-33.7); 29.2 (25.7-33.5) Bilateral oophorectomy (n): CEE versus placebo: 1938/5310; 2111/5429 Age at hysterectomy (y, n): CEE versus placebo: <40: 2100/5310; 2148/5429 40-49: 2280/5310; 2275/5429 50-54: 501/5310; 566/5429 ≥55: 401/ 5310; 404/5429 Results Fractures from overall study population in the intervention phase for both CEE and CEE+MPA trials (hazard ratios with 95% confidence intervals) Vertebral fracture: CEE versus placebo: HR 0.64 (95%CI 0.44-0.93) CEE+MPA versus placebo: HR 0.68 (95%CI 0.48-0.96) All fracture: CEE versus placebo: HR 0.72 (95%CI 0.64-0.80) CEE+MPA versus placebo: HR 0.76 (95%CI 0.69-0.83) Fractures from overall study population in the post intervention phase for both CEE and CEE+MPA trials (hazard ratios with 95% confidence intervals) Hip fracture: CEE versus placebo: HR 1.16 (95%CI 0.85-1.58) CEE+MPA versus placebo: HR 0.88 (95%CI 0.72-1.08) Fractures from overall study population (combined intervention and post intervention phase) for both CEE and CEE+MPA trials (hazard ratios with 95% confidence	(systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-No (only about 81% surviving participants of WHI trials consented to extension pahse participation) A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-No Level of risk- High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were

Study details	Study design	Comparison	Results	Other
and extended poststopping phases of the Women's Health Initiative randomized trials, JAMA, 310, 1353-1368, 2013 Ref Id 294268 Study type Randomised controlled trial followed by observational study Source of funding National Heart, Lung and Blood Institute National Institutes of Health US Department of Health and Human Services Country/ies where the study was carried out USA (multicentre) Study dates Recruitment of participants: 1993-1998 Early			intervals) Hip fracture: CEE versus placebo: HR 0.91 (95%CI 0.72-1.15) CEE+MPA versus placebo: HR 0.81 (95%CI 0.68-0.97) Fractures from overall study (intervention phase), stratified by age for both trials: Hip fracture: 50-59 years: CEE versus placebo: HR 5.01 (95%CI 0.59- 42.91) CEE+MPA versus placebo: HR 0.17 (95%CI 0.02-1.45) 60-69 years: CEE versus placebo: HR 0.47 (95%CI 0.22-1.04) CEE+MPA versus placebo: HR 0.70 (95%CI 0.38-1.27) Fractures as secondary endpoints (stratified by age) for both trials: Vertebral fractures: 50-59 years: CEE versus placebo: HR 0.50 (95%CI 0.17-1.47) CEE+MPA versus placebo: HR 0.38 (95%CI 0.15-0.97) 60-69 years: CEE versus placebo: HR 0.48 (95%CI 0.26-0.89) CEE+MPA versus placebo: HR 0.47 (95%CI 0.26-0.85) All fractures: 50-59 years: CEE versus placebo: HR 0.90 (95%CI 0.72-1.11) CEE+MPA versus placebo: HR 0.82 (95%CI 0.68-1.00) 60-69 years: CEE versus placebo: HR 0.63 (95%CI 0.53-0.75) CEE+MPA versus placebo: HR 0.70 (95%CI 0.61-0.81)	kept 'blind' to treatment allocation-N/a Level of risk: N/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-Not reported C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Unclear D. Detection bias (bias in how outcomes are ascertained, diagnosed or

<table>
<tr><th>Study details</th><th>Study design</th><th>Comparison</th><th>Results</th><th>Other</th></tr>
<tr><td>termination of intervention phase: 2004
Post-interventional follow-up: through September 2010</td><td></td><td></td><td></td><td>verified)
D.1 The study had an appropriate length of follow-up-Yes
D.2 The study used a precise definition of outcome-Yes
D.3 A valid and reliable method was used to determine the outcome-Yes
D.4 Investigators were kept 'blind' to participants' exposure to the intervention-No
D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-No
Level of bias: High</td></tr>
<tr><td>Full citation
Prentice,R.L., Manson,J.E., Langer,R.D., Anderson,G.L., Pettinger,M., Jackson,R.D., Johnson,K.C., Kuller,L.H., Lane,D.S., Wactawski-Wende,J., Brzyski,R., Allison,M., Ockene,J., Sarto,G., Rossouw,J.E., Benefits and risks of postmenopausal hormone therapy when</td><td>Aim of the study
To analyse the effects of CEE and CEE/MPA (particularly longer-term effects), when initiated soon after menopause, on a range of clinical outcomes, including the global index. The analyses used both WHI clinical trial data and combined WHI clinical trial and observational study data.
Inclusion criteria
-To enhance comparablility with the clinical trial eligibility criteria, women</td><td>Details
CEE (0.625mg/daily)
CEE/MPA (0.625mg/daily CEE plus 2.5mg/daily MPA)
placebo/no use of HRT/no prior use of HRT
Methods
Details
-As reported under Anderson et al. 2004 and Manson et al. 2003 with regard to the RCT components;
-In the observational cohort, clinical outcomes were also reported semiannually. Medical record documentation of self-reported outcomes was obtained and diagnoses were confirmed at WHI clinical centres.

Statistical methods:
-"Time from WHI enrollment was the "basic time variable" in Cox regression analyses that stratified data on cohort (clinical trials vs. observational study) and baseline age.
-Confounding in the observational study was addressed by including standard risk factors for each outcome in Cox regression models. The set of risk factors to include was the same as previous reports for CVD and breast cancer and otherwise based on the knowledge and experience of the investigator group, prior to data analysis. They included age, BMI, education, smoking, physical functioning construct, history of treated diabetes, family history of cancer, cholesterol etc.</td><td>Characteristics
Distribution of subjects from both the clinical trials and observational studies, by prior use of HRT and gap time from menopause to first use of HRT among HRT users, 1993-2004

<table><tr><th>Gap time, years</th><th></th><th></th><th></th><th></th><th></th></tr><tr><td>Use of CEE</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>Clinical trials</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td></td><td>No prior HT</td><td>Prior HT</td><td></td><td></td><td></td></tr><tr><td></td><td><5 yr</td><td>5-14 yr</td><td>>=15</td><td><5 yr</td><td>5-14 yr</td></tr><tr><td>No. women (%)</td><td>198 (10%)</td><td>618 (32%)</td><td>1136 (84%)</td><td>2129 (84%)</td><td>294 (12%)</td></tr><tr><td>No. of cases</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>CHD</td><td>2</td><td>22</td><td>59</td><td>76</td><td>8</td></tr></table></td><td>Other information
-According to this study, the effects of CEE and CEE/MPA did not depend significantly on gap time from menopause to first use of HRT for most clinical outcomes considered, either in further analyses of clinical trial data or in combined clinical trail and observational study data analyses.
-The interpretation of these hazard ratios by years from HT initiation among women with or without prior use of HT should be interpreted with caution: there is multiple testing isue. One would expect approximately 3 of the 95% confidence intervals to exclude 1 by chance</td></tr>
</table>

Study details	Study design	Comparison	Results						Other
it is initiated soon after menopause, American Journal of Epidemiology , 170, 12-23, 2009 Ref Id 230128 Study type randomised controlled trial Source of funding NIH Country/ies where the study was carried out USA Study dates 1993-1998 to 2004	from the observational subcohort were required to be without a personal history of breast cancer and to have had a mammogram within 2 years prior to enrollment. -To have a known age at first use of HRT use. Exclusion criteria -As reported under Anderson et al. 2004 and Manson et al. 2003 as the same in/exclusion criteria were used for clinical trials and observtional study at baseline in WHI (besides that the observational cohort was comprised of clinical trial screenees who were either ineligible or unwilling to participate in the clinical trial).	-"Prior hormone therapy" use in the clinical trials and in non-hormone-therapy group in the observational study was defined relative to th time of WHI enrollment. -Prior use for hormone therapy users in the observational study was defined relative to the beginning of the hormone therapy episode that was ongoing at enrollment. Going back in time, a change in hormone regimen or usage gap of 1 year or longer defined a new hormone therapy episode. -Nominal 95% CIs are presented for hazard ratio parameters; Follow-up -As reported under Anderson et al. 2004 and Manson et al. 2003 with regard to the RCT components; -For the observational study, the cohorts were followed through Dec 15, 2004 (CEE) AND Feb 28, 2003 (CEE+MPA), an average follow-up periods of 7.1 yrs and 5.5 yrs, respectively. Sample size CEE clinical trial: Active CEE group: 4493; placebo: 4636 CEE/MPA trial: Active CEE/MPA group: 7679; placebo: 7509 Observational study (women with intact uterus): CEE/MPA group: 6756; No hormone therapy group: 24, 186							alone. Another limitation of the current analyses was that hazard ratio pertaining to 5 or more years from HRT initiation were derived mainly from the observational study. Limitations Study quality NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-Yes (observational study subjects were those who were unwilling to or unsuitable to participate in the clinical trials of WHI, although all participants across studies were selected from the same population) A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes (confounders in the observational study were controlled for in analyses, as reported by the
			Stroke	3	19	46	3	3	
			Global index	15	68	202	308	22	
			Observational study						
				No prior HT	Prior HT				
				<5 yr	5-14 yr	>=15	<5 yr	5-14 yr	
			No. women (%)	6626 (76%)	1454 (17%)	597 (7%)	1662 (87%)	213 (11%)	
			No. of cases						
			CHD	104	28	15	31	6	
			Stroke	119	39	13	42	7	
			Global index	689	164	75	203	29	
			Gap time, years						
			Use of CEE/MPA						
			Clinical trials						
				No prior HT	Prior HT				
				<5 yr	5-14 yr	>=15	<5 yr	5-14 yr	
			No. women (%)	952 (17%)	2338 (43%)	2160 (40%)	1864 (84%)	302 (14%)	
			No. of cases						
			CHD	10	35	71	43	5	
			Stroke	6	37	53	28	3	
			Global	54	205	281	171	29	

Study details	Study design	Comparison	Results	Other
			index Observational study No prior HT: <5 yr: No. women (%) 4257 (75%); 5-14 yr: 1115 (20%); >=15: 338 (6%) Prior HT: <5 yr: 916 (88%); 5-14 yr: 113 (11%) No. of cases CHD: 30; 13; 7; 8; 2 Stroke: 27; 7; 3; 8; 0 88: 340; 88; 41; 85; 13 Results Risk of hip fracture in relation to use of CEE, HR (95%CI): By time from menopause to first use of HT: Hip fracture: < 5 years: No prior HT: N/a Prior HT: 0.54 (0.30-0.99) >5 years (just for information giving in evidence table): No prior HT: 0.87 (0.48-1.60) Prior HT: N/a P for gap time interaction: 0.58 Risk of hip fracture in relation to use of CEE/MPA, HR (95%CI): By time from menopause to first use of HT: Hip fracture: < 5 years: No prior HT: N/a Prior HT: 0.25 (0.09-0.74) >5 years (just for information giving in evidence table): No prior HT: 0.81 (0.53-1.24) Prior HT: N/a P for gap time interaction: 0.04 Risk of hip fracture in relation to use of CEE and	authors) A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Unclear Level of risk-High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/a B.2 Participants receiving care were kept 'blind' to treatment allocation-N/a B.3 Individuals administering care were kept 'blind' to treatment allocation-N/a Level of risk: n/a C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-No, slight differences across trials and observationl study with regard to early-stopped times) C.2a How many participants did not complete treatment in

Study details	Study design	Comparison	Results	Other
			CEE/MPA (among women who began HRT immediately following menopause), from combined analysis of clinical trial and observational study data, HR (95%CI): (subjects the following analyses were limited to those who adhered to their hormone therapy regime from both the clinical trials and observational studies, because of the high drop-out rates in trials and the data from the observational study was combined) By year from HT initiation among women with no prior use of HT: Hip fracture: <2 years: CEE: 0.46 (0.04-4.88) CEE/MPA: 0.35 (0.10-1.17) 2-4 years: CEE: 0.53 (0.11-2.51) CEE/MPA: 0.33 (0.10-1.10) >=5 years (just for information giving in the evidence table) CEE: 0.69 (0.19-2.56) CEE/MPA: 0.22 (0.07-0.71) By year from "current" HT episode among women with prior use of HT: Hip fracture: <2 years: CEE: 0.60 (0.11-3.24) CEE/MPA: 0.26 (0.05-1.25) 2-4 years: CEE: 0.13 (0.02-1.08) CEE/MPA: 0.26 (0.05-1.25) >=5 years: CEE: 0.54 (0.16-1.76) CEE/MPA: 0.43 (0.09-2.07)	each group?- High drop-out in the clinical trials as reported previously under Anderson et al. 2004 and Manson et al. 2003; for the observational cohort, drop-out rate was not reported in the current analysis) C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-Unclear (reasons not investigated) C.3a For how many participants in each group were no outcome data available?- As reported in Anderson et al. 2004 and Manson et al. 2003 with regard to clinical trials; for the observational study, data not reported) C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-Yes Level of risk: High D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of

Study details	Study design	Comparison	Results	Other
				follow-up-Unclear (all subcohorts were stopped early due to ethical reasons) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-Yes D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-Unclear (details about the observational study not reported) Level of bias: Unclear Indirectness Does the study match the review protocol in terms of; Population: Yes Outcome: Yes Indirectness: Some
Full citation Heiss,G., Wallace,R., Anderson,G. L., Aragaki,A., Beresford,S. A.A., Brzyski,R., Chlebowski, R.T., Gass,M., Lacroix,A., Manson,J.E., Prentice,R.L.	Aim of the study To report health outcomes at three years (mean 2.4 years of follow-up) after intervention was stopped Inclusion criteria Post-menopausal women aged 50-79 with an intact uterus, who gave written informed consent	Details CEE+MPA (0.625mg combined equine oestrogen+ 2.5mg medroxyprogesterone acetate) Placebo Methods Intervention phase: Women were randomly assigned to receive HRT or placebo and were followed up for 5.6 years. Semi-annual telephone contact by the clinic or annual visit to the WHI clinic using a standardised form was collected on symptoms, adverse events, adherence to study pills, and potential trial clinical outcomes. Potential outcomes were verified by obtaining medical records and death certificates and reviewed by a physician who was blinded to the treatment assignment.	Characteristics Age at baseline (mean, SD), years: CEE+MPA: 63.1 (7.1) Placebo: 63.3 (7.1) BMI (n): <25: CEE+MPA: 2430; placebo: 2373 25-<30: CEE+MPA: 2826; placebo: 2689 ≥30: CEE+MPA: 2760; placebo:2568 Hypertension (n): CEE+MPA: 2851; placebo: 2772 Years since menopause (n): <5 years: CEE+MPA: 1268; placebo: 1167 5-<10 years: CEE+MPA: 1405; placebo:1432 10-<15 years: CEE+MPA: 1545; placebo: 1494 ≥15 years: CEE+MPA: 3066; placebo: 3027	Other information Limitations Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. Unclear. Attempts were made within the design or analysis to balance the comparison groups for potential confounders.

Study details	Study design	Comparison	Results	Other
, Rossouw,J., Stefanick,M. L., Health risks and benefits 3 years after stopping randomized treatment with estrogen and progestin, JAMA - Journal of the American Medical Association, 299, 1036-1045, 2008 Ref Id 295998 Study type Cohort study (From WHI randomised controlled trial CEE+MPA vs placebo Source of funding National Heart, Lung, and Blood Institute, NIH, Department of Health and Human Services Country/ies where the study was carried out USA (multicentre)	Exclusion criteria Reported in previous reports from WHI	Analysis of the outcomes was performed at 5.2 years. Post-intervention phase: Intervention was terminated early (July 2002). Pre-defined end of trial was March 2005. (2002-2005 defines post-intervention phase). Data was collected semi-annually, with annual mammography surveillance. Statistical analysis: Baseline characteristics of women in CEE+MPA versus placebo trial with any post-intervention data were compared by X2 or t test. Annualised rates of events in intervention and post intervention phase, and overall were estimated by dividing the number of events by the corresponding survival time in each phase. ITT and time to event was applied. Hazard ratios (HR) were estimated from Cox proportional hazard analyses stratified by age, prior disease if appropriate, and randomisation assignment in the dietary modification trial. A formal test of whether HR in the clinical trial was equal to HR in the post intervention phase. Sensitivity analysis was performed to assess risk among women who had been adherent to study medication (≥80%) during intervention phase of the trial. For comparison, participants adherent at end of intervention phase were included in the post intervention HR estimation using inverse of the participants estimated adherence probability as a weighting factor. The probabilities were estimated by logistic regression including baseline variables of age, ethnicity, education, BMI, smoking, self-reported general health, night sweats, hot flashes, breast tenderness and treatment assignment (at year 1). Sample size Number (n) alive at follow-up: CEE+MPA: 8052 Placebo: 7678	HRT usage status (n): Never used: CEE+MPA: 5929; placebo: 5710 Past user: CEE+MPA: 1589; placebo: 1492 Current user: CEE+MPA: 530; placebo: 473 HRT duration (n): < 5 years: CEE+MPA: 1468; placebo: 1394 5-<10 years: CEE+MPA: 405; placebo: 329 ≥10 years: CEE+MPA: 250; placebo:244 Results During clinical trial phase, N: 16,608 All fractures CEE+MPA: 741/8506; placebo:903/8102; HR: 0.76 (95%CI 0.69-0.83) Hip fractures CEE+MPA:53/8506; placebo:75/8102; HR: 0.67 (95%CI 0.47-0.95) Vertebral fractures CEE+MPA:56/8506; placebo:78/8102; HR: 0.68 (95%CI 0.48-0.96) Other osteoporotic fractures CEE+MPA:650/8506; placebo:800/8102; HR: 0.75 (95%CI 0.68-0.83) During post intervention phase, N: 15,730 All fractures CEE+MPA:337/8052; placebo:346/7678; HR: 0.91 (95%CI 0.78-1.06) Hip fractures CEE+MPA: 54/8052; placebo:57/7678; HR: 0.92 (95%CI 0.64-1.34) Vertebral fractures CEE+MPA:46/8052; placebo:47/7678; HR: 0.96 (95%CI 0.64-1.44) Other osteoporotic fractures CEE+MPA:267/8052; placebo:285/7678; HR 0.87 (95%CI 0.74-1.03) Overall combined phases All fractures CEE+MPA:1078/8506; placebo:1249/8102; HR: 0.80 (95%CI 0.73-0.86) Hip fractures CEE+MPA:107/8506; placebo:132/8102; HR: 0.78 (95%CI 0.60-1.00) Vertebral fractures CEE+MPA:102/8506; placebo:125/8102; HR: 0.78 (95%CI 0.60-1.01)	Yes. The groups were comparable at baseline, including all major confounding and prognostic factors. Unclear - only reported as fracture cases compared to non-fracture cases, rather than HRT use compared to no HRT use. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias

Study details	Study design	Comparison	Results	Other
Study dates Recruitment of participants:1993-1998 Post-intervention commenced: 2002			Other osteoporotic fractures CEE+MPA:917/8506:placebo:1085/8102; HR:0.78 (0.72-0.85)	The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Unclear. Investigators were kept 'blind' to participants' exposure to the intervention. No. Investigators were kept 'blind' to other important confounding and prognostic factors. No.
Full citation Effects of hormone therapy on bone mineral density: results from the postmenopausal estrogen/progestin interventions (PEPI) trial. The Writing Group for the PEPI, JAMA, 276, 1389-1396, 1996 Ref Id 294605 Study type Randomized controlled trial. Source of funding Research	Aim of the study To assess the effects of hormone replacement therapy on bone mineral density at the spine and hip of postmenopausal women. Inclusion criteria Surgically or naturally menopausal women (longer than 1 year, but less than 10 years since LMP) aged 45 to 64. Not taking oestrogens or progestins for at least 2 months prior to the first screening visit (> 4 months before randomization).	Details Participants were assigned to one of the following regimes in 28 day cycles: 1. placebo 2. active treatment arms, which included four separate regimes: • conjugated equine estrogens (CEE) 0.625mg/day • CEE 0.625mg/day plus medroxyprogesterone acetate (MPA) 10mg/day for days 1 to 12 • CEE 0.625mg/day plus MPA 2.5mg/day • CEE 0.625mg/day plus micronized progesterone 200mg/day for day 1 to 12 For the purposes of this analysis data for the four active treatment arms were combined. Methods After the first randomization visit, participants returned 3 times during the first year and biannually for the remaining 2 years. Symptoms, occurrence of vaginal bleeding, medications used, adherence to medications, adverse experiences (including fractures), blood pressure, weight and height were assessed at each visit. Sample size N = 875 n = 174 placebo group n = 701 active treatment group	Characteristics Average age 56.1 years No significant differences in prior menopausal hormone use, smoking status, ethnicity, physical activity or baseline bone mineral density between the groups. Results Risk of any fracture in HRT groups compared to placebo groups unadjusted RR (95% CI): 0.66 (0.31 to 1.40)	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Unclear. There was adequate concealment of allocation. Unclear. The groups were comparable at baseline. Yes. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Yes. Individuals administering care were kept 'blind' to treatment allocation. Yes. Attrition bias

Study details	Study design	Comparison	Results	Other
grants from the National Heart, Lung and Blood Institute; the National Institute of Child Health and Human Development; the National Institute of Arthritis and Musculoskel etal and Skin Diseases; the National Institute of Diabetes and Digestive and Kidney Diseases and the National Institute on Aging. Support was also provided by General Clinical Research Center Grants (University of California, Los Angeles; University of California, SanDiego and University of Iowa). Study medications were provided by	If treated with thyroid hormone replacement, to have been on a stable dose for at least 3 months prior to initial screening. Exclusion criteria Extreme hyperlipidaemia, marked obesity, severe hypertension, recent myocardial infarction, congestive heart failure, stroke or TIA, anti-arrythmia medication use, diabetes mellitus requiring insulin, prior breast or endometrial cancer, melanoma, any non-basal cell skin cancer in the previous five years, an elevated thyroid stimulating hormone concentration, a history of trauma to the lower spine or hip fracture, chronic steroid use and severe menopausal symptoms.			All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 11 placebo group, n = 28 HRT groups. The groups were comparable for treatment completion. Yes. For how many participants in each group were outcome data not available? n = 11 placebo group, n = 28 HRT groups. The groups were comparable with respect to the availability of outcome data. Yes. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.

Study details	Study design	Comparison	Results	Other
Wyerth-Ayerst Laboratories, Philadelphia, Pa (conjugated equine estrogens), The Upjohn Company, Kalamazoo, Mich (medroxyprogesterone acetate) and Schering-Plough Research Institute, Kenilworth, NJ (micronized progesterone). Country/ies where the study was carried out USA Study dates Randomization occurred between December 1989 and February 1991. Trial duration was for three years.				
Full citation Bagger,Y.Z., Tanko,L.B., Alexanderse n,P., Hansen,H.B.,	Aim of the study To clarify whether 2 to 3 years of HRT administered in the early postmenopausal	Details Women who completed 2 to 3 years of treatment with HRT (during the original RCTs) and then discontinued treatment were compared to those who were assigned to placebo for the original studies. Time since cessation is unclear in the article, but presumably	Characteristics Characteristics at time of follow up: Short term HRT group: Age, years (mean ± SD): 65.2 (3.7) BMI, kg/m² (mean ± SD): 26.3 (4.4) Placebo group:	Other information Limitations Study quality Selection bias The method of allocation to treatment groups was

Study details	Study design	Comparison	Results	Other
Mollgaard,A., Ravn,P., Qvist,P., Kanis,J.A., Christiansen, C., Two to three years of hormone replacement treatment in healthy women have long-term preventive effects on bone mass and osteoporotic fractures: the PERF study, Bone, 34, 728-735, 2004 Ref Id 230899 Study type Prospective cohort study (observational follow up of participants in previous RCTs). Source of funding Not reported. Country/ies where the study was carried out Denmark Study dates Original RCTs conducted	years provide long-term benefits in terms of preventing bone loss and osteoporotic fractures. Inclusion criteria Older than 45 years of age, passed a natural menopause at least 6 months previously, and had normal bone mineral content or bone mineral density. Exclusion criteria Prior treatment with estrogens or other drugs. Chronic disease known to influence bone metabolism.	was at least 7 years (RCTs conducted until 1993 at the latest, follow up commenced in 2000). Methods At follow up, lateral X-rays of the thoracic and lumbar spine were taken. Digital measurements of morphological changes were taken to determine radiographic vertebral fractures. Information on the incidence of non-vertebral fractures was collected at follow up. Sample size N = 263 n = 155 short term HRT use n = 108 no HRT use	Age, years (mean ± SD): 64.5 (3.3) BMI, kg/m² (mean ± SD): 25.8 (4.1) Results Risk of vertebral fracture in women who took short term HRT compared to women who took placebo: Adjusted OR (95% CI): 0.47 0.24 to 0.93) Risk of nonvertebral fracture in women who took short term HRT compared to women who took placebo: Adjusted OR (95% CI): 0.68 (0.30 to 1.60) Risk of any fracture in women who took short term HRT compared to women who took placebo: Adjusted OR (95% CI): 0.48 (0.26 to 0.88) Adjusted for age, baseline forearm bone mineral content and spine bone mineral density.	unrelated to potential confounding factors. Yes. Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline, including all major confounding and prognostic factors. Yes. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Unclear. Individuals administering care were kept 'blind' to treatment allocation. Unclear. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? Not reported. The groups were

Study details	Study design	Comparison	Results	Other
between 1977 and 1993. Follow up conducted during 2000 and 2001. Study duration up to 24 years.				comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Banks,E., Beral,V., Reeves,G., Balkwill,A., Barnes,I., Fracture Incidence in Relation to the Pattern of Use of Hormone Therapy in Postmenopausal Women, Journal of the American Medical Association, 291, 2212-2220, 2004 Ref Id 295564	Aim of the study To investigate the effects of different patterns of hormone therapy use on fracture incidence. Inclusion criteria Postmenopausal women aged 50 to 69 years. Exclusion criteria Not reported.	Details Comparison was made between women who reported use of HRT baseline and those reporting no use of HRT at baseline. Methods Women completed a baseline questionnaire regarding use of HRT at recruitment. The follow up questionnaire included questions on incident fractures over the follow up period. Sample size N = 138737 n = 5197 with fracture n = 133540 with no fracture	Characteristics Women sustaining a fracture Age 50-54 (%): 22.3 Age 55-59 (%): 36.3 Age 60 to 64 (%): 37.2 Age 65 to 69 (%): 4.2 BMI < 25 (%): 46.6 Women not sustaining a fracture Age 50-54 (%): 26.3 Age 55-59 (%): 38.0 Age 60 to 64 (%): 32.4 Age 65 to 69 (%): 3.3 BMI < 25 (%): 48.1 Results Risk of fracture in current users of HRT compared with never users Adjusted relative risk (95% CI): 0.62 (0.58 to 0.66) Risk of fracture in past users of HRT compared with never users (during the first year of the study) Adjusted relative risk (95% CI): 1.07 (0.95 to 1.22)	Other information Limitations Use of HRT was only reported in the baseline questionnaire, not the follow up, therefore "current" and "never" users of HRT may have changed status by the time of follow up. Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. Unclear. Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes.

Study details	Study design	Comparison	Results	Other
Study type Prospective cohort study. Source of funding UK Medical Research Council Cancer Research UK UK National Health Service Breast Screening Programme Country/ies where the study was carried out UK Study dates Recruitment from June 1996 to March 1998. Follow up for 1.9 to 3.9 years.			Duration of use of HRT: Risk of fracture in current users of HRT for less than 1 year, compared with never users Adjusted relative risk (95% CI): 0.75 (0.60 to 0.93) Risk of fracture in current users of HRT for 1 to 4 years, compared with never users Adjusted relative risk (95% CI): 0.66 (0.60 to 0.74) Risk of fracture in current users of HRT for 5 to 9 years, compared with never users Adjusted relative risk (95% CI): 0.58 (0.53 to 0.65) Risk of fracture in current users of HRT for ≥ 10 years, compared with never users Adjusted relative risk (95% CI): 0.57 (0.50 to 0.66) Recent use of HRT: Risk of fracture in past users of HRT, ceasing use within the past year, compared with never users Adjusted relative risk (95% CI): 1.09 (0.91 to 1.30) Risk of fracture in past users of HRT, ceasing use between 1 and 2 years ago, compared with never users Adjusted relative risk (95% CI): 0.96 (0.85 to 1.10) Risk of fracture in past users of HRT, ceasing use between 3 and 4 years ago, compared with never users Adjusted relative risk (95% CI): 1.09 (0.93 to 1.28) Risk of fracture in past users of HRT, ceasing use 5 or more years ago, compared with never users Adjusted relative risk (95% CI): 1.10 (0.97 to 1.23) Adjusted for age, region, socioeconomic status, time since menopause, BMI and physical activity.	The groups were comparable at baseline, including all major confounding and prognostic factors. Unclear - only reported as fracture cases compared to non-fracture cases, rather than HRT use compared to no HRT use. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an

Study details	Study design	Comparison	Results	Other
				appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Unclear. Investigators were kept 'blind' to participants' exposure to the intervention. No. Investigators were kept 'blind' to other important confounding and prognostic factors. No.
Full citation Barrett-Connor,E., Wehren,L.E., Siris,E.S., Miller,P., Chen,Y.T., Abbott,3rd.T. A., Berger,M.L., Santora,A.C., Sherwood,L. M., Recency and duration of postmenopausal hormone therapy: effects on bone mineral density and fracture risk in the National Osteoporosis Risk Assessment (NORA)	Aim of the study To evaluate bone mineral density and 1 year fracture risk in postmenopausal women stratified by duration and recency of HRT. Inclusion criteria Postmenopausal women aged 50 years or older. At least 6 months postmenopausal. Exclusion criteria Previous diagnosis of osteoporosis, BMD testing in the preceding 12 months or current use of bone-specific medications.	Details Current use of HRT, and past use of HRT was compared to never use of HRT with regard to fracture risk. Methods Information regarding HRT use was collected by standard self-administered questionnaire. One year incident fractures of the wrist, rib, spine and hip were identified from follow up questionnaires. Participants reporting four or more new fractures (likely to reflect major trauma) were excluded from analyses. Sample size N = 170852 n = 68258 never used HRT n = 79569 current users of HRT n = 22755 previous users of HRT	Characteristics Median age 63 years Mean BMI 27.7 ± 5.9 kg/m² Mean number of years since menopause 18.1 ± 11.1 Mean T score -0.86 ± 1.15 Results Current use and duration of use: Risk of osteoporotic fracture in current users of HRT for ≤ 5 years compared to never users adjusted OR (95% CI): 0.75 (0.65 to 0.88) Risk of osteoporotic fracture in current users of HRT for 6 to 10 years compared to never users adjusted OR (95% CI): 0.71 (0.59 to 0.84) Risk of osteoporotic fracture in current users of HRT for ≥ 10 years compared to never users adjusted OR (95% CI): 0.75 (0.66 to 0.85) Previous use and duration of use Risk of osteoporotic fracture in previous users of HRT for ≤ 5 years (stopped ≤ 5 years ago) compared to never users adjusted OR (95% CI): 0.90 (0.71 to 1.15) Risk of osteoporotic fracture in previous users of HRT for 6 to 10 years (stopped ≤ 5 years ago) compared to never users adjusted OR (95% CI): 0.98 (0.61 to 1.57) Risk of osteoporotic fracture in previous users of HRT for ≥ 10 years (stopped ≤ 5 years ago) compared to never users adjusted OR (95% CI): 1.32 (0.93 to 1.87)	Other information Limitations Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. Unclear. Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline, including all major confounding and prognostic factors. No - differences were noted in BMI, years postmenopausal, exercise, alcohol intake, caffeine intake, diuretic use, previous fracture, calcium cupplements and family history of osteoporosis.

Study details	Study design	Comparison	Results	Other
study, Menopause (New York, N.Y.), 10, 412-419, 2003 Ref Id 295578 Study type Prospective cohort study. Source of funding Not reported. Country/ies where the study was carried out USA Study dates Cohort identified in 1997. Study duration 1 year.			Risk of osteoporotic fracture in previous users of HRT for ≤ 5 years (stopped > 5 years ago) compared to never users adjusted OR (95% CI): 1.09 (0.92 to 1.29) Risk of osteoporotic fracture in previous users of HRT for 6 to 10 years (stopped > 5 years ago) compared to never users adjusted OR (95% CI): 1.39 (0.99 to 1.94) Risk of osteoporotic fracture in previous users of HRT for ≥ 10 years (stopped > 5 years ago) compared to never users adjusted OR (95% CI): 1.06 (0.72 to 1.56) Adjusted for age, previous fracture, health status, maternal history of fracture and cortisone use.	Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Unclear. Investigators were kept

Study details	Study design	Comparison	Results	Other
				'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Bjarnason,N. H., Christiansen, C., Early response in biochemical markers predicts long-term response in bone mass during hormone replacement therapy in early postmenopausal women, Bone, 26, 561-569, 2000 Ref Id 266115 Study type Randomised controlled trial. Source of funding Schering AG. Country/ies where the study was carried out Denmark Study dates Not reported.	Aim of the study To investigate the effect of short term and low dose HRT. Inclusion criteria Healthy women within 1 to 6 years of menopause, with an intact uterus. Exclusion criteria Treatment with medication known to affect bone metabolism, clinical or laboratory evidence of confounding diseases.	Details Fracture rates in women taking HRT were compared to those in women taking placebo. Methods Women were randomised to daily oral treatment with either 2mg estradiol sequentially combined with 25µg gestodene, 2mg estradiol sequentially combined with 50µg gestodene, 1mg estradiol sequentially combined with 25µg gestodene, 1mg estradiol continuously combined with 25µg gestodene, or placebo. For the purposes of this analysis all four HRT treatment groups were combined. The trial duration was 3 years. Sample size N = 278 n = 222 HRT n = 56 placebo	Characteristics HRT group: Age, years (mean): 53.5 BMD spine, g/m² (mean): 0.966 Placebo group: Age, years (mean): 53.6 BMD spine, g/m² (mean): 0.952 Results Taken from data supplied by the authors to Torgerson and Bell-Syer for their meta-analysis (Torgerson and Bell-Syer 2001). Data only includes women who completed the trial, therefore per-protocol analysis, not intention to treat. Risk of non-vertebral fracture in women taking HRT compared to those taking placebo: unadjusted relative risk (95% CI): 1.46 (0.17 to 12.72)	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Unclear. There was adequate concealment of allocation. Unclear. The groups were comparable at baseline. Yes. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Yes. Individuals administering care were kept 'blind' to treatment allocation. Yes. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 15 placebo, n = 110

Study details	Study design	Comparison	Results	Other
Trial duration 3 years.				HRT group. The groups were comparable for treatment completion. No - fewer drop-outs in placebo group. For how many participants in each group were outcome data not available? n = 15 placebo, n = 110 HRT group, but not included in risk analysis. The groups were comparable with respect to the availability of outcome data. No - fewer drop-outs in placebo group. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Unclear. A valid and reliable method was used to determine the outcome. Unclear. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Cauley,J.A., Robbins,J., Chen,Z., Cummings,S. R., Jackson,R.D.	Aim of the study To determine the effects of treatment with oestrogen alone, or oestrogen plus progesterone on	Details Fracture rates were compared in women taking oestrogen only preparations or oestrogen plus progestin preparations and those taking placebo. Methods Two parallel trials were conducted - one in hysterectomized women, and the other in women with an intact uterus.	Characteristics Oestrogen plus progestin arm: HRT group: Age, years (mean ± SD): 63.2 ± 7.10 BMI, kg/m² (mean ± SD): 28.5 ± 5.80 Previous use of HRT (%): 26.2 < 10 years since menopause (%): 36.23	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used to allocate participants to

Study details	Study design	Comparison	Results	Other
, LaCroix,A.Z., LeBoff,M., Lewis,C.E., McGowan,J., Neuner,J., Pettinger,M., Stefanick,M. L., Wactawski-Wende,J., Watts,N.B., Effects of estrogen plus progestin on risk of fracture and bone mineral density: the Women's Health Initiative randomized trial, JAMA : the journal of the American Medical Association, 290, 1729-1738, 2003 Ref Id 295677 Study type Randomised controlled trial, followed by period of observational follow up post-intervention. Source of funding National Heart, Lung and Blood	a variety of important chronic diseases of older women. Inclusion criteria Oestrogen only arm: Postmenopausal women with prior hysterectomy, aged 50 to 79 years. Oestrogen plus progestin arm: Postmenopausal women with an intact uterus, aged 50 to 79 years. Exclusion criteria Use of tamoxifen. Women who used postmenopausal hormones required a three month washout period prior to study entry.	Women with an intact uterus were randomised to treatment with either placebo, or conjugated equine oestrogen 0.625mg/day and medroxyprogesterone acetate 2.5mg/day as a single tablet. Follow up was for an average of 5.6 years. Women with a previous hysterectomy were randomised to treatment with either placebo or conjugated equine oestrogens 0.625mg/day. Follow up was for an average of 7.1 years. Both trials were terminated prematurely under the advice of the trial steering commitee. However, participants have been followed up as part of a subsequent observational study to assess the longer term effects of treatment after stopping hormones. Sample size Oestrogen plus progestin arm: N = 16608 n = 8506 HRT n = 8102 placebo Oestrogen alone arm: N = 10739 n = 5310 HRT n = 5429 placebo	Placebo group: Age, years (mean ± SD): 63.3 ± 7.10 BMI, kg/m² (mean ± SD): 28.5 ± 5.90 Previous use of HRT (%): 25.7 < 10 years since menopause (%): 36.12 Oestrogen alone arm: HRT group: Age, years (mean ± SD): 63.6 ± 7.3 BMI, kg/m² (mean ± SD): 30.1 ± 6.1 Previous use of HRT (%): 47.8 < 10 years since menopause (%): 18.4 Placebo group: Age, years (mean ± SD): 63.6 ± 7.3 BMI, kg/m² (mean ± SD): 30.1 ± 6.2 Previous use of HRT (%): 49 < 10 years since menopause (%): 17.6 Results Fracture risks during treatment Oestrogen plus progesterone arm: Risk of hip fracture in HRT group compared to placebo adjusted hazard ratio (95% CI): 0.67 (0.47 to 0.96) Risk of wrist fracture in HRT group compared to placebo adjusted hazard ratio (95% CI): 0.71 (0.59 to 0.85) Risk of vertebral fracture in HRT group compared to placebo adjusted hazard ratio (95% CI): 0.68 (0.48 to 0.96) Risk of any fracture in HRT group compared to placebo adjusted hazard ratio (95% CI): 0.76 (0.69 to 0.83) Risk of non-vertebral fracture in HRT group compared to placebo unadjusted relative risk (95% CI): 0.79 (0.72 to 0.86) Risk of hip fracture in women aged 50 to 59 years in HRT group compared to placebo adjusted hazard ratio (95% CI): 0.17 (0.02 to 1.45) Risk of hip fracture in women aged 60 to 69 years in HRT group compared to placebo adjusted hazard ratio (95% CI): 0.70 (0.38 to 1.27) Risk of hip fracture in women aged 70 to 79 years in HRT group compared to placebo adjusted hazard ratio (95% CI): 0.71 (0.46 to 1.12) Oestrogen alone arm: Risk of hip fracture in HRT group compared to placebo adjusted hazard ratio (95% CI): 0.65 (0.45 to 0.94) Risk of wrist fracture in HRT group compared to placebo adjusted hazard ratio (95% CI): 0.58 (0.47 to 0.72)	treatment groups. Yes. There was adequate concealment of allocation. Yes. The groups were comparable at baseline. Yes. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Yes. Individuals administering care were kept 'blind' to treatment allocation. Yes. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). No. The study was stopped earlier than the pre-specified end date of the intervention. How many participants did not complete treatment in each group? 544 in CEE+MPA group; 482 in placebo group. The groups were comparable for treatment completion. No - fewer drop-outs in placebo group. For how many participants in each group were outcome data not available? 544 in treatment group; 482 in placebo group The groups were

Study details	Study design	Comparison	Results	Other
Institute. Drug treatment and placebo tablets were provided by Wyeth. Country/ies where the study was carried out USA Study dates Trial recruitment began in September 1993. Trial intervention was terminated on July 7th 2002, but longitudinal observational follow up continues (as a cohort study).			Risk of vertebral fracture in HRT group compared to placebo adjusted hazard ratio (95% CI): 0.64 (0.44 to 0.93) Risk of any fracture in HRT group compared to placebo adjusted hazard ratio (95% CI): 0.71 (0.64 to 0.80) Risk of non-vertebral fracture in HRT group compared to placebo unadjusted relative risk (95% CI): 0.73 (0.66 to 0.82) Risk of hip fracture in women aged 50 to 59 years in HRT group compared to placebo adjusted hazard ratio (95% CI): 5.01 (0.59 to 42.91) Risk of hip fracture in women aged 60 to 69 years in HRT group compared to placebo adjusted hazard ratio (95% CI): 0.47 (0.22 to 1.04) Risk of hip fracture in women aged 70 to 79 years in HRT group compared to placebo adjusted hazard ratio (95% CI): 0.65 (0.42 to 1.00) Data obtained from a series of publications originating from the WHI trial.	comparable with respect to the availability of outcome data. Yes. Detection bias The study had an appropriate length of follow up. Yes The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. No. Investigators were kept 'blind' to other important confounding and prognostic factors. No.
Full citation Cherry,N., Gilmour,K., Hannaford,P., Heagerty,A., Khan,M.A., Kitchener,H., McNamee,R., Elstein,M., Kay,C., Seif,M., Buckley,H., ESPRIT team., Oestrogen therapy for	Aim of the study To assess the effect of unopposed oestradiol valerate on risk of another cardiac event or death in postmenopausal women who had just survived their first myocardial infarction. Inclusion criteria Women aged 50 to 69 years admitted to	Details Outcomes were compared between women taking HRT and those taking placebo tablets. Methods Women were randomly allocated to receive either 2mg oestradiol valerate or placebo, taken as one tablet daily for 2 years. Participants and investigators were blinded to treatment allocation. Fracture dated was collected by questionnaires sent to family doctors as an adverse event. Sample size N = 1017 n = 513 HRT n = 504 placebo	Characteristics HRT group Age at admission to hospital, years (mean ± SD): 62.3 ± 5.2 BMI, kg/m² (mean ± SD): 26.8 ± 5.1 Previous fracture in last 10 years (%): 14% Placebo group Age at admission to hospital, years (mean ± SD): 62.9 ± 4.9 BMI, kg/m² (mean ± SD): 26.7 ± 5.3 Previous fracture in last 10 years (%): 19% Results Risk of any fracture in HRT group compared to placebo group: unadjusted relative risk (95% CI): 0.60 (0.29 to 1.26)	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Yes. There was adequate concealment of allocation. Yes. The groups were comparable at baseline. Yes. Performance bias The comparison groups received the same care

Study details	Study design	Comparison	Results	Other
prevention of reinfarction in postmenopausal women: a randomised placebo controlled trial, Lancet, 360, 2001-2008, 2002 Ref Id 229092 Study type Randomised controlled trial. Source of funding UK National Health Service Research and Development Programme on Cardiovascular Disease and Stroke. University of Manchester. Schering Health Care Ltd. Country/ies where the study was carried out England and Wales Study dates July 1996 and February 2000. Trial duration	coronary care units or general medical wards with a diagnosis of myocardial infarction, in participating hospitals for the duration of the study. Discharged alive from hospital within 31 days of admission. Exclusion criteria Previous myocardial infarction (prior to the index event). Use of HRT or vaginal bleeding in the 12 months prior to admission. History of breast, ovarian or endometrial carcinoma. Active thrombophlebitis, or a history of deep vein thrombosis or pulmonary embolus. Acute or chronic liver disease, Rotor syndrome, Dubin-Johnson syndrome or severe renal disease.			apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Yes. Individuals administering care were kept 'blind' to treatment allocation. Yes. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 184 placebo, n = 294 HRT. The groups were comparable for treatment completion. No - more women in the HRT group did not comply with treatment, due to vaginal bleeding. For how many participants in each group were outcome data not available? None. The groups were comparable with respect to the availability of outcome data. Yes. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome.

Study details	Study design	Comparison	Results	Other
2 years.				Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Yes. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Delmas,P.D., Confavreux, E., Garnero,P., Fardellone,P., De Vernejoul,M. C., Cormier,C., Arce,J.C., A combination of low doses of 17 beta-estradiol and norethisterone acetate prevents bone loss and normalizes bone turnover in postmenopausal women, Osteoporosis International, 11, 177-187, 2000 Ref Id 231349 Study type Randomised controlled trial. Source of	Aim of the study To investigate the effect of 17β oestradiol in combination with low doses of norethisterone acetate on bone mineral density at the lumbar spine. Inclusion criteria Aged 45 to 65 years with a lumbar spine BMD T score between -2 and +2 (within 2 SD of the mean value for healthy young adult women). Postmenopausal, as defined by cessation of menstrual bleeding for at least 1 year with oestradiol levels ≤ 30 pg/ml and FSH levels > 40 IU/l. Exclusion criteria Endometrial thickness > 4mm. Known or suspected past history of breast cancer or	Details BMD and fracture incidence was compared between the placebo group and those taking HRT. Methods Women were randomly assigned to one of three treatment groups: placebo, oestradiol 1mg with norethisterone acetate 0.25mg daily, or oestradiol 1mg with norethisterone 0.5mg daily. All women received a daily calcium supplement of 500mg. Trial duration was 2 years. Method of identification of vertebral fractures unclear, as data obtained from meta-analysis (see results section). Sample size N = 135 n = 90 HRT n = 45 placebo	Characteristics Age, years (range): 58 (47 to 65) Mean time from last menses: 9 years Results Risk of non-vertebral fracture in HRT group compared to placebo group unadjusted relative risk (95% CI): 0.65 (0.02 to 2.68) N.B. fracture data obtained from existing meta-analysis of HRT and nonvertebral fractures (Torgerson and Bell-Syer, 2001) - data obtained for this meta-analysis by direct contact with the authors, rather than published data.	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Unclear. There was adequate concealment of allocation. Unclear. The groups were comparable at baseline. Yes. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Yes. Individuals administering care were kept 'blind' to treatment allocation. Unclear. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants

Study details	Study design	Comparison	Results	Other
funding Novo Nordisk. Country/ies where the study was carried out France Study dates Not reported. Trial duration 2 years.	oestrogen dependent cancer. Liver diseases, active or past history of VTE, thromboembolic disorders or cerebrovascular accidents, abnormal vaginal bleeding of unknown aetiology, pituitary tumour, diabetes mellitus, unstable thyroid diseases, congestive heart failure, angina pectoris, arrythmia, myocardial infarction, systolic blood pressure > 170 mmHg and/or diastolic blood pressure > 100mmHg, renal failure, oestrogen/progestogen treatment within the last 6 months, fluoride treatment for more than 6 months (or less than 6 months duration but within the past 6 months), more than 2 courses of bisphosphonate treatment and/or washout of less than 6 months, chronic systemic			did not complete treatment in each group? n = 12 placebo, n = 32 HRT group. The groups were comparable for treatment completion. Yes. For how many participants in each group were outcome data not available? n = 12 placebo, n = 32 HRT group. The groups were comparable with respect to the availability of outcome data. Yes. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Unclear. A valid and reliable method was used to determine the outcome. Unclear. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.

Study details	Study design	Comparison	Results	Other
	corticosteroid treatment with washout of less than 6 months, osteoporotic fractures, Paget's disease of bone, primary hyperparathyroidism, osteomalacia, known lumbar arthrosis with or without lumbar scoliosis, porphyria, current liver enzyme inducing medication, known alcohol or drug abuse, heavy tobacco consumption or participation in other studies involving investigational products within the previous 3 months.			
Full citation Engel,P., Fabre,A., Fournier,A., Mesrine,S., Boutron-Ruault,M.C., Clavel-Chapelon,F., Risk of osteoporotic fractures after discontinuation of menopausal hormone	Aim of the study To identify the risk of osteoporotic fracture in women who had discontinued HRT. Inclusion criteria Women born between 1925 and 1950. Exclusion criteria Not reported.	Details All comparisons used a reference point from women who had never used HRT. Comparisons were made between women who had ever used HRT and those who currently used HRT. For past users, comparisons were made between those who had stopped within the last 5 years, and those who had stopped more than 5 years ago. For current users and previous users, duration of use was considered (total use < 2 years, 2 - 4.9 years and ≥ 5 years). For previous users, risk of fracture was also stratified according to duration of use and time since stopping HRT. Methods Occurrence of fractures was self reported on each follow up questionnaire. Confirmation of fractures through radiography, surgery or practitioner reports was not possible. Available data on reimbursed radiographic examinations were provided by the	Characteristics Baseline characteristics Never users of HRT Year of birth (% of participants) 1925 to 1929 14.6 1930 to 1934 18.1 1935 to 1939 17.1 1940 to 1944 18.6 1945 to 1949 31.6 BMI (kg/m², % of participants) < 20 11.4 20 to 25 55.3 > 25 33.3 Ever users of HRT Year of birth (% of participants) 1925 to 1929 4.1	Other information Limitations Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. Unclear. Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline,

Study details	Study design	Comparison	Results	Other
therapy: results from the E3N cohort, American Journal of Epidemiology, 174, 12-21, 2011 Ref Id 231459 Study type Prospective cohort study. Source of funding French League Against Cancer European Community Mutuelle Générale de l'Education Nationale Institut Gustave Roussy Institut Nationale de la Santé et de la Recherche Médicale French National Cancer Institute Country/ies where the study was carried out France Study dates 1990 to		medical insurance company and showed very good agreement between self reports and examinations performed during a 2 months interval after osteoporotic fracture occurrence. Osteoporotic fractures were considered to be any low energy fracture which occurred after menopause, excluding those of the ribs, fingers and face. Women reporting multiple fractures were assigned to only 1 relevant site according to the following hierarchy: proximal femur first, then spine, shoulder, leg, foot, ankle, wrist and arm. Sample size N = 70182 n = 18651 never users of HRT n = 51531 "ever" users of HRT	1930 to 1934 10.5 1935 to 1939 21.1 1940 to 1944 29.7 1945 to 1949 34.6 BMI (kg/m², % of participants) < 20 14.1 20 to 25 65.4 > 25 20.5 Results Any use of HRT Current use of HRT compared to never use of HRT Adjusted hazard ratio for osteoporotic fracture (95% CI): 0.78 (0.73 to 0.83) Past use of HRT compared to never use of HRT Adjusted hazard ratio for osteoporotic fracture (95% CI): 0.99 (0.92 to 1.06) Past use of HRT and time since last use Past use of HRT within the past 5 years compared to never use of HRT Adjusted hazard ratio for osteoporotic fracture (95% CI): 0.92 (0.83 to 1.01) Past use of HRT more than 5 years ago compared to never use of HRT Adjusted hazard ratio for osteoporotic fracture (95% CI): 1.05 (0.96 to 1.14) Past use of HRT and duration of use Past use of HRT for < 2 years compared to never use of HRT Adjusted hazard ratio for osteoporotic fracture (95% CI): 1.04 (0.94 to 1.15) Past use of HRT for 2 to 4.9 years compared to never use of HRT Adjusted hazard ratio for osteoporotic fracture (95% CI): 0.99 (0.88 to 1.11) Past use of HRT for ≥ 5 years compared to never use of HRT Adjusted hazard ratio for osteoporotic fracture (95% CI): 0.89 (0.80 to 0.99) Past use of HRT, including duration of use and time since stopping Past use of HRT for < 2 years and stopped < 5 years ago, compared to never use of HRT	including all major confounding and prognostic factors. Not reported. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable

Study details	Study design	Comparison	Results	Other
2008. Study duration 18 years.			Adjusted hazard ratio for osteoporotic fracture (95% CI): 0.95 (0.83 to 1.09) Past use of HRT for 2 to 4.9 years and stopped < 5 years ago, compared to never use of HRT Adjusted hazard ratio for osteoporotic fracture (95% CI): 0.93 (0.79 to 1.09) Past use of HRT for ≥ 5 years and stopped < 5 years ago, compared to never use of HRT Adjusted hazard ratio for osteoporotic fracture (95% CI): 0.79 (0.66 to 0.95) Past use of HRT for < 2 years and stopped ≥ 5 years ago, compared to never use of HRT Adjusted hazard ratio for osteoporotic fracture (95% CI): 1.14 (1.00 to 1.30) Past use of HRT for 2 to 4.9 years and stopped ≥ 5 years ago, compared to never use of HRT Adjusted hazard ratio for osteoporotic fracture (95% CI): 1.06 (0.91 to 1.24) Past use of HRT for ≥ 5 years and stopped ≥ 5 years ago, compared to never use of HRT Adjusted hazard ratio for osteoporotic fracture (95% CI): 0.95 (0.85 to 1.07) Adjusted for BMI, physical activity, age at menopause, parity, previous use of oral contraceptives, previous use of calcium supplements and educational level.	method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Genant,H.K., Lucas,J., Weiss,S., Akin,M., Emkey,R., Naney-Flint,H., Downs,R., Mortola,J., Watts,N., Yang,H.M., Banav,N., Brennan,J.J., Nolan,J.C., Low-dose esterified estrogen therapy:	Aim of the study To determine the effect of three doses of esterified oestrogens in preventing bone loss in postmenopausal women. Inclusion criteria Naturally or surgically postmenopausal women. Final menstrual period at least 6 months, and within 4 years of the start of the	Details Fracture rates in women taking one of the three different HRT doses was compared to that in women taking placebo. Methods Subjects were randomly assigned to one of four treatment groups: placebo, 0.3mg esterified oestrogens, 0.625mg esterified oestrogens or 1.25mg esterified oestrogens. The study drug was administered continuously and no progestin was given. Sample size N = 406 n = 303 HRT n = 103 placebo	Characteristics HRT group Age, years (mean): 51.6 BMI, kg/m² (mean): 25.7 Previous HRT use (%): 29 Placebo group Age, years (mean): 51.3 BMI, kg/m² (mean): 25.6 Previous HRT use (%): 33 Results N.B. fracture data not reported in this article, but obtained directly from the authors in the meta-analysis by Torgerson and Bell-Syer (Torgerson and Bell-Syer 2001). Risk of fracture in HRT group compared to placebo group: unadjusted relative risk (95% CI): 0.50 (0.09 to 2.98)	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Unclear. There was adequate concealment of allocation. Unclear. The groups were comparable at baseline. Yes. Performance bias The comparison groups received the same care apart from the

Study details	Study design	Comparison	Results	Other
effects on bone, plasma estradiol concentrations, endometrium, and lipid levels. Estratab/Osteoporosis Study Group, Archives of Internal Medicine, 157, 2609-2615, 1997 Ref Id 294866 Study type Randomised controlled trial. Source of funding Solvay Pharmaceuticals, Inc. Country/ies where the study was carried out USA Study dates Not reported. Trial duration 2 years.	study. FSH level < 50IU/L, no use of HRT within 8 weeks of the start of the trial, baseline lumbar spine BMD within 2.0 SD of mean peak bone mass. Women who had not had a hysterectomy were required to have a baseline endometrial biopsy that indicated an atrophic, mildly proliferative or moderately proliferative endometrium. Exclusion criteria Smokers. Women taking drugs that would affect bone mineral metabolism (e.g. bisphosphonates, calcitonin or androgens).			intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Yes. Individuals administering care were kept 'blind' to treatment allocation. Yes. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 41 placebo, n = 147 HRT. The groups were comparable for treatment completion. No - more women discontinued in the HRT group (many due to endometrial hyperplasia). For how many participants in each group were outcome data not available? n = 41 placebo, n = 147 HRT. The groups were comparable with respect to the availability of outcome data. No - as above. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Unclear. A valid and reliable method was used to

Study details	Study design	Comparison	Results	Other
				determine the outcome. Unclear. Investigators were kept 'blind' to participants' exposure to the intervention. Yes. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Hoidrup,S., Gronbaek,M., Pedersen,A. T., Lauritzen,J.B., Gottschau,A., Schroll,M., Hormone replacement therapy and hip fracture risk: effect modification by tobacco smoking, alcohol intake, physical activity, and body mass index, American Journal of Epidemiology, 150, 1085-1093, 1999 Ref Id 294939 Study type Prospective cohort study. Source of	Aim of the study To evaluate the overall effect of HRT on hip fracture risk. Inclusion criteria Participants in the Copenhagen City Heart Study (overall age 20 to 92). Postmenopausal women. Exclusion criteria Previous hip fracture before entrance into the study.	Details Current users of HRT at baseline were compared with non-users. Methods A self administered questionnaire was conducted with detailed questions regarding behavioural habits and other health related items. Women were asked if their periods had stopped, and at what age this happened. Postmenopausal women were asked whether they currently received hormone replacement therapy. Follow up was until the time of first hip fracture, death, disappearance, emigration or end of follow up (December 31 1993), whichever came first. Sample size N = 6146 n = 1314 HRT users n = 4832 non-users of HRT	Characteristics HRT users: Age, years (mean ± SD): 54.8 ± 5.8 Age at menopause, years (mean ± SD): 46.7 ± 5.4 BMI, kg/m² (mean ± SD): 24.4 ± 4.2 Non-users of HRT: Age, years (mean ± SD): 59.5 ± 8.0 Age at menopause, years (mean ± SD): 47.4 ± 5.4 BMI, kg/m² (mean ± SD): 25.3 ± 4.6 Results Comparison of HRT users (at baseline) to non-users of HRT: adjusted RR (95% CI): 0.71 (0.50 to 1.01) Adjusted for age, BMI, physical activity, smoking, alcohol intake, cohabitation, marital status, school education, age at menopause and parity.	Other information Limitations Study uses baseline data only to inform use of HRT. Possibility that women who were not using HRT at baseline may have commenced therapy at some time during the follow up period, or current users may discontinue, which would tend to reduce the effect size for HRT. Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. Unclear. Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline, including all major confounding and prognostic factors. Unclear. Performance bias The comparison groups

Study details	Study design	Comparison	Results	Other
funding The Copenhagen Hospital Corporation The Research Academy The Health Insurance Fund The Danish Medical Research Foundation The Danish Medical Research Council The Danish National Board of Health. Country/ies where the study was carried out Denmark Study dates Baseline examination in 1976 to 1978. Study duration 17 years.				received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the

Study details	Study design	Comparison	Results	Other
				intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Honkanen,R. J., Honkanen,K. , Kroger,H., Alhava,E., Tuppurainen, M., Saarikoski,S. , Risk factors for perimenopausal distal forearm fracture, Osteoporosis International, 11, 265-270, 2000 Ref Id 231884 Study type Prospective cohort study. Source of funding The European Foundation for Osteoporosis . Kuopio University Hospital. The Yrjö Jahnsson Foundation. Country/ies where the	Aim of the study To examine prospectively which factors predict peri- and early post-menopausal distal forearm fracture. Inclusion criteria Women aged 47 to 56 and resident in Kuopio Province, Finland. Exclusion criteria Not reported.	Details Women who used HRT continuously during the five year follow up period were compared to those who did not use HRT during the follow up. Methods The baseline postal inquiry included questions about risk factors. The five-year inquiry included questions about fractures and HRT use during follow up. Reported follow up fractures were validated against radiographic reports in the patient records. Only validated follow up fracture was used as an endpoint event. Sample size N = 11798 n = 4837 HRT users during follow up n = 6961 no HRT use during follow up	Characteristics Women who sustained a wrist fracture: Age, years (mean ± SD): 53.2 ± 2.9 BMI, kg/m² (mean ± SD): 25.2 ± 3.9 HRT use during follow up, %: 30 Previous fracture history, %: 26.9 Women who did not sustain a wrist fracture: Age, years (mean ± SD): 52.3 ± 2.9 BMI, kg/m² (mean ± SD): 26.3 ± 4.3 HRT use during follow up, %: 41.4 Previous fracture history: 16.7 Results Risk of wrist fracture in women who used HRT during follow up compared to those who did not use HRT during follow up: adjusted hazard ratio (95% CI): 0.37 (0.23 to 0.61) Adjusted for age, menopausal state, BMI, calcium intake, wrist fracture history and parity.	Other information Limitations Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. Unclear. Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline, including all major confounding and prognostic factors. Unclear. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes.

Study details	Study design	Comparison	Results	Other
study was carried out Finland Study dates Baseline inquiry carried out in May 1989, follow up in May 1994. Study duration 5 years.				How many participants did not complete treatment in each group? Not reported. N = 1302 women who responded to the baseline questionnaire but not the follow up. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? N = 1302 women who responded to the baseline questionnaire but not the follow up. The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Hosking,D., Chilvers,C.E.	Aim of the study To compare the efficacy, safety	Details Occurrence of traumatic non-vertebral fractures was compared in the HRT group and those taking placebo.	Characteristics HRT group: Age, years (mean ± SD): 53 ± 4	Other information Limitations Study quality

Study details	Study design	Comparison	Results	Other
, Christiansen, C., Ravn,P., Wasnich,R., Ross,P., McClung,M., Balske,A., Thompson,D., Daley,M., Yates,A.J., Prevention of bone loss with alendronate in postmenopausal women under 60 years of age. Early Postmenopausal Intervention Cohort Study Group, New England Journal of Medicine, 338, 485-492, 1998 Ref Id 231894 Study type Randomised controlled trial. Source of funding Merck Research Laboratories. Country/ies where the study was carried out UK,	and tolerability of alendronate with those of a combination of oestrogen and progestin. Inclusion criteria Aged 45 to 59 years and in good health. Postmenopausal for at least 6 months (confirmed by a high serum FSH). Exclusion criteria No clinical or laboratory evidence of systemic disease. Abnormal renal function, history of cancer, peptic ulcer or oesophageal disease requiring prescription medication within the past 5 years, previous treatment with a bisphosphonate or fluoride, regular therapy with a phosphate binding antacid, oestrogen replacment therapy within the previous 3 months and therapy with any other drug that affects the skeleton.	Methods Women were randomly assigned to receive placebo, 2.5mg alendronate, 5 mg alendronate or open label oestrogen-progestin. In the United States, the oestrogen-progestin were given as conjugated oestrogens (Premarin 0.625mg daily) and medroxyprogesterone acetate (Provera, 5mg daily). In Europe the oestrogen and progestins were given in a cyclical regimen (Trisequens) of 2mg of micronized oestrogen daily for 22 days, 1mg of norethindrone acetate per day on days 13 to 22, and 1mg of estradiol per day on days 23 to 28. Women were questioned about adverse effects (including fractures) at clinic visits every 3 months. Follow up was for 2 years. Sample size N = 563 n = 102 HRT n = 461 placebo (additional 897 women randomised to alendronate, but not included for this analysis).	BMI, kg/m² (mean ± SD): 25 ± 3 Years since menopause (mean ± SD): 4 ± 3 BMD at lumbar spine, g/cm² (mean ± SD): 0.93 ± 0.12 Placebo group: Age, years (mean ± SD): 53 ± 4 BMI, kg/m² (mean ± SD): 25 ± 4 Years since menopause (mean ± SD): 6 ± 5 BMD at lumbar spine, g/cm² (mean ± SD): 0.94 ± 0.12 Results Risk of any non-vertebral fracture in HRT treatment compared to placebo group: unadjusted relative risk (95% CI): 0.98 (0.29 to 3.34)	Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Unclear. There was adequate concealment of allocation. Unclear. The groups were comparable at baseline. Yes. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No - estrogen-progestin was provided as an open label preparation. Individuals administering care were kept 'blind' to treatment allocation. No - as above. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 93 placebo, n = 19 HRT group. The groups were comparable for treatment completion. Yes. For how many participants in each group were outcome data not

Study details	Study design	Comparison	Results	Other
Denmark, and USA. Study dates Not reported. Trial duration 2 years.				available? n = 10 placebo, n = 4 HRT group. The groups were comparable with respect to the availability of outcome data. Yes. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Hundrup,Y.A., Hoidrup,S., Ekholm,O., Davidsen,M., Obel,E.B., Risk of low-energy hip, wrist, and upper arm fractures among current and previous users of hormone replacement therapy: The Danish	Aim of the study To examine the effect of oestrogen alone and oestrogen plus progestin on the risk of low energy hip, wrist and upper arm fractures. Examination of to what extent duration of use, previous use and recency of discontinuation of HRT influences the fracture risk. Inclusion criteria	Details Current users of HRT were compared to never users. Duration of use of HRT and how recently HRT was used were also taken into account. Methods Detailed information on the use of HRT was obtained in the baseline questionnaire (current and previous use). Sample size N = 7082 n = 1936 current users of HRT n = 922 previous users of HRT n = 4019 never users of HRT	Characteristics Current users of HRT Age range 50 - 59 years (%): 79 Age range 60 - 69 years (%): 21 Age at menopause < 45 years (%): 11 Age at menopause 45 - 55 years (%): 66 Age at menopause > 55 years (%): 4 BMI < 18.5 (%): 2 BMI 18.5 - 24 (%): 75 BMI 25 - 29 (%): 19 BMI > 30 (%): 3 Previous users of HRT Age range 50 - 59 years (%): 56 Age range 60 - 69 years (%): 44 Age at menopause < 45 years (%): 16 Age at menopause 45 - 55 years (%): 68 Age at menopause > 55 years (%): 2	Other information Limitations Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. Unclear. Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline, including all major confounding and

Study details	Study design	Comparison	Results	Other
Nurse Cohort Study, European Journal of Epidemiology, 19, 1089-1095, 2004 Ref Id 294159 Study type Prospective cohort study. Source of funding Not reported. Country/ies where the study was carried out Denmark Study dates Cohort recruited in 1993. Follow up in 1999. Study duration 6 years.	Female members of the Danish Nurses' Organisation aged 45 years and over. Exclusion criteria Premenopausal women. Fracture prior to 1993, or previous fracture but year of fracture not reported. Aged less than 50 or more that 69 at the baseline evaluation.		BMI < 18.5 (%): 2 BMI 18.5 - 24 (%): 65 BMI 25 - 29 (%): 27 BMI > 30 (%): 6 Never users of HRT Age range 50 - 59 years (%): 67 Age range 60 - 69 years (%): 33 Age at menopause < 45 years (%): 6 Age at menopause 45 - 55 years (%): 73 Age at menopause > 55 years (%): 5 BMI < 18.5 (%): 2 BMI 18.5 - 24 (%): 66 BMI 25 - 29 (%): 25 BMI > 30 (%): 6 Results How recently HRT was used use Risk of low-energy non-spinal fractures in current users of HRT compared to never users of HRT adjusted hazard ratio (95% CI): 0.50 (0.35 to 0.71) Risk of low-energy non-spinal fractures in previous users of HRT compared to never users of HRT adjusted hazard ratio (95% CI): 1.23 (0.89 to 1.70) How recently HRT was used: past users Risk of low-energy non-spinal fractures in past users of HRT discontinued < 5 years compared to never users of HRT adjusted hazard ratio (95% CI): 1.05 (0.63 to 1.73) Risk of low-energy non-spinal fractures in past users of HRT discontinued 5 to 10 years compared to never users of HRT adjusted hazard ratio (95% CI): 0.85 (0.45 to 1.61) Risk of low-energy non-spinal fractures in past users of HRT discontinued ≥ 10 years compared to never users of HRT adjusted hazard ratio (95% CI): 2.03 (1.25 to 3.29) Duration of use: current users Risk of low-energy non-spinal fractures in users of HRT for < 5 years compared to never users of HRT adjusted hazard ratio (95% CI): 0.65 (0.37 to 1.14) Risk of low-energy non-spinal fractures in users of HRT for 5 to 10 years compared to never users of HRT adjusted hazard ratio (95% CI): 0.62 (0.36 to 1.07) Risk of low-energy non-spinal fractures in users of HRT	prognostic factors. Unclear. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome.

Study details	Study design	Comparison	Results	Other
			for ≥ 10 years compared to never users of HRT adjusted hazard ratio (95% CI): 0.32 (0.16 to 0.64) Duration of use: Previous users Risk of low-energy non-spinal fractures in users of HRT for < 5 years compared to never users of HRT adjusted hazard ratio (95% CI): 1.41 (0.97 to 2.05) Risk of low-energy non-spinal fractures in users of HRT for > 5 years compared to never users of HRT adjusted hazard ratio (95% CI): 0.94 (0.54 to 1.64) Recency and duration of use Risk of low-energy non-spinal fractures in users of HRT for < 5 years and stopped within the past 5 years compared to never users of HRT adjusted hazard ratio (95% CI): 1.03 (0.52 to 2.04) Risk of low-energy non-spinal fractures in users of HRT for > 5 years and stopped within the past 5 years compared to never users of HRT adjusted hazard ratio (95% CI): 1.11 (0.54 to 2.27) Risk of low-energy non-spinal fractures in users of HRT for < 5 years and stopped more than 5 years ago compared to never users of HRT adjusted hazard ratio (95% CI): 1.65 (1.07 to 2.53) Risk of low-energy non-spinal fractures in users of HRT for > 5 years and stopped more than 5 years ago compared to never users of HRT adjusted hazard ratio (95% CI): 0.84 (0.36 to 1.92) Adjusted for family history, BMI and age at menopause.	Unclear. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Huopio,J., Kroger,H., Honkanen,R., Saarikoski,S., Alhava,E., Risk factors for perimenopausal fractures: a prospective study, Osteoporosis International,	Aim of the study To evaluate the risk factors for perimenopausal fractures among Finnish women. Inclusion criteria Women aged between 47 and 56 years residing in Kuopio Province, Eastern Finland in 1989. Exclusion criteria Not reported.	Details Women who were using HRT at the time of the baseline study were compared to those who were not using HRT. Methods Follow up questionnaires were sent in 1990-1 and 1994. The first fracture during the follow up period was taken to be the endpoint event. All self reported fractures were validated by cross-checking radiological reports from medical records. Fractures due to road traffic accidents were excluded. Sample size N = 3068 n = 799 HRT users n = 2269 non-HRT users	Characteristics Comparison between fracture cases and those without fractures at follow up only: Fracture cases: Age, years (mean ± 95% CI): 53.5 (53.1 to 53.9) HRT use (%): 18.7 Nonfracture cases: Age, years (mean ± 95% CI): 53.4 (53.3 to 53.5) HRT use (%): 26.7 Results Risk of any fracture in women taking HRT at baseline, compared to those not taking HRT at baseline: adjusted RR (95% CI): 0.66 (0.46 to 0.94)	Other information Limitations Data on HRT only obtained during baseline questionnaire, therefore women not taking HRT at baseline may have started HRT over the course of follow up, potentially reducing the effect size. Study quality Selection bias The method of allocation to treatment groups was

Study details	Study design	Comparison	Results	Other
11, 219-227, 2000 Ref Id 294954 Study type Prospective cohort study. Source of funding Academy of Finland The Yrjö Jahnsson Foundation The Sigrid Juselius Foundation Country/ies where the study was carried out Finland Study dates Baseline inquiry in 1990 to 1991, folllow up in May 1994. Study duration 3.6 years.			Adjusted for age, weight, height, menopausal status, BMD, previous fracture history, maternal hip fracture, use of HRT, smoking, calcium intake, and multiple chronic health disorders. (risk in HRT non-users compared to users in the article, therefore reciprocals taken for this analysis).	unrelated to potential confounding factors. Unclear. Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline, including all major confounding and prognostic factors. Unclear. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? Not reported. The groups were

Study details	Study design	Comparison	Results	Other
				comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Jackson,R.D., Wactawski-Wende,J., LaCroix,A.Z., Pettinger,M., Yood,R.A., Watts,N.B., Robbins,J.A., Lewis,C.E., Beresford,S.A., Ko,M.G., Naughton,M.J., Satterfield,S., Bassford,T., Women's Health Initiative Investigators., Effects of conjugated equine	Aim of the study To assess the effects on major disease incidence rates of oestrogen alone and oestrogen plus progestin HRT. Inclusion criteria Oestrogen plus progesterone arm: Postmenopausal women with an intact uterus, aged 50 to 79 years at randomization. Oestrogen alone arm: Postmenopausal women with a	Details Fracture rates were compared between women enrolled in the oestrogen plus progestin group and those taking placebo. Similar comparison was made between women in the oestrogen alone arm and those taking placebo. Time-to-event analyses were conducted based on the intention-to-treat principle. Fracture incidence rates were compared using hazards ratios, nominal 95% CIs and Wald statistic p values from Cox proportional hazards models stratified by age, prior fracture history and randomization status in the dietary modification trial (subgroup of WHI). Methods Women with an intact uterus were randomly assigned to treatment with either 0.625mg conjugated equine oestrogens plus 2.5mg medroxyprogesterone acetate daily, or placebo. Women with a previous hysterectomy were randomly assigned to treatment with 0.625mg conjugated equine oestrogens daily, or placebo. Reports of hip, clinical vertebral, wrist/lower arm and other osteoporotic fractures (excluding chest/sternum, ribs, skull/face, fingers, toes and cervical vertebrae) were ascertained by semiannual questionnaire. All reported fractures were confirmed	Characteristics Oestrogen plus progestin arm: Average age, years (mean ± SD): 63.2 ± 7.10 Average BMI, kg/m² (mean ± SD): 28.5 ± 5.80 Oestrogen alone arm: Average age, years (mean ± SD): 63.6 ± 7.3 Average BMI, kg/m² (mean ± SD): 30.1 ± 6.1 Results N.B. multiple publications have arisen from the same trial, therefore relevant results from a number of different publications are included here. Current use Current use of oestrogen plus progestin HRT (Cauley et al., 2003) Hip fracture in current oestrogen plus progestin users compared to placebo group Hazard ratio (95% CI): 0.67 (0.47 to 0.96) Wrist fracture in current oestrogen plus progestin users compared to placebo group Hazard ratio (95% CI): 0.71 (0.59 to 0.85) Vertebral fracture in current oestrogen plus progestin users compared to placebo group Hazard ratio (95% CI): 0.65 (0.46 to 0.92)	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Yes. There was adequate concealment of allocation. Yes. The groups were comparable at baseline. Yes. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation.

Study details	Study design	Comparison	Results	Other
estrogen on risk of fractures and BMD in postmenopausal women with hysterectomy: results from the women's health initiative randomized trial, Journal of Bone and Mineral Research, 21, 817-828, 2006 Ref Id 231983 Study type Randomised controlled trial. After discontinuation of the trial, participants were followed up as an observational cohort study. Source of funding National Heart, Lung and Blood Institute, U.S. Department of Health and Human Services. Active study	prior hysterectomy. 50 to 79 years at randomization. Likely to reside in the area for 3 years. Exclusion criteria Medical conditions likely to be associated with a predicted survival of < 3 years, previous breast cancer, other cancer within the last 10 years (except for non-melanoma skin cancer), alcoholism, dementia, transportation problems.	by review of the radiology reports by centrally trained local adjudicators who were blinded to treatment assignment. Hip fractures underwent a second central adjudication. Sample size Oestrogen plus progestin arm: N = 16608 n = 8506 oestrogen plus progestin group n = 8102 placebo group Oestrogen alone arm: N = 10739 n = 5310 oestrogen group n = 5429 placebo group	Any fracture in current oestrogen plus progestin users compared to placebo group Hazard ratio (95% CI): 0.76 (0.69 to 0.83) Hip fracture in current oestrogen plus progestin users aged 50 to 59 compared to placebo group Hazard ratio (95% CI): 0.17 (0.02 to 1.43) Hip fracture in current oestrogen plus progestin users aged 60 to 69 compared to placebo group Hazard ratio (95% CI): 0.76 (0.41 to 1.39) Any fracture in current oestrogen plus progestin users aged 50 to 54 compared to placebo group Hazard ratio (95% CI): 0.68 (0.49 to 0.93) Any fracture in current oestrogen plus progestin users aged 55 to 59 compared to placebo group Hazard ratio (95% CI): 0.91 (0.71 to 1.16) Any fracture in current oestrogen plus progestin users aged 60 to 64 compared to placebo group Hazard ratio (95% CI): 0.80 (0.65 to 0.98) Any fracture in current oestrogen plus progestin users aged 65 to 69 compared to placebo group Hazard ratio (95% CI): 0.68 (0.49 to 0.93) Current use of oestrogen alone HRT (Jackson et al., 2006) Hip fracture in current oestrogen only users compared to placebo group Hazard ratio (95% CI): 0.65 (0.45 to 0.94) Wrist fracture in current oestrogen only users compared to placebo group Hazard ratio (95% CI): 0.58 (0.47 to 0.72) Vertebral fracture in current oestrogen only users compared to placebo group Hazard ratio (95% CI): 0.64 (0.44 to 0.93) Any fracture in current oestrogen only users compared to placebo group Hazard ratio (95% CI): 0.71 (0.64 to 0.80) Hip fracture in current oestrogen only users aged 50 to 59 compared to placebo group Hazard ratio (95% CI): 5.02 (0.59 to 43.02) Hip fracture in current oestrogen only users aged 60 to 69 compared to placebo group Hazard ratio (95% CI): 0.47 (0.22 to 1.04) Any fracture in current oestrogen only users aged 50 to 59	Unclear. Individuals administering care were kept 'blind' to treatment allocation. Unclear. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors.

Study details	Study design	Comparison	Results	Other
drug and placebo were supplied by Wyeth (Radnor P.A.) Country/ies where the study was carried out USA Study dates Recruitment began in 1993. Trial suspended in July 2002 (oestrogen plus progesterone arm) and February 2004 (oestrogen only arm). Median intervention duration 5.2 years in combined therapy arm, 7.2 years for oestrogen only arm.			compared to placebo group Hazard ratio (95% CI): 0.90 (0.72 to 1.12) Any fracture in current oestrogen only users aged 60 to 69 compared to placebo group Hazard ratio (95% CI): 0.63 (0.53 to 0.75) Previous use Past use of oestrogen plus progestin HRT (median duration of treatment 5.2 years), discontinued a mean of 2.4 years ago (Heiss et al., 2008) Hip fracture in past oestrogen plus progestin users compared to placebo group Hazard ratio (95% CI): 0.78 (0.60 to 1.00) Vertebral fracture in past oestrogen plus progestin users compared to placebo group Hazard ratio (95% CI): 0.78 (0.60 to 1.01) Any fracture in past oestrogen plus progestin users compared to placebo group Hazard ratio (95% CI): 0.80 (0.73 to 0.86) Past use of oestrogen only HRT (mean duration of treatment 7.2 years), discontinued a mean of 3.9 years ago (LaCroix et al., 2011) Hip fracture in past oestrogen only users compared to placebo group Hazard ratio (95% CI): 0.92 (0.71 to 1.18) Hip fracture in past oestrogen only users aged 50 to 59 compared to placebo group Hazard ratio (95% CI): 1.55 (0.51 to 4.75) Hip fracture in past oestrogen only users aged 60 to 69 compared to placebo group Hazard ratio (95% CI): 0.87 (0.57 to 1.35) Past use of oestrogen plus progestin HRT (median duration of treatment 5.2 years), discontinued a median of 8.2 years ago (Manson et al., 2013) Hip fracture in past oestrogen plus progestin users compared to placebo group Hazard ratio (95% CI): 0.81 (0.68 to 0.97) Hip fracture in past oestrogen plus progestin users aged 50 to 59 compared to placebo group Hazard ratio (95% CI): 0.57 (0.31 to 1.04) Hip fracture in past oestrogen plus progestin users aged 60 to 69 compared to placebo group Hazard ratio (95% CI): 0.94 (0.71 to 1.24)	Unclear.

Study details	Study design	Comparison	Results	Other
			Past use of oestrogen only HRT (median duration of treatment 7.2 years), discontinued a median of 6.6 years ago (Manson et al., 2013) Hip fracture in past oestrogen only users compared to placebo group Hazard ratio (95% CI): 0.91 (0.72 to 1.15) Hip fracture in past oestrogen only users aged 50 to 59 compared to placebo group Hazard ratio (95% CI): 0.88 (0.36 to 2.17) Hip fracture in past oestrogen only users aged 60 to 69 compared to placebo group Hazard ratio (95% CI): 0.95 (0.64 to 1.43)	
Full citation Komulainen, M.H., Kroger,H., Tuppurainen, M.T., Heikkinen,A. M., Alhava,E., Honkanen,R. , Saarikoski,S. , HRT and Vit D in prevention of non-vertebral fractures in postmenopausal women; a 5 year randomized trial.[Reprint in Maturitas. 2008 Sep-Oct;61(1-2):85-94; PMID: 19434882], Maturitas, 31, 45-54, 1998 Ref Id 232124 Study type	Aim of the study To identify the effect of HRT and low-dose vitamin D on the BMD in non-osteoporotic early postmenopausal women. Inclusion criteria Postmenopausal women aged 47 to 56. Within 6 to 24 months of their last menstrual period. Exclusion criteria History of breast or endometrial cancer, thromboembolic diseases and medication resistant hypertension.	Details Fracture incidence in women taking HRT was compared to that in women taking placebo. Methods Women were randomized to treatment with HRT (2mg estradiol valerate day [1 to 21] and 1 mg cyproterone acetate [days 12 to 21] followed by a treatment-free interval [days 22 to 28]) or placebo. Other participants were treated with vitamin D alone, or vitamin D plus HRT, but are not included for the purposes of this analysis. Sample size N = 232 n = 116 HRT n = 116 placebo	Characteristics HRT group Age, years (mean + 95% CI): 52.9 (52.5 to 53.3) BMI, kg/m² (mean + 95% CI): 26.4 (25.7 to 27.2) Previous fracture during the last 15 years, %: 14 Lumbar spine BMD g/cm² (mean + 95% CI): 1.132 (1.104 to 1.160) Placebo group Age, years (mean + 95% CI): 52.6 (52.2 to 53.0) BMI, kg/m² (mean + 95% CI): 26.1 (25.3 to 26.8) Previous fracture during the last 15 years, %: 13 Lumbar spine BMD g/cm² (mean + 95% CI): 1.151 (1.122 to 1.179) Results N.B.relative risk presented in article uses per-protocol analysis, rather than intention to treat. Also combines data from HRT+vitamin D group with HRT alone. For the purposes of this analysis results from the intention to treat analysis were used, and only participants in the HRT only or placebo group were included. Risk of non-vertebral fracture in women using HRT compared to those using placebo: relative risk (95% CI): 0.32 (0.13 to 0.76) Risk of wrist fracture in women using HRT compared to those using placebo: relative risk (95% CI): 0.29 (0.06 to 1.35)	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Yes. There was adequate concealment of allocation. Yes. The groups were comparable at baseline. Yes. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No - open label design. Individuals administering care were kept 'blind' to treatment allocation. No - open label design. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of

Study details	Study design	Comparison	Results	Other
Randomised controlled trial. Source of funding Leiras Oy. Schering AG. Country/ies where the study was carried out Finland Study dates Recruitment in 1990 to 1991. Trial duration 5 years.				follow up). Yes. How many participants did not complete treatment in each group? n = 11 placebo, n = 42 HRT. The groups were comparable for treatment completion. No - more women in the HRT group did not comply with treatment. For how many participants in each group were outcome data not available? n = 3 placebo, n = 11 HRT group. The groups were comparable with respect to the availability of outcome data. Yes. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Lafferty,F.W., Fiske,M.E., Postmenopa	Aim of the study To assess the long-term effects of oestrogen	Details Women using oestrogen replacement therapy were compared to those who remained untreated. Methods	Characteristics HRT users Age, years (mean ± SD): 52.6 ± 4.8 Years of menopause before entry to study (mean ± SD):	Other information Limitations Study quality Selection bias

Study details	Study design	Comparison	Results	Other
usal estrogen replacement: a long-term cohort study, American Journal of Medicine, 97, 66-77, 1994 Ref Id 229713 Study type Prospective cohort study. Source of funding University Hospitals, Cleveland, Ohio. Country/ies where the study was carried out USA Study dates Cohort identified from 1964 to 1983. Average follow up 12 years.	replacement therapy in postmenopausal women. Inclusion criteria Postmenopausal women (at least 12 months of amenorrhoea) aged between 43 and 60 years of age. For women with a previous hysterectomy, postmenopause was taken as the time of onset of hot flushes, or upon reaching 55 years of age. Healthy, ambulatory, white women with no abnormality by physical examination, ECG, haematological or biochemical abnormalities. Exclusion criteria Past or present history of major disease, including cancer, severe hypertension or cardiovascular disease, osteoporosis, diabetes mellitus, alcoholism, COPD, ulcerative colitis, depression, rheumatoid arthritis.	Women were treated with 0.625mg conjugated equine oestrogen for the first 25 days of each month from 1964 until 1983. After this time, women with an intact uterus also received 5mg medroxyprogesterone acetate from day 14 until day 25 of every 6th month. Subjects were followed up prospectively with annual or biennial physical examinations. Peripheral fractures were verified by radiological reports and letters from the subjects orthopaedic surgeons. Fractures of the phalanges and facial bones were not included. Vertebral fractures were detected on lateral views of the thoracic spine by chest x-rays taken every 3 years, or at the onset of unusual back pain. Sample size N = 157 n = 81 HRT group n = 76 no treatment group	4.7 ± 4.6 BMI, kg/m² (mean ± SD): 22.3 ± 3.2 No treatment group Age, years (mean ± SD): 54.7 ± 3.8 Years of menopause before entry to study (mean ± SD): 5.1 ± 5.3 BMI, kg/m² (mean ± SD): 24.4 ± 3.4 Results Risk of vertebral fracture in HRT group compared to no treatment group: adjusted relative risk (95% CI): 0.27 (0.12 to 0.60) Risk of non-vertebral fracture in HRT group compared to no treatment group: adjusted relative risk (95% CI): 0.23 (0.06 to 0.97) Risk of any fracture in HRT group compared to no treatment group: adjusted relative risk (95% CI): 0.28 (0.09 to 0.89) Adjusted for age	The method of allocation to treatment groups was unrelated to potential confounding factors. Unclear. Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline, including all major confounding and prognostic factors. Yes. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept ‘blind’ to treatment allocation. No. Individuals administering care were kept ‘blind’ to treatment allocation. No. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? Not reported.

Study details	Study design	Comparison	Results	Other
				The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Unclear. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Lees,B., Stevenson,J. C., The prevention of osteoporosis using sequential low-dose hormone replacement therapy with estradiol-17 beta and dydrogesterone, Osteoporosis International, 12, 251-258, 2001 Ref Id 232214	Aim of the study To investigate the efficacy of sequential regimens of either 1mg or 2mg of 17β oestradiol in the prevention of postmenopausal osteoporosis. Inclusion criteria Women aged between 44 and 65 years. No previous hysterectomy. Naturally postmenopausal (amenorrhoeic for at least 6 months) with serum FSH > 20 IU/l in all	Details Fractures were recorded as adverse events. Rate of fracture in women taking HRT was compared to that in women taking placebo tablets. Methods Participants were randomly allocated into one of five groups to receive either placebo or one of four different HRT preparations (estradiol 1mg daily plus 5mg dydrogesterone from day 15 to 28, estradiol 1mg daily plus dydrogesterone 10mg from day 15 to 28, estradiol 2mg daily plus 10mg dydrogesterone from day 15 to 28 or estradiol 2mg daily plus 20mg dydrogesterone from day 15 to 28). For the purposes of this analysis data from all HRT arms were combined. Sample size N = 579 n = 466 HRT n = 113 placebo	Characteristics Age, years (mean ± SD): 55.6 ± 4.6 Weight, kg (mean ± SD): 66.4 ± 9.9 Amenorrhoea, months (mean ± SD): 70.4 ± 57.8 Results Risk of any non-vertebral fracture in HRT group compared to placebo group: unadjusted relative risk (95% CI): 0.79 (0.22 to 2.81)	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Yes. There was adequate concealment of allocation. Yes. The groups were comparable at baseline. Yes. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to

Study details	Study design	Comparison	Results	Other
Study type Randomised controlled trial. Source of funding The Heart Disease and Diabetes Research Trust. Solvay Pharmaceuticals. Country/ies where the study was carried out UK and Canada Study dates Not reported. Trial duration 2 years.	cases. Baseline endometrial biopsy confirmed no endometrial hyperplasia or neoplasia. BMD measurements at least 0.80g/cm² in the lumbar spine and 0.65g/cm² in the femoral neck for Lunar instruments and 0.70g/cm² in the lumbar spine and 0.52g/cm² in the femoral neck for Holologic instruments. Exclusion criteria Ever use of HRT by implant, or use of other types of HRT in the previous 6 months. Ever use of bisphosphonates or fluoride. Evidence of cancer, renal, liver or cardiovascular disease, hypertension or diabetes. More than 25% heavier than ideal body weight. Evidence of alcohol or drug abuse.			treatment allocation. Yes. Individuals administering care were kept 'blind' to treatment allocation. Unclear. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 227 total (data for individual groups not provided). The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? None. The groups were comparable with respect to the availability of outcome data. Yes. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Unclear. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important

Study details	Study design	Comparison	Results	Other
				confounding and prognostic factors. Unclear.
Full citation Liu,J.H., Muse,K.N., The effects of progestins on bone density and bone metabolism in postmenopausal women: a randomized controlled trial, American Journal of Obstetrics and Gynecology, 192, 1316-1323, 2005 Ref Id 232278 Study type Randomised controlled trial. Source of funding The National Institutes of Aging, National Institutes of Health. Country/ies where the study was carried out	Aim of the study To explore the role of progestins in bone metabolism in early postmenopausal women. Inclusion criteria Healthy, postmenopausal women aged 45 to 60. Less than 5 years from menopause, FSH level > 40 IU/L, bone density T-score less than -2 on baseline BMD, normal mammogram and normal cervical smear within the past 6 months. Exclusion criteria Severe vasomotor symptoms, hypertension, bone disease, vertebral fracture, any medical contraindications to taking oestrogen, serious psychiatric disorder, hypertriglyceridaemia > 300mg/dL, previous	Details Fracture rates in women taking progestins were compared with those taking placebo for the duration of the trial. Methods Women were randomised to one of 6 treatment groups: micronized progesterones 300mg/day, medroxyprogesterone acetate 10mg/day, norethindrone 1mg/day, micronized oestradiol 1mg/day, oestradiol 1mg/day + medroxyprogesterone acetate 1mg/day and placebo. Treatment duration was 2 years. Sample size N = 132 n = 65 progestin only preparations n = 21 combined oestrogen/progestin HRT n = 23 oestrogen alone HRT n = 23 placebo	Characteristics Progestin only group: Age, years (mean): 52.7 BMI, kg/m² (mean): 27.8 Combined HRT group: Age, years (mean): 52.9 BMI, kg/m² (mean): 25.6 Oestrogen alone HRT group: Age, years (mean): 52.0 BMI, kg/m² (mean): 28.2 Placebo group: Age, years (mean): 52.6 BMI, kg/m² (mean): 27.3 Results No vertebral or hip fractures were sustained in any group, therefore unable to calculate relative risk.	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Unclear. There was adequate concealment of allocation. Unclear. The groups were comparable at baseline. Yes. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Yes. Individuals administering care were kept 'blind' to treatment allocation. Yes. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 3 placebo group, n = 15 progestin group, n = 1 combined HRT group, n = 4 oestrogen only HRT

Study details	Study design	Comparison	Results	Other
USA Study dates Recruitment between 1995 and 1999. Trial duration 2 years.	treatment with a bisphosphonate or fluoride, use of any steroid medications within the past 3 months.			group. The groups were comparable for treatment completion. Yes. For how many participants in each group were outcome data not available? n = 3 placebo group, n = 15 progestin group, n = 1 combined HRT group, n = 4 oestrogen only HRT group. The groups were comparable with respect to the availability of outcome data. Yes. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Unclear. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Lufkin,E.G., Wahner,H.W., O'Fallon,W.M., Hodgson,S.F., Kotowicz,M.	Aim of the study To assess the effect of transdermal oestrogen in the treatment of established osteoporosis. Inclusion criteria	Details Fracture rates in the HRT group were compared to the placebo group. Methods Women were randomly assigned to treatment with oestrogen (0.1mg estradiol daily delivered as a transdermal patch) and medroxyprogesterone acetate (10mg/day orally for days 11 to 21) or placebo. Trial duration was for one year.	Characteristics HRT group Age, years (median and range): 65.5 (54.6 to 72.1) Time since menopause, years (median and range): 16.6 (5.7 to 27.6) Number of previous vertebral fractures (median and range): 4 (1 to 9.3) BMD at lumbar spine, g/cm² (median and range): 0.79 (0.65 to 0.91)	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Unclear.

Study details	Study design	Comparison	Results	Other
A., Lane,A.W., Judd,H.L., Caplan,R.H., Riggs,B.L., Treatment of postmenopausal osteoporosis with transdermal estrogen, Annals of Internal Medicine, 117, 1-9, 1992 Ref Id 232295 Study type Randomised controlled trial. Source of funding Ciba-Geighy Corporation. Country/ies where the study was carried out USA Study dates Not reported. Trial duration 1 year.	Fully ambulatory, postmenopausal, white women aged 47 to 75 years of age. Documented osteoporosis but no evidence of an associated disease or a history of use of any drug known to cause osteoporosis or to affect calcium levels. Osteoporosis defined as BMD at lumbar spine and proximal femur below the 10th percentile of normal premenopausal women and one or more vertebral fractures (defined as a decrease in vertebral height of more than 15%). Exclusion criteria Ever use of sodium fluoride or bisphosphonate.	Vertebral fracture was assessed using lateral radiographs of the thoracic and lumabr spine at baseline and after 1 year. Sample size N = 75 n = 36 HRT n = 39 placebo	Placebo group Age, years (median and range): 64.1 (55.1 to 70.4) Time since menopause, years (median and range): 14.0 (5.0 to 25.0) Number of previous vertebral fractures (median and range): 4 (2 to 9) BMD at lumbar spine, g/cm² (median and range): 0.77 (0.65 to 1.03) Results Risk of new vertebral fracture in HRT group compared to placebo group: unadjusted relative risk (95% CI): 0.63 (0.28 to 1.43)	There was adequate concealment of allocation. Unclear. The groups were comparable at baseline. Yes. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Yes. Individuals administering care were kept 'blind' to treatment allocation. Yes. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 5 placebo, n = 5 HRT group. The groups were comparable for treatment completion. Yes. For how many participants in each group were outcome data not available? n = 5 placebo, n = 5 HRT group. The groups were comparable with respect to the availability of outcome data. Yes. Detection bias The study had an appropriate length of follow up. Unclear.

Study details	Study design	Comparison	Results	Other
				The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Maxim,P., Ettinger,B., Spitalny,G.M., Fracture protection provided by long-term estrogen treatment, Osteoporosis International, 5, 23-29, 1995 Ref Id 232383 Study type Prospective cohort study. Source of funding The Northern California Kaiser Foundation Hospitals, Inc. Community Service Program.	Aim of the study To quantify the protective effect of long-term oestrogen replacement therapy on vertebral, wrist and hip fracture while adjusting for age and other covariates. Inclusion criteria White postmenopausal women (last period at least 6 months ago, or bilateral oophorectomy), within 3 years of menopause. Exclusion criteria Use of thyroid medication in excess of 2 grains (sic) daily. Use of anticonvulsants or glucocorticoids. Chronic	Details Risk of fracture in users of oestrogen at baseline were compared to those who were not using oestrogen at baseline. Methods Demographic data were recorded during the baseline medical record review. In 1992, medical records were reviewed again to determine the year, site and associated trauma for all fractures sustained in the follow up period. Fractures occurring within 5 years of menopause and any fractures sustained during road traffic accidents were not included. In the case of vertebral fractures whcih were not symptomatic a radiographic report was accepted as evidence of a new fracture. Sample size N = 490 n = 245 oestrogen users n = 245 non-users of oestrogen	Characteristics Oestrogen users: Age at menopause, years (mean ± SD): 50.8 ± 3.3 BMI, kg/m² (mean ± SD): 24.0 ± 3.6 Non-users of oestrogen: Age at menopause, years (mean ± SD): 49.8 ± 3.5 BMI, kg/m² (mean ± SD): 24.7 ± 4.2 Results Risk of wrist fracture in oestrogen users compared to non-users adjusted relative risk (95% CI): 0.44 (0.23 to 0.84) Risk of vertebral fracture in oestrogen users compared to non-users adjusted relative risk (95% CI): 0.60 (0.36 to 0.99) Risk of hip fracture in oestrogen users compared to non-users adjusted relative risk (95% CI): 1.31 (0.55 to 3.12) Adjusted for age at menopause, BMI and smoking history.	Other information Limitations Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. Unclear. Attempts were made within the design or analysis to balance the comparison groups for potenial confounders. Yes. The groups were comparable at baseline, including all major confounding and prognostic factors. No - oestrogen users were more liekly to be white, current smokers and nulliparous and were 1 year older at menopause. Performance bias The comparison groups received the same care apart from the intervention(s) studied.

Study details	Study design	Comparison	Results	Other
Country/ies where the study was carried out USA Study dates Cohort identified in 1980, using records from 1968 to 1971. Study duration 25.4 years.	alcoholism, chronic renal or hepatic disease, hyper- or hypo-parathyroidism, diabetes mellitus, hyperthyroidism, other conditions known to affect skeletal integrity (immobilization, malnutrition or severe debilitating chronic disease of any sort).			Yes. Participants receiving care were kept 'blind' to treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important

Study details	Study design	Comparison	Results	Other
				confounding and prognostic factors. Unclear.
Full citation Melton,L.J.,III, Crowson,C.S., Malkasian,G.D., O'Fallon,W.M., Fracture risk following bilateral oophorectomy, Journal of Clinical Epidemiology, 49, 1111-1115, 1996 Ref Id 308135 Study type Prospective cohort study. Source of funding National Institutes of Health, US Public Health Service. Country/ies where the study was carried out USA Study dates Cohort identified from 1959 to 1979. Study duration 30 years.	Aim of the study To estimate the risk of fractures of the hip, spine and distal forearm among an inception cohort of premenopausal women who had bilateral oophorectomy for a benign ovarian condition. Inclusion criteria Women who underwent oophorectomy from 1959 to 1979 at the Mayo Clinic. Premenopausal at the time of surgery. Exclusion criteria Surgery due to a malignant condition.	Details Women who had ever taken oestrogen replacement therapy (for > 3 months in total) were compared to those who did not take HRT. Methods Participants were followed through their records in the community until death, or the date of the last medical record entry. Follow up was complete to death in 12% (median 8.5 years of follow up per person) and was for a median of 15.1 years for survivors. Only fractures that occurred after the date of oophorectomy were considered for this analysis. The records contained the clinical history and the radiologists report of each fracture, but the original X-rays were not available for review. Ascertainment of the fractures of interest is believed to be complete except for vertebral fractures, some of which are never diagnosed. Sample size N = 463 n = 259 users of HRT n = 204 non-users of HRT	Characteristics Median age at surgery 43.8 years (range 18 to 56 years). Ever use of HRT: 56% Results Ever treatment with HRT Risk of hip fracture in women treated with HRT for at least 3 months, compared to those never treated with HRT adjusted relative risk (95% CI): 0.8 (0.2 to 2.6) Risk of vertebral fracture in women treated with HRT for at least 3 months, compared to those never treated with HRT adjusted relative risk (95% CI): 0.8 (0.4 to 1.9) Risk of wrist fracture in women treated with HRT for at least 3 months, compared to those never treated with HRT adjusted relative risk (95% CI): 1.6 (0.8 to 3.2) Duration of treatment with HRT Risk of vertebral fracture per 5 years of HRT therapy compared to no treatment adjusted odds ratio (95% CI): 0.4 (0.2 to 0.97) Risk of wrist fracture per 5 years of HRT therapy compared to no treatment adjusted odds ratio (95% CI): 0.7 (0.4 to 1.2) Risk of hip fracture per 5 years of HRT therapy compared to no treatment adjusted odds ratio (95% CI): 0.8 (0.3 to 2.0)	Other information Limitations Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. Unclear. Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline, including all major confounding and prognostic factors. Unclear. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group?

Study details	Study design	Comparison	Results	Other
				Not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Middleton,E. T., Steel,S.A., The effects of short-term hormone replacement therapy on long-term bone mineral density, Climacteric, 10, 257-263,	Aim of the study To investigate whether women who take short-term HRT around the time of the menopause have long-term gains in their bone mineral density as compared to those who take no treatment. Inclusion criteria	Details Women considered at risk of osteoporosis at baseline (due to a BMD in the lowest quartile for their age matched population) were recommended treatment with HRT. Those women considered at risk, and an equal number of randomly selected women not recommended for treatment were invited back for repeated assessment 2, 5 and 9 years later. Methods All women who were followed up for 9 years as part of a screening program were included. Women were allocated to one of three groups: • no HRT • 24 to 48 months of HRT prior to the 5 years visit (i.e. followed by 4 years without HRT)	Characteristics No HRT group: Age, mean years (95% CI): 52.5 (1.4) Weight mean kg (95% CI): 67.1 (10.6) Age at menopause, mean years (95% CI): 49.3 (4.7) Short term HRT group: Age, mean years (95% CI): 52.5 (1.33) Weight mean kg (95% CI): 63.5 (9.6) Age at menopause, mean years (95% CI): 49.1 (3.6) Results Risk of any fracture in short-term HRT group, compared to no HRT group (2 to 4 years HRT treatment, followed by 5 years without treatment): relative risk (95% CI) : 0.46 (0.14 to 1.57)	Other information Limitations Study results subject to bias, as women taking HRT in this study were known to be osteopenic at baseline, as compared to women not taking HRT. Therefore, the fracture risk in women taking HRT is likely to have been increased as compared with the fracture risk in non-users

Study details	Study design	Comparison	Results	Other
2007 Ref Id 232444 Study type Prospective cohort study. Source of funding National Osteoporosis Society part funded the follow up visits. Country/ies where the study was carried out UK Study dates Recruitment during 1990s. Study duration 9 years.	Women aged 50 to 54 years at baseline. Exclusion criteria Terminal illness, with in excess of 125kg or physical inability to comply with the standard DXA scanning technique. Use of bisphosphonates or raloxifene before or during the follow up period.	• HRT use for at least 8.5 years Fracture data is reported for the first two groups only. Sample size N = 400 (excluding patients taking long term HRT as no fracture data available) n = 340 no HRT n = 60 short term HRT	Adjusted for baseline BMD.	at baseline. However, study results do adjust for baseline BMD. Furthermore, women taking HRT were made aware of their risk of osteoporosis, therefore may have taken other steps to reduce their risk of fracture. Any beneficial effect of HRT may therefore be confounded by other lifestyle modifications (calcium intake, exercise etc.) Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. No. Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline, including all major confounding and prognostic factors. No. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept ‘blind’ to treatment allocation. No. Individuals administering care were kept ‘blind’ to treatment allocation. No. Attrition bias All groups were followed

Study details	Study design	Comparison	Results	Other
				up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Unclear. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Mosekilde,L., Beck-Nielsen,H., Sorensen,O. H.,	Aim of the study To study the fracture reducing potential of HRT in recent postmenopausal	Details Comparison was made between women who were treated with HRT and those who were given placebo (within the RCT arm). Comparison was also made between women who were treated/not treated with HRT through their own choice, but no risk adjustment was made to account for confounders, therefore	Characteristics Randomised to HRT group: Age, years (mean ± SD): 49.5 ± 2.7 BMI kg/m² (mean ± SD): 25.3 ± 4.3 Previous fracture (%): 21	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used

Study details	Study design	Comparison	Results	Other
Nielsen,S.P., Charles,P., Vestergaard, P., Hermann,A.P ., Gram,J., Hansen,T.B., Abrahamsen, B., Ebbesen,E.N ., Stilgren,L., Jensen,L.B., Brot,C., Hansen,B., Tofteng,C.L., Eiken,P., Kolthoff,N., Hormonal replacement therapy reduces forearm fracture incidence in recent postmenopa usal women - results of the Danish Osteoporosis Prevention Study, Maturitas, 36, 181-193, 2000 Ref Id 232505 Study type Randomised controlled trial and prospective cohort study. Source of funding Karen Elise	women in a primary preventive scenario. Inclusion criteria Women with a uterus aged 45 to 58 years old, within 3 to 34 months since their last menstrual period, or experiencing perimenopausal symptoms combined with elevated serum FSH levels. Hysterectomised women aged 45 to 52 years old with elevated FSH. Exclusion criteria Metabolic bone disease (including osteoporosis, defined as non-traumatic vertebral fractures on X-ray). Current oestrogen use, or oestrogen use within the past 3 months. Current or past treatment with glucocorticoids for over 6 months. Current or past malignancy. Newly diagnosed or uncontrolled chronic disease. Alcohol or drug addiction.	these data were not used for this analysis. Methods Women were recruited to the study and asked whether they agreed to being randomised to HRT or no HRT. Those who accepted randomisation were block randomised in groups of ten by the envelope method to HRT treatment (sequential combined HRT for women with a uterus [2mg oestradiol for 12 days, 2mg oestradiol plus 1mg norethisterone acetate for 10 days, then 1mg oestradiol for 6 days] or oestrogen only for women with a previous hysterectomy [2mg oestradiol daily]). Treatment was not blinded. If a change of HRT type was required, a number of alternatives were available. Women were followed up for a duration of 5 years. X-rays of the spine (T4 to L5) were obtained at baseline and after 5 years. A fracture was defined as more than 20% reduction in the height of a vertebrae, compared to the highest vertical distance of that vertebrae. Sample size N = 1006 n = 502 randomised to HRT n = 504 randomised to no treatment (additional women participated in cohort study, but not included in this analysis)	Time since menopause, years (mean ± SD): 0.7 ± 0.6 BMD of lumbar spine g/cm² (mean ± SD): 1.041 ± 0.141 Randomised to no treatment group: Age, years (mean ± SD): 50.0 ± 2.8 BMI kg/m² (mean ± SD): 25.2 ± 4.5 Previous fracture (%): 21 Time since menopause, years (mean ± SD): 0.7 ± 0.6 BMD of lumbar spine g/cm² (mean ± SD): 1.016 ± 0.127 Results Randomised arm of study: Risk of any fracture in HRT treated group compared to untreated group unadjusted relative risk (95% CI): 0.82 (0.53 to 1.29) Risk of vertebral fracture in HRT treated group compared to untreated group unadjusted relative risk (95% CI): 2.00 (0.62 to 6.49) Risk of hip fracture in HRT treated group compared to untreated group unadjusted relative risk (95% CI): 3.01 (0.12 to 73.76)	to allocate participants to treatment groups. Yes. There was adequate concealment of allocation. Yes. The groups were comparable at baseline. Yes. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No - open label design. Individuals administering care were kept 'blind' to treatment allocation. No - open label design. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 55 no treatment group, n = 54 HRT group. The groups were comparable for treatment completion. Yes. For how many participants in each group were outcome data not available? n = 55 no treatment group, n = 54 HRT group. The groups were comparable with respect to the availability of

Study details	Study design	Comparison	Results	Other
Jensen's Foundation. Danish Medical Research Council. Novo Nordisk Denmark, Novartis Denmark and Leo Denmark provided the study medication free of charge. Country/ies where the study was carried out Denmark Study dates November 1990 to March 1993. Trial duration 5 years.				outcome data. Yes. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Paganini-Hill,A., Atchison,K.A., Gornbein,J.A., Nattiv,A., Service,S.K., White,S.C., Menstrual and reproductive factors and fracture risk: the Leisure World Cohort Study, Journal of Women's	Aim of the study To investigate the potential associations of oestrogen exposure and the risk of osteoporotic fracture in a large, population based, prospective cohort study of older women. Inclusion criteria Residents of a California retirement community. Exclusion criteria	Details Comparison of fracture risk in women who had ever used HRT, compared to those who had never used HRT. Also compared fracture risk according to duration of oestrogen therapy and years since last oestrogen therapy. Methods A baseline postal survey was completed at recruitment. Follow up surveys were used to identify incident fractures in 1983, 1985, 1992 and 1998. Follow up was from 1981 to 2002. Follow up time was calculated as the time from the initial survey to the first fracture of interest, or censoring. Sample size N = 8850 n = 4987 ever users of HRT n = 3863 never users of HRT	Characteristics Baseline characteristics: Age, years (mean ± SD): 73 ± 7.4 BMI, kg/m² (mean ± SD): 23 ± 3.5 Ever use of postmenopausal oestrogens (%): 56 Results Ever use of HRT compared to never use of HRT Risk of wrist fracture in ever users of HRT compared to never users: adjusted hazard ratio (p value): 0.95 (NS) Risk of vertebral fracture in ever users of HRT compared to never users: adjusted hazard ratio (p value): 0.95 (NS) Duration of use of HRT, compared to never use of HRT Risk of wrist fracture in users of HRT for < 3 years compared to never users: adjusted hazard ratio (p value): 1.15 (NS) Risk of vertebral fracture in users of HRT for < 3 years	Other information Limitations Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. Unclear. Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline, including all major confounding and

Study details	Study design	Comparison	Results	Other
Health, 14, 808-819, 2005 Ref Id 232655 Study type Prospective cohort study. Source of funding National Institutes of Health. Earl Carroll Trust Fund. Wyerth-Ayerst Laboratories. Country/ies where the study was carried out USA Study dates Recruitment took place from 1981. Study duration was for 21 years.	Not reported.		compared to never users: adjusted hazard ratio (p value): 0.79 (NS) Risk of wrist fracture in users of HRT for 3 to 14 years compared to never users: adjusted hazard ratio (p value): 0.85 (NS) Risk of vertebral fracture in users of HRT for 3 to 14 years compared to never users: adjusted hazard ratio (p value): 1.01 (NS) Risk of wrist fracture in users of HRT for ≥ 15 years compared to never users: adjusted hazard ratio (p value): 0.85 (NS) Risk of vertebral fracture in users of HRT for ≥ 15 years compared to never users: adjusted hazard ratio (p value): 0.93 (NS) Length of time since last oestrogen therapy, compared to never use Risk of wrist fracture in users of HRT who discontinued ≥ 15 years ago, compared to never users: adjusted hazard ratio (p value): 1.30 (NS) Risk of vertebral fracture in users of HRT who discontinued ≥ 15 years ago, compared to never users: adjusted hazard ratio (p value): 0.86 (NS) Risk of wrist fracture in users of HRT who discontinued 2 to 14 years ago, compared to never users: adjusted hazard ratio (p value): 0.90 (NS) Risk of vertebral fracture in users of HRT who discontinued 2 to 14 years ago, compared to never users: adjusted hazard ratio (p value): 1.05 (NS) Risk of wrist fracture in users of HRT who discontinued ≤ 1 year ago, compared to never users: adjusted hazard ratio (p value): 0.60 (p = 0.05) Risk of vertebral fracture in users of HRT who discontinued ≤ 1 year ago, compared to never users: adjusted hazard ratio (p value): 0.82 (NS) Adjusted for history of fracture, BMI, heart attack, alcohol consumption, vitamin A supplement use, cola intake and hysterectomy (for wrist fracture) and for history of fracture, BMI, blood pressure medication, non-prescription pain medication, smoking, exercise and attitude (for vertebral fracture).	prognostic factors. Unclear. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome.

Study details	Study design	Comparison	Results	Other
			Article does not report 95% confidence intervals, only p values for comparisons. NS: not significant Data for hip fracture also reported, but more robust data presented in Paganini-Hill et al 1991, therefore these data were used.	Unclear. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Paganini-Hill,A., Chao,A., Ross,R.K., Henderson,B.E., Exercise and other factors in the prevention of hip fracture: the Leisure World study, Epidemiology, 2, 16-25, 1991 Ref Id 295180 Study type Prospective cohort study. Source of funding The National Cancer Institute, National Institutes of Health. Country/ies where the study was carried out USA Study dates Recruitment	Aim of the study To assess the association between postmenopausal hip fractures and a variety of health and lifestyle factors. Inclusion criteria Residents of Leisure World retirement community near Los Angeles, California. Exclusion criteria Not reported.	Details Comparison was made between participants who took any oestrogen and those who did not. Analysis was also given depending on the duration of oestrogen use and recency of use. Methods A detailed baseline questionnaire was completed by all participants. Follow up questionnaires were sent in 1983 and 1985. Sample size N = 8600 n = 332 with hip fracture n = 8268 without hip fracture	Characteristics Median age 73 years. Other characteristics not reported. Results Risk of hip fracture in ever users of oestrogen compared to never users adjusted relative risk (95% CI): 1.02 (0.81 to 1.27) Duration of oestrogen use Risk of hip fracture in ever users of oestrogen for ≤ 3 years compared to never users adjusted relative risk (95% CI): 1.19 (0.89 to 1.60) Risk of hip fracture in ever users of oestrogen for 4 to 14 years compared to never users adjusted relative risk (95% CI): 0.89 (0.63 to 1.23) Risk of hip fracture in ever users of oestrogen for ≥ 15 years compared to never users adjusted relative risk (95% CI): 0.88 (0.63 to 1.24) Recency of oestrogen use Risk of hip fracture in users of oestrogen who discontinued 0 to 1 year ago, compared to never users adjusted relative risk (95% CI): 0.80 (0.53 to 1.21) Risk of hip fracture in users of oestrogen who discontinued 2 to 14 years ago, compared to never users adjusted relative risk (95% CI): 0.88 (0.63 to 1.23) Risk of hip fracture in users of oestrogen who discontinued ≥ 15 years ago, compared to never users adjusted relative risk (95% CI): 1.15 (0.88 to 1.50) Duration of use and time since stopping Risk of hip fracture in users of oestrogen for ≤ 3 years who discontinued 0 to 1 years ago, compared to never users adjusted relative risk (95% CI): 0.87 (0.28 to 2.73) Risk of hip fracture in users of oestrogen for ≤ 3 years who discontinued 2 to 14 years ago, compared to never users	Other information Although median age of participants was 73, data on "ever use" compared to "never use" are repoted, as well as data on time since stopping HRT, and total duration of treatment, which would be relevant to women under 65. Limitations Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. Unclear. Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline, including all major confounding and prognostic factors. Unclear. Performance bias The comparison groups received the same care apart from the intervention(s) studied.

Study details	Study design	Comparison	Results	Other
began in June 1981. Follow up for this analysis was until April 1 1988. Study duration 7 years.			adjusted relative risk (95% CI): 0.79 (0.38 to 1.60) Risk of hip fracture in users of oestrogen for ≤ 3 years who discontinued ≥ 15 years ago, compared to never users adjusted relative risk (95% CI): 1.33 (0.97 to 1.82) Risk of hip fracture in users of oestrogen for 4 to 14 years who discontinued 0 to 1 years ago, compared to never users adjusted relative risk (95% CI): 0.72 (0.31 to 1.64) Risk of hip fracture in users of oestrogen for 4 to 14 years who discontinued 2 to 14 years ago, compared to never users adjusted relative risk (95% CI): 0.86 (0.52 to 1.42) Risk of hip fracture in users of oestrogen for 4 to 14 years who discontinued ≥ 15 years ago, compared to never users adjusted relative risk (95% CI): 0.95 (0.61 to 1.49) Risk of hip fracture in users of oestrogen for ≥ 15 years who discontinued 0 to 1 years ago, compared to never users adjusted relative risk (95% CI): 0.85 (0.53 to 1.38) Risk of hip fracture in users of oestrogen for ≥ 15 years who discontinued 2 to 14 years ago, compared to never users adjusted relative risk (95% CI): 0.97 (0.61 to 1.53) Risk of hip fracture in users of oestrogen for ≥ 15 years who discontinued ≥ 15 years ago, compared to never users adjusted relative risk (95% CI): 0.57 (0.18 to 1.79)	Yes. Participants receiving care were kept 'blind' to treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important

Study details	Study design	Comparison	Results	Other
				confounding and prognostic factors. Unclear.
Full citation Randell,K.M., Honkanen,R. J., Kroger,H., Saarikoski,S. , Does hormone-replacement therapy prevent fractures in early postmenopausal women?, Journal of Bone and Mineral Research, 17, 528-533, 2002 Ref Id 232807 Study type Prospective cohort study. Source of funding European Foundation for Osteoporosis Yrjö Jahnsson Foundation The Ministry of Health and Social Affairs The Academy of Finland Country/ies where the	Aim of the study To evaluate the effect of HRT on clinically diagnosed bone fractures in early postmenopausal women. Inclusion criteria Women aged 47 to 56 years residing in Kuopio Province Eastern Finland in May 1989. Post menopausal (≥ 6 months since last natural menstruation). Exclusion criteria Women whose menopause could not be defined because of a hysterectomy performed before menopause.	Details Risk of any fracture was compared between women who had used HRT in the past (> 5 years ago, before the baseline inquiry), women who were current uers of HRT for at least 4.5 years and never users of HRT. Methods Postal inquiries were sent to all participants at baseline, and again 5 years later. Women were grouped into those who had never used HRT, those who had reported past use at the baseline inquiry but no further use, and those who had reported continuous use during the 5 years follow up (> 4.5 years). Analysis was also performed on those women who had used HRT for some of the time during the 5 years follow up. Sample size N = 7217 n = 3335 never use of HRT n = 130 past use of HRT (before baseline inquiry) n = 1335 continuous use of HRT during follow up Remainder were part-time users of HRT during the period of the study (n = 1335). These participants were excluded from this analysis.	Characteristics Age, years (mean ± SD): 53.3 ± 2.7 Time since menopause, years (mean ± SD): 4.05 ± 4.07 BMI, kg/m² (mean ± SD): 26.3 ± 4.3 Menopause status > 5 years ago (%): 30.8 Results Risk of any fracture in past users of HRT (discontinued ≥ 5 years ago) compared to never users of HRT adjusted relative risk (95% CI): 1.02 (0.82 to 1.26) Risk of wrist fracture in past users of HRT (discontinued ≥ 5 years ago) compared to never users of HRT adjusted relative risk (95% CI): 1.44 (1.06 to 1.95) Risk of any fracture in current users of HRT (> 4.5 years of use in the past 5 years) compared to never users of HRT adjusted relative risk (95% CI): 0.62 (0.48 to 0.79) Risk of wrist fracture in current users of HRT (> 4.5 years of use in the past 5 years) compared to never users of HRT adjusted relative risk (95% CI): 0.41 (0.26 to 0.67) Adjusted for age,, time since menopause, BMI, number of chronic health disorders and history of previous fractures.	Other information Limitations Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. Unclear. Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline, including all major confounding and prognostic factors. No, therewere significant differences in age, time since menopause, heigh, weight, BMI, dietary calcium intake, history of oophorectomy, history of hysterectomy, smoking status, physical activity, number of health disorders and use of calcium supplements. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept ‘blind’ to treatment allocation. No. Individuals administering care were kept ‘blind’ to treatment allocation. No.

Study details	Study design	Comparison	Results	Other
study was carried out Finland Study dates Recruitment took place in May 1989. 5 year follow up occurred in May 1994.				Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Ravn,P., Bidstrup,M., Wasnich,R.D	Aim of the study To compare the effects of alendronate,	Details Women were randomised to treatment with 5mg oral alendronate, 2.5mg oral alendronate, placebo or HRT. Methods	Characteristics HRT group Age, years (mean ± SD): 55 ± 3 Time since menopause, years (mean ± SD): 5 ± 3	Other information Limitations Study quality Selection bias

Study details	Study design	Comparison	Results	Other
., Davis,J.W., McClung,M. R., Balske,A., Coupland,C., Sahota,O., Kaur,A., Daley,M., Cizza,G., Alendronate and estrogen-progestin in the long-term prevention of bone loss: four-year results from the early postmenopausal intervention cohort study. A randomized, controlled trial, Annals of Internal Medicine, 131, 935-942, 1999 Ref Id 232820 Study type Randomised controlled trial. Source of funding Merck Research Laboratories. Country/ies where the study was carried out	placebo and HRT on bone mass and bone turnover. Inclusion criteria Healthy women aged 45 to 59 years. At least 6 months post menopausal at baseline. Exclusion criteria Not reported.	In the USA, conjugated equine oestrogens 0.625mg plus 5mg medroxyprogesterone acetate were used as the HRT preparation. In Europe a cyclic combined regimen of estradiol 2mg/d for 22 days, norethisterone acetate 1mg/d on days 13 to 22 and estradiol 1mg/d on day 23 to 28 was used. All patients were reviewed every 3 months. Total follow up was for 4 years of treatment. Sample size N = 612 n = 110 HRT n = 502 placebo (additional participants were randomised to alendronate, but are not included in this analysis)	BMI, kg/m² (mean ± SD): 25 ± 4 BMD at lumbar spine g/cm² (mean ± SD): 0.98 ± 0.12 Placebo group Age, years (mean ± SD): 55 ± 4 Time since menopause, years (mean ± SD): 8 ± 5 BMI, kg/m² (mean ± SD): 25 ± 4 BMD at lumbar spine g/cm² (mean ± SD): 0.92 ± 0.12 Results Risk of any fracture in HRT group compared to placebo group: relative risk (95% CI): 0.59 (0.24 to 1.45)	An appropriate method of randomisation was used to allocate participants to treatment groups. Unclear. There was adequate concealment of allocation. Unclear. The groups were comparable at baseline. No - women in the HRT group had experienced menopause more recently (5 ± 3 years) than those in the placebo group (8 ± 5 years). Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No, HRT was administered as an open label preparation. Individuals administering care were kept 'blind' to treatment allocation. No. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 134 placebo group, n = 28 HRT group. The groups were comparable for treatment

Study details	Study design	Comparison	Results	Other
USA, UK, Denmark. Study dates Not reported. Trial duration 4 years.				completion. Yes. For how many participants in each group were outcome data not available? n = 134 placebo group, n = 28 HRT group. The groups were comparable with respect to the availability of outcome data. Yes. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Unclear. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Reid,I.R., Eastell,R., Fogelman,I., Adachi,J.D., Rosen,A., Netelenbos, C., Watts,N.B., Seeman,E., Ciaccia,A.V., Draper,M.W., A comparison of the effects	Aim of the study To compare the long term lipid and skeletal effects of raloxifene and oestrogen. Inclusion criteria Postmenopausal women aged 40 to 60 years. Previous hysterectomy (no more than 15 years before the	Details Women were assigned to one of four treatment groups: 60mg/d raloxifene, 150mg/d raloxifene, 0.625mg/d conjugated equine oestrogens or placebo. All women were also given a daily supplement of 400 to 600mg of elemental calcium. Methods Study visits occurred every 3 months for 24 months, and then every 6 months for a further year (total of 3 years follow up). Lateral spine radiographs were performed at baseline and at 3 years and fractures were assessed semi-quantitively. Sample size N = 310 n = 158 HRT n = 152 placebo	Characteristics HRT group: Age, years (mean ± SD): 52.7 ± 4.7 Time since menopause, years (mean ± SD): 6.5 ± 6.0 BMI, kg/m² (mean ± SD): 27.1 ± 5.1 Placebo group: Age, years (mean ± SD): 53.0 ± 4.7 Time since menopause, years (mean ± SD): 6.0 ± 5.0 BMI, kg/m² (mean ± SD): 27.5 ± 4.7 Results Risk of vertebral fracture in women receiving HRT compared to placebo: unadjusted relative risk (95% CI): 0.96 (0.06 to 15.24)[1]	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Yes. There was adequate concealment of allocation. Yes. The groups were comparable at baseline. Yes. Performance bias

Study details	Study design	Comparison	Results	Other
of raloxifene and conjugated equine estrogen on bone and lipids in healthy postmenopausal women, Archives of Internal Medicine, 164, 871-879, 2004 Ref Id 254776 Study type Randomised controlled trial. Source of funding Lilly Research Laboratories. Country/ies where the study was carried out Europe, North America, Australasia and South Africa. Study dates Not reported. Trial duration 3 years.	start of the study). Serum oestradiol < 73 pmol/L. FSH level of ≥ 40 mIU/mL. Lumbar spine BMD between 2.5 SDs below and 2.0 SDs above the mean value for normal premenopausal women. Exclusion criteria History of breast cancer or oestrogen dependent tumours. Use of oestrogen, progestin, androgen, calcitonin or systemic corticosteroids within the previous 6 months. Ever use of bisphosphonate or fluoride. Current use of anti-epileptics, pharmacological doses of vitamin D or lipid lowering drugs. History of thromboembolic disorders, diabetes mellitus of other endrocrine disorders requiring therapy (except thyroid hormone	(additional women included in raloxifene treatment groups, but not included for this analysis.)	[1] Calculated by the NCC-WCH technical team from data reported in the article.	The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Unclear - presumed not blinded. Individuals administering care were kept 'blind' to treatment allocation. Unclear - presumed not blinded. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 62 placebo, n = 56 HRT group. The groups were comparable for treatment completion. Yes. For how many participants in each group were outcome data not available? n = 62 placebo, n = 56 HRT group. The groups were comparable with respect to the availability of outcome data. Yes. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome.

Study details	Study design	Comparison	Results	Other
	therapy). Abnormal renal or hepatic function. Serious postmenopausal symptoms. Consumption of more than 4 alcoholic drinks per day.			Yes. A valid and reliable method was used to determine the outcome. Unclear. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear - presumed not blinded. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Tuppurainen, M., Kroger,H., Honkanen,R., Puntila,E., Huopio,J., Saarikoski,S., Alhava,E., Risks of perimenopausal fractures--a prospective population-based study, Acta Obstetricia et Gynecologica Scandinavica, 74, 624-628, 1995 Ref Id 295400 Study type Prospective cohort study. Source of	Aim of the study To examine the associations between potential risk factors, including gynaecological and behavioural variables, and fractures. Inclusion criteria Women aged 47 to 56 years old at baseline, residing in Kuopio Province, Eastern Finland. Exclusion criteria Not reported.	Details Characteristics were compared between women with and without a history of fractures. Methods Information on the occurrence of fractures, time and site of fracture, causes and treatment and the place of treatment were obtained in a postal enquiry in December 1992. All reported fractures were verified by examination of the patients' medical records, but X-ray films were not checked. BMD measurements were taken at the lumbar spine and femoral neck in 1990 to 1991, and only fracture data reported after the BMD measurement were taken into account. Fractures resulting from a fall from standing height or less were classified as low energy fractures. A few rib fractures were diagnosed only on clinical examination. All vertebral fractures were based on x-ray examination. Fractures resulting from car accidents of other high energy accidents were excluded. The mean observation time was 2.4 years (range 2 days to 3.4 years). In fracture patients the duration of HRT was calculated as the treatment time up to the occurence of the first fracture. In non-fracture participants the respective time interval was until the end of 1992. Sample size N = 3140 n = 157 sustained a fracture n = 2983 no fracture	Characteristics Fracture group Age, years (mean ± SD): 53.7 ± 2.9 BMI, kg/m² (mean ± SD): 26.0 ± 4.9 Lumbar spine BMD, g/cm² (mean ± SD): 1.063 ± 0.160 Non-fracture group Age, years (mean ± SD): 53.4 ± 2.8 BMI, kg/m² (mean ± SD): 26.1 ± 4.3 Lumbar spine BMD, g/cm² (mean ± SD): 1.131 ± 0.158 Results Risk of fracture in past or present users of HRT, compared to never users: Adjusted odds ratio (95% CI): 0.70 (0.50 to 0.96) Adjusted for age	Other information Limitations Study quality Selection bias The method of allocation to treatment groups was unrelated to potential confounding factors. Unclear. Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline, including all major confounding and prognostic factors. Unclear - baseline characteristics only reported fro fracture cases versus no fracture cases. Performance bias The comparison groups received the same care apart from the

Study details	Study design	Comparison	Results	Other
funding University of Kuopio Yrjö Jahnsson Foundation Country/ies where the study was carried out Finland. Study dates Recruitment during 1989. Duration of study 2.4 years.				intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept

Study details	Study design	Comparison	Results	Other
				'blind' to other important confounding and prognostic factors. Unclear.
Full citation Veerus,P., Hovi,S.L., Fischer,K., Rahu,M., Hakama,M., Hemminki,E., Results from the Estonian postmenopausal hormone therapy trial [ISRCTN35338757], Maturitas, 55, 162-173, 2006 Ref Id 230596 Study type Randomised controlled trial. Source of funding Academy of Finland. STAKES (National Research and Development Centre for Welfare and Health) The Estonian ministry of Education and Research. Trial	Aim of the study To ascertain harms and benefits of combined continuous hormone therapy. Inclusion criteria Women aged 50 to 64 years old. Postmenopausal. Exclusion criteria Medical contraindication to hormone therapy.	Details Women were randomised into 4 groups: HRT (blinded to treatment allocation) Placebo (blinded to treatment allocation) HRT (aware of treatment allocation) Control (aware of treatment allocation) Methods The HRT preparation use comprised 0.625mg conjugated oestrogens and 2.5mg medroxyprogesterone acetate. Women within 3 years of their last menstrual period were given 5.0mg medroxyprogesterone acetate instead of 2.5mg. Sample size N = 1778 n = 494 open label HRT n = 507 control n = 404 blind HRT n = 373 placebo	Characteristics Open label HRT group Age, years (mean ± SD): 58.6 ± 4.0 Age at menopause, years (mean ± SD): 50.2 ± 3.9 BMI, kg/m² (mean ± SD): 27.2 ± 4.5 Control group Age, years (mean ± SD): 58.9 ± 4.0 Age at menopause, years (mean ± SD): 50.5 ± 4.0 BMI, kg/m² (mean ± SD): 26.9 ± 4.6 Blind HRT group Age, years (mean ± SD): 58.5 ± 3.9 Age at menopause, years (mean ± SD): 50.4 ± 3.8 BMI, kg/m² (mean ± SD): 27.0 ± 4.8 Placebo group Age, years (mean ± SD): 59.0 ± 3.9 Age at menopause, years (mean ± SD): 50.3 ± 3.9 BMI, kg/m² (mean ± SD): 26.9 ± 4.2 Results Risk of any fracture in HRT groups (open label and blinded combined) compared to no HRT adjusted hazard ratio (95% CI): 0.61 (0.42 to 0.89)	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Yes. There was adequate concealment of allocation. Yes. The groups were comparable at baseline. Yes. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Trial included a 'blind' arm and a 'non-blind' arm. Individuals administering care were kept 'blind' to treatment allocation. Unclear. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? None. The groups were comparable for treatment

Study details	Study design	Comparison	Results	Other
medications were provided by Wyeth Ayerst. Country/ies where the study was carried out Estonia Study dates Recruitment in January 1999 to December 2001. Follow up for 2 to 5 years.				completion. Yes. For how many participants in each group were outcome data not available? None. The groups were comparable with respect to the availability of outcome data. Yes. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Vickers,M.R., MacLennan, A.H., Lawton,B., Ford,D., Martin,J., Meredith,S.K ., DeStavola,B. L., Rose,S., Dowell,A., Wilkes,H.C., Darbyshire,J. H., Meade,T.W., WISDOM	Aim of the study To assess the long term risks and benefits of HRT. Inclusion criteria Postmenopausal women aged 50 to 69 (no menstrual period in the last 12 months, or had undergone hysterectomy). Exclusion criteria History of breast cancer, any other	Details Three treatment arms were included:- 1. Combined HRT (0.625mg conjugated equine oestrogens plus 2.5mg or 5.0mg medroxyprogesterone acetate daily). 5.0mg dose of MPA was used for women with a uterus and within 3 years of their last period, those aged 50-53, and older women with unacceptable breakthrough bleeding. Women with a uterus who experienced unacceptable spotting or bleeding with the 5.0mg dose were offered open label Premarin 0.625mg orally daily plus MPA 10mg orally for the last 14 days of a 28 days cycle. 2. Oestrogen alone HRT (0.625mg conjugated equine oestrogens daily) 3. Placebo For the purpose of this review, only data from the combined HRT versus placebo arm was included (oestrogen alone preparation was only compared to oestrogen plus progesterone, not to	Characteristics Mean age: 62.9 ± 4.8 years Use of HRT at screening: 1175/5692 (21%) Ever use of HRT at screening: 3144/5692 (55%) Mean BMI: 28.0 ± 5.0 kg/m² Results Comparison of combined HRT to placebo. Any osteoporotic fracture Hazard ratio (95% CI): 0.69 (0.46 to 1.03) Hip fracture Relative risk (95% CI): 0.66 (0.11 to 3.97)[1] [1] Calculated by the NCC-WCH technical team from data provided in the article.	Other information Trial stopped prematurely due to publication of WHI data. Limitations As far as possible the trial was conducted in a double-blind manner. However, this was not possible when vaginal bleeding triggered a code break and investigation for possible pathology. Study quality Selection bias An appropriate method of randomisation was used

Study details	Study design	Comparison	Results	Other
group., Main morbidities recorded in the women's international study of long duration oestrogen after menopause (WISDOM): a randomised controlled trial of hormone replacement therapy in postmenopausal women, BMJ, 335, 239-, 2007 Ref Id 230610 Study type Randomised, double blind, placebo controlled trial. Source of funding UK Medical Research Council, British Heart Foundation, Department of Health for England, Scottish Office, Welsh Office, Department of Health and Social Services for	cancer in the past 10 years (except basal and squamous cell skin cancer), endometriosis or endometrial hyperplasia, venous thromboembolism, gall bladder disease in women who had not had a cholecystectomy, myocardial infarction, unstable angina, cerebrovascular accident, subarachnoid haemorrhage, transient ischaemic attack. Use of HRT within the last 6 months. Women taking HRT at screening who were prepared to enter the study agreed to stop the therapy for three months before the run-in phase. During run-in all participants took placebo, so that at randomisation they had not taken HRT for 6 months.	placebo, and the numbers of fractures sustained are unclear, due to duplicate data entry). Methods Treatment was randomly allocated centrally with a computer based, stratified block randomisation program. Stratification was based on hysterectomy status and intended use of HRT. Women with a uterus or previous subtotal hysterectomy were randomised to combined oestrogen plus progestin or to placebo using a block size of 16. Women with no uterus and unwilling to take placebo were randomised to either oestrogen alone or combined oestrogen and progestin therapy using a block size of 16. Women with no uterus willing to enter a placebo controlled comparison were randomised to oestrogen alone, combined oestrogen plus progestin or placebo using a block size of 24. Outcome data were collected at each follow up visit. A member of the study team confirmed any data needed to verify a clinical event with the GP, hospital or coroner. 10% of fractures were reviewed by indenpendent assessors. Sample size N = 5692 total n = 2196 combined oestrogen and progesterone n = 2189 placebo (Remaining women allocated to comparison of oestrogen alone therapy to oestrogen and progestin HRT).		to allocate participants to treatment groups. Yes. There was adequate concealment of allocation. Yes. The groups were comparable at baseline. Yes. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Yes. Individuals administering care were kept 'blind' to treatment allocation. Yes. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 415 HRT, n = 200 placebo. The groups were comparable for treatment completion. No - more women withdrew from the HRT arm than placebo. For how many participants in each group were outcome data not available? 5 women in total (data for individual groups not reported). The groups were comparable with respect to the availability of

Study details	Study design	Comparison	Results	Other
Northern Ireland, Royal Australian and New Zealand College of Obstetricians and Gynaecologists, Australasian Menopause Society, National Health and Medical Research Council, National Heart Foundation of Australia, The Cancer Council of South Australia, The Cancer Society of New Zealand, NHS R&D Executive. Wyeth Ayerst provided active drugs and matched placebo but had no other involvement in the trial. Country/ies where the study was carried out UK, Australia				outcome data. Yes. Detection bias The study had an appropriate length of follow up. No - trial terminated prematurely. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Yes. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.

Study details	Study design	Comparison	Results	Other
and New Zealand. Study dates 1999 to 2002. Trial terminated prematurely after median follow up 11.9 months (planned treatment duration 10 years).				
Full citation Weiss,S.R., Ellman,H., Dolker,M., A randomized controlled trial of four doses of transdermal estradiol for preventing postmenopausal bone loss. Transdermal Estradiol Investigator Group, Obstetrics and Gynecology, 94, 330-336, 1999 Ref Id 233468 Study type Randomised controlled trial. Source of funding	Aim of the study To investigate the efficacy of different doses of a transdermal oestradiol delivery system for the prevention of bone loss in postmenopausal women. Inclusion criteria Women with a previous hysterectomy. If no previous oophorectomy: at least 45 years old and with ovarian failure, as evidenced by vasomotor symptoms for at least 1 to 5 years prior to enrollment. If previous oophorectomy: at least 40 years old, and 4 weeks to 5 years post	Details Women treated with transdermal oestradiol were compared to those treated with placebo. Methods Eligible women were randomly assigned to receive placebo or one of four doses of a 17β transdermal estradiol system. Participants and investigators were blinded to the treatment allocation. Treatment was continued for 26 four-week cycles (2 years). Sample size N = 175 n = 129 transdermal estradiol (four different doses combined) n = 46 placebo	Characteristics Mean age: 51.2 years Results Risk of any non-vertebral fracture in HRT group compared to placebo group: Relative risk (95% CI): 1.07 (0.11 to 10.03)	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Yes. There was adequate concealment of allocation. Yes. The groups were comparable at baseline. Yes. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Yes. Individuals administering care were kept 'blind' to treatment allocation. Yes. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for

Study details	Study design	Comparison	Results	Other
Berlex Laboratories. Country/ies where the study was carried out USA Study dates Not reported. Trial duration 2 years.	oophorectomy. Serum E2 level of ≤ 20 pg/mL, FSH of ≥ 50 U/L and fasting serum cholesterol of ≤ 300mg/dL, triglycerides of ≤ 300mg/dL and glucose of ≤ 140mg/dL. Baseline BMD of L2-L4 of ≥ 0.09g/cm² (Lunar) or ≥ 0.086g/cm² (Hologic). Exclusion criteria Known or suspected bone disease, hypo or hypercalcaemia, vitamin D deficiency, bone fracture within 6 months, immobilization for 2 or more of the preceding 6 months, hot flashes requiring hormone therapy or a history of skin irritation caused by transdermal drug-delivery systems. Women were also excluded if they had ever recived bisphosphonates, fluoride or calcitonin, were receiving chronic treatment with corticosteroids or agents that affect			differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported - only report total of 78 women withdrew from the study. The groups were comparable for treatment completion. Unclear. For how many participants in each group were outcome data not available? 78 women in total. The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Unclear. Investigators were kept 'blind' to participants' exposure to the intervention. Yes. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.

Study details	Study design	Comparison	Results	Other
	bone metabolism, had had recent oestrogen replacement therapy or treatment with lipid lowering drugs, or had participated in another clinical trial within 3 months.			
Full citation Wimalawansa,S.J., A four-year randomized controlled trial of hormone replacement and bisphosphonate, alone or in combination, in women with postmenopausal osteoporosis, American Journal of Medicine, 104, 219-226, 1998 Ref Id 233482 Study type Randomised controlled trial. Source of funding Not reported. Country/ies	Aim of the study To compare whether there is an additional benefit to BMD when HRT is combined with cyclical etidronate in patients with established osteoporosis. Inclusion criteria Postmenopausal Caucasian women with established osteoporosis (defined as at least 1, but not more than 4, radiographically demonstrable atraumatic thoracic vertebral crush fractures and spine BMD 2.0 SD below the reference range for normal healthy women aged 35 years). Exclusion criteria Surgical	Details Comparison was made in fracture risk between women allocated to HRT and those allocated to no treatment. Methods Patients were randomly allocated into one of two treatment groups: control group (no treatment) and HRT (premarin 0.625mg daily and norgestrel 150µg for 12 days each month). All participants were also given a daily supplement of calcium and vitamin D. Other women were recruited and allocated to different treatment groups (etidronate or HRT plus etidronate) but are excluded from analysis for the purposes of this review. Lateral radiographs of the thoracic and lumbar spine were obtained at the beginning of the study and after 4 years of treatment. Sample size N = 36 n = 18 HRT n = 18 no treatment	Characteristics HRT group: Age, years (mean ± SD): 64.0 ± 0.86 Time since menopause, years (mean ± SD): 15.2 ± 0.74 BMI, kg/m² (mean ± SD): 24.5 ± 0.78 BMD lumbar spine g/cm² (mean ± SD): 0.82 ± 0.01 No treatment group: Age, years (mean ± SD): 65.7 ± 0.83 Time since menopause, years (mean ± SD): 14.9 ± 0.68 BMI, kg/m² (mean ± SD): 25.4 ± 0.83 BMD lumbar spine g/cm² (mean ± SD): 0.82 ±0.02 Results Risk of non-vertebral fracture in HRT group compared to no treatment group: unadjusted relative risk (95% CI): 1.00 (0.07 to 14.79) Risk of vertebral fracture in HRT group compared to no treatment group: unadjusted relative risk (95% CI): 0.40 (0.09 to 1.80)	Other information Limitations Study quality Selection bias An appropriate method of randomisation was used to allocate participants to treatment groups. Yes. There was adequate concealment of allocation. Yes. The groups were comparable at baseline. Yes. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. Unclear - presumed not blinded. Individuals administering care were kept 'blind' to treatment allocation. Unclear - presumed not blinded. Attrition bias All groups were followed up for an equal length of time (or analysis was

Study details	Study design	Comparison	Results	Other
where the study was carried out UK Study dates Not reported. Trial duration 4 years.	menopause, secondary osteoporosis, other medical conditions that can affect the skeleton, taking medications that affect calcium metabolism within the previous 3 years. Patients treated with HRT, anabolic steroids, glucocorticoids, calcitonin, fluoride or bisphosphonates at any time since the menopause were also excluded.			adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? n = 4 no treatment group, n = 3 HRT group. The groups were comparable for treatment completion. Yes. For how many participants in each group were outcome data not available? None. The groups were comparable with respect to the availability of outcome data. Yes. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear - presumed not blinded. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.
Full citation Yates,J., Barrett-Connor,E., Barlas,S., Chen,Y.T.,	Aim of the study To assess the association between the cessation of postmenopausal	Details Duration of HRT and recency of treatment were assessed and compared to women who had never taken HRT. Methods Participants were asked to complete a follow up questionnaire approximately 12 months after the baseline evaluation. This	Characteristics Age, years (mean ± SD): 63.8 ± 8.97 BMI, g/cm² (mean ± SD): 27.7 ± 5.9 BMD T score (mean ± SD): -0.82 ± 1.13 Results Current/ever use compared to never use	Other information Limitations Study quality Selection bias The method of allocation to treatment groups was

Study details	Study design	Comparison	Results	Other
Miller,P.D., Siris,E.S., Rapid loss of hip fracture protection after estrogen cessation: evidence from the National Osteoporosis Risk Assessment, Obstetrics and Gynecology, 103, 440-446, 2004 Ref Id 233518 Study type Prospective cohort study. Source of funding Merck and Company, Inc. International Society of Clinical Densitometry. Country/ies where the study was carried out USA Study dates Recruitment commenced in 1997. Study duration 12 months.	oestrogen therapy and hip fracture risk. Inclusion criteria Postmenopausal women aged at least 50 years. Exclusion criteria Previous diagnosis of osteoporosis, bone mineral density testing within the past 12 months or use of osteoporosis specific medications.	included information on the occurrence and sites of new fractures. Participants reporting four or more fractures were excluded as multiple fractures were likely to have been the result of trauma. Telephone contact was used to confirm the reported occurrence of any hip fracture. Sample size N = 140,582 n = 86,845 ever users of HRT n = 53,737 never users of HRT	Risk of hip fracture in current users of HRT compared to never users: adjusted OR (95% CI): 0.60 (0.44 to 0.82) Risk of hip fracture in previous users (stopped ≤ 5 years ago) of HRT compared to never users: adjusted OR (95% CI): 1.65 (1.05 to 2.59) Risk of hip fracture in previous users of HRT (stopped > 5 years ago) compared to never users: adjusted OR (95% CI): 0.93 (0.63 to 1.38) Duration of current treatment Risk of hip fracture in current users of HRT (duration 0 to 5 years) compared to never users: adjusted OR (95% CI): 0.35 (0.18 to 0.67) Risk of hip fracture in current users of HRT (duration 6 to 10 years) compared to never users: adjusted OR (95% CI): 0.71 (0.41 to 1.23) Risk of hip fracture in current users of HRT (duration > 10 years) compared to never users: adjusted OR (95% CI): 0.66 (0.46 to 0.95) Duration of previous treatment Risk of hip fracture in previous users of HRT (duration 0 to 5 years) compared to never users: adjusted OR (95% CI): 1.00 (0.68 to 1.48) Risk of hip fracture in previous users of HRT (duration 6 to 10 years) compared to never users: adjusted OR (95% CI): 1.69 (0.91 to 3.12) Risk of hip fracture in previous users of HRT (duration > 10 years) compared to never users: adjusted OR (95% CI): 1.24 (0.67 to 2.30) Adjusted for age, BMI, previous fracture, health status, maternal history of fracture and cortisone use.	unrelated to potential confounding factors. Unclear. Attempts were made within the design or analysis to balance the comparison groups for potential confounders. Yes. The groups were comparable at baseline, including all major confounding and prognostic factors. No - significant differences in age, T-score, BMI, health status, prior fracture, maternal history of fracture and cortisone use. Performance bias The comparison groups received the same care apart from the intervention(s) studied. Yes. Participants receiving care were kept 'blind' to treatment allocation. No. Individuals administering care were kept 'blind' to treatment allocation. No. Attrition bias All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow up). Yes. How many participants did not complete treatment in each group? Not reported. The groups were comparable for treatment completion. Unclear.

Study details	Study design	Comparison	Results	Other
				For how many participants in each group were outcome data not available? Not reported. The groups were comparable with respect to the availability of outcome data. Unclear. Detection bias The study had an appropriate length of follow up. Yes. The study used a precise definition of outcome. Yes. A valid and reliable method was used to determine the outcome. Yes. Investigators were kept 'blind' to participants' exposure to the intervention. Unclear. Investigators were kept 'blind' to other important confounding and prognostic factors. Unclear.

H.8.7 Dementia

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Full citation Shao,H., Breitner,J.C., Whitmer,R.A., Wang,J., Hayden,K., Wengreen,H., Corcoran,C., Tschanz,J., Norton,M., Munger,R., Welsh-Bohmer,K., Zandi,P.P., Cache,County,I, Hormone therapy and Alzheimer disease dementia: new findings from the Cache County Study, Neurology, 79, 1846-1852, 2012	Sample size n=5677 Characteristics Age at baseline (mean y, SD): HRT group=73.4 (SD5.6) No HRT group=76.7 (SD6.9) Years of education (mean y, SD): HRT group=13.1 (SD 2.2) No HRT group =12.7	Interventions Any HRT No HRT use	Details Eligible participants from Cache county, Utah participated at baseline assessement and screened for dementia (APOE genotyping and completion of detailed questionnaire on potential risk factors and protective factors for dementia). Participants at baseline without dementia were followed up again at year	Results Cox proportional hazard models of association with incident Ad by timing, duration, and type of HT (Hr, 95%CI) Model 1 Adjusted for baseline age, APOE status, years of education No HT =1.0 Any HT =0.78(0.57,1.06) Adjusted for baseline age, APOE status, years of education, and decile propensity score No HT=1.0 Any HT=0.80 (1.58,1.09)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Ref Id 300732 Country/ies where the study was carried out USA Study type Cohort study Aim of the study To examine whether the association of HT with AD varies with timing or type of HT use Study dates 1995- 2006 Source of funding National institutes of health	(SD 2.2) Age at menopause (mean y, SD) HRT group=47.3 (SD 6.8) No HRT group=48.2 (SD 6.3) No. of years form menopause to baseline (mean y, SD) HRT group=26.0 (SD 8.8) No HRT group=28.4 (SD 9.5) Hypertension (Yes or no) HRT group=492 yes, 611 no No HRT group=307 yes, 353 no Stroke (yes or no) HRT group=69 yes, 1032 no No HRT group=39 yes, 623 no Family history of AD (yes or no) HRT group=271 yes, 704 no No HRT group=150 yes, 414 no History of smoking (yes or no) HRT group=226 yes, 876 no No HRT group=135 yes, 527 no Inclusion criteria Women from the Cache county study who provided a detailed history on age at menopause and use of HRT.		3, 6, and 9. All participants consented and next of kin consented for participants who were unable to provide it. Dementia was evaluated at baseline and follow-up by using the modified mini-mental state examination (3MS) or the Informant questionnaire for cognitive decline in the elderly. Participants showing cognitive decline were given a clinical assessment, physical examination and a one hour battery of neuropsychological tests. Covariate assessments were evaluated by the Women's health questionnaire via telephone between baseline and year 3 of follow-up. Women who completed the questionnaire were included in the analysis. Statistical analysis: X2 Tests were used to compare characteristics of HRT users and non HRT users. Cox proportionaly hazard models were generated to evaluate association between HRT and incident AD. Participants were followed from their age at the entry of the study to the time of AD onset or last assessment. Participants without AD were censored at onset of dementia.	Model 2 Adjusted for baseline age, APOE status, years of education No HT=1.0 HT (any type) initiated within 5 years of menopause=0.69(0.49, 0.98) HT initiated >5 years after menopause=0.70(0.49,0.99) Adjusted for baseline age, APOE status, years of education, and decile propensity score No HT=1.0 HT (any type) initiated within 5 years of menopause=0.96(0.64,1.34) HT initiated >5 years after menopause=1.03(0.68,1.55)	outcome(s) under study)- No A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Yes Level of risk-Low B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/A B.2 Participants receiving care were kept 'blind' to treatment allocation-N/A B.3 Individuals administering care were kept 'blind' to treatment allocation-N/A Level of risk: Low C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A (less than 10%) C.2b The groups were comparable for treatment completion (that is, there were

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Women using any form of HRT. Exclusion criteria Not reported		Hazard ratios and 95% confidence intervals were estimated from unadjusted models and from 2 sets of adjusted models.		no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes (7-year follow-up) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/A D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/A Level of bias: Low Indirectness Does the study match the review protocol in terms of; Population: Unclear (the participants were not representative of the general population) Outcome: Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Indirectness: Some
Full citation Petitti,D.B., Crooks,V.C., Chiu,V., Buckwalter,J.G., Chui,H.C., Incidence of dementia in long-term hormone users, American Journal of Epidemiology, 167, 692-700, 2008 Ref Id 300771 Country/ies where the study was carried out USA Study type Cohort study Aim of the study To investigate the incidence of dementia in long-term hormone users Study dates 1998 Source of funding National institute of ageing	Sample size N=2906 Characteristics At baseline: Age (number of women) 75-79 years=1999 80-84 years=732 ≥85 years=175 Education (number of women) Less than high school=331 High school graduation=781 Some college/trade school=1098 College degree or more=691 Refused/didn't know=5 Race/ethnicity (number of women) Non-hispanic/white=2583 Hispanic=97 African-American=122 Asian/Pacific Islander=43 Other/unknown=61 Stroke (number of women, yes or no) Yes=133 No=2763 Myocardial infarction (number of women, yes or no) Yes=247 No=2646 Hypertension (number of women, yes or no) Yes=1518 No=1370 Diabetes (number of	Interventions Oestrogen use (hormone therapy users) No oestrogen use (non users)	Details 3681 women were eligible for the study and were assessed by interview (Telephone Interview of Cognitive Status-modified) at baseline in 1999. 636 women were not contactable and were excluded from the study. Women who were classifed as having dementia at baseline were also excluded from the study (140 women). 2906 women were dementia-free and were included in the analysis. Annual telephone interviews were attempted for the 2906 women until they died or were classified as having dementia, or until follow-up. Proxy interviews for women who could not be interviewed by telephone were attempted and were asked to identify people they saw at least once a month who knew them well. Woman-years of follow-up were calculated from the date of the baseline interview to the date of teh interview that resulted in dementia classification. Classification of cognitive status was assessed at each annual follow-up by a neurologist and	Results Adjusted hazard ratios for dementia in oestrogenor oestrogen+progestin users, and incidence of dementia (1999-2003) Adjusted for age and education (95%CI) No hormone use by prescription or self report (n=879; incidence of dementia=24.8/1000)=1.00 (referent) Oestrogen use by both prescription and self report (n=1011; incidence of dementia=26.0/1000)=1.01 (0.76,1.36) Oestrogen/progestin use by both prescription and self-report (n=410; incidence of dementia=31.4/1000)=1.32 (0.92, 1.89) Oestrogen or oestrogen/progestin use by prescription but neither by self-report (n=98; incidence of dementia=44.1/1000)=1.64 (0.94,2.87) Oestrogen or oestrogen/progestin use by self-report but neither by prescription (n=493; incidence of dementia=20.8/1000)=0.81 (0.55,1.19) Adjusted for age, education, and medical risk factors (95%CI) No hormone use by prescription or self report (n=879)=1.00 (referent) Oestrogen use by both prescription and self report (n=1011)=1.07 (0.79, 1.44) Oestrogen/progestin use by both prescription and self-report (n=410)=1.32 (0.91, 1.91) Oestrogen or oestrogen/progestin use by prescription but neither by self-report (n=98)=1.64(0.94,2.88) Oestrogen or oestrogen/progestin use by self-report but neither by prescription (n=493)=0.80 (0.54,1.19) Adjusted hazard ratios for dementia according to self-reported hormone use, by timing of the start of hormone use in relation to menopause (1999-2003) Adjusted for age and education (95%CI)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)- Yes A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Yes Level of risk-Low B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/A B.2 Participants receiving care were kept 'blind' to treatment allocation-N/A B.3 Individuals administering care were kept 'blind' to treatment allocation-N/A Level of risk: Unclear C. Attrition bias (systematic

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	women, yes or no) Yes=214 No=2690 Parkinson's disease (number of women, yes or no) Yes=20 No=2885 Horomone use by prescription (number of women) Not a hormone user=1387 Prescription oestrogen user=1072 Prescription oestrogen/progestin user=447 Inclusion criteria Women aged ≥75 years in 1998 who had been continuously enrolled in the health plan from 1992 to 1998. Hormone therapy users were defined as women who had filled at least one prescription for oral oestrogen at a health plan pharmacy in every calendar year from 1992 to 1998. Non users were defined as women without any oestrogen prescriptions from 1992 to 1998. Exclusion criteria Women who had intermittent prescriptions from 1992 to 1998		neuropsychological testing. The dementia outcome was classified as 1) no cognitive impairment, or minimal impairment; 2) Cognitive impairment without definitive dementia 3) dementia with the gold standard. Women with dementia were censored in the analysis. Sensitivity in comparing dementia with no dementia using the gold standard was 0.83 and specificity was 1.0. Statistical analyses were generated for demographic and self-reported medical condition variables (Stroke, myocardial infarction, diabetes, hypertension, and Parkinson's disease). Chi squared tests were done for statistical significance in the analysis of no response. Kaplan-Meier was used to estimate probability of dementia-free survival by hormone therapy use. The log rank test was used to assess the statistical significance of differences in dementia-free survival. Cox proportional hazards model was used to estimate crude and age-adjusted hazard ratios, and hazard ratios were adjusted for other confounders. The	Never use of hormones (baseline, n=977)=1.00 (referent) Hormone use (within 10 years of menopause) Current hormone user (baseline, n=957)=0.93 (0.70,1.24) Former hormone user (baseline, n=346)=0.89 (0.59,1.34) Hormone (after 10 years of menopause) Current hormone user (baseline, n=313)=0.85 (0.56,1.30) Former hormone user (baseline, n=48)=0.21(0.03,1.50) Adjusted for age, education, and medical risk factors Never use of hormones (baseline, n=977)=1.00 (referent) Hormone use (within 10 years of menopause) Current hormone user (baseline, n=957)=0.95 (0.71,1.28) Former hormone user (baseline, n=346)=0.84 (0.55,1.28) Hormone (after 10 years of menopause) Current hormone user (baseline, n=313)=0.90 (0.59,1.38) Former hormone user (baseline, n=48)=0.22 (0.03,1.55)	differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A (about less than 10% of the cohort did not have ERT use data in this study) C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes (5-year follow-up) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
			regression models included self-reported variables found to be strongly related to dementia in the literature (age and education) and other available variables that were associated in the data set. The variables in the final, fully adjusted model were forced. Exact 95% confidence intervals were calculated for all hazard ratio estimates. A p value of less than 0.05 was considered statistically significant. The main analyses inlcuded information on hormone therapy use as determined by prescription. Non-users were the reference group. Analyses were carried out taking both prescription information and self-reported information on hormone therapy use at baseline. Age at menopause was defined as the self-reported age at which menstrual periods stopped and association of initiation of hormone use near menopause with risk of dementia was assessed.		D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/A D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/A Level of bias: Low Indirectness Does the study match the review protocol in terms of; Population: No (the participants were not representative of the general population) Outcome: Yes Indirectness: Some
Full citation Ryan,J., Carriere,I., Scali,J., Ritchie,K., Ancelin,M.L., Life-time estrogen exposure and cognitive functioning in later life, Psychoneuroendocrinology,	Sample size n=996 Characteristics Age (mean years, SD)=72.8 (SD 5.5) Age at menopause (mean years,	Interventions HRT (past or current) No HRT	Details The ESPRIT study recruited participants over a 2 year period from 1999 to 2001 by random selection. At baseline participants	Results Association between lifetime outcomes and decline in cognitive performance in 4 year follow-up period (adjusted for age, educational level and baseline cognitive test score) Global function (MMSE<-2) (OR,95%CI)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
34, 287-298, 2009 Ref Id 300838 Country/ies where the study was carried out France Study type Cohort study (ESPRIT study) Aim of the study To examine whether factors related to oestrogen exposure across the life-time were associated with cognitive function in postmenopausal women Study dates Participants recruited from 1999 to 2001 Source of funding Regional government of Languedoc-Roussillon Agence nationale de la recherche Novartis France Alzheimer grant	SD)=49.5 (SD 5.4) ≥12 years of education (%)=28.6 Hormone treatment (%): Never=65.8 Past=19.4 Current=14.8 Duration of hormone treatment (%): Never=65.8 0-9 years of past use=11.8 ≥10 years of past use=7.6 0-9 years of current use=3.7 ≥10 years of current use=11.0 Surgical menopause (%)=18.7 Current smoker (more than 10 packets per year) (%)=3.7 Carrier of APOE4 allele (%)=17.8 Inclusion criteria Women aged 65 years and older Non-institutionalised Exclusion criteria Diagnosed with possible or probable dementia If they were deceased Lost to follow-up 4 year period Incomplete data relating to cognitive tests administered at baseline or follow-up Missing at least some data concerning covariates included in the multivariate analysis		were administered a number of standard questionnaires by trained staff and also underwent clinical examinations. Cognitive assessment was administered by trained staff at baseline and at each year of follow-up. Tests included verbal memory, the Benton's visual retention test, Trail making tests A and B, and the mini mental state examination for global measure of cognitive function. At baseline and each follow-up all participants were assessed by a neurologist and a standard clinical protocol was used to identify cases of dementia using the DSM-IV criteria. All inicdent cases were further validated by a group of neurologiccal experts and when dementia was diagnosed, the date of onset was recorded as the date of the follow-up assessment. Reproductive characteristics were assessed by administering a questionnaire specific for reproductive lifetime events and hormonal exposure was administered as part of a general clinical examination. Duration of hormone treatment and oral contraceptives was also assessed.	Never HT user: 1 Past HT user: 0.93 (0.61, 1.43) Verbal fluency (Isaacs ≤6) (OR, 95%CI) Never HT user: 1 Past HT user:0.96 (0.62,1.50) Visual memory (Benton ≤ -2) (OR, 95%CI) Never HT user:1 Past HT user:0.81 (0.52,1.27) Verbal memory (Word recall ≤ -2) (OR, 95%CI) Never HT user:1 Past HT user:0.92 (0.57,1.50) Psychoomotor speed (Trail making A ≥15) (OR,95%CI) Never HT User:1 Past HT user:0.82 (0.52,1.29) Executive function (Trail making B ≥35) (OR, 95%CI) Never HT user:1 Past HT user:0.74 (0.47,1.19) Duration of HT (OR, 95%CI) Never HT user:1 0-9 years of past use: 0.70 (0.40-1.22) ≥ 10 years of past use: 1.37 (0.77,2.45) 0-9 years of current use: 0.75 (0.28, 2.02) ≥ 10 years of current use: 1.20 (0.70, 2.06)	comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)- N/A A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Yes Level of risk-Low B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/A B.2 Participants receiving care were kept 'blind' to treatment allocation-N/A B.3 Individuals administering care were kept 'blind' to treatment allocation-N/A Level of risk: Low C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
			Potential covariates that may influence cognitive performance and potentially linked to use of HRT or other reproductive markers included activities of daily living, depressive symptoms (depression scale), regular smoking, alcohol consumption, BMI, vascular diseases, chronic illnesses, anticholinergic medication, diagnosis of cancer within the last two years, and carriers of the APOE4 allele. Statistical analyses included Chi-squared tests to determine bivariate associations between baseline characteristics and cognitive function. Horomonal characteristics associated with cognitive performance at 20% significance were considered simultaneously in logistic models adjusted for age, education level, marital status, depressive symptoms, high caffeine intake, physical incapacities and comorbidity. The final multivariate models contained the hormonal variables that remained significantly associated with cognitive function after inclusion of all of the potential confounders. Multivariate logistic		of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A (less than 10%) C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes (4-year follow-up) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/A D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/A Level of bias: Low

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
			analysis was used to determine whether baseline hormone-related factors were associated with the risk of cognitive decline over the 4 year follow-up, while adjusting for the potential confounders and their baseline cognitive scores. Cox proportional hazards models with delayed entry were developed to determine which reproductive factors were associated with the incidence of dementia during the follow-up period. All statistical significance was <0.05.		Indirectness Does the study match the review protocol in terms of; Population: Yes Outcome: Yes Indirectness:none Reporting bias: The authors do not report the participant numbers for outcomes. For duration no information on participants was reported. Other information Retrospective study Bias due to exclusion of some participants. Participants with extreme cognitive problems were excluded from the analyses and may reduce power to detect significant associations if they were present. Differential recall of hormone use by participants.
Full citation Henderson,V.W., Benke,K.S., Green,R.C., Cupples,L.A., Farrer,L.A., MIRAGE Study Group., Postmenopausal hormone therapy and Alzheimer's disease risk: interaction with age, Journal of Neurology, Neurosurgery and Psychiatry, 76, 103-105, 2005 Ref Id 301077 Country/ies where the study was carried out USA Study type Case control study Aim of the study To evaluate the relation between HT and AD in postmenopausal women	Sample size N=1694 Characteristics Age (years (SD)) AD= 71.1 (8.1) controls=65 (8.6) Oestrogen use >6 months (%) AD= 87/426 (21%) Control=192/545 (35%) History of hysterectomy or oophorectomy (%) AD=141/426 (35%) Controls=231/545 (42%) Inclusion criteria MIRAGE participants who were	Interventions HT No HT	Details MIRAGE probands were included to meet criteria for probable or definite AD. Controls were first degree relatives or spouses of probands. Consent from controls were provided, participants who were not able to provide consent gave proxy informed consent. Risk factor data were collected from AD participants or from secondary informants, or medical records where possible. Controls without dementia provided their own risk factor information.	Results Age stratified risk of AD associated with prior use of hormone therapy (Odds ratio, 95%CI) Age 50-63 years No HT+AD=58 HT+AD=17 No HT+control=135 HT+control=112 Adjusted OR (95% CI)=0.35 (0.19, 0.66) HT vs No HT Age 64-71 years No HT+AD=105 HT+AD=28 No HT+control=127 HT+control=52 Adjusted OR (95% CI)=0.86 (0.50, 1.5) HT vs No HT	Limitations Section 1: Internal validity 1.1 The study addresses an appropriate and clearly focused question-yes Selection 1.2 The cases and controls are taken from comparable populations-no. The control group was not representative of the population, they were spouses or first degree relatives 1.3 The same exclusion criteria are used for both cases and controls-Unclear 1.4 What was the participation rate for each group (cases and controls)? 532/1694 cases, 819/1694 controls (obtained from abstract of cited paper) 1.5 Participants and non-

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
aged 65 years and older Study dates Not reported Source of funding National institutes of health Merit award from the veterans administration	postmenopausal, or if unsure of menopausal status, were at least 60 years of age. Used oestrogen replacement therapy or oestrogen medication for birth control, menopausal symptoms, osteoporosis on a daily basis for 6months Initiated HT at least one year prior to dementia onset/censored age or failed to specify a start date for HT MIRAGE probands had probable or definite AD Controls were first degree relatives or spouses of probands Exclusion criteria Birth control medication when used before age 36 Women who reported birth control use after age 35 but could not specify type of oestrogen		Potential interactions between oestrogen and APOE4 genotype was evaluate, and oestrogen use, age, education, ethnicity and APOE4 allele were used to limit the number of participants in the analysis. Other confounding factors including alcohol use, cigarette smoking, daily use of NSAIDs for more than 6 months, prior hysterectomy or oophorectomy were adjusted for effects of HRT use and risk of AD. Statistical analysis: Comparisons of patients compared with controls were made using the Wilcoxon rank sum tests for continuous measures and Chi squared tests for dichotomous measures. Odds ratios were calculated (crude and adjusted) to evaluate potential confounders. Multivariate analyses were also generated for correlations among subjects within families. Odds ratios were adjusted for age, education, and ethnicity.		participants are compared to establish their similarities or differences-yes 1.6 Cases are clearly defined and differentiated from controls- yes 1.7 It is clearly established that controls are not cases-yes Risk of bias: high Assessment 1.8 Measures were taken to prevent knowledge of primary exposure from influencing case ascertainment-unclear 1.9 Exposure status is measured in a standard, valid and reliable way-yes Risk of bias: high Confounding 1.10 The main potential confounders are identified and taken into account in the design and analysis-yes (for age, education, ethnicity) Risk of bias: low Statistical analysis 1.11 Have confidence intervals been provided? Yes Risk of bias: Low Section 2: Description of study 2.1 How many people participated in the study:1694 2.2 What are the main characteristics of the study population? Age 65 and above, education (12 years or more), ethnicity (African American), Oestrogen use for more than 6 months, history of hysterectomy or oophorectomy 2.3 What environmental or prognostic factor is being investigated? AD 2.4 What comparisons are made? No HRT vs HRT in AD

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					or no AD cases 2.5 For how long are participants followed up? Unclear 2.6 What outcome measure(s) is/are used? Risk of AD as odds ratio 2.7 What size of effect is identified? Adjusted OR at 50-56 years=0.35 (0.19, 0.66) 2.8 How was the study funded? National institutes of health 2.9 Does this study help to answer your guideline review question? No, there is bias due to control group selection Risk of bias:high Indirectness Population: Yes Outcome:Yes Indirectness: Some, control group is not truly representative of the population Other information study design leads to selection bias no information on progestin use, unable to distinguish effects of opposed oestrogen from oestrogen+progestin HT exposure was not validated against pharmacy or prescription records Use of proxy informant for cases but not for controls could have led to misclassification sons and brothers were less reliable in reporting HT use 48 cases with brother or son informants were excluded and could have modified the oestrogen effect on AD risk by age

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					In analyses adjusting for age, education, and race, HT was associated with a 30% reduction in AD risk In analyses stratified by age, HT was significantly associated with reduced risk in the 50-63 years age stratum
Full citation Whitmer,R.A., Quesenberry,Jr, Zhou,J., Yaffe,K., Timing of hormone therapy and dementia: The critical window theory revisited, Annals of Neurology, 69, 163-169, 2011 Ref Id 301458 Country/ies where the study was carried out USA Study type Cohort study Aim of the study To compare HT use in mid-life with that in late life on risk of dementia in postmenopausal women Study dates 1994-1998 Source of funding National institutes of health	Sample size n=5504 Characteristics Age at midlife survey (y, mean, SD): No HRT group=49.0 (SD 4.2) Mid-life HRT group=49.0 (SD 3.9) Late HRT group=47.3 (SD 4.5) Race/ethnicity (number, %): Asian= No HRT:90 (3.7); Mid-life: 26 (1.9); Late-life: 27 (4.0) Black=No HRT:587 (23.9); Mid-life:283 (20.5); Late-life: 94 (14.0) White=No HRT: 1659 (67.6); Mid-life:1033 (74.6); late-life:518 (77.0) Other=No HRT:117 (4.8); Mid-life:42 (3.0); Late-life:34 (5.1) Education (number, %): Trade school or college No HRT=556 (32.4) Mid-life=323 (32.99) Late-life=198 (39.13) High school No HRT=804 (32.8)	Interventions Both HT in mid-life HT in late -life No HT	Details The analytical sample included women who self-reported as being postmenopausal at the time of the MHC exam, who were alive and health plan members in 1994 and without a diagnosis of dementia prior to 1999. Midlife data collection: Data was collected through interviews for information on demographics, lifestyle, and medical history (menopausal status, medical conditions, medication use). Women were considered to be taking mid-life HRT if they aswered 'yes' to taking hormones and did not have a self report of endocrine diseases. Latelife hormone therapy: KPNC pharmacy databases were searched for HRT prescriptions. Thoses with two or more prescriptions or refills of HRT during 4 years were considered as late-life HRT users. Each prescription was a 100 day prescription, thus two or more prescriptions	Results Frequency of dementia cases by hormone therapy status stratified by median age in 1999 Age <80.4 years No dementia No HT=914 (78.3) Mid-life HT=458(79.1) Late-lfe HT=33(76.9) Both=427(78.8) Dementia No HT=253(21.6) Mid-life HT=121(20.9) Late-lfe HT=99(23.1) Both=115(21.2) Age ≥80.4 years No dementia No HT=841(65.3) Mid-life HT=550(68.3) Late-lfe HT=155(63.5) Both=305(67.6) Dementia No HT=446(34.6) Mid-life HT=255(31.6) Late-lfe HT=89(36.5) Both=146(32.4) Cox proportional hazard models of hormone use and risk of dementia Timing of hormone use Unadjusted (for age as the timescale) No HT=10. Mid-life HT=0.86(0.72,1.03) Late-lfe HT=1.30(1.04,1.63) Both=1.00(0.82, 1.22)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)- No A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Yes Level of risk-moderate B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-Unclear B.2 Participants receiving care

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Mid-life=523 (37.8) Late-life=208 (30.9) Grade school No HRT=432 (17.6) Mid-life=246 (17.8) Late-life=82 (12.2) Diabetes (number, %) No HRT=490 (12.0) Mid-life=261 (18.9) Late-life=115 (17.1) Hypertension (number, %) No HRT=1809 (73.7) Mid-life=1005 (72.6) Late-life=529 (78.6) Hyperlipidaemia (number, %) No HRT=880 (35.9) Mid-life=502 (36.3) Late-life=296 (44.0) Stroke (number, %) No HRT=556 (22.7) Mid-life=324 (23.4) Late-life=187 (27.8) Hysterectomy		was considered as equal to 6 months of HRT use. Dementia diagnosis: Dementia was ascertained through medical records from a database containing diagnoses from all outpatient and inpatient cases at KP medical centres and clinics. Participants were considered to have dementia of they had any of the ICD code diagnoses. Diagnoses were ascertained when the participants were aged 75 and 84 years at the start of the study, and between 84 years and 93 years of age at the completion of the study. Late-life comorbidities and mortality Stroke was recorded from hospital discharge diagnoses (ICD 9 codes) from 1971 to end of study (2008). Late life diabetes was ascertained from the diabetes registry. Hypertension and hyperlipidaemia were recorded from outpatient databases from 1994 to 2008. Mortality was recorded through the end of 2007. Statistical analysis Preliminary Chi squared tests and t tests were performed to determine if	Adjusted for education, race, BMI, number of children No HT=1.0 Mid-life HT=0.75(0.59,0.95) Late-lfe HT=1.54(1.15,2.06) Both=1.13(0.86, 1.47) Additionally adjusted for diabetes, hypertension, hyperlipidaemia, stroke No HT=1.0 Mid-life HT=0.74(0.58,0.94) Late-lfe HT=1.48(1.10,1.98) Both=1.02(0.78,1.34)	were kept 'blind' to treatment allocation-Unclear B.3 Individuals administering care were kept 'blind' to treatment allocation-Unclear Level of risk: High C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-Unclear C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-Unclear C.3a For how many participants in each group were no outcome data available?-Unclear C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-Unclear Level of risk: high D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes (9-year follow-up)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	(number, %) No HRT=81 (3.3) Mid-life=76 (5.49) Late-life=52 (7.73) Inclusion criteria Women who self-reported as being postmenopausal at the time of the multiphasic health checkup (MHC), who were alive and health plan members in 1994. For mid-life data collection, women were on mid-life HT. For late-life data, all HT oral or patch were included, and those with two or more prescriptions equated to approximately 6 months of HT use. Exclusion criteria Thyroid hormone or endocrine diseases Those with diagnoses of dementia, cognitive impairment or general memory complaints prior to commencement of dementia ascertainment in 1999		demographic and clinical characteristics were significantly different by HRT group. The frequency of dementia cases stratified by median age in 1999 was examined in women over 80 years age as dementia cases occurred mostly in this group. Kaplan Meier survival curves (unadjusted for age) of dementiarisk were conducted to examine the likelihood of dementia over age and time in different HRT groups. Cox proportional hazards models with age (mid-life or late-life) as time scale was investigated for HRT use and risk of dementia. Models were adjusted for age, education, ethnicity, mid-life BMI, diabetes, hypertension, hyperlipidaemia, stroke and hysterectomy status. A sensitivity analysis was performed of HRT and dementia risk stratified by stroke status.		D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-No D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-No Level of bias: Moderate Indirectness Does the study match the review protocol in terms of; Population: yes Outcome: Yes Indirectness: none Other information Retrospective study Bias due to exclusion of some participants. Participants with extreme cognitive problems were excluded from the analyses and may reduce power to detect significant associations if they were present. Differential recall of hormone use by participants.
Full citation Baldereschi,M., Di,Carlo A., Lepore,V., Bracco,L., Maggi,S., Grigoletto,F., Scarlato,G., Amaducci,L., Estrogen-replacement therapy and Alzheimer's disease in the Italian Longitudinal Study on	Sample size n=2816 enrolled n=2046 assessed for oestrogen replacement therapy Characteristics Age (y, mean, SD): Never users=74.7 (SD 5.8)	Interventions Oestrogen replacement therapy (ever use) No oestrogen replacement therapy (never use)	Details Participant and covariate information The Italian longitudinal study on ageing (ILSA) participants completed the mini mental state examination at baseline for diagnosis of dementia	Results Risk of AD in oestrogen ever users and oestrogen never users: Cases of AD+never use=89/1382, OR=1.00 Cases of AD+ever use=3/186 Cases of non-AD+never use=1293/1382 Cases of non-AD+ever use=183/186 OR=0.24 (95%CI 0.07 to 0.77)	Limitations Section 1: Internal validity 1.1 The study addresses an appropriate and clearly focused question-yes Selection 1.2 The cases and controls are taken from comparable populations-yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Aging, Neurology, 50, 996-1002, 1998 Ref Id 313561 Country/ies where the study was carried out Italy Study type Case control study Aim of the study To study the association of oestrogen replacement therapy andother oestrogen-related variables with AD in postmenopausal women. Study dates 1992-1993 Source of funding Italian national research council	Ever users:73.2 (SD 5.4) Education (y, mean, SD): Never users=5.1(SD 3.8) Ever users=6.1 (SD 4.4) Hypertension (%): Never users=68.3 Ever users=70.6 Diabetes (%): Never users=14.5 Ever users=10.2 Body weight at age 50 years (kg, mean, SD): Never users=63.3 (SD 11.7) Ever users=62.8 (SD 11.4) Age at menarche (y, mean, SD): Never users=13.2 (SD 1.8) Ever users=13.2 (SD 1.7) Age at menopause (y, mean, SD): Never users=48.4 (SD 5.4) Ever users=47.9 (SD 5.7) Ever smokers (%): Never users=16.4 Ever users=21.1 Ever drinkers (%): Never users=67.1 Ever users=74.6		(cutoff score 23/24). A history of oestrogen use was obtained by interviewing the participant or by proxy if the participant was not able to provide the information. For women who took oestrogen therapy, their age at menopause, age at initiation of treatment and age when treatment was stopped was ascertained. During home interviews, boxes of pills were examined to ascertain current use of HRT. Confounding factors were also recorded and included education, smoking and alcohol habits, other medical conditions such as diabetes and hypertension. Statistical analyses Chi squared tests were carried out for age-specific comparisons. Student's t test and Chi squared tests were used for demographic and medical comparisons (continuous and dichotomous variables respectively). AD was measured by the odds ratio with 95% confidence intervals. Multivariate regression was used to estimate the risk of AD as a function of all oestrogen-related variables in the study.		1.3 The same exclusion criteria are used for both cases and controls-Not reported 1.4 What was the participation rate for each group (cases and controls)? AD group=92; controls=1476 1.5 Participants and non-participants are compared to establish their similarities or differences-yes 1.6 Cases are clearly defined and differentiated from controls- yes 1.7 It is clearly established that controls are not cases-yes Risk of bias:low Assessment 1.8 Measures were taken to prevent knowledge of primary exposure from influencing case ascertainment-Not reported 1.9 Exposure status is measured in a standard, valid and reliable way-yes Risk of bias: low Confounding 1.10 The main potential confounders are identified and taken into account in the design and analysis-yes, but which variables accounted for in analysis not reported Risk of bias: high Statistical analysis 1.11 Have confidence intervals been provided? Yes Risk of bias: Low Section 2: Description of study 2.1 How many people participated in the study:2816 2.2 What are the main characteristics of the study population? Age 65-84 years,

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Inclusion criteria Population was from ILSA cohort study Women aged 65 to 84 years Women screened positive for AD Exclusion criteria Not reported				education (5 years or more), age at menopause 47 years and above 2.3 What environmental or prognostic factor is being investigated? AD 2.4 What comparisons are made? No HRT vs HRT in AD or no AD cases 2.5 For how long are participants followed up? Not reported 2.6 What outcome measure(s) is/are used? Risk of AD as odds ratio 2.7 What size of effect is identified? OR=0.24 (0..07 to 0.77) 2.8 How was the study funded? Italian national research council 2.9 Does this study help to answer your guideline review question? Yes, but only for overall risk of AD with HRT use Risk of bias:low Indirectness Population: Yes Outcome:Yes Indirectness: None
Full citation Kang,J.H., Weuve,J., Grodstein,F., Postmenopausal hormone therapy and risk of cognitive decline in community-dwelling aging women, Neurology, 63, 101-107, 2004 Ref Id 314410 Country/ies where the study was carried out USA Study type	Sample size n=15, 646 women Non users n=4258 Past users n=4611 Current oestrogen+progestin users n=1358 Current oestrogen users only n=3580 Current oestrogen users only (recent initiators, hormone use 5 years prior to baseline cogntive	Interventions Oestrogen alone Oestrogen+progestin no hormone therapy	Details The NHS included 121, 700 female registered nurses. Participants completed mailed questionnaires twice a year to update information on lifestyle and medical history (>90% follow-up maintained). For cognitive function, participants aged 70 years and older were selected who were free of	Results Substantial decline in cognitive performance over 2 years in relation to postmenopausal hormone use and duration TICS Total decline, n (at least 2 SD of the baseline score) ≥5 points; multivariate adjusted RR (95%CI): Never users=4258 (202); adjusted RR (95%CI)=1.0 Past hormone user=4611 (249); adjusted	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)- No

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Cohort study Aim of the study To investigate the relation of postmenopausal hormone therapy to cognitive decline Study dates Study start:1976 1995-2001: eligible women contacted for baseline telephone cognitive assessment 2003: second cognitive assessment Source of funding National institutes of health Ellison medical foundation	testing) n=282 Characteristics Age (y, mean, SD): Non users=74.0 (SD 2.2) Past users=74.4 (SD 2.3) Current users of oestrogen and progestin=73.9 (SD 2.2) Current users of oestrogen only=74.0 (SD 2.2) Current uses of oestrogen only-recent initiators=73.8 (SD 2.2) Education (masters/doctorate degree, %): Non users=6 Past users=6 Current users of oestrogen and progestin=7 Current users of oestrogen only=6 Current uses of oestrogen only-recent initiators=6 Hypertension (%): Non users=54 Past users=55 Current users of oestrogen and progestin=49 Current users of oestrogen only=56 Current uses of oestrogen only-recent initiators=53 Diabetes (%): Non users=10 Past users=9		diagnosed stroke. Baseline cognitive assessments were carried out, and the study analysis included assessments with complete information on two assessments. Only women with natural menopause or bilateral oophorectomy were included for analysis of hormone therapy at menopause and hormone initiation at older ages as age at menopause was difficult to determine in other groups. Informed consent was obtained from all participants. Cognitive function assessment: At baseline, the telephone interview for cognitive status (TICs) was used. Five other tests were asded to the battery and participant rates were similar across the tests. The tests included immediate and delayed recall of the East Boston memory test, Category fluency, delayed recall of TICs, digit span backwards, and verbal memory. The results of these scores was combined to produce a composite score of verbal memory by normalising results of each test using z scores and average of the four z scores. For validity and reliability of telophone	RR (95%CI)=1.07(0.87,1.30) Current use, oestrogen only=3580 (181); adjusted RR (95%CI)= 1.06 (0.85, 1.32) Current use, oestrogen+20 years=1134 (55); adjusted RR (95%CI)= 0.95 (0.69, 1.32) Current use, oestrogen+progestin=1358 (82);adjusted RR (95%CI)= 1.27(0.97, 1.68) Current use, oestrogen+progestin 10+ years=732 (48);adjusted RR (95%CI)=1.36(0.97, 1.92) Verbal memory Total decline, n (at least 2 SD of the baseline score) ≥1.38 points; multivariate adjusted RR (95%CI): Never users=3696 (75); adjusted RR (95%CI)=1.0 Past hormone user=3967 (93); adjusted RR (95%CI)=1.0(0.79,1.51) Current use, oestrogen only=3106 (69); adjusted RR (95%CI)= 1.10 (0.76, 1.57) Current use, oestrogen+20 years=956 (26); adjusted RR (95%CI)=1.25(0.76, 2.06) Current use, oestrogen+progestin=1191(34);adjusted RR (95%CI)= 1.41(0.91, 1.68) Current use, oestrogen+progestin 10+ years=732 (48);adjusted RR (95%CI)=1.72 (1.03,2.88) Category fluency Total decline, n (at least 2 SD of the baseline score) ≥9 points; multivariate adjusted RR (95%CI): Never users=4060 (114); adjusted RR (95%CI)=1.0 Past hormone user=4405 (146); adjusted RR (95%CI)=1.20 (0.91,1.518) Current use, oestrogen only=3448 (111); adjusted RR (95%CI)= 1.18 (0.88, 1.59) Current use, oestrogen+20 years=1087(36); adjusted RR	A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes (but only age and education, age at menopause or hormone use were adujsted for in analyses) A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Yes Level of risk-High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/A B.2 Participants receiving care were kept 'blind' to treatment allocation-N/A B.3 Individuals administering care were kept 'blind' to treatment allocation-N/A Level of risk: Unclear C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A (about less than 10% of the cohort did not

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Current users of oestrogen and progestin=5 Current users of oestrogen only=7 Current uses of oestrogen only-recent initiators=10 Age at menopause (y, mean, SD): Non users=50 Past users=48 Current users of oestrogen and progestin=50 Current users of oestrogen only=49 Current uses of oestrogen only-recent initiators=49 Current smoking (%): Non users=9 Past users=9 Current users of oestrogen and progestin=7 Current users of oestrogen only=6 Current uses of oestrogen only-recent initiators=6 Inclusion criteria Women aged 70 years and older who were free of diagnosed stroke Exclusion criteria Women who did not have detailed information on age, education, age at menopause, or hormone use		assessments, a comparable population was given the telephone assessment to compare with the participant group. Validity was assessed by administering two tests at an interval of one month in both the participant group and the comparable population. Postmenopausal hormone use was ascertained by the twice yearly questionnaire which asked women about hormone use after menopause. Information on duration of hormone use was collected by self-reporting, and were validated by comparing with medical records. Use of hormones at menopause was defined as any use occurring within 2 years of the reported age at menopause, and first use at older ages was defined as initiation during the 5 years prior to the baseline cognitive test. Statistical analysis: Chane in cognitive function over time was assessed by using multiple linear regression to estimate the adjusted mean differences in decline across various categories of hormone use. Logistic regression was used to calculate	(95%CI)=1.37(0.89, 2.11) Current use, oestrogen+progestin=1315(52);adjusted RR (95%CI)= 1.68 (1.07, 2.64) Current use, oestrogen+progestin 10+ years=712(30);adjusted RR (95%CI)=1.72 (1.03,2.88) Digital span backwards Total decline, n (at least 2 SD of the baseline score) ≥5 points; multivariate adjusted RR (95%CI): Never users=3698 (134); adjusted RR (95%CI)=1.0 Past hormone user=3970 (139); adjusted RR (95%CI)=1.00 (0.77, 1.32) Current use, oestrogen only=3110 (121); adjusted RR (95%CI)= 1.180 (0.82, 1.46) Current use, oestrogen+20 years=959(46); adjusted RR (95%CI)=1.48(0.99, 2.22) Current use, oestrogen+progestin=1191(39);adjusted RR (95%CI)= 0.92 (0.62, 1.38) Current use, oestrogen+progestin 10+ years=643(20);adjusted RR (95%CI)=0.93 (0.55, 1.57) Substantial decline in cognitive performance over 2 years in relation to timing of initiating postmenopausal hormone therapy (subset of population (80%) who reported age at natural menopauseor bilateral oophorectomy) TICS score Total decline, n (at least 2 SD of the baseline score) ≥5 points; multivariate adjusted RR (95%CI): Never user=3615 (169); adjusted RR (95%CI)=1.0 Initiation at menopause (within 2 years of menopause)=3814 (196); adjusted RR (95%CI)=1.10 (0.88, 1.38) Recent initiation of oestrogen alone (during 5 years prior to baseline cognitive testing)=282 (22); adjusted RR (95%CI)= 1.74 (1.08, 2.81)	have ERT use data in this study) C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes (2-year follow-up) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/A D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/A Level of bias: Low Indirectness Does the study match the review protocol in terms of;

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Women reporting heart disease Women who were unreachable or refused , or had died Women with incomplete cognitive assessment		adjusted relative risks of clincally meaningful cognitive decline. In all analyses, data on hormone use and on potential confounders were updates through the questionnaire immediately prior to the baseline cognitive assessment.	Total decline, n (at least 2 SD of the baseline score) ≥ 1.38 points; multivariate adjusted RR (95%CI): Never user=3127 (64); adjusted RR (95%CI)=1.0 Initiation at menopause (within 2 years of menopause)=3258 (81); adjusted RR (95%CI)=1.27 (0.89, 1.82) Recent initiation of oestrogen alone (during 5 years prior to baseline cognitive testing)=254 (5); adjusted RR (95%CI)= 1.11 (1.43, 2.88) Category fluency Total decline, n (at least 2 SD of the baseline score) ≥9 points; multivariate adjusted RR (95%CI): Never user=3456 (95); adjusted RR (95%CI)=1.0 Initiation at menopause (within 2 years of menopause)=3651 (129); adjusted RR (95%CI)=1.38 (1.02, 1.86) Recent initiation of oestrogen alone (during 5 years prior to baseline cognitive testing)=275 (8); adjusted RR (95%CI)= 1.12 (0.52, 2.42) Digits backward Total decline, n (at least 2 SD of the baseline score) ≥5 points; multivariate adjusted RR (95%CI): Never user=3129(112); adjusted RR (95%CI)=1.0 Initiation at menopause (within 2 years of menopause)=3258 (121); adjusted RR (95%CI)=1.13 (0.84, 1.53) Recent initiation of oestrogen alone (during 5 years prior to baseline cognitive testing)=255 (8); adjusted RR (95%CI)= 1.11 (0.50, 2.45)	Population: No (the participants were not representative of the general population) Outcome: Yes Indirectness: Some Participants all registered nurses (indirectness) Information on hormone use was self-reported Telephone assessment of cognition subject to misclassification Loss to follow-up=8% Confounding unknown factors affecting results Possible differences in cognitive decline between hormone users and non users small and difficult to detect, possibly owing to insufficient follow-up time of 2 years (between cognitive interviews) Other information Authors found little association between postmenopausal hormone use, eithe of oestrogen alone or combined with progestin, and decline in cognitive performance over 2 years
Full citation Kawas,C., Resnick,S.,	Sample size N= 472 (514 subjects	Interventions Oral or transdermal	Details Consent:	Results Adjusted RR (95% CI):	Limitations NICE guidelines manual 2012:

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Morrison,A., Brookmeyer,R., Corrada,M., Zonderman,A., Bacal,C., Lingle,D.D., Metter,E., A prospective study of estrogen replacement therapy and the risk of developing Alzheimer's disease: the Baltimore Longitudinal Study of Aging.[Erratum appears in Neurology 1998 Aug;51(2):654], Neurology, 48, 1517-1521, 1997 Ref Id 314433 Country/ies where the study was carried out US Study type prospective study Aim of the study To investigate the use of estrogen replacement therapy and the risk of developing Alzheimer's disease (AD) in a prospective multidisciplinary study of normal aging conducted by the National Institute of Aging. Study dates 1978-1994 (16 years follow-up) Source of funding National Institute on Aging, US	were enrolled, 472 had ERT data) Characteristics Age at enrolment in years, mean (range): 61.5 (28-94) Education level, %: College or graduate degress: 63% Some college: 24% High school education or less: 14% Age of menopause, mean (SD): 46.4 (6.5) Age of menarche, mean (SD): 12.7 (1.5) Ethnicity, % White: 92% Hysterectomy, % Yes: 29% Inclusion criteria -514 post or perimenopausal women who had been followed up to 16 years in the Baltimore Longitudinal Study of Aging were eligible for the study; Exclusion criteria Not reported	estrogens;	Not reported Setting: Research centres Methods: -The BLSA has been collecting ERT data since enrolment of women began in 1978. Use of ERT was documented every 2 years. Every 2 years, subjects returned to the research centre for 2.5 days of multidisciplinary evaluations that included medical history, medication useage (including estrogen), physical and neurological examinations, neuropsychological and functional assessment. -Women who had ever used oral or transdermal estrogens were considered ERT users. Women who had used only estrogen creams were included in the nonuser group because this form of therapy generally does not significantly increase circulating levels of estrogens. Use of ERT was documented every 2 years. -Information on past and presnt duration of ERT use was reported by subjects via categorical assignment (i.e., <6 months, 7 months to 1 year, etc) rather than total months of ERT use.	ERT vs. nonusers: Non users: 1 (reference group) ERT users: 0.457 (0.209-0.997) (only age and educated adjusted in the model) Duration of use categories: 0 year: 1 (reference group) >0-5 years: 0.44 (0.13-1.51), p=0.19 >5-10 years: 0.338 (0.05-2.5), p=0.29 >10 years: 0.50 (0.50-0.17), p=0.21 (only age and education adjusted in the model)	Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)- No A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes (but only age and education were adujsted for in analyses) A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Yes Level of risk-High B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/A B.2 Participants receiving care were kept 'blind' to treatment allocation-N/A B.3 Individuals administering care were kept 'blind' to treatment allocation-N/A Level of risk: Unclear C. Attrition bias (systematic differences between the

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
			Midpoint of the interval was taken as the duration of ERT exposure. -Dementia was diagnosed by neurologic examination and appropriate laboratory and imaging studies. All AD subjects met DSM-III_R criteria for dementia. Statistical methods: -A cox proportional hazards regression analysis was chosen as the method of analysis. Chronologic age was used as the time scale, thus enabling the analysis to control for age; -The model compares each case of AD with all subjects in the study who are alive and free of AD at the age when the AD case was diagnosed. -Education was also included in the model as a binary variable; other variables examined individually included age at menopause, age at menarche, years of natural cyclic estrogen exposure, duration of menopause, and surgical menopause. Follow-up: 16 years -		comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A (about less than 10% of the cohort did not have ERT use data in this study) C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes (16-year follow-up) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-No. Authors report Cox

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					regression but no KM graph. Information on duration is expressed as RR and not HR, misleading reporting. Not all information reported on participant numbers. D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/A D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/A Level of bias: High Indirectness Does the study match the review protocol in terms of; Population: No. Some of the participants were perimenopausal as well as postmenopausal. Proportions of either group not clear. Outcome: Yes Indirectness: Some Other information -In this observational study, estrogen use showed a protective effect in the development of Ad, but the effect was not related to duration of the therapy. -The study was published in 1997 (before 2000), before WHI data was out; -The BLSA is not representative of the general population in terms of education, SES status, and estrogen usage. Also, the authors cannot evaluate the effect of individual esrogen components and routes of delivery because subjects used a variety of oral

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					formulations and few subjects used estrogen pathces.
Full citation Khoo,S.K., O'Neill,S., Byrne,G., King,R., Travers,C., Tripcony,L., Postmenopausal hormone therapy and cognition: effects of timing and treatment type, Climacteric, 13, 259-264, 2010 Ref Id 314467 Country/ies where the study was carried out Australia Study type Cohort study Aim of the study To determine the effects of oestrogen only and oestrogen + progestogen preparations on cognitive performance (cognitive status, general and working memory) whoen taken early and late from onset of menopause Study dates Not reported. The study was published in 2010. Source of funding Royal Brisbane and Women's Hospital Foundation National Health and Medical Council of Australia	Sample size n=410 women from the longitudinal assessment of ageing in women study (LAW) Characteristics Age (years, mean, 95%CI): Never users=56.9 (55.3-58.6) Early starters=59.7 (58.6-60.8) Late starters=64.7 (62.2-67.2) Physical activity (h/week, number): 1-2: Never users=72 Early starters=45 Late starters=12 3-4: Never users=105 Early starters=88 Late starters=23 5+: Never users=32 Early starters=24 Late starters=2 Smoking (number): Never: Never users=111 Early starters=88 Late starters=23 Current: Never users=31 Early starters=9 Late starters=5 Past: Never users=71 Early starters=61 Late starters=0	Interventions Oestrogen Oestrogen+progestogen	Details Participants: Participants were derived from a cohort who had participated in the Longitudinal assessment of Ageing in Women study (LAW study). Written consent was provided by each participant. Women were assessed by physical examination with a qualified medical practitioner and provided a complete sociodemographic history (marital status, years of education, employment status, and socioeconomic status). Information on menopause was ascertained (age of onset, natural or surgical, use of hormone therapy, type of preparation, duration, and timing of initiation of therapy in relation to menopause) as well as information on lifestyle factors (smoking history, amount of physical activity, alcohol consumption). Women who could not recall required information were excluded from the study. Each participant was assessed on two occasions, 5 years apart. The psychometric test battery was administered by a registered psychologist, using a pre-determined	Results Cognitive decline by the Mini-mental state examination (proportion with≥10% decrease in score, HR and 95%CI) Never users (n=213): 1.00 Early start, oestrogen only (n=68):0.28 (0.08, 0.97) Early start, oestrogen+progestogen (n=90): 0.85 (0.38, 1.88) Cognitive decline by the Wechsler memory scale version 3 (proportion with ≥10% decrease in score, HR and 95%CI) Never users (n=213):1.00 Early start, oestrogen only (n=68): 1.01 (0.57, 1.79) Early start, oestrogen+progestogen (n=900: 0.89 (0.53, 1.52) Cognitive decline by the Wechsler memory scale version 3 general memory index vs hormone(proportion with ≥10% decrease in score, HR and 95%CI) Never users (n=213):1.00 Early start, oestrogen only (n=68): 2.80 (0.88, 8.92) Early start, oestrogen+progestogen (n=90): 3.44 (1.21, 9.81)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)- yes A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Yes Level of risk-Low B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/A B.2 Participants receiving care were kept 'blind' to treatment allocation-N/A B.3 Individuals administering care were kept 'blind' to treatment allocation-N/A Level of risk: Low

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Inclusion criteria Women aged 40-60 Women who could recall information on menopause, and information in relation to lifestyle factors Exclusion criteria Women who could not recall information on menopause, and information in relation to lifestyle factors		set of instruments. Cognitive function tests: The mini-mental state examination (MMSE) and National adult reading test (NART) were used to determine cognitive function. Memory was tested using the Wechsler memory scale 3 (WMS-3) and adjusted for age. The general memory index was used to ascertain a global measure of memory ability across both verbal and visual domains, and data was adjusted for age. Statistical analysis: Only women who had used hormone therapy for at least 12 months and at any time during the observation period of the study were considered users. Users of hormone therapy of less than 12 months and past users were excluded from the study. Early starters were defined as ever-users who commenced therapy within 3 years of onset of menopause. Late starters were defined as ever-users who commenced therapy more than 3 years following menopause. A logistic regression model controlling for lifestyle factors, including age, BMI, physical activity, smoking and alcohol intake was generated. All tests were		C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A (less than 10%) C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes (5-year follow-up) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
			two-sided with a p value of 0.05 being significant. A multivariate analysis was performed to evaluate independent effect of each variable on cognitive scores, controlling for age, and other lifestyle factors.		'blind' to participants' exposure to the intervention-N/A D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/A Level of bias: Low Indirectness Does the study match the review protocol in terms of; Population: yes Outcome: Yes Indirectness: None Other information Other information Variation in dose/duration of therapy Study design was cohort
Full citation Rasgon,N.L., Geist,C.L., Kenna,H.A., Wroolie,T.E., Williams,K.E., Silverman,D.H., Prospective randomized trial to assess effects of continuing hormone therapy on cerebral function in postmenopausal women at risk for dementia, PLoS ONE [Electronic Resource], 9, e89095-, 2014 Ref Id 315033 Country/ies where the study was carried out USA Study type RCT Aim of the study To examine effects of oestrogen-based hormone therapy on regional cerebral metabolism in postmenopausal women at risk of development of dementia.	Sample size n=64 Characteristics Age (y, mean, SD): HRT continuers=58..3 (SD 4.5) HRT discontinuers=57.7 (SD 5.6) Years of education (y, mean, SD): HRT continuers=16.0 (SD 1.9) HRT discontinuers=16.6 (SD 2.0) Duration of HRT use (y, mean, SD): HRT continuers=10.5 (SD 4.9) HRT discontinuers=9.4 (SD 6.2) Age at menopause (y, mean, SD): HRT continuers=46.1	Interventions Continued HT use Discontinued HT use	Details Participants All participants were recruited between 2004 and 2007, and two year follow-up assessments occurred between 2006 and 2009. A target sample size of 64 subjects (32 randomised to continue HRT and 32 to discontinue HRT) completing all procedures at 2 years follow-up was establised. Participants were recruited according to the criteria for menopause (Stages of reproductive ageing workshop) and were taking continuous HRT> Screening for the eligibility included willingness to sign consent for all study procedures and to undergo randomisation to continue or discontinue	Results Cerebral metabolism change between randomisation groups (two year change) Medial prefrontal cortex: Continuing users (HT+, n=28) vs discontinuing users (HT-, n=14), greater decline in metabolism in HT- group (t=4.14, P<0.001) Lateral frontal and parietal cortex: Greater decline in HT- group vs HT+ group (t=5.46, P<0.0005) Left frontopareital area: Greater decline in HT- group vs HT+ group (t=5.28, P<0.0005) Oestrogen type and differences in HT randomisation groups Medial cortical area 17bE- discontinuing group (n=13): greater decline in right side precuneus/posterior cingulate than left side (t=4.77, P<0.0005) 17bE+ continuing group (n=16): no significant change in either hemisphere	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A Selection bias A1 - Was there appropriate randomisation - No. Participants were aware of which group they had been randomised to A2 - Was there adequate concealment - No. A3 - Were groups comparable at baseline - Yes Level of bias: Very High B Performance bias B1 - Did groups get same level of care - Yes B2 - no B3 - Were individuals administering care blinded to treatment allocation- Yes Level of bias: High C Attrition bias C1 - Was follow-up equal for

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Study dates 2004-2007 Follow-up two years later between 2006-2009 Source of funding National institute of ageing National centre for research resource, national institutes of health	(SD 7.9) HRT discontinuers=47.5 (SD 4.8) Years of endogenous oestrogen exposure (y, mean, SD): HRT continuers=32.7 (SD 7.5) HRT discontinuers=33.9 (SD 4.6) Inclusion criteria Age 50-65 years of age at time of recruitment ≥1 year current HT use ≥1 year post-complete cessation of menses ≥8 years of education Elevated at risk for dementia (ApoE-allele) Exclusion criteria History of TIAs Carotid bruits on auscultation Lacunes on MRI Evidence of Parkinson's disease Current depression History of drug or alcohol abuse Contraindication for MRI History of mental illness Significant cognitive impairment MI within previous year or unstable cardiac disease Significant cerebrovascular disease		current HRT, psychiatric, physical, and neurological examination, and laboratory blood measures. Eligible participants underwent interim assessments every 3 months to monitor cognition and mood. If a participants 'cognition or mood was determined to have declined, then a referral was made to treating physician for medication management in order to assure mood stabilisation and prevent negative effects on brain metabolism and cognition. At the end of 2 years, participants repeated all baseline assessments, including PET and neuropsychological testing. Self-reported information from participants was confirmed by documentation from primary health care providers whenever possible. 32 participants were randomised to continue HRT and 32 participants were randomised to discontinue HRT. Participants were aware of their randomisation condition (HRT+ vs HRT-). Two group t tests and Chi squared tests were used to assess any potential	CEE+continuing group (n=12): significant bilateral decline in precuneus/posterior regions (left:-4,-20,30, t=6.48, P<0.0005; right: 16, -56, 26, t=4.71, P<0.0005) Progestin use and differences in HT randomisation groups (two year change) 17bE Opposed discontinuation group (n=6) vs opposed discontinuation group 17bE (n=7): Significant difference in metabolic change in posterior cingulate (t=3.95, P<0.001) between both groups 17bE + concurrent progestin continuing group (n=12):significant decline in left parietotemporal and posterior cingulate cortex(P<0.0005) 17bE+concurrent progestin discontinuing group: significant decline in medial frontal gyrus (P<0.0005) 17bE discontinuing unapposed group (n=7): significant decline in precuneus and posterior dorsofrontal cortex (P<0.001).	both groups - Yes C2 - Were groups comparable for dropout - No (more participants dropped out in the discontinued hormone therapy arm) C3 - Were groups comparable for missing data - n/a Level of bias: High D Detection bias D1 - Was follow-up appropriate length - yes (2 years) D2 - Were outcomes defined precisely - Yes D3 - Was a valid and reliable method used to assess outcome - Yes D4 - Were investigators blinded to intervention - yes D5 - Were investigators blinded to confounding factors - unclear Level of bias: Low Indirectness Does the study match the review protocol in terms of Population: yes Intervention: yes Outcomes: yes Indirectness: Some. The authors report that participants were aware of their randomisation condition (HRT or no HRT) Other information

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	Uncontrolled hypertension History of significant liver or pulmonary disease Diabetes Cancer Dementia or other condition that could be expected to produce cognitive deterioration Ue of drugs with potential to significantly affect psychometric test results Parkinsonian medication or phytoestrogen-containing products that could produce oestrogenergic agonist and antagonist effects		differences in clinical or demographic variables in the two treatment groups. PET analysis PET data was analysed by registering and reorientating images into a standardised coordinate system in which data was smoothed, and normalised to mean global activity. The set of pooled data was assessed with the t-statistic on a voxel-by-voxel basis, to identify the profile of voxels that significantly differed between subject groups. The bilateral precuneus/posterir cingulate areas, parietotemporal cortex, and medial prefronatl cortex was decided before the analysis as these areas of the brain show age-related metabolic decline. The medial temporal including the hippocampal area, inferior lateral temporal, and dorsolateral prefrontal cortex were analysed as they have a role in cognitive processes vulnerable to early decline in ageing individuals. A Bonferroni type correction was applied to 12 pre-specified regions, and gorup difference in those regions were noted if P<0.05 after correction. Differences in other regions were described if P<0.0005		

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
			before adjustment		
Full citation Roberts,R.O., Cha,R.H., Knopman,D.S., Petersen,R.C., Rocca,W.A., Postmenopausal estrogen therapy and Alzheimer disease: overall negative findings, Alzheimer Disease and Associated Disorders, 20, 141-146, 2006 Ref Id 315087 Country/ies where the study was carried out USA Study type Cohort study Aim of the study To identify women in Rochester-MN who developed Alzheimer's disease (AD) and the inverse association between AD and Oestrogen therapy (ET). Study dates January 1st, 1985 and December 21st, 1989 Source of funding NR	Sample size N=528 AD cases: n=245 Controls: n=245 Characteristics Not reported Inclusion criteria Women resident in Rochester MN identified by medical records-linkage system. Exclusion criteria Non DA living outside Rochester MN	Interventions NR	Details All medical records from any community care-provider were abstracted for information relevant to the diagnosis of dementia or AD. DSM-IV was used to define diagnosis, and cases were confirmed by a neurologist. Women in the control group had no record of cognitive impairment before the index year. Women with oral or parenteral ET (≥6 months) were contrasted with women who used ET ≤6 months or never. E-creams or E-suppositories were considered non-users. Odds ratios, 95% CIs and p-values (2-tailed test. x=0.05) using conditional logic regression. All regression models included type of menopause. Possible confounders were examined using multi-variable models. Efect modification of variables was evaluated indirectly in stratified analyses to determine significant differences across strata, and directly in multivariable models. For these analyses, matching was ignored to reduce the loss of statistical power caused by missing data (and included age in tertiles in all logistic regression models.	Results n(%) ET use - n(%): <6 months or never: Cases: 216(88.2); Controls: 216(88.2) ≥6 months or ever: Cases: 28(11.4); Controls: 26(10.6) Duration in years: Never: Cases: 216(88.2); Controls: 216(88.2) 0.5-3: Cases: 14(5.7); Controls: 12(4.9) >3: Cases: 14(5.7); Controls: 14(5.7) Age at initiation: Never: Cases: 216(88.2); 216(88.2); ≤49.5: Cases: 17(6.9); Controls: 10(4.1) >49.5: Cases: 11(4.5); Controls: 16(6.5)	Limitations Because this was not a RCT, the samples were not randomised. It is unclear how the controls were matched to the cases during the group-allocation stage. Section 1: Internal validity 1.1 The study addresses an appropriate and clearly focused question-yes Selection 1.2 The cases and controls are taken from comparable populations-yes 1.3 The same exclusion criteria are used for both cases and controls-yes 1.4 What was the participation rate for each group (cases and controls)? n=143 for AD group;n=92 for control group 1.5 Participants and non-participants are compared to establish their similarities or differences 1.6 Cases are clearly defined and differentiated from controls- yes 1.7 It is clearly established that controls are not cases-yes Risk of bias:low Assessment 1.8 Measures were taken to prevent knowledge of primary exposure from influencing case ascertainment-unclear, not reported 1.9 Exposure status is measured in a standard, valid and reliable way-yes Risk of bias: high Confounding 1.10 The main potential confounders are identified and

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					taken into account in the design and analysis-yes (but adjusted only for age and education) Risk of bias: low Statistical analysis 1.11 Have confidence intervals been provided? Yes Risk of bias: Low Section 2: Description of study 2.1 How many people participated in the study 235 (controls) and cases 2.2 What are the main characteristics of the study population? Age, education, symptom duration, MMSE score 2.3 What environmental or prognostic factor is being investigated? AD 2.4 What comparisons are made? AD vs no AD, oestrogen replacement vs no oestrogen replacement 2.5 For how long are participants followed up? Not reported 2.6 What outcome measure(s) is/are used? MMSE score 2.7 What size of effect is identified? MMSE score in oestrogen therapy group wih AD=14.9 (SD 8.1); No oestrogen therapy group with AD=6.5 (AD7.6) 2.8 How was the study funded? Not reported 2.9 Does this study help to answer your guideline review question? Yes Risk of bias:low Indirectness Population: Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Outcome:Yes Indirectness: None Other information
Full citation Seshadri,S., Zornberg,G.L., Derby,L.E., Myers,M.W., Jick,H., Drachman,D.A., Postmenopausal estrogen replacement therapy and the risk of Alzheimer disease, Archives of Neurology, 58, 435-440, 2001 Ref Id 315196 Country/ies where the study was carried out UK Study type Cohort study (nested case control study) Aim of the study To determine whether exposure to ERT is associated with a reduced risk of AD Study dates 1990-1998 Source of funding National institute of ageing, national institutes of health, Stirling Morton charitable trust, Stanley and Harriet Friedman research fund	Sample size N=280 Characteristics Age (y, mean): Cases=66.7 Controls=65.2 Oestrogen exposure (y, mean) Cases=4.2 Controls=4.5 Hypercholesterolaemia (number, %) Cases=3 (5.1) Controls=7 (3.2) Diabetes (number, %) Cases=1 (1.7) Controls=6 (2.7) Hypertension (number, %) Cases= 14 (23.7) Controls=47 (21.3) Inclusion criteria All women who had received at least one prescription for a systemic (oral or transdermal) oestrogen preperation between 1990 and 1998. Women aged 59 to older than 80 years Diagnosis of AD Exclusion criteria Vascular dementias Non-Alzheimer disease degenerative dementia Metabolic conditions (hypothyroidism, metastatic carcinoma, COPD) Other neurological conditions (head injury	Interventions ERT No ERT	Details Participants: Women were identified in the population who were born before January 1 1950 and had received at least one prescription for a systemic oestrogen preparation between 1990 and 1998. Matched controls who had not received any oestrogen at any recorded time were included. AD identification and validation: All women with AD, senile dementia, or presenile dementia between 1992 and 1998 were identified through computer records of the base cohorts of oestrogen therapy users and non-users, without knowledge of their use of oestrogen therapy. Diagnosis was based on the criteria for probable AD (NINCDS-ADRDA). Participants were required to have evidence of dementia (defined as impairment of memory with deficits in at least 2 other domains of cognitive function) by history and clinical examination, and documented progression for at least 6 months. Exposure to oestrogens: Current users were classified as women who had received oestrogen	Results Relative risk of incident AD associated with duration of use of current ERT in postmenopausal women (adjusted for BMI, and cigarette smoking) Oestrogen use non user cases=44/59 non user controls=168/221 Current user cases=15/59 Current user controls=53/221 Adjusted relative risk (95%CI): non user=1.00; current user=1.18 (0.59, 2.37) Duration of oestrogen use (months) Months: 0: cases=44/59; controls=168/221; Adjusted relative risk=1.00 12-35: cases=6/59; controls=14/221; Adjusted relative risk=1.68 (0.60, 4.69) 36-59: cases=5/59; controls=19/221; Adjusted relative risk=0.89 (0.29, 3.44) ≥60: cases=4/59; controls=20/221; Adjusted relative risk=1.05 (0.32, 3.44)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: case control studies Section 1: Internal validity 1.1 The study addresses an appropriate and clearly focused question-yes Selection 1.2 The cases and controls are taken from comparable populations-yes 1.3 The same exclusion criteria are used for both cases and controls-yes 1.4 What was the participation rate for each group (cases and controls)? n=59 for AD group;n=221 for control group, no, there is imbalance in the case group 1.5 Participants and non-participants are compared to establish their similarities or differences-yes 1.6 Cases are clearly defined and differentiated from controls- yes 1.7 It is clearly established that controls are not cases-yes Risk of bias:high Assessment 1.8 Measures were taken to prevent knowledge of primary exposure from influencing case ascertainment-unclear, not reported 1.9 Exposure status is measured in a standard, valid and reliable way-yes Risk of bias: high Confounding 1.10 The main potential confounders are identified and

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	etc.) Depressive disorder with pseudodementia Uncertain cause No documentation of dementia progression		for at least one year and had their last prescription within one year before the index date of diagnosis of AD and the same date in controls were classified as current users. Women who used oestrogen were further classified as combined users of oestrogen and progestin and oral or transdermal formulations. Duration of oestrogen treatment was determined from prescriptions. Use of oestrogen was pre-specified to include those women who had used oestrogen for at least one year. Statistical analysis: A matched analysis was conducted using conditional logistic regression to calculate relative risk estimates (odds ratios) and 95% confidence intervals of developing AD, adjusted for smoking and BMI.		taken into account in the design and analysis-yes (but adjusted only for smoking and BMI) Risk of bias: low Statistical analysis 1.11 Have confidence intervals been provided? Yes Risk of bias: Low Section 2: Description of study 2.1 How many people participated in the study :280 participants 2.2 What are the main characteristics of the study population? Age, use of hormone therapy by prescription, smoking and BMI 2.3 What environmental or prognostic factor is being investigated? AD 2.4 What comparisons are made? AD vs no AD, oestrogen replacement vs no oestrogen replacement, and combination of oestrogen and progestin 2.5 For how long are participants followed up? 5.34 years 2.6 What outcome measure(s) is/are used? Duration of use of oestrogen therapy 2.7 What size of effect is identified? AD risk estimate comparing all current oestrogen users with non users was 1.18 (95%CI 0.59-2.37) 2.8 How was the study funded? National institutes on ageing , national institutes of health 2.9 Does this study help to answer your guideline review

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					question? Yes Risk of bias:low Indirectness Population: Yes Outcome:Yes Indirectness: None Indirectness Does the study match the review protocol in terms of; Population: Yes, but there are fewer cases compared to controls Outcome: Yes Indirectness: None Other information Negative results were probably due to selection bias Number of recorded past ERT users was small, and the primary analysis was restricted to current oestrogen users Authors did not examine other risk factors for AD Study was limited in size due to restrictions of study population to incident rather than prevalent cases, and because of the relative youth and health of ERT users in the study population No evidence was found that current ERT use in postmenopausal women reduced the risk of developing AS. The risk estimate comparing all ERT users vs non users =1.8 (95%CI 0.59, 2.37) women using ERT for more than 5 years vs non users the risk estimate=1.05 (95%CI 0.32, 3.44) Odds ratios were similar in women who used unoppposed

<table>
<tr><th>Study details</th><th>Participants</th><th>Interventions</th><th>Methods</th><th>Outcomes and Results</th><th>Comments</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>oestrogens and for those using progestins</td></tr>
<tr><td>Full citation
Tang,M.X., Jacobs,D., Stern,Y., Marder,K., Schofield,P., Gurland,B., Andrews,H., Mayeux,R., Effect of oestrogen during menopause on risk and age at onset of Alzheimer's disease, Lancet, 348, 429-432, 1996
Ref Id
311731
Country/ies where the study was carried out
USA
Study type
Cohort study
Aim of the study
To examine the effect of previous oestrogen use on the development of AD among elderly women
Study dates
Not reported
Source of funding
Federal grants
Charles S Robertson memorial gift for AD research from the Banbury fund</td><td>Sample size
n=1124 women free of AD, PD, and stroke
Characteristics
Age (y, mean, SD)=74.2 (SD 7.0)
Duration of education (y, mean, SD)=9.2 (SD 4.6)
Ethnicity (number, %)=400 (36) African American, 431 (38) Hispanic, 293 (26) Caucasian.
AD at follow-up 1-5 years (number, %)=167 (14.9)
Age at menopause similar in AD and non-AD groups
Duration of oestrogen use (y, mean, range)=6.8 (range 2 months to 49 years)
HRT use for >1 year in women who had hysterectomy vs natural menopause (number, %)=23/227 (10.1) vs 35/897 (4.0)
Inclusion criteria
No evidence of cognitive impairment at initial interview
No history of stroke or PD
At least one subsequent annual follow-up assessment
Exclusion criteria
Not reported</td><td>Interventions
No oestrogen use
oestrongen use</td><td>Details
Participants:
Participants were selected from a random sample of medicare recipients of the health care financing administration.
Each participant underwent a 90 minute face to face interview followed by a standard assessment, which included a medical history, physical and neurological examination, and a brief battery of neuropsychological tests.
A standard history of oral oestrogen use was obtained from all women at start of study by a trained interviewer as part of the risk-factor questionnaire.
Dementia diagnosis was ascertained by medical records and imaging studies as well as data from the initial and follow-up study examinations. Diagnosis was established by consensus among an independent group of physicians and neuropsychologists from information provided. The group was blinded to the process.
Chi squared tests were used to compare demographic characteristics and history of oestrogen use in women who developed</td><td>Results
Mean age of participating women=74.2 years (SD 7.0)
167/1124 women developed AD and wer older than those women who did not develop AD (78.5 (7.7) vs 73.7 (6.6) years, P=0.001)
156/1124 women reported using oestrogen at onset of menopause
Average duration of oestrogen use=6.8 years (2months to 49 years)
Women who took oestrogen had an earlier onset of menopause (age 45.4 (8.1) years vs 47.0 (7.7) years, P=0.06)
Oestrogen use lower in women who developed AD vs women remaining free of AD (P=0.0006)

Relative risk of incident AD associated with use of oestrogen during postmenopausal period
<table><tr><th></th><th>At risk</th><th>AD*</th><th>Healthy</th><th>Relative risk (95 %CI)</th></tr><tr><td>No oestrogen use</td><td>968</td><td>158</td><td>810</td><td>1.0</td></tr><tr><td>Oestrogen use</td><td>156</td><td>9</td><td>147</td><td>0.4 (0.22, 0.85), p=0.01</td></tr><tr><td>Total</td><td>1124</td><td>167</td><td>957</td><td></td></tr></table>*Cumulative incidence of AD over whole study period</td><td>Limitations
NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies
A. Selection bias (systematic differences between the comparison groups)
A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)- yes
A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-No. The authors did not report information
A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-No. The authors did not report information
Level of risk-High

B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation)
B.1 The comparison groups received the same care apart from the intervention(s) studied-N/A
B.2 Participants receiving care were kept 'blind' to treatment allocation-N/A
B.3 Individuals administering care were kept 'blind' to treatment allocation-N/A</td></tr>
</table>

<table>
<tr><th>Study details</th><th>Participants</th><th>Interventions</th><th>Methods</th><th>Outcomes and Results</th><th>Comments</th></tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AD and those who did not develop AD. ANOVA was used for continuous variables.
Age. ethnic origin, and education were compared in women with and without AD.
The analysis was stratified by median age at baseline because older women entering the study had a higher probability of developing AD than younger women.
Martingale methods were used to check proportional hazards.</td>
<td>Duration of oestrogen use
<table>
<tr><th>Oestrogen use</th><th>At risk</th><th>AD*</th><th>Healthy</th><th>Relative risk (95% CI)</th></tr>
<tr><td>None</td><td>968</td><td>158</td><td>810</td><td>1.0</td></tr>
<tr><td>unknown</td><td>31</td><td>3</td><td>28</td><td>1.3 (0.4, 4.20)</td></tr>
<tr><td>≤ one year</td><td>67</td><td>5</td><td>62</td><td>0.47 (0.20, 1.10)</td></tr>
<tr><td>> one year</td><td>58</td><td>1</td><td>57</td><td>0.13 (0.02, 0.92), p<0.01</td></tr>
</table>
*Cumulative incidene of AD over whole study period</td>
<td>Level of risk: Low

C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants
C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes
C.2a How many participants did not complete treatment in each group?-N/A (less than 10%)
C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A
C.3a For how many participants in each group were no outcome data available?-N/A
C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A
Level of risk: Low

D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified)
D.1 The study had an appropriate length of follow-up-No. Authors did not report information
D.2 The study used a precise definition of outcome-Yes
D.3 A valid and reliable</td>
</tr>
</table>

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/A D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/A Level of bias: Low Indirectness Does the study match the review protocol in terms of; Population: Yes Outcome: Yes Indirectness: None Other information Observational study design Oestrogen was assessed by history Oestrogen use was less common in African-American women and more likely among better educated women Bias could have resulted from unidentified exposure or lifestyle characteristic and could account for results observed
Full citation Zandi,P.P., Carlson,M.C., Plassman,B.L., Welsh-Bohmer,K.A., Mayer,L.S., Steffens,D.C., Breitner,J.C.S., Hormone replacement therapy and incidence of Alzheimer disease in older women: The Cache County Study, Journal of the American Medical Association, 288, 2123-2129, 2002 Ref Id 315595 Country/ies where the study	Sample size N=3246 Characteristics Age (y, mean, SD): No HRT use=76.2 (SD 7.0) Any HRT use=73.1 (SD 5.8) Years of education (y, mean, SD): No HRT use=12.7 (SD 2.3) Any HRT use=13.1 (SD 2.2) AD (number, % yes or no):	Interventions HRT users HRT non-users	Details Participants were screened using the mini-mental state examination followed by the dementia questionnaire to monitor cognitive decline. Results of those women suggesting cognitive change were clinically assessed by specialist trained nurses and psychometric technicians administered a 1 hour battery of neuropsychological	Results Relative hazards of Alzheimer's disease (AD) in women with different degrees of duration and recency of HRT use (estimates from discrete time logistic regression models) Overall HRT use Former =0.33(0.15, 0.65) (n=490, 9 with AD, age=74.5 (sd5.9)) Current =1.08(0.59, 1.91) (n=576,17 with AD, age=71.9 (sd5.4)) HRT use stratified by use duration (y) Former <3 years=0.58 (0.22, 1.27)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)-No. The selected participants

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
was carried out Utah, USA Study type Prospective cohort study. Aim of the study To examine the relationship between use of HRT and risk of Alzheimer's disease (AD) among elderly women. Study dates First assessment in 1995-1997 (Follow-up conducted in 1998-2000). Source of funding NIH grant R01-AG-11380.	No HRT use=yes:58 (7.3); no:742 (92.8) Any HRT use=yes:26 (2.4); no:1040 (97.6) Inclusion criteria Not reported Exclusion criteria 88 women with missing HRT use data		tests. A psychiatrist and neuropsychologist then reviewed the results and assigned diagnosis of dementia. Exposure assessment Women were asked if they had ever taken HRT and for how long. Information on prior use of any medication including HRT was also ascertained. All participants provided their own exposure information. HRT was classified according to report of lifetime use, categorising participants as exposed if they endorsed ever having taken HRT or if HRT was among their current medication. Exposed HRT users were classed as current users or former users. Among current users 72 % were taking unopposed oral oestrogen preparation. Statistical analysis: Characteristics of HRT users and non users were compared using Chi squared tests for dichotomous data and 2-sample t tests for continuous data. Risks of incident AD among HRT users and non users were compared using discrete time survival analysis. Hazard ratios were estimated by odds ratios in logistic models accomodating for multiple covariates.	(n=252, 6 AD, age=73.8(sd5.7)) 3-10 years=0.32 (0.08, 0.68) (n=146, 1 AD, age=74.9 (sd6.0)) >10 years=0.17 (0.01, 0.80) (n=83, 1 AD, age=75.4 (sd6.3)) Current <3 years= 2.41 (0.70, 6.34) (n=58, 4 AD, age 73 (sd 6.2)) 3-10 years=2.12 (0.83, 4.71) (n=173, 7 AD, age 70.9 (sd5.0)) >10 years= 0.55 (0.21, 1.23) (n=344, 6 AD, age 72.1 (sd5.3)	from the screening process were elderly and were classed as definite, probable or possible for AD. This could have an effect on the outcome for risk of dementia A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes, they accounted for age, education, APOE alleles A.3 The groups were comparable at baseline, including all major confounding and prognostic factors- Unclear. Only characteristics for participants who completed wave I and II were reported Level of risk-high B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/A. B.2 Participants receiving care were kept 'blind' to treatment allocation-N/A B.3 Individuals administering care were kept 'blind' to treatment allocation-N/A Level of risk: Low C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					(or analysis was adjusted to allow for differences in length of follow-up)-Yes, those women who completed both assessments were included C.2a How many participants did not complete treatment in each group?-N/A (less than 10%) C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-there was missing information for HRT use for 23 participants (with and without AD) C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-No. There were 1066 participants with any HRT use, and 800 participants without HRT use (difference=266) Level of risk: High D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes (2-year follow-up) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-No. Not reported D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-No. Not reported Level of bias: High Indirectness Does the study match the review protocol in terms of; Population: Yes Outcome: Yes Indirectness: None
Full citation Zucchella,C., Sinforiani,E., Citterio,A., Giarracca,V., Bono,G., Mauri,M., Reproductive life events and Alzheimer's disease in Italian women: a retrospective study, Neuropsychiatric Disease and Treatment, 8, 555-560, 2012 Ref Id 315637 Country/ies where the study was carried out Italy Study type Case-control study Aim of the study To investigate the relationship between major reproductive life events in women with AD. Study dates Women were referred to an Alzheimer assessment unit for diagnosis of AD between 2007 and 2010.	Sample size N=551 AD=275 Controls=276 Characteristics Age (y, mean, SD): AD patients=77.6 (SD 6.3) Controls=76.7 (SD 7.5) Schooling (years): AD patients=6.1 (SD 2.9) Controls=.67 (SD 3.2) Family history for dementia (yes/no): AD patients=98/177 Controls=61/215 Age at disease onset (years): AD patients=74.7 (SD 6.2) Early-onset AD (≤65 years, n, %): AD patients=18 (6.5) Late-onset AD (>65 years, n, %):	Interventions HRT No HRT	Details Diagnosis of dementia: Diagnostic evaluation involved an objective neurological examination, a neuropsychological examination, and neuroimaging (MRI or computed tomography). Control sample was composed of women aged 50 or more who were referred as outpatients to the same hospitals for non-cognitive neurological complaints, including peripheral nervous system diseases, motor disturbances, anxiety, and headache. Controls and AD patients showed the same social and geographical distribution. All participants were menopausal.	Results HRT use AD+HRT+=6/275 AD+HRT-=269/275 AD-HRT+=32/276 AD-HRT-=244/276 X2 test: 17.568 (df=1), P=0.001	Limitations Section 1: Internal validity 1.1 The study addresses an appropriate and clearly focused question-yes Selection 1.2 The cases and controls are taken from comparable populations-yes 1.3 The same exclusion criteria are used for both cases and controls-Not reported 1.4 What was the participation rate for each group (cases and controls)? AD group=275; controls=276 1.5 Participants and non-participants are compared to establish their similarities or differences-yes 1.6 Cases are clearly defined and differentiated from controls- yes 1.7 It is clearly established that controls are not cases-yes Risk of bias:low Assessment

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Source of funding Not reported	AD patients=257 (93.5) Disease duration (years, mean, SD): AD patients=2.9 (SD 1.6) Inclusion criteria Not reported Exclusion criteria Patients with Parkinson's disease or cerebrovascular lesions		All participants completed a structured interview for the collection of demographic and clinical characteristics. Patient data was collected and caregivers participted to provide data when required. All participants were administered the mini-mental state examination to obtain a global cognitive evaluation. AD patients were also examined by the activities of daily living scale (basic everyday activities, higher score=higher autonomy level (range 0-6)), instrumental activities of daily living scale (to evaluate advanced complex activities, range 0-8, higher score=higher autonomy), neuropsychiatric inventory to evaluate presence and severity of behavioural disturbances (range 0-144, higher score=worse), clinical dementia rating to evaluate disease severity (range 0-3, higher score=worse). Statistical analysis: Chi squared test was used for univariate comparison of discrete variables and ANOVA for continuous variables. A multivariate comparison was performed with a regression model, including all the personnel and clinical variables for reproductive life events).		1.8 Measures were taken to prevent knowledge of primary exposure from influencing case ascertainment-Not reported 1.9 Exposure status is measured in a standard, valid and reliable way-yes Risk of bias: low Confounding 1.10 The main potential confounders are identified and taken into account in the design and analysis-yes, but which variables accounted for in analysis not reported Risk of bias: high Statistical analysis 1.11 Have confidence intervals been provided? no Risk of bias: high Section 2: Description of study 2.1 How many people participated in the study:551 2.2 What are the main characteristics of the study population? Mean age 76 (SD 6.3) and above in AD group and 76.7 (SD7.5) in control group, education (4 years or more), age at disease onset 74.7 (SD6.2) in AD group 2.3 What environmental or prognostic factor is being investigated? AD 2.4 What comparisons are made? No HRT vs HRT in AD or no AD cases 2.5 For how long are participants followed up? Not reported 2.6 What outcome measure(s) is/are used? ANOVA chi squared test, univariate and multivariate

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					2.7 What size of effect is identified? Chi squared test=17.568 (1 df), P=0.001 2.8 How was the study funded? Not reported 2.9 Does this study help to answer your guideline review question? Yes, but only for overall risk of AD with HRT use Risk of bias:high Indirectness Population: Yes Outcome:Yes Indirectness: None
Full citation Bove,R., Secor,E., Chibnik,L.B., Barnes,L.L., Schneider,J.A., Bennett,D.A., De Jager,P.L., Age at surgical menopause influences cognitive decline and Alzheimer pathology in older women, Neurology, 82, 222-229, 2014 Ref Id 320209 Country/ies where the study was carried out USA Study type Cohort study Aim of the study To determine the association between age at surgical menopause and both cognitive decline and AD pathology in two longitudinal cohorts Study dates Religious orders study (ROS) start=1994 Memory and ageing project (MAP) start=1997 Study end=2012 Source of funding	Sample size n=1884 (ROS+MAP) Characteristics Age at baseline (y, mean, SD): Natural menopause=78.3 (SD 8.0) Surgical menopause=77.4 (SD 7.7) Race (%caucasian): Natural menopause=93 Surgical menopause=86 Ethnicity (%hispanic): Natural menopause=6 Surgical menopause=6 Age at menopause (y, mean, SD): Natural menopause=49.1 (SD 5.3) Surgical menopause=42.7 (SD 7.2) Duration of reproductive period (y, mean, SD):	Interventions HRT No HRT	Details Participants were from two longitudinal studies of cognitive decline: the Religious Order Study (ROS). which started in 1994, and the Memory and Ageing Project (MAP), which started in 1997. Participants (men and women) agreed to annual clinical evaluations and signed both an informed consent. Both cohorts shared a large coer of identical phenotypic data, allowing efficient merging for joint analyses. The baseline evaluation was completed between 2004 and 2012. Analyses were based on 1884 women who completed the baseline evaluation. The clinical evaluation was repeated annually for up to 18 years with examiners blinded to previously collected data. It included a	Results Non HRT users=1252 All HRT users=632 Inverse association between age at surgical menopause and risk of neurological outcomes pathologic AD diagnosis (adjusted for age at death, education (years), smoking, and study (ROS vs MAP) OR (95%CI)= 0.957 (0.92, 1.00), P=0.053 Clinical AD diagnosis (n=592, adjusted for age at enrollment, education (years), smoking, and study (ROS vs MAP)) Hazard ratio (95%CI)= 0.988 (0.98, 1.00) Assooociation between duration of HRT exposure, when administered within a 5-year window of surgical menopause, and outcomes pathologic AD diagnosis (adjusted for age at death, education (years), smoking, and study (ROS vs MAP) HRT use for 10 years or more vs <10 years: OR(95%CI)=1.053 (0.356, 3.114), P=0.9252 Duration of HRT use (y): OR (95%CI)=1.014 (0.980, 1.049) Clinical AD diagnosis (n=592, adjusted for age at enrollment, education (years), smoking, and study (ROS vs MAP)) HRT use for 10 years or more vs <10	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)- No A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Yes Level of risk-Low B. Performance bias (systematic differences between groups in the care provided, apart from the

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
National institutes of health grants	Natural menopause=36.1 (SD 5.5) Surgical menopause=29.9 (SD 7.4) Hormone replacement therapy use Ever use (%): Within 5 years of menopause: Natural menopause=17.2; surgical menopause=41.6 No HRT: Natural menopause=72.5; surgical menopause=46.3 Current users of HRT (n, %): natural menopause=99 (28); surgical menopause=108 (34) Duration of HRT use (y, mean, SD) Within 5 years of menopause: Natural menopause=12.7 (12.2); surgical menopause=18.6 (15.1) Inclusion criteria Participants free of known dementia at enrollment Exclusion criteria Age at menopause <20 or >60 years age Age of menarch >30 years		medical history, neurologic examination, and cognitive function assessment. Hormonal variables Participants were asked about exogenous hormone use at baseline, dates of use, age at menarche and menopause, and whether menopause had occurred naturally or been induced surgically. Current hormone replacement therapy use was verified by inventory of prescription bottles during participant interviews, with an agreement of 93%. Total duration of HRT use was calculated but was censored in current HRT users at study entry. Cognitive function measures A battery of 19 tests was administered annually to each participant by trained examiners. the mini-mental state examination was used for descriptive purposes. The remaining 17 tests were combined to form a global function cognition score and categorised into 5 domains: 1) Episodic memory 2) Semantic memory 3) Working memory 4) Perceptual memory 5) Visuospatial memory	years: Hazard ratio= 0.917 (0.744, 1.131), P=0.4188 Duration of HRT use (y): Hazard ratio= 0.999 (0.988, 1.009), P=0.8053	intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/A B.2 Participants receiving care were kept 'blind' to treatment allocation-N/A B.3 Individuals administering care were kept 'blind' to treatment allocation-N/A Level of risk: Low C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A (less than 10%) C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
			Dementia and AD classification Clinical diagnosis was made by an expert clinician based on the Joint Working Group of the National Institute of Neurologic and Communicative Disorders and Stroke/AD and Related Disorders Association following a detailed clinical evaluation. The diagnosis of clinical AD was confirmed pathologically in 90% of autopsied participants. Participants meeting criteria for dementia at the baseline clinical evaluation were excluded from the analyses. Statistical measures Demographic and reproductive characteristics of women undergoing natural and surgical menopause were compared using 2 independent sample t tests, Chi squared tests, and Fisher exact test when required. The primary analysis examined the association between age at menopause and longitudinal decline in the global cognition composite score. Adjustments for age at enrollment, years of education, study (ROS		Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes (Up to 18-years) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/A D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/A Level of bias: Low Indirectness Does the study match the review protocol in terms of; Population: Yes Outcome: Yes Indirectness: None Other information

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
			vs MAP) and smoking were made in analyses. Association of age at menopause and AD-related neuropathologic outcomes using multivariate linear regression adjusted for age at death, years of education, smoking, and study. Association of HRT and cognitive decline was assessed as well as duration of use of HRT for 10 years or more compared with less than 10 years of HRT use.		
Full citation Fillenbaum,G.G., Hanlon,J.T., Landerman,L.R., Schmader,K.E., Impact of estrogen use on decline in cognitive function in a representative sample of older community-resident women, American Journal of Epidemiology, 153, 137-144, 2001 Ref Id 320337 Country/ies where the study was carried out USA Study type Cohort study Aim of the study To examine the impact of oestrogen use after menopause on the future level of cognitive function Study dates Enrollment=1986-1987 Assessed=3-6 years later Source of funding National institute on ageing	Sample size n=2705 enrolled n=1907 assessed Characteristics Age=72.78, ranging from 64-100 years All African American women Inclusion criteria Level of cognition unimpaired at baseline according to the Short Portable Mental Status Questionnaire (SPMSQ) Exclusion criteria Not reported	Interventions Past use of oestrogen No use of oestrogen recent use of oestrogen Continuous or intermittent use of oestrogen	Details Participants: The sample was derived from the Duke Established Populations for Epidemiologic Studies of the Elderly (EPESE) programme and were randomly stratified. The participants for the study were women whose cognitive function level was unimpaired at baseline, assessed by the Short Portable Mental Status Questionnaire (SPMSQ) and who survived at 3 years follow-up and were tracked to 6 years follow-up. Data collection: Participants were contacted once a year to complete the SPMSQ as well as face to face interviews to gather information on demographic characteristics, health	Results Oestrogen use and cognitive impairment (multivariable model) (Model 1 and 2 at stage 3 adjusted for majority covariates) model 1 Recent user (n=1826): OR=0.94 (0.42,2.15) past user (n=1826): OR=1.17 (0.76, 1.79) Model 2 continuous user (n=1823):OR =0.68 (0.23, 1.99) intermittent user (n=1823): OR=1.16 (0.76,1.75)	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)- yes Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Yes Level of risk-Low B. Performance bias (systematic differences

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
			condition and health status, and health behaviours. At baseline, information on hormone use was ascertained through interviews. Cognitive function assessment: Cognitive function was assessed by the SPMSQ by introducing two variables: an increase in errors resulting in transition, across a scoring threshold, to impaired cognitive function and an increase of two or more errors on the SPMSQ which predicted decline in functional status. Oestrogen exposure: Exposure to oestrogen was determined from participants' records, especially prescriptions drug data and was defined as recent use, past use and non-use. Duration of use was defined as continuous use or intermittent use. Those women who never used oestrogen were the reference group. Control variables: Potential confounding variables were adjusted and measured at baseline and included age, education, race, marital status, number of natural children, health-related behaviours, smoking status, and alcohol consumption, medications that may influence		between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/A B.2 Participants receiving care were kept 'blind' to treatment allocation-N/A B.3 Individuals administering care were kept 'blind' to treatment allocation-N/A Level of risk: Low C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A (less than 10%) C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
			cognitive impairment, or other self-reported conditions (stroke, diabetes, hip fracture, arthritis, heart attack, hypertension, self-rated health, physical health status, activities of daily living, and depression. Statistical methods: Data for those participants with incomplete information was not included in the analyses. Data was firstly summarised as percentages or means for covariates, follwoed by a univarate analysis to determine associations with cognitive function. Three-stage multivariable models including controls for baseline SPMSQ score at stage 1, then demographic characteristics at stage 2, and health/health related behaviours and medications at stage 3. Discrete-time hazards models were used for the longitudinal analysis for cognitive decline among participants who were not impaired at baseline. In the analysis, respondents who died during the course of the study were removed from the models estimating risk of cognitive impairment and decline.		outcome data were not available)-N/A Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes (3-6 years follow-up) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/A D.5 Investigators were kept 'blind' to other important confounding and prognostic factors-N/A Level of bias: Low Indirectness Does the study match the review protocol in terms of; Population: Yes Outcome: Yes Indirectness: Some. The authors reported that 80% of the sampled participants were women, but do not clarify the other 20% Other information
Full citation Mitchell,J.L., Cruickshanks,K.J.,	Sample size N=1462 Characteristics	Interventions Current HT use Past HT use	Details Participants and data collection:	Results Association of HT with cognitive impairment (OR, 95% CI)	Limitations NICE guidelines manual 2012: Appendix D: Methodology

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Klein,B.E., Palta,M., Nondahl,D.M., Postmenopausal hormone therapy and its association with cognitive impairment, Archives of Internal Medicine, 163, 2485-2490, 2003 Ref Id 229917 Country/ies where the study was carried out USA Study type Cohort study Aim of the study To investigate the association between HT use and cognitive impairment Study dates Initiation of study=1987-1988 5 year follow-up=1993-1995 10 year follow-up=1998-2000 Source of funding Department of veterans affairs women's health fellowship National institites of health	Age (y, mean): Current users=61.5 Past or never users=71.8 High school graduate (%): Current users=91 Past or never users=78 Currently working (%): Current users=46 Past or never users=27 Hysterectomy (%): Current users=61 Past or never users=36 Bilateral oophorectomy (%): Current users=33 Past or never users=17 Alcoholic drink weekly (%): Current users=23 Past or never users=22 Currently smoking (%): Current users=8 Past or never users=10 Weekly vigorous exercise (%): Current users=45 Past or never users=22 BMI (mean)(kg/height in metres): Current users=28.7 Past or never users=29.7 Inclusion criteria Postmenopausal women aged 43-84 Exclusion criteria Women who did not	Previous HT use No HT use	All participants gave written informed consent. Postmenopausal women who participated in the 5 year follow-up for the Epidemiology of Hearing Loss Study (EHLS) were eligible for the study. Participants had to be residents of Beaver Dam, and have a nage of 43-84 years in 1987-1988, and participation in the Beaver Dam Eye study (BDES) in 1988-1990 baseline examination. The follow-up times for the EHLS were 5 years and 10 years for the BDES. and assessments for cognitive function were measured using the mini-mental state examination (MMSE) and SF-36 at baseline and 5 years. As part of the BDES at baseline, 5 years and 10 years, trained interviewers administered detailed questionnaires to ascertain information on reproductive history, current and past use of HRT, and past medical history (including diagnosis of AD). HRT use was confirmed by a physical inventory of prescription bottles or products participants had brought with them to the visit. Current HRT use was defined as use at the 1998-2000 visit. Post menopausal status was defined as a history of surgical menopause	(adjusted for age and education) Current HT use vs past use or never used (n=1460):0.6 (0.2, 1.3) past HT use only vs never used (n=1420):1.0 (0.6, 1.8) Previous HT use vs no previous use (n=1303):0.7 (0.3, 1.8) Duration of HT use vs continuous model (n=1402):0.9(0.8,.1) HT use of ≥ 5years vs never used (n=1402):0.7(0.4,1.4) Age ≥65 years and current HT use vs past or never used (n=934): 0.6(0.2,1.5)	checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A.1 The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)- N/A A.2 Attempts were made within the design or analysis to balance the comparison groups for potential confounders-Yes A.3 The groups were comparable at baseline, including all major confounding and prognostic factors-Yes Level of risk-Low B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B.1 The comparison groups received the same care apart from the intervention(s) studied-N/A B.2 Participants receiving care were kept 'blind' to treatment allocation-N/A B.3 Individuals administering care were kept 'blind' to treatment allocation-N/A Level of risk: Low C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C.1 All groups were followed

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	answer questions on current HT use or did not complete the MMSE		(bilateral oophorectomy), natural menopause (≥ovary, an intact uterus, and cessation of menses for 6 or more months), or hysterectomy if they were older than 56 years. Past HRT use was defined as any past use, exclusive of current use. Information at baseline, 5 years and 10 years was used to calculate duration of HRT use. Statistical analysis: Two-tailed unpaired t tests were used to test differences in characteristics (continuous) of participants. Chi-squared tests were used for dichotomous associations. Odds ratios were obtained from multiple logistic regression analyses for presence of cognitive impairment in current HRT users compared with non-current users. Covariates were added to the analysis in a step-wise manner, and interactions between HRT use, age, education and measures of mental health were also assessed. This analysis was repeated for duration of HRT use and past use of HRT. Analyses were also repeated using current HRT use as determined by the 5 year follow-up examination, and covariate data from		up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)-Yes C.2a How many participants did not complete treatment in each group?-N/A (less than 10%) C.2b The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)-N/A C.3a For how many participants in each group were no outcome data available?-N/A C.3b The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)-N/A Level of risk: Low D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D.1 The study had an appropriate length of follow-up-Yes (10-year follow-up) D.2 The study used a precise definition of outcome-Yes D.3 A valid and reliable method was used to determine the outcome-Yes D.4 Investigators were kept 'blind' to participants' exposure to the intervention-N/A D.5 Investigators were kept 'blind' to other important confounding and prognostic

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
			the 5 year follow-up visit was used. Repeated analyses were carried out excluding history of AD because data would be unreliable. Surgical menopause was also excluded from a repeated analysis because it would have a different impact on the relationship between HRT use and impaired cognition. Participants with bilateral oophorectomy or depression were also excluded from repeated analyses due to different impact on HRT use and cognitive function.		factors-N/A Level of bias: Low Indirectness Does the study match the review protocol in terms of; Population: Yes Outcome: Yes Indirectness: None Other information Study did not find a significant association between postmenopausal HT use and impaired cognition after adjustment of age and education

H.8.8 Loss of muscle mass (sarcopenia)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Full citation Sipila,S., Taaffe,D.R., Cheng,S., Puolakka,J., Toivanen,J., Suominen,H., Effects of hormone replacement therapy and high-impact physical exercise on skeletal muscle in post-menopausal women: a randomized placebo-controlled study, Clinical Science, 101, 147-157, 2001 Ref Id 288718 Country/ies where the study was carried out Finland Study type Randomized, placebo-	Sample size N=80 Exercise group: 20 HRT group: 20 Exercise+HRT group: 20 Control group: 20 Characteristics Postmenpausal women aged 50-55 years; were within 5 years of onset of menopause Body mass (kg)/mean (SD) HRT group: 69.9 (10.7) Control group: 68.3 (11.7) Lean body mass (kg)/mean (SD) HRT group: 45.8 (4.4) Control group: 47.4 (5.1) Body fat (%)/mean (SD)	Interventions Combined oestradiol (2mg) and noretisterone acetate (1mg) administered continuously, one tablet per day, for 1 year Exercise group participated in a 1-year progressive physical training programme that included a supervised circuit training session twice a week and a series of home exercises on 4 days per week. Control group were instructed to continue their daily routines and not to change their physical activity levels.	Details Subjects randomly assigned to one of 4 groups: Exercise; HRT; exercise + HRT; and control Randomisation carried out manually by drawing lots HRT carried out double-blind. Muscle perfomance measured using Maximal isometric knee extension force. Cross-sectional area (CSA) and lean tissue CSA (LCSA) measured in the quadriceps femoris	Results Muscle strength Assessed by maximal isometric muscle torque (knee extension torque, KEt) Muscle mass Assessed by quadriceps and lower leg muscle CSA and LCSA 6 months measurements (number of participants who completed) HRT group: 17 Control group:17	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Yes A3. The groups were comparable at baseline including all majorconfounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
controlled trial Aim of the study Investigated the effect of HRT and high-impact physical exercise on muscle performance, muscle cross-sectional area, and muscle composition in postmenopausal women. Study dates Not reported. Study published in 2001. Source of funding Not reported.	HRT group: 33.9 (6.5) Control group: 29.7 (6.0) Inclusion criteria Participants had to have no serious medical conditions, no current or previous (unless for no longer than 6 months in duration and at least 2 years prior to screening) use of medications including oestrogen, fluoride, calcitonin, biophosphonates or steroids, their menstruation at least 0.5 years but not more than 5 years ago, FSH > 30 i.u./L, and no contrainidications for exercise and HRT. Exclusion criteria Not specifically reported. See above.		and lower leg muscles (ie. ankle flexors and extensors). Measuements made at 6 and 12 months. There were 6 and 12 months treatment groups	12 month measurements (number of participants who completed) HRT group: 15 Control group:15 MUSCLE STRENGTH KEt, mean (SD) change at 6 months (Nm) HRT group: baseline: 9.6 (16.1) Control group: baseline: -5.1 (17.3) KEt, mean (SD) change at 12 months (Nm) HRT group: baseline: -1.1 (13.7) Control group: baseline: -10.8 (18.5) MUSCLE MASS Quadriceps muscle CSA, mean (SD) change at 6 months (cm^2) HRT group: baseline: 1.6 (4.7) Control group: baseline: 0.1 (4.6) Quadriceps muscle CSA, mean (SD) change at 12 months (cm^2) HRT group: baseline: 2.7 (4.9) Control	between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - 25% in each treatment group did not complete treatment C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				group: baseline: 0.4 (4.7) Quadriceps muscle LCSA, mean (SD) change at 6 months (cm²) HRT group: baseline: 1.5 (4.6) Control group: baseline: -0.2 (4.4) Quadriceps muscle LCSA, mean (SD) change at 12 months (cm²) HRT group: baseline: 2.6 (4.7) Control group: baseline: 0.2 (4.6) Lower leg muscle CSA, mean (SD) change at 6 months (cm²) HRT group: baseline: 2.3 (4.3) Control group: baseline: 1.6 (5.9) Lower leg muscle CSA, mean (SD) change at 12 months (cm²) HRT group: baseline: 3.6 (4.2) Control group: baseline: 2.0 (5.8) Lower leg muscle LCSA, mean (SD) change at 6 months	D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - N/A Low risk of bias Other information For the purposes of the review question, only results for the HRT and control groups were presented.

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				(cm²) HRT group: baseline: 2.5 (4.1) Control group: baseline: 1.7 (5.7) Lower leg muscle LCSA, mean (SD) change at 12 months (cm²) HRT group: baseline: 3.6 (4.1) Control group: baseline: 2.1 (5.5)	
Full citation Armstrong,A.L., Oborne,J., Coupland,C.A., Macpherson,M.B., Bassey,E.J., Wallace,W.A., Effects of hormone replacement therapy on muscle performance and balance in post-menopausal women, Clinical Science, 91, 685-690, 1996 Ref Id 294639 Country/ies where the study was carried out UK Study type Randomised, double-blind controlled trial Aim of the study To evaluate the effect of oral HRT plus calcium versus calcium alone on balance, muscle performance and falls over 48 weeks in postmenopausal women. Study dates Not reported.	Sample size N=116 HRT and calcium group=57 Calcium group=59 Characteristics Age, mean (SD) years HRT and calcium group: 60.5 (6.3) Calcium group: 61.3 (5.8) Post-menopausal years, mean (SD) years HRT and calcium group: 11.7 (7.6) Calcium group: 13.7 (7.3) Weight, mean (SD) kg HRT and calcium group: 63.7 (12.6) Calcium group: 67.8 (9.3) Inclusion criteria Caucasian post-menopausal women who had suffered a wrist fracture within the previous 7 weeks. No contra-indication to HRT Exclusion criteria 1. Overt neurological or neuromuscular condition that	Interventions Prempak C or Premarin 0.625 mg depending on uterine status Both test and control group given 1000 mg/day elemental calcium	Details Blocked randomisation and stratified by age and time out of the fracture treatment device. Measurements were made blind to treatment group Isometric hand grip strength measured using a calibrated electronic dynamometer All measurements were made every 12 weeks for 24 weeks. Hand grip strength assessed over 48 weeks.	Results Muscle strength Isometric hand grip strength Muscle mass Not evaluated MUSCLE STRENGTH Hand grip strength, mean (SD) change over 48 weeks, kg HRT and calcium group: 0.64 (3.51) Calcium group: 1.01 (2.69) NS	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Unclear A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Source of funding Wishbone Trust and the Special Trustees for the Nottingham Hospitals	might impair strength, balance or mobility. 2. Use of drugs that affect balance				Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - 21% in test group and 7% in control group C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Yes Low risk of bias Indirectness Does the study match the review protocol in

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious
Full citation Kenny,A.M., Kleppinger,A., Wang,Y., Prestwood,K.M., Effects of ultra-low-dose estrogen therapy on muscle and physical function in older women, Journal of the American Geriatrics Society, 53, 1973-1977, 2005 Ref Id 320065 Country/ies where the study was carried out USA Study type Double-blind, placebo-controlled trial Aim of the study To determine the effects of ultra-low-dose hormone therapy on muscle mass and physical function in community-dwelling women. Study dates Not reported. Source of funding Claude Pepper Older Americans Independence Center General Clinical Research Center Paul Beeson Physician Faculty Scholars in Aging Research Program	Sample size N=167 Estrogen group=83 Placebo grroup=84 Characteristics Healthy community-dwelling women aged 65 years and older Age, mean (SD) years Estrogen group: 73.9 (0.6) Placebo group: 74.7 (0.6) BMI, mean (SD) kg/m² Estrogen group: 28.0 (0.5) Placebo group: 28.3 (0.5) Appendicular skeletal muscle mass (ASM), mean (SD) kg Estrogen group: 15.7 (0.2) Placebo group: 15.7 (0.2) ASM/height², mean (SD) kg/m² Estrogen group: 6.4 (0.9) Placebo group: 6.4 (0.9) Inclusion criteria Healthy, community-dwelling women older than 65 years. Exclusion criteria 1. Diseases ormedications affecting bone metabolism. 2. Use of estrogen or calcitonin within the past 6 months 3. Ever use of bisphosphonates of fluoride 4. History of breast or endometrial cancer within the past 5 years 5. Baseline endometrial thickness greater than 5 mm.	Interventions 0.25 mg 17-beta estradiol or placebo for 36 months. All women (estradiol or placebo) with an intact uterus received micronized progesterone 100 mg/d for 2 weeks every 6 months. All women received 1,300 mg elemental calcium with 1,000 IU vitamin D per day.	Details Randomisation to treatment with estradiol or placebo using a computer-generated list. Staff and participants were blinded to treatment group. Appendicular skeletal muscle mass deermined by combining the lean tissue mass of the regions of the arms and legs	Results Muscle strength Not evaluated Muscle mass Appendicular skeletal muscle mass Sarcopenia Defined as ASM/height² 2 standard deviations or less than young, healthy reference population mean Sarcopenia was present in 13% of population at baseline MUSCLE MASS ASM, mean (SD) change over 3 years, kg Estrogen group: -0.2 (0.13) Placebo group: -0.4 (0.13) NS changes ASM/height², mean (SD) change over 3 years, kg/m² Estrogen group: -0.1 (0.57) Placebo group: -0.1 (0.57) NS changes	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Unclear A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Low risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - 12 in estrogen

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	6. Any thromboembolic event within 6 months 7. Bome mineral density t score less than -4 8. Symptomatic vertebral fracture within the past year or past history of low trauma hip fracture.				group and 16 in placebo group C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious
Full citation Skelton,D.A., Phillips,S.K., Bruce,S.A., Naylor,C.H., Woledge,R.C., Hormone replacement therapy increases isometric	Sample size N = 102 HRT group = 50 Control group = 52 Characteristics Age, mean (SD) years	Interventions Prempak-C (Cyclical HRT preparation containing conjugated oestrogens (0.625 mg taken each day) with norgestrel (0.15 mg taken 12	Details Open-label design. Subjects randomly assigned to control or HRT group. Adductor pollicis	Results OUTCOMES Muscle strength Adductor pollicis muscle MVF	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups)

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
muscle strength of adductor pollicis in post-menopausal women, Clinical Science, 96, 357-364, 1999 Ref Id 320097 Country/ies where the study was carried out United Kingdom Study type Open-label randomized trial Aim of the study To assess the change in adductor pollicis (AP) muscle strength and/or muscle cross-sectional area during 1 year's HRT treatment. Study dates 1993 to 1997 Source of funding Not reported.	HRT group: 60.9 (3.2) Control group: 60.6 (3.3) Body weight, mean (SD) kg HRT group: 65.8 (9.3) Control group: 64.4 (9.1) Maximal voluntary force (MVF) of AP, mean (SD) N HRT group: 59.3 (7.7) Control group: 57.7 (7.8) Cross-sectional area (CSA) of AP, mean (SD) mm² HRT group: 59.3 (7.7) Control group: 57.7 (7.8) Inclusion criteria Generally healthy women 5-15 pears post-menopause, with a serum oestradiol level below 150 pmol/l and a body mass index of 20-29 kg/m². Exclusion criteria 1. Pain or stiffness of the thumb 2. Evidence of wasting of hand muscles or generalised cardiovascular or neuromuscular disease 3. Were regularly using any medication likely to affect muscle function or motivation. 4. Hysterectomy, undiagnosed genital bleeding, chronic renal or hepatic disease, stroke or transient ischaemic attack, gall bladder disease. 5. Known or suspected estrogen-dependent neoplasia, any other malignancy, known hypersensitivity to oestrogens or progestins 6. Use in the previous 12 months of oestrogen-containing preparations or tibolone 7. Use within the previous 3 years of oestrogen implants	consecutive days during each 28 day cycle).	MVF and CSA measured at baseline and at 2, 4, 6, 13, 26, 39, and 52 weeks.	Muscle mass Adductor pollicis CSA MUSCLE STRENGTH Adductor pollicis muscle MVF, mean (SE) percentage change HRT group: 12.4 (1.0) Control group: -2.9 (0.9) mean (SE) percentage difference between the two groups: 15.4 (1.3) *Significant increase in muscle strength in HRT group compared to control group. MUSCLE MASS Adductor pollicis muscle CSA No significant changes in both groups. Results of follow-up study 2-3 years after trial (which is reported in Onambele et al. study id: 320079) Adductor pollicis muscle MVF Muscle strength was maintained in HRT group. Adductor pollicis muscle CSA	A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Unclear A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - No A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes High risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - No B3. Individuals administering care were kept 'blind' to treatment allocation - No High risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - 13 in treatment group and 4 in control group C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - No C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
	8. History of glucocorticoid use 9. Blood-clotting disorders, malasorpton, alcohol or drug abuse, or use of any medications that would influence the metabolism of oestrogen.			No significant changes in both groups.	respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes High risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - No D5. Investigators were kept 'blind' to other important confounding and prognostic factors - No High risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious
Full citation Ribom,E.L., Piehl-Aulin,K., Ljunghall,S., Ljunggren,O., Naessen,T., Six months of hormone replacement therapy does not influence muscle strength in postmenopausal women, Maturitas, 42, 225-231, 2002 Ref Id 294406 Country/ies where the study was carried out Sweden Study type Double blinded,	Sample size N=40 HRT group=20 Placebo group=20 Characteristics Postmenopausal women aged 60-78 years. Age, mean (SD) years HRT group: 67.5 (1.2) Placebo group: 67.0 (0.9) BMI, mean (SD) kg/m² HRT group: 67.5 (1.2) Placebo group: 67.0 (0.9) Inclusion criteria 1. 60 years of age or older	Interventions Menorest 50 µg/24 hr (estradiol 4.3 mg) and Gestapuran 2.5 mg (medroxyprogesteron) daily or placebo	Details Randomisation was stratified. Hand grip strength (maximal voluntary contraction, MVC) measured using a JAMAR hydraulic hand dynamometer. Isokinetic knee flexion and extension strength measured using a Cybex II dynamometer.	Results Muscle strength 1. Hand grip strength (MVC) 2. Isokinetic knee flexion and extension strength (MVC) Muscle mass Not evaluated MUSCLE STRENGTH Right knee flexion strength, mean (SD) Nm change at 6 months	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Yes A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Unclear A3. The groups were comparable at baseline including all major confounding and prognostic

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
prospective and placebo controlled trial. Aim of the study To evaluate the effect of 6 months of HRT on muscle strength in postmenopausal women, older than 60 years of age. Study dates Not reported. Source of funding Swedish National Centre for Research in Sports and the Swedish Society of Medicine (No. 99-02-0248)	2. Free of diseases that could interfere with results of study 3. Not haven taken any HRT for at least the last 6 months Exclusion criteria See above.			HRT group: 0.7 (9.8) Placebo group: -0.1 (12.3) NS Left knee flexion strength, mean (SD) Nm change at 6 months HRT group: 3.7 (12.5) Placebo group: -1.1 (9.4) NS Right knee extension strength, mean (SD) Nm change at 6 months HRT group: 5.6 (16.0) Placebo group: 4.2 (12.1) NS Left knee extension strength, mean (SD) Nm change at 6 months HRT group: 6.4 (14.6) Placebo group: -2.1 (13.9) P=0.0 Right hand grip strength, mean (SD) kg change at 6 months HRT group: 1.8 (1.6) Placebo group: 1.9 (2.7) NS Left hand grip strength, mean (SD)	factors - Yes Unclear risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - 3 participants in each treatment group C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? -None C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
				kg change at 6 months HRT group: 2.4 (3.4) Placebo group: 0.8 (2.3) P=0.1	D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Yes Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious
Full citation Maddalozzo,G.F., Cardinal,B.J., Li,F., Snow,C.M., The association between hormone therapy use and changes in strength and body composition in early postmenopausal women, Menopause, 11, 438-446, 2004 Ref Id 320166 Country/ies where the study was carried out USA Study type Prospective, non-randomized, 1-year comparative cohort study. Aim of the study To prospectively examine potential differences in upper- and lower-body muscle strength in early postmenopausal women on and not on HRT. Study dates Not reported. Source of funding	Sample size N=136 HRT group=67 Non-HRT group=59 Characteristics Postmenopausal women Age, mean (SD) years HRT group: 50.9 (3.0) Non-HRT group: 51.3 (3.0) Time past menopause, mean (SD) months HRT group: 15.2 (10.1) Non-HRT group: 12.6 (1.1) Weight, mean (SD) kg HRT group: 66.0 (9.3) Non-HRT group: 68.6 (1.4) Inclusion criteria 1. Women who had experienced menopause within the previous 36 months from the time of baseline testing. 2. Period-free for 12 months without being pregnant 3. FSH levels of 40 mIU/ml or higher 4. BMI (19-30 kg/m²) 5. Diagnosed as	Interventions HRT (0.625 mg conjugated equine estrogen, brand name Premarin) or non-HRT group.	Details Measurements taken at baseline and at 12 months. Muscle strength of hip abductors, knee extensors and flexors, chest and upper back assessed by isokinetic dynamometry.	Results Muscle strength 1. Muscle strength of quadriceps, hamstring, hip abduction, pectoral (chest) and latissimus dorsi (upper back) 2. Mean total strength composite score of five strength variables Muscle mass Not evaluated. MUSCLE STRENGTH Individual strength measures No between group differences of individual muscle groups Total muscle strength score, mean (SD) change from baseline, N	Limitations NICE guidelines manual 2012: Appendix D: Methodology checklist: cohort studies A. Selection bias (systematic differences between the comparison groups) A1. The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome(s) under study)- No A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders - No A3. The groups were comparable at baseline, including all major confounding and prognostic factors - Yes High risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Unclear B2. Participants receiving care were kept 'blind' to treatment allocation - No B3. Individuals administering care were kept 'blind' to treatment allocation - No High risk of bias

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
Not reported.	postmenopausal by a physician for 36 months or less 6. Participants taking HRT (0.625 mgconjugated equine estrogen, brand name Premarin). Exclusion criteria 1. Non-HRT users who had taken HRT for 12 consecutive months before applying to the study. 2. Hypertension 3. Metabolic diseases that may affect bone or muscle metabolism [including diabetes, thyroid disease, hypercholesterolemia (with statin medication) and multiple sclerosis] 4. Any musculoskeletal disorders that prevented participation in the study.			HRT group: 5.95 (9.66) Non-HRT group: 6.47 (9.72) P=0.52	C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - None C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - None C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - No D5. Investigators were kept 'blind' to other important confounding and prognostic factors - No High risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					Other information SD change calculated from [(SDbaseline² + SDfinal²) - (2*correlation coefficient*SDbaseline*SDfinal)]½
Full citation Taaffe,D.R., Sipila,S., Cheng,S., Puolakka,J., Toivanen,J., Suominen,H., The effect of hormone replacement therapy and/or exercise on skeletal muscle attenuation in postmenopausal women: a yearlong intervention, Clinical Physiology and Functional Imaging, 25, 297-304, 2005 Ref Id 320173 Country/ies where the study was carried out Finland Study type Double-blind randomised placebo controlled trial. Aim of the study To evaluate whether the hormonal and metabolic effects of HRT would preserve or enhance the attenuation of skeletal muscle Study dates Not reported. Source of funding Academy of Finland. Ministry oF Education.	Sample size N=80 HRT group=20 Exercise=20 HRT+exercise=20 Control=20 Characteristics Height, mean (SD) cm HRT: 159.8 (6.7) Control: 163.4 (5.3) Body weight, mean (SD) kg HRT: 69.2 (10.8) Control: 68.3 (11.7) Inclusion criteria 1. Healthy postmenopausal women aged 50-57 years. 2. No serious cardiovascular or locomotor conditions 3. Not currently or previously (no longer than 6 months and at least 2 years prior to screening) taking medications including oestrogen, fluoride, calcitonin, bisphosphonates or steroids 4. Last menstruation at least 0.5 years but not more than 5 years ago 5. BMI < 33 kg/m² 6. Willingness to participate Exclusion criteria See above	Interventions Daily (one tablet) combined oestradiol (2 mg) and norethisterone acetate (1 mg) or placebo for 1 year	Details Participants randomised in a double-blind fashion. Cross-sectional area (CSA) of quadriceps and posterior muscles derived from CT analysis. Isometric knee extension strength assessed in a custom-made dynamometer chair.	Results Muscle strength Isometric knee extension strength Muscle mass 1. Quadriceps muscles CSA 2. Posterior muscles CSA MUSCLE STRENGTH Knee extensor strength, mean (SD) change over 1 year, Nm HRT: 6.5 (39.0) Control: -21.6 (60.6) MUSCLE MASS Quadriceps muscles CSA, mean (SD) change over 1 year, cm² HRT: 2.6 (4.7) Control: 0.2 (4.6) Posterior muscles CSA, mean (SD) change over 1 year, cm² HRT: 3.0 (3.8) Control: 1.0 (3.7)	Limitations NICE guidelines manual 2012: Appendix C: Methodology checklist: randomised controlled trials A. Selection bias (systematic differences between the comparison groups) A1. An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) - Unclear A2. There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) - Unclear A3. The groups were comparable at baseline including all major confounding and prognostic factors - Yes Unclear risk of bias B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation) B1. The comparison groups received the same care apart from the intervention(s) studied - Yes B2. Participants receiving care were kept 'blind' to treatment allocation - Yes B3. Individuals administering care were kept 'blind' to treatment allocation - Yes Low risk of bias C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) - Yes C2a. How many participants did not complete treatment in each group? - 6 in HRT group and 5 in control group did not complete treatment

Study details	Participants	Interventions	Methods	Outcomes and Results	Comments
					C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) - Yes C3a. For how many participants in each group were no outcome data available? - Outcome data was available for those who completed treatment. C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available) - Yes Low risk of bias D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified) D1. The study had an appropriate length of follow-up - Yes D2. The study used a precise definition of outcome - Yes D3. A valid and reliable method was used to determine the outcome - Yes D4. Investigators were kept 'blind' to participants' exposure to the intervention - Yes D5. Investigators were kept 'blind' to other important confounding and prognostic factors - Unclear Low risk of bias Indirectness Does the study match the review protocol in terms of Population: Yes Intervention: Yes Outcomes: Yes Indirectness: No serious Other information For the purposes of the review question, only results for the HRT and placebo group have been reported.

H.9 Premature ovarian insufficienty

H.9.1 Diagnosis of premature ovarian insufficiency

Study details	Participants	Tests	Methods	Outcomes and Results	Comments
Full citation Jadoul,P., Anckaert,E., Dewandeleer,A., Steffens,M., Dolmans,M.M., Vermylen,C., Smitz,J., Donnez,J., Maiter,D., Clinical and biologic evaluation of ovarian function in women treated by bone marrow transplantation for various indications during childhood or adolescence, Fertility and Sterility, 96, 126-133, 2011 Ref Id 267224 Country/ies where the study was carried out Belgium Source of funding Belgian National Fund for Scientific Research. Fondation Saint Luc. Unrestricted grant from Novo-Nordisk. Study dates Not reported. Study type Cross-sectional observational study. Aim of the study	Sample size N = 33 • n = 12 ongoing ovarian function • n = 21 ovarian failure Characteristics Mean age at time of BMT = 9.8 ± 5.2 years (range 1.2 - 19.0) Mean age at time of evaluation = 25.3 ± 7.2 years (range 16.6 to 46.4) Number receiving BMT for a benign disease = 12 (34%) Number receiving BMT following chemotherapy for malignant disease = 23 (66%) Inclusion criteria Female patients aged ≥ 16 years who had undergone BMT before the age of 19 years and had been in complete remission for	Tests FSH, estradiol and AMH were measured at the time of the study and related to ovarian function 10 years after BMT. The last documented FSH level prior to starting hormonal therapy was also reported. Definitions used Evidence of ovarian function: Presence and progression of pubertal development, occurence of menstrual cycles in the absence of hormonal treatment, or pregnancy. Ovarian failure: Absent pubertal development or progression, secondary amenorrhoea confirmed by the observation of menopausal FSH levels.	Methods Patients attended the clinic for a single evaluation. Assessment of gonadal function was based on a complete clinical history (pubertal development, menstruation patterns, occurence of pregnancy, fertility work-up, menopausal symptoms and hormone use), retrospective analysis of hormone levels before estrogen-progesterone therapy and measurement of hormone levels at the time of the study (FSH, estradiol and AMH).	Results 76% of women were taking either HRT or OCP when the following measurements were taken. AMH Cut-off ≤ 0.5 µg/L to diagnose POI Sensitivity, % (95% CI): 52.6 (29 to 76)[1] Specificity, % (95% CI): 75 (43 to 95)[1] Positive likelihood ratio, (95% CI): 2.11 (0.72 to 6.13)[1] Negative likelihood ratio, (95% CI): 0.63 (0.36 to 1.12)[1] AMH Cut-off ≤ 1.12 µg/L to diagnose POI (= 8pmol/L) Sensitivity, % (95% CI): 100 (82 to 100)[1] Specificity, % (95% CI): 33 (10 to 65)[1] Positive likelihood ratio, (95% CI): 1.50 (1.01 to 2.24)[1] Negative likelihood ratio, (95% CI): 0.00 (NC)[3]	Limitations All current hormone measurements were taken whilst the majority of participants were taking hormonal medication (either HRT or OCP) which will have affected the hormone levels. It is unclear how evidence of ongoing ovarian function at the time of the study was established, as the majority of participants were taking hormonal medication which will have stimulated a menstrual cycle even in the absence of underlying ovarian function. Further, "evidence of ongoing ovarian function 10 years after BMT" is reported, however 4 participants are reported as being within 10 years of BMT. The timing of measurement of "last FSH values without treatment" is not described in any individual woman. Other information

Study details	Participants	Tests	Methods	Outcomes and Results	Comments
To evaluate ovarian function in young women several years after bone marrow transplantation (BMT) and compare the impact of different pretransplantation conditioning regimes. Also to investigate whether primary pathology, age and pubertal status at BMT, or time elapsed since BMT may influence the effect on ovarian function.	≥ 3 years. Exclusion criteria Not reported.			FSH cut-off > 30 mIU/mL to diagnose POI Sensitivity, % (95% CI): 38 (18 to 62)[1] Specificity, % (95% CI): 100 (74 to 100)[1] Positive likelihood ratio, (95% CI): ∞ (NC)[2] Negative likelihood ratio, (95% CI): 0.62 (0.44 to 0.87)[1] Estradiol cut off < 50 pg/mL to diagnose POI Sensitivity, % (95% CI): 52 (30 to 74)[1] Specificity, % (95% CI): 33 (10 to 65)[1] Positive likelihood ratio, (95% CI): 0.79 (0.44 to 1.39)[1] Negative likelihood ratio, (95% CI): 1.43 (0.57 to 3.58)[1] Using the final FSH measurement before treatment was started to diagnose POI gives FSH cut-off > 30 mIU/mL to diagnose POI Sensitivity, % (95% CI): 100.0 (84 to 100)[1] Specificity, % (95% CI): 100 (69 to 100)[1] Positive likelihood ratio, (95% CI): ∞ (NC)[2] Negative likelihood ratio, (95% CI): 0.00 (NC)[3] 1 Point estimate and 95% CI calculated by the NCC-WCH technical team from data reported in the article 2 Specificity = 100% therefore +LR = ∞ and 95%	

Study details	Participants	Tests	Methods	Outcomes and Results	Comments
				CI not calculable. Calculated by the NCC-WCH technical team from data reported in the article. 3 Sensitivity = 100% therefore -LR = 0 and 95% CI not calculable. Calculated by the NCC-WCH technical team from data reported in the article.	
Full citation Giuseppe,L., Attilio,G., Edoardo,D.N., Loredana,G., Cristina,L., Vincenzo,L., Ovarian function after cancer treatment in young women affected by Hodgkin disease (HD), Hematology, 12, 141-147, 2007 Ref Id 266903 Country/ies where the study was carried out Italy Source of funding Not reported. Study dates Not reported. Study type Observational case series. Aim of the study To evalulate the best method of assessing ovarian reserve in 29 women with Hodgkin's disease treated with chemotherapy (and to assess the ovarian protective effect of GnRH-analogues).	Sample size N = 29 • n = 21 normal cycles • n = 8 amenorrhoeic Characteristics Age, years (mean, SD) = 28.5 ± 7.3 Mean time between end of chemotherapy and present observation, years (mean, SD) = 4.2 ± 2.8 Inclusion criteria Patients treated for Hodgkin's lymphoma between 1996 and 2002. Exclusion criteria Not described.	Tests Transvaginal ovarian follicle count was conducted on day three of the menstrual cycle, in addition to serum levels of FSH, LH, inhibin B and AMH. In amenorrhoeic patients, clinical and laboratory evaluations were performed at first visit, or after three months suspension of hormonal replcament therapy, if any. Definitions used Menstrual cycle present: normal cycles or oligomenorrhoeic. Menstrual cycle absent: amenorrhoea.	Methods FSH level was measured using recombinant immunoassay. Normal values were considered as < 10 mIU/mL Inhibin B was measured in duplicate using ELISA. Normal values were considered as ≥ 60 pg/mL AMH was measured using ELISA. Normal values were considered as ≥ 2 pmol/L Ovarian ultrasound was conducted with a 5MHz transvaginal probe or, whenever impossible, a transabdominal full bladder examination with a 3.5MHz probe. After localization of the ovaries, scanning was performed from the outer to the inner margin. Round or oval echo-free structures, ranging from 4 to 10mm in the ovaries were regarded as follicles and were counted and measured. The number of follicles in both ovaries was added to give the total antral follicle count. All transvaginal ultrasound measurements were performed by the same observer.	Results FSH level (cut-off not described, assumed ≥ 10 mIU/mL) Sensitivity, % (95% CI) 55 (24 to 84)[1] Specificity, % (95% CI) 85 (64 to 95)[1] Positive likelihood ratio (95% CI) 3.66 (1.11 to 12.12)[2] Negative likelihood ratio (95% CI) 0.53 (0.24 to 1.16)[2] Inhibin B level (cut-off not described, assumed < 60 pg/mL) Sensitivity, % (95% CI) 57 (24 to 84)[1] Specificity, % (95% CI) 77 (58 to 92)[1] Positive likelihood ratio (95% CI) 2.47 (0.92 to 6.65)[2] Negative likelihood ratio (95% CI) 0.56 (0.24 to 1.28)[2] AMH level (cut-off not described, assumed < 2 pmol/L) Sensitivity, % (95% CI) 73 (35 to 91)[1] Specificity, % (95% CI) 77 (58 to 92)[1]	Limitations Cut points for diagnostic tests not fully described. No cut point for AFC given, but thresholds for serum markers assumed to be when outside the normal range (reported in the article). No diagnostic testing for POI performed, ovarian reserve based on presence/absence of menstrual cycles alone. Other information

Study details	Participants	Tests	Methods	Outcomes and Results	Comments
				Positive likelihood ratio (95% CI) 3.17 (1.30 to 7.72)[2] Negative likelihood ratio (95% CI) 0.35 (0.11 to 1.12)[2] AFC (cut-off not described) Sensitivity, % (95% CI) 83 (47 to 97)[1] Specificity, % (95% CI) 74 (53 to 89)[1] Positive likelihood ratio (95% CI) 3.13 (1.44 to 6.86)[2] Negative likelihood ratio (95% CI) 0.23 (0.05 to 1.09)[2] FSH level + AMH level Sensitivity, % (95% CI) 55 (24 to 84)[1] Specificity, % (95% CI) 89 (70 to 97)[1] Positive likelihood ratio (95% CI) 4.91 (1.26 to 19.09)[2] Negative likelihood ratio (95% CI) 0.51 (0.23 to 1.11)[2] AFC + AMH level Sensitivity, % (95% CI) 83 (47 to 97)[1] Specificity, % (95% CI) 88 (70 to 97)[1] Positive likelihood ratio (95% CI) 7.03 (2.10 to 23.60)[2] Negative likelihood ratio (95% CI) 0.19 (0.04 to 0.90)[2] AFC + inhibin B level Sensitivity, % (95% CI) 83 (47 to 97)[1] Specificity, % (95% CI) 87	

Study details	Participants	Tests	Methods	Outcomes and Results	Comments
				(70 to 97)[1] Positive likelihood ratio (95% CI) 6.38 (2.02 to 20.16)[2] Negative likelihood ratio (95% CI) 0.20 (0.04 to 0.91)[2] [1] Point estimate provided, 95% CI calculated by the NCC-WCH technical team from data reported in the article. [2] Point estimate and 95% CI calculated by the NCC-WCH technical team from data reported in the article.	
Full citation Hagen,C.P., Aksglaede,L., Sorensen,K., Main,K.M., Boas,M., Cleemann,L., Holm,K., Gravholt,C.H., Andersson,A.M., Pedersen,A.T., Petersen,J.H., Linneberg,A., Kjaergaard,S., Juul,A., Serum levels of anti-Mullerian hormone as a marker of ovarian function in 926 healthy females from birth to adulthood and in 172 Turner syndrome patients, Journal of Clinical Endocrinology and Metabolism, 95, 5003-5010, 2010 Ref Id 267023 Country/ies where the study was carried out Denmark Source of funding Kirsten and Freddy Johansen Foundation. AMH kits were supplied by Beckman Coulter. Study dates Not reported. Study type Cross sectional study. Aim of the study	Sample size N = 67 • n = 53 Turner Syndrome with POI. • n = 14 Turner Syndrome with ongoing ovarian function. Characteristics Aged 12 to 25 years Inclusion criteria Diagnosis of Turner syndrome was confirmed by routine G-band karyotyping. All subjects had participated in one of three Danish cohort studies. Exclusion criteria Not reported.	Tests Serum AMH levels were determined using an enzyme immunometric assay, with a sensitivity of 2.0pmol/L. Definitions used POI: absent spontaneous puberty, or spontaneous puberty with cessation of ovarian function subsequently treated with estrogen due to lack of pubertal progression or secondary amenorrhoea. No POI: spontaneous puberty with ongoing ovarian function and ongoing pubertal progression or regular spontaneous menstrual bleeding.	Methods Non-fasting blood samples were drawn between 0800 and 1700 from an antecubital vein, clotted, centrifuged and serum was stored at -20°C until hormone analyses were performed. All samples were analysed after a maximum of 4 years of storage in the freezer at -20°C.	Results AMH level, cut-point of 8 pmol/L (to distinguish Turner Syndrome patients with POI from Turner Syndrome patients without POI): Sensitivity, % (95% CI): 96 (87 to 100)[1] Specificity, % (95% CI): 86 (57 to 98)[1] Positive likelihood ratio (95% CI): 6.74 (1.86 to 24.33)[2] Negative likelihood ratio (95% CI): 0.04 (0.01 to 0.17)[2] 1 Point estimate provided in the article. 95% CI calculated by the NCC-WCH technical team. 2 Point estimate and 95% CI calculated by the NCC-WCH technical team from data reported in the article.	Limitations Other information

Study details	Participants	Tests	Methods	Outcomes and Results	Comments
To determine normative data for circulating AMH levels in females, including longitudinal values in infancy. In addition, AMH levels in patients with Turner Syndrome are reported, according to their age, karyotype and ovarian function. Data used for this review considered whether AMH could be used in patients with Turners syndrome in order to distinguish those with POI from those with ongoing ovarian function.					

H.9.2 Management of premature ovarian insufficiency

Study details	Study design	Intervention	Results	Quality checklist	Other information
Full citation Langrish,J.P., Mills,N.L., Bath,L.E., Warner,P., Webb,D.J., Kelnar,C.J., Critchley,H.O., Newby,D.E., Wallace,W.H., Cardiovascular effects of physiological and standard sex steroid replacement regimens in premature ovarian failure, Hypertension, 53, 805-811, 2009 Ref Id 287559 Source of funding CLIC Sargent Wellcome Trust British Heart Foundation	Study type Open label, randomized, controlled cross-over trial. After an initial 2 month washout period, participants were randomized to the intervention or comparator treatment for a total of 12 months. This was followed by a further 2 month washout period before participants were switched to the alternative treatment for the final 12 months. Inclusion criteria Premature ovarian insufficiency attributed to chemotherapy or radiotherapy, idiopathic or surgical treatment of Turner syndrome. Diagnostic criteria for	Interventions HRT regimen ("Physiological sex steroid replacment"), comprising transdermal Estradiol 100µg daily for week one, and 150µg daily for weeks two to four (Estraderm TTS patches, Novartis Pharmaceuticals UK Ltd.). This was combined with 200mg progesterone pessaries twice daily in weeks three to four (Cyclogest, Actavis UK Ltd.). Some women used oral progesterone in preference to vaginal pessaries (dydrogesterone 10mg twice daily; Duphaston, Solvay Healthcare Ltd.). Comparator OCP regimen ("Standard hormone replacment") of ethinylestradiol 30µg and noresthisterone 1.5mg daily for weeks one to three, followed by seven "pill-free" days (Loestrin 30, Galen	Results Blood pressure and arterial stiffness At 12 months: Mean difference in systolic blood pressure (mmHg) on HRT (compared to OCP) = -7.3 (95% CI -2.5 to -12.0) Mean difference in diastolic blood pressure (mmHg) on HRT (compared to OCP) = -7.4 (95% CI -3.9 to -11.0) Statistically significant differences were seen at 3 ($P < 0.05$), 6 ($P < 0.05$) and 12 months ($P < 0.01$). There were no differences in carotid-radial pulse wave velocity or 24 hour mean heart rate through the study period. Renal and humoral factors HRT reduced plasma angiotensin II levels ($P = 0.007$) and serum creatinine concentration ($P = 0.015$) as compared with OCP. However, plasma renin activity, serum urea nitrogen, sodium, potassium and aldosterone concentrations were unchanged. Body Mass Index (BMI) There were no changes in BMI throughout the study. Discontinuation rate	A1 - An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups) Yes A2 - There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) Yes A3 - The groups were comparable at baseline, including all major confounding and	Other information All data on bone mineral density, bone markers and uterine indices obtained from secondary publications Crofton et al. 2010 and O'Donnell et al. 2012 (see excluded studies list for full citation). Limitations Participants for whom outcome data were not available are not described, therefore it is unclear whether there are any systematic differences between these women and those in whom data were obtained. Participants were aware of treatment allocation as this was an open label trial.

Study details	Study design	Intervention	Results	Quality checklist	Other information
Study dates February 2002 to November 2006 Country/ies where the study was carried out UK	POI were not described in the paper. Exclusion criteria Not reported. Method of blinding Open label study. Calculation of cardiovascular, renal and humoral measures was performed by investigators blind to treatment allocation. Investigators were blinded to treatment allocation until all bone outcome measurements were complete. The radiologist performing measurements of uterine volume, endometrial thickness and uterine blood flow was aware of the aetiology of POI for each patient, but was not aware of the treatment received. Randomization Equal 1:1 randomization was performed separately for each aetiology in balanced blocks of 10 by opaque multipart assignment "envelopes" produced at the Medical Statistics Unit, University of Edinburgh. Power calculation Not reported.	Ltd.). Sample size N = 42 3 withdrawals prior to washout period, 5 withdrawals during washout period. Therefore N = 34 randomized. n = 16 randomized to physiological treatment followed by standard treatment. n = 18 randomized to standard treatment followed by physiological treatment.	HRT: n = 9/16 during first treatment phase • 2 = patch reaction • 1 = patch reaction and migraine/hormonal symptoms • 1 = time off work and patch reaction • 1 = difficulty attending appointments and migraines • 1 = unable to attend • 1 = ovarian cyst needing intervention • 1 = IVF treatment • 1 = abdominal pain n = 1/13 during second treatment phase • 1 = blood pressure not controlled and stress of forthcoming cataract operation OCP: n = 5/18 during first treatment phase • 1 = personal reasons and coping with intervention • 1 = personal reasons and lack of childcare • 1 = could not attend appointments • 1 = migraine and wish less intervention • 1 = impossible to cannulate n = 0/6 during second treatment phase n = 1 during 2 month washout period between treatment phases (not coping with washout symptoms). Bone mineral density (Data all obtained from secondary publication in excluded studies list, Crofton et al. 2010) Mean difference in lumbar spine BMD z-score on HRT (compared to OCP) = +0.09 (95% CI -0.06 to +0.25) (P = 0.2) BMD measurement / HRT / OCP Lumbar spine BMD, g/cm² / +0.019* (+0.008 to +0.029) / +0.01 (-0.002 to +0.022) Lumbar spine BMD, z-score / +0.17* (+0.07 to +0.27) / +0.07 (-0.03 to +0.18) Femoral neck BMD, g/cm² / +0.012 (-0.007 to +0.030) / +0.011 (-0.005 to +0.027) Femoral neck BMD, z-score / +0.12 (-0.05 to +0.29) / +0.11 (-0.04 to +0.25)	prognostic factors Yes B1 - The comparison groups received the same care apart from the intervention(s) studied Yes B2 - Participants receiving care were kept 'blind' to treatment allocation No B3 - Individuals administering care were kept 'blind' to treatment allocation Unclear C1 - All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) Yes C2a - How many participants did not complete treatment in each group? 16 withdrawals occurred over the course of the study. 10 women discontinued treatment whilst taking HRT, and 5 women discontinued whilst taking OCP (1 withdrew during the 2 month washout period between treatments). C2b - The groups were comparable for treatment	Whether individuals administering care were kept blind to treatment is not clear, but investigators were reported as being blinded. Differences were noted between women who completed and those who withdrew from the study. Amongst women completing the study were more women with Turner syndrome, more women with prepubertal onset of premature ovarian insufficiency and more women randomised to oral contraceptive pill as first treatment. Due to the cross-over nature of the trial, participants who completed the trial contributed data to both the intervention and comparator arms. Follow up was for one year for the intervention and comparator treatments. Whether this is sufficient to detect longer term cardiovascular or bone density changes is unclear.

Study details	Study design	Intervention	Results	Quality checklist	Other information
			Total hip BMD, g/cm² -0.009 (-0.051 to +0.034) +0.005 (-0.007 to +0.017) Total hip BMD, z-score -0.04 (-0.16 to +0.08) +0.03 (-0.08 to +0.13) Data are expressed as mean (95% CI mean) * P < 0.01 versus baseline BMD. No statistically significant difference between the two treatments for any BMD outcomes. Bone ALP and PINP increased from baseline in response to HRT, but decreased in response to OCP. Responses at 3, 6 and 12 months were different between treatments in terms of percentage change versus postwashout baseline (bone ALP P < 0.001 at all time points, PINP P < 0.001, < 0.001 and 0.03, respectively). Responses were also different in terms of absolute values (bone ALP P ≤ 0.001 at all time points, PINP P < 0.001, < 0.001 and 0.006, respectively). Both treatments suppressed CrossLaps, although suppression was less pronounced for HRT than for OCP. Significant differences between the two treatments were noted at 3 months (P = 0.01 for percentage changes and for absolute values) and 6 months (P = 0.02 for percentage changes, P = 0.003 for absolute values) but not at 12 months. Uterine volume, endometrial thickness and blood flow (Data all obtained from secondary publication in excluded studies list, O'Donnell et al. 2012) n = 29 eligible participants (5 participants had previously undergone hysterectomy). n = 25 completed at least one assessment on treatment (continued to three month assessment for first treatment period) therefore contributed data to analysis of treatment effect. n = 17 completed full 28 months study period. Endometrial thickness: Mean difference of +1.8mm (95% CI +0.7 to +2.8mm) when treated with HRT as compared with OCP (p = 0.002). Uterine volume: Mean difference of +4.2cm³ (95% CI -0.4 to +8.7cm³) when treated with HRT as compared with OCP (p = 0.07).	completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) No C3a - For how many participants in each group were no outcome data available? Data were available for 25 participants for uterine indices (although only 17 completed the full treatment period), 17 participants for blood pressure readings, 13 participants for renal and humoral measurements and 18 participants for bone mineral density and bone marker measurements. However, due to the cross-over nature of the trial all women will contribute data to both treatment arms. Data on discontinuation were available for all participants, and reported for all participants who commenced treatment. C3b - The groups were comparable	

Study details	Study design	Intervention	Results	Quality checklist	Other information
			Uterine artery resistance index: Mean difference of -0.01 (95% CI -0.03 to +0.01) when treated with HRT as compared with OCP (p = 0.39). Uterine artery pulsatility index: Mean difference of -0.20 (95% CI -0.56 to +0.17) when treated with HRT as compared with OCP (p = 0.27)	with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available). Unclear D1 - The study had an appropriate length of follow-up Unclear D2 - The study used a precise definition of outcome Yes D3 - A valid and reliable method was used to determine the outcome Yes D4 - Investigators were kept 'blind' to participants' exposure to the intervention Yes D5 - Investigators were kept 'blind' to other important confounding and prognostic factors Unclear	
Full citation Guttmann,H., Weiner,Z., Nikolski,E., Ish-Shalom,S., Itskovitz-Eldor,J., Aviram,M., Reisner,S., Hochberg,Z., Choosing an	Study type Randomised controlled trial with crossover design. Inclusion criteria Women with Turner Syndrome who were otherwise healthy. Exclusion criteria BMI > 30kg/m².	Interventions Each participant undertook a 4-6 month washout period of no treatment at the start of the trial. This was followed by 6 months of treatment with one study regimen, then 6 months of treatment with the other. Sequential conjugated	Results Outcome / HRT / OCP / Significance Fasting glucose (mmol/l) 4.1 ± 0.3 4.1 ± 0.5 NS Insulin (nmol/l) 61 ± 40 66 ± 20 NS Triglyceride (mmol/l) 1.45 ± 0.55 1.55 ± 0.65 NS Cholesterol (mmol/l) 4.53 ± 0.93 4.81 ± 0.93 P < 0.05	A1 - An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across	Other information Limitations Study was not blinded. Small sample size. No washout period was conducted between trial interventions, and no analysis was conducted to assess

Study details	Study design	Intervention	Results				Quality checklist	Other information
oestrogen replacement therapy in young adult women with Turner syndrome, Clinical Endocrinology, 54, 159-164, 2001 Ref Id 301721 Source of funding Not reported. Study dates Not reported. Country/ies where the study was carried out Israel	Method of blinding Unblinded study. Randomization Method not described. Power calculation Not reported.	oestrogen (0.625mg) was given for 14 days, followed by conjugated oestrogen (0.625mg) and medroxyprogesterone acetate (5mg) for the following 14 days (Premaril Plus MP®, Dexxon). Treatment duration was 6 months. Comparator Ethinyloestradiol 30µg plus gestodene 75µg was given for 6 months. Sample size N = 17.	HDL cholesterol (mmol/l)	1.19 ± 0.65	1.16 ± 0.57	NS	groups) Unclear A2 - There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation) Unclear A3 - The groups were comparable at baseline, including all major confounding and prognostic factors Yes B1 - The comparison groups received the same care apart from the intervention(s) studied Yes B2 - Participants receiving care were kept 'blind' to treatment allocation No B3 - Individuals administering care were kept 'blind' to treatment allocation Unclear C1 - All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) Yes C2a - How many participants did not	any treatment order effect.
			LDL cholesterol (mmol/l)	2.40 ± 1.06	2.95 ± 0.94	NS		
			ALP (U/l)	127 ± 41	92 ± 29	P < 0.0005		
			25OHD (µg/l)	16 ± 12	20 ± 14	NS		
			1,25(OH)2D3 (ng/l)	38 ± 14	41 ± 12	NS		
			Osteocalcin (µg/l)	13.6 ± 4.6	9.1 ± 3.3	NS		
			Deoxypyridinoline (µmol/mol Cr)	12.6 ± 3.9	11.2 ± 5.9	NS		
			Endometrial thickness (mm)	4.0 ± 0.6	3.7 ± 0.5	NS		
			Uterine pulsatility index*	2.6 ± 1.0	2.6 ± 1.2	NS		
			Data shown represents mean value ± standard deviation. Significance reflects comparison of the two treatment arms. * Described as resistance index in article, but methods specify calculation of pulsatility index.					

Study details	Study design	Intervention	Results	Quality checklist	Other information
				complete treatment in each group? None. C2b - The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment) Yes C3a - For how many participants in each group were no outcome data available? None. C3b - The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available). Not applicable D1 - The study had an appropriate length of follow-up Yes D2 - The study used a precise definition of outcome Yes D3 - A valid and reliable method was used to determine the outcome	

Study details	Study design	Intervention	Results	Quality checklist	Other information
				Yes D4 - Investigators were kept 'blind' to participants' exposure to the intervention No D5 - Investigators were kept 'blind' to other important confounding and prognostic factors Unclear	

H.10 Economic evidence

Study	Limitations	Applicability	Other comments	Incremental			Uncertainty
				Costs	Effects	ICER	
Botteman 2004	Transition probabilities for vasomotor symptoms derived from a trial with a small sample size Did not account for long-term clinical or economic aspects	Partially applicable (US study)	Study used a Markov decision-analytic model with a 1-year time horizon Research sponsored in part by Pfizer	NA/EE vs no therapy $680.84 CEE/MPA vs no therapy $847.93	NA/EE vs no therapy 0.110 QALYs CEE/MPA vs no 0.104 QALYs	NA/EE dominates CEE/MPA NA/EE vs no therapy $6,200 per QALY CEE/MPA v no therapy $8,200 per QALY	Univariate, bivariate, threshold and probabilistic sensitivity analysis
Brown 2006	Hot flushes used as proxy for presence and severity of postmenopausal symptoms	Partially applicable (Canadian study)	Study employed a Markov decision-analytic model with a 5-year time horizon	Patch vs oral $296 Patch vs no therapy $654-665	Patch vs oral 0.00 QALYs Patch vs no therapy 0.02-0.08 QALYs	• Oral dominates patch • Patch compared to no therapy for moderate	One-way and probabilistic sensitivity analysis undertaken

Study	Limitations	Applicability	Other comments	Incremental			Uncertainty
				Costs	Effects	ICER	
						($32,300 per QALY) and severe ($8,300 per QALY)	
Coyle 2003	Hot flushes used as proxy for menopausal symptoms No probabilistic sensitivity analysis conducted	Partially applicable (Canadian study)	Study employed a Markov decision-analytic model with a 5-year time horizon Study funded by Pfizer inc.	NA/EE vs CEE/MPA $600-400 NA/EE vs no therapy $700-400	NA/EE vs CEE/MPA 0.02-0.03 QALYs NA/EE vs no therapy 0.33-0.39 QALYs	• NA/EE vs CEE/MPA • 1st line: $20,300 per QALY • 2nd line: $16,400 per QALY	One-way and threshold sensitivity analysis undertaken
Lekander 2009[a]	No comparison with alternative treatment No probabilistic sensitivity analysis conducted	Directly applicable (UK study)	Study employed a Markov decision analytic model with a lifetime horizon Study funded and conducted by consultants for Wyeth	HRT vs No therapy £252-£677	HRT vs No therapy 1.17-1.23 QALYs	HRT v no therapy £205-£580 per QALY	Univariate and threshold sensitivity analysis undertaken
Lekander 2009[b]	No comparison with alternative treatment • No probabilistic sensitivity analysis conducted • Study conducted from a societal perspective	Partially applicable (US study)	Study employed a Markov decision analytic model with a lifetime horizon Study funded and conducted by consultants for Wyeth	HRT vs No therapy $358-$3224	HRT vs No therapy 1.15-1.21 QALYs	HRT v no therapy $295-$2803 per QALY	Univariate and threshold sensitivity analysis undertaken

Study	Limitations	Applicability	Other comments	Incremental			Uncertainty
				Costs	Effects	ICER	
Swift 2005	Model structure and type presented unclearly. Utilities on menopausal symptom severity only included	Directly applicable (UK study)	Study developed an economic model over a one-year time horizon Study funded and conducted by consultants for Wyeth	Low-dose vs high dose CE/MPA • -£1,443	Low-dose vs high dose CE/MPA 0.62-1.49 QALYs	Low dose dominates high dose CE/MPA	Probabilistic sensitivity analysis undertaken
Yilkangas 2007	No probabilistic sensitivity analysis conducted	Partially applicable (Finnish study	Study conducted a trial-based economic evaluation over a 9-year time horizon Study was funded by Orion Pharma	ccHRT vs gen population €101	ccHRT vs gen population 0.022 QALYs	• ccHRT vs gen population • €4613 per QALY	Univariate sensitivity analysis undertaken
Zethraeus 2005	Study conducted from a societal perspective No probabilistic sensitivity analysis undertaken	Partially applicable (Swedish study)	Study employed a Markov decision analytic model with a lifetime horizon Funding for this study was provided by Wyeth Lederle	Intact uterus HRT vs No HRT SEK 15,242 Hysterectomised HRT vs No HRT SEK 10,107	Intact uterus HRT vs No HRT 1.19 QALYs Hysterectomised HRT vs No HRT 1.22 QALYs	Intact uterus HRT vs No HRT SEK 12,807 per QALY Hysterectomised HRT vs No HRT SEK 8,266 per QALY	Univariate sensitivity analysis undertaken
Diaby 2007	Assumptions made concerning utility of reduction of symptoms No probabilistic sensitivity	Partially applicable (Canadian study)	Study employed a Markov decision-analytic model with a 3-year time horizon	Tibolone (2.5mg) vs ccHRT (CEE/MPA 0.625/2.5mg) $253	Tibolone (2.5mg) vs ccHRT (CEE/MPA 0.625/2.5mg) 0.03 QALYs	Tibolone (2.5mg) vs ccHRT (CEE/MPA 0.625/2.5mg) $9,198	Univariate and bivariate sensitivity analysis undertaken

Study	Limitations	Applicability	Other comments	Incremental			Uncertainty
				Costs	Effects	ICER	
	analysis						