National Institute for Health and Care Excellence

Draft

Heavy menstrual bleeding (update)

Evidence reviews for diagnostic test accuracy in investigation for women presenting with heavy menstrual bleeding

NICE guideline TBC

Evidence reviews

August 2017

Draft for Consultation

These evidence reviews were developed by National Guideline Alliance, hosted by the Royal College of

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN:

Contents

Abbreviations

Abbreviation	Definition
2D	Two-dimensional
2D-TAUS	Two-dimensional transabdominal ultrasound scan
2D-TVUS	Two-dimensional transvaginal ultrasound scan
3D	Three-dimensional
3D-TAUS	Three-dimensional transabdominal ultrasound scan
3D-TVUS	Three-dimensional transvaginal ultrasound scan
AUC	Area under the curve
BSGE	British Society for Gynaecological Endoscopy
CI	Confidence interval
D&C	Dilatation and curettage
EBx	Endometrial biopsy
EQ-5D	EuroQol five dimensions questionnaire
FN	False negative
FP	False positive
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HMB	Heavy menstrual bleeding
HRQoL	Health-related quality of life
HSROC	Hierarchical summary receiver operating characteristic
LNG-IUS	Levonorgestrel-releasing intrauterine system
LR+	Positive likelihood ratio
LR-	Negative likelihood ratio
MRI	Magnetic resonance imaging
N/A	Not applicable
NGA	National Guideline Alliance
NHS	National Health Service
NICE	National Institute of Health and Care Excellence
NMA	Network meta-analysis
NMB	Net monetary benefit
OPH	Outpatient hysteroscopy
PSA	Probabilistic sensitivity analysis
QALY	Quality-adjusted life year
QUADAS-2	A quality assessment tool for diagnostic accuracy studies
RCOG	Royal College of Obstetricians and Gynaecology
RCT	Randomised controlled trial
RPOC	Retained products of conception
Sens	Sensitivity
SIS	Saline infusion sonography
Spec	Specificity
TAUS	Transabdominal ultrasound scan
TN	True negative
TP	True positive
TVUS	Transvaginal ultrasound scan

1

2

1 Diagnosis of heavy menstrual bleeding

2 Review questions

3 Review question 1. What is the diagnostic accuracy of ultrasound and hysteroscopy for
4 investigation of women presenting with heavy menstrual bleeding?

5 Review question 2. What is the most clinically effective imaging strategy for diagnosing
6 adenomyosis in women with heavy menstrual bleeding?

7 Introduction

8 Many women presenting to primary care with heavy menstrual bleeding (HMB) will only ever
9 require simple treatment without the need for further investigations. However, some women
10 with HMB may have an underlying pathology, which is often not apparent from the woman's
11 history or examination and further investigations to identify the cause may be needed. The
12 aim of investigation is to identify structural abnormalities such as submucosal fibroids,
13 endometrial polyps, larger fibroids, or adenomyosis, and to also detect rare and potentially
14 serious pathologies such as endometrial hyperplasia or carcinoma.

15 The identification of structural, focal or histological abnormalities enables targeted treatment,
16 which should result in more effective management. Whilst the cause of HMB will not always
17 be apparent from the investigation, women may also gain reassurance in knowing that no
18 serious pathology is causing their symptoms, and that no focal treatable lesion was missed
19 that make simple treatments less effective.

20 Review question 1. What is the diagnostic accuracy of 21 ultrasound and hysteroscopy for investigation of women 22 presenting with heavy menstrual bleeding?

23 Introduction

24 Ultrasound scan has been the technique most commonly used to investigate causes of
25 heavy menstrual bleeding. Over time, diagnostic hysteroscopy techniques have improved,
26 which has also enabled the development of 'see-and-treat' services. Furthermore, new
27 evidence is available on the clinical and cost-effectiveness of different diagnostic techniques
28 in the investigation of women presenting with heavy menstrual bleeding.

29 The aim of this review was to determine the diagnostic accuracy of hysteroscopy and
30 ultrasound in investigation of women presenting with heavy menstrual bleeding, also taking
31 into account patient acceptability and satisfaction as well as the cost-effectiveness of these
32 approaches.

33 Summary of the protocol

34 Please see Table 1 for a summary of the characteristics of this review including population,
35 index tests, reference standard, outcomes and target conditions.

36 **Table 1: Summary of the protocol**

Population	Premenopausal women with heavy menstrual bleeding (HMB). At least 66% of the population should be women with heavy menstrual bleeding.

Index test(s)	• Transvaginal ultrasound scan (TVUS)
	  ○ Two-dimensional (2D)
	  ○ Three-dimensional (3D)
	• Transabdominal ultrasound scan (TAUS)
	  ○ 2D
	  ○ 3D
	• Hysteroscopy
	  ○ Inpatient/day case (under general anaesthesia/sedation or spinal/epidural anaesthesia)
	  ○ Outpatient vaginoscopy[1]
	  ○ Outpatient all other cases
	• Transvaginal ultrasound followed by hysteroscopy if needed

Reference standard	• Histopathology • Ultrasound scan (when compared with hysteroscopy) • Inpatient hysteroscopy (when compared with outpatient hysteroscopy)
Outcomes	• Sensitivity (sens) • Specificity (spec) • Positive likelihood ratio (LR+) • Negative likelihood ratio (LR-) • Area under the curve (AUC) if meta-analysis can be conducted • Patient satisfaction and acceptability
Target condition	Examples: Sub-mucous fibroids, polyps, hyperplasia, cancer, abnormal uterine thickness, synechiae, endometritis, retained products of conception (RPOC), fibroids less than 3 cm in diameter, fibroids larger than 3 cm in diameter, congenital abnormalities.

2D: two-dimensional; 3D: three-dimensional; AUC: area under the curve; HMB: heavy menstrual bleeding; LR+: positive likelihood ratio; LR-: negative likelihood ratio; RPOC: retained products of conception; sens: sensitivity; spec: specificity; TAUS: transabdominal ultrasound scan; TVUS: transvaginal ultrasound scan
1 Vaginoscopy approach to hysteroscopy is a technique where the hysteroscope is inserted into the vagina and through the cervical canal and into the uterine cavity without a vaginal speculum or cervical instrumentation.

For full details see review protocol in Appendix A – Review protocols.

Clinical evidence

Histopathology was considered as the reference standard to which the diagnostic accuracy of ultrasound and hysteroscopy were compared. Histopathology was derived from surgery specimen from hysterectomy, hysteroscopy-guided biopsy, or dilatation and curettage (D&C). No evidence was found comparing the diagnostic accuracy of hysteroscopy to ultrasound as a reference standards, however, studies that compared the diagnostic accuracy of ultrasound to hysteroscopy as the reference standard were considered instead.

The target conditions in which the diagnostic accuracy of these tests were studied included endometrial polyps, fibroids, endometrial hyperplasia, endometrial carcinoma, and any abnormal finding.

A combined literature search was undertaken for this review question together with the review question on the diagnostic accuracy of imaging techniques in detecting adenomyosis in women presenting with heavy menstrual bleeding.

In addition, relevant studies included in NICE guideline on Heavy menstrual bleeding: assessment and management (CG44) published in 2007 were assessed and included if they fitted the inclusion criteria in the current protocol.

Due to the limited evidence available when the population was restricted only to women presenting with heavy menstrual bleeding, the committee agreed that one third of the population could be women without heavy menstrual bleeding and these results could be extrapolated to women presenting with heavy menstrual bleeding as long as the indirectness of the population was accounted for when assessing the quality of the evidence. Since most studies did not report the exact proportion of women presenting with heavy menstrual bleeding, studies which included premenopausal women with abnormal uterine bleeding were still included with the assumption that heavy menstrual bleeding is likely to be a common presentation of abnormal uterine bleeding in these women. Again, the indirectness of the population was accounted for when assessing the quality of evidence and the confidence in the evidence using Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology.

No evidence was found on the diagnostic accuracy of three-dimensional transvaginal ultrasound scan (3D-TVUS) or transabdominal ultrasound scan (TAUS).

Meta-analysis was conducted when appropriate with a minimum of 5 studies. In the presence of heterogeneity, potential reasons for heterogeneity were explored and subgroup analyses were conducted when possible according to pre-specified groups in the protocol.

Test sensitivity or specificity was considered high when sensitivity or specificity was 90% or higher, and moderate when sensitivity or specificity was between 75% and 89%. A positive likelihood ratio (LR+) of 10 or higher was considered to indicate that the test was very useful in 'ruling in' the condition, and a LR+ between 5 and 9.9 was considered moderately useful in 'ruling in' the condition. A negative likelihood ratio (LR-) of 0.1 or lower was considered to indicate that the test was very useful in 'ruling out' the condition, and a LR- between 0.1 and 0.2 was considered to indicate that the test was moderate useful in 'ruling out' the condition.

Included studies

Nineteen studies (n=3501) were included in the review (Abd Elkhalek 2016; Abe 2008; Alborzi 2007; Cicinelli 1995; Critchley 2004; Dasgupta 2011a; Dasgupta 2011b; Dueholm 2001a; Erdem 2007; Fakhar and Mahmud 2010; Krampl 2001; Mukhopadhayay 2007; Najeeb 2010; Nanda 2002; Soguktas 2012; Taylor 2001; Vercellini 1997; Williams and Marshburn 1998; Yildiz 2009). Of these 1 study investigated patient acceptability and satisfaction of ultrasound and hysteroscopy (Critchley 2004), the other studies investigated the diagnostic accuracy of these tests.

Fifteen studies provided evidence on the diagnostic accuracy of two-dimensional transvaginal ultrasound scan (2D-TVUS) (Abe 2008; Alborzi 2007; Cicinelli 1995; Dasgupta 2011a; Dasgupta 2011b; Dueholm 2001a; Erdem 2007; Krampl 2001; Mukhopadhayay 2007; Najeeb 2010; Nanda 2002; Soguktas 2012; Taylor 2001; Vercellini 1997; Williams and Marshburn 1998). Six studies provided evidence on the diagnostic accuracy of hysteroscopy (Abd Elkhalek 2016; Cicinelli 1995; Fakhar and Mahmud, 2010; Mukhopadhayay 2007; Soguktas 2012; Yildiz 2009).

The studies came from the following countries: Denmark (Dueholm 2001a); Egypt (AbdElkhalek 2016); India (Dasgupta 2011a; Dasgupta 2011b; Mukhopadhayay 2007; Nanda 2002); Iran (Alborzi 2007); Italy (Cicinelli 1995; Vercellini 1997); Japan (Abe 2008); Norway (Krampl 2001); Pakistan (Fakhar and Mahmud 2010; Najeeb 2010); Turkey (Erdem 2007; Soguktas 2012; Yildiz 2009); UK (Critchley 2004); and US (Williams and Marshburn 1998).

Evidence from these are presented in Appendix G – GRADE tables.

See also Appendix D – Clinical evidence study selection, Appendix H – Forest plots and hierarchical summary receiver operating characteristic (HSROC) plots, and Appendix F – Clinical evidence tables.

1 Excluded studies

2 Studies not included in this review with reasons for their exclusions are provided in Appendix
3 I – Excluded studies.

4 Summary of clinical studies included in the evidence review

5 Table 2 and Table 3 provide a brief summary of the included studies.

6 **Table 2: Summary of included studies on diagnostic accuracy**

Study	Population	Index test	Reference standard	Target condition (prevalence)
Abd Elkhalek 2016 (Egypt)	N=50 patients with abnormal uterine bleeding and normal endometrial lining on 2D-TVUS, age range 25-45 years Proportion with HMB not reported. Exclusions: bleeding secondary to obvious pelvic infection; cervical and adnexal pathologies	Hysteroscopy (under general anaesthetic) following 2D-TVUS	Histopathology (hysteroscopy-guided biopsy)	• Polyp or fibroid (64%)
Abe 2008 (Japan)	N=213 premenopausal women with abnormal uterine bleeding, less than 50 years of age Proportion with HMB not reported. Exclusions: presence of cervical polyps or neoplasm; hormone replacement therapy; cases with TVUS with indeterminate results	2D-TVUS	Histopathology (vacuum biopsy)	• Any abnormal finding (69%)
Alborzi 2007 (Iran)	N=81 women with abnormal uterine bleeding Proportion with HMB not reported. Exclusions: none specified	2D-TVUS	Histopathology (hysteroscopy-guided biopsy)	• Polyp (40%) • Fibroid (28%)
Cicinelli 1995 (Italy)	N=52 premenopausal women who	• 2D-TVUS • Hysteroscopy (outpatient)	Histopathology (hysterectomy)	• Polyp (2%) • Fibroid (19%)

Study	Population	Index test	Reference standard	Target condition (prevalence)
	underwent hysterectomy 67% with HMB Exclusions: none specified			
Dasgupta 2011a (India)	N=274 women with abnormal uterine bleeding for at least 3 months, age range 40-50 years of which N=252 included in final analysis Proportion with HMB not reported. Exclusions: uterus larger than that at 12 weeks' gestation; hormone therapy within the last 6 months; previous abnormal endometrial biopsy; positive pregnancy test; cervical pathology on speculum examination; abnormal cervical Pap smear; active pelvic infection	2D-TVUS	Histopathology (hysteroscopy-guided biopsy)	• Polyp (12%) • Submucosal fibroid (18%)
Dasgupta 2011b (India)	N=100 perimenopausal (40-55 years old) women with abnormal uterine bleeding who had been on oral progesterone therapy for at least 10 days of which n=83 were included in final analysis Proportion with HMB not reported. Exclusions: uterus larger than that at 12 weeks' gestation; previous abnormal endometrial biopsy; cervical lesion on	2D-TVUS	Histopathology (hysteroscopy-guided biopsy)	• Any abnormal finding (65%) • Polyp (13%) • Fibroid (16%)

Study	Population	Index test	Reference standard	Target condition (prevalence)
	speculum examination; abnormal Pap smear; active pelvic infection; adnexal mass on clinical examination or during ultrasound scan; positive pregnancy test			
Dueholm 2001a (Denmark)	N=452 women less than 55 years of age with abnormal uterine bleeding Proportion with HMB not reported. Exclusions: hormone replacement therapy and indefinite menopausal status with duration of hormone replacement therapy less than 3 years	2D-TVUS	Histopathology (hysteroscopy-guided biopsy or hysterectomy)	• Polyp or fibroid (62%)
Erdem 2007 (Turkey)	N=133 women with abnormal uterine bleeding of which n=122 in final analysis 78% premenopausal, 22% postmenopausal Proportion with HMB not reported. Exclusions: bleeding due to pregnancy or pelvic infections; premenopausal women who had received hormonal therapy	2D-TVUS	Histopathology (hysterectomy, hysteroscopy-guided biopsy, or D&C)	• Polyp (50%) • Submucosal fibroid (16%) • Abnormal endometrial thickness/endometrial hyperplasia (3%)
Fakhar and Mahmud 2010 (Pakistan)	N=290 women 35 years old or more with abnormal uterine bleeding of which n=223 included in final analysis 66.2% with HMB.	Hysteroscopy (mostly outpatient)	Histopathology (hysteroscopy-guided biopsy)	• Adenocarcinoma (1%) • Retained products of conception (2%) • Polyp (9%) • Hyperplasia (12%) • Endometritis (20%)

Study	Population	Index test	Reference standard	Target condition (prevalence)
	Exclusions: refusal of procedure; incomplete follow-up; positive pregnancy test; recent cervicitis, vaginitis, endometritis, pelvic infection, or uterine perforation; non-availability of histopathology reports			
Krampl 2001 (Norway)	N=100 women with abnormal uterine bleeding 89% premenopausal Proportion with HMB not reported. Exclusions: an endometrial biopsy within the past 1 year; large multiple fibroid causing discomfort; patients considered medically unfit for general or spinal anaesthesia	• 2D-TVUS • Hysteroscopy (under spinal or general anaesthetic)	Histopathology (hysteroscopy-guided biopsy)	• Abnormal endometrial thickness (double-layer endometrium 12 mm or thicker and single-layer endometrium 6 mm or thicker) (10%) • Focal pathology (polyp or fibroid) (24%)
Mukhopadhyay 2007 (India)	N=85 perimenopausal women aged 40-55 years with abnormal uterine bleeding Proportion with HMB not reported. Exclusions: active bleeding per vagina; atrophic vaginitis; carcinoma cervix; cervical polyp; bleeding following trauma; varicose vein	• 2D-TVUS • Hysteroscopy (under general anaesthetic)	Histopathology (hysteroscopy-guided biopsy or D&C)	• Polyp (17%) • Hyperplasia (17%)
Najeeb 2010 (Pakistan)	N=141 perimenopausal women with abnormal uterine bleeding aged 40-47 years	2D-TVUS	Histopathology (D&C)	• Polyp (23%) • Fibroid (4%)

Study	Population	Index test	Reference standard	Target condition (prevalence)
	Proportion with HMB not reported. Exclusions: any form of hormonal treatment; known gynaecological malignancy; endocrinological disorder			
Nanda 2002 (India)	N=50 women with abnormal uterine bleeding, aged 30-50 years, undergoing hysterectomy Proportion with HMB not reported. Exclusions: none specified	2D-TVUS	Histopathology (hysterectomy)	• Submucosal fibroid (38%) • Polyp (6%)
Soguktas 2012 (Turkey)	N=93 premenopausal women with abnormal uterine bleeding related to intracavitary pathology of which n=89 were included in final analysis Proportion with HMB not reported. Exclusions: pelvic infection; pregnancy; abnormal uterine bleeding without intracavitary pathology	• 2D-TVUS • Hysteroscopy (under general anaesthetic)	Histopathology (hysteroscopy-guided biopsy or D&C)	• Polyp (38%) • Submucosal fibroid (5%) • Hyperplasia (8%) • Endometrial carcinoma (2%) • Any abnormal finding (53%)
Taylor 2001 (UK)	N=264 premenopausal women with abnormal uterine bleeding Proportion with HMB not reported. Exclusions: none specified	2D-TVUS	Hysteroscopy or hysteroscopy-guided biopsy	• Polyp (17%) • Suspicious focal thickening (3%)
Vercellini 1997 (Italy)	N=793 women with HMB Exclusions: women with IUD; hormonal treatment in the last 3 months (6 months for GnRh); women	2D-TVUS	Hysteroscopy	• Any abnormal finding (58%)

Study	Population	Index test	Reference standard	Target condition (prevalence)
	who had already undergone D&C or diagnostic or operative hysteroscopy			
Williams and Marshburn 1998 (US)	N=47 women with abnormal uterine bleeding that had not responded to medical treatment of which n=39 included in final analysis 92% premenopausal Proportion with HMB not reported. Exclusions: inability to undergo endovaginal ultrasonography; refusal to undergo hysteroscopy; interval pregnancy; suspected current cervical, uterine or tubal infection; suspected anovulatory (dysfunctional) bleeding; active menstrual bleeding	2D-TVUS	Hysteroscopy or histopathology from hysterectomy	• Any abnormal finding (31%)
Yildiz 2009 (Turkey)	N=86 women with abnormal uterine bleeding 84% premenopausal 65.1% of total sample with menometrorrhagia Exclusions: genital malignancy; pregnancy	Hysteroscopy (under general, local or no anaesthetic)	Histopathology (hysteroscopy-guided biopsy or hysterectomy)	• Any abnormal finding (81%)

2D-TVUS: two-dimensional transvaginal ultrasound scan; 3D-TVUS: three-dimensional transvaginal ultrasound scan; D&C: dilatation and curettage; GA: general anaesthesia; HMB: heavy menstrual bleeding;

Table 3: Summary of included studies on patient satisfaction and acceptability

Study	Population	Test(s)	Outcomes	Comments
Critchley 2004 (UK)	n=326 premenopausal women aged 40 years or more, or	• Blind endometrial biopsy	• Found procedure	• Unblinded trial (due to the nature of investigations)

Study	Population	Test(s)	Outcomes	Comments
	aged less than 40 years with risk factors, 68% with HMB (Total N=683 women with abnormal uterine bleeding in three groups based on level of risk of endometrial cancer: High risk group: n=200 postmenopausal women (not of interest to this review) Moderate risk group: n=326 premenopausal women aged 40 years or more, or aged less than 40 years with risk factors, 68% with HMB Low risk group: n=157 premenopausal women aged less than 40 years without risk factors, 57% with HMB (not of interest to this review))	• Hysteroscopy • TVUS or TAUS	'markedly unpleasant' • Abdominal discomfort • Experiences about the clinic visit and health 1 day, 10 months, and 24 months after investigation	• Only descriptive reporting of evidence on patient satisfaction/acceptability • All women underwent biopsy

HMB: heavy menstrual bleeding; TAUS: transabdominal ultrasound scan; TVUS: transvaginal ultrasound scan

See Appendix F – Clinical evidence tables for full evidence tables.

Quality assessment of clinical studies included in the evidence review

Studies providing evidence on the diagnostic accuracy of ultrasound and hysteroscopy in investigating women presenting with heavy menstrual bleeding were individually assessed using the QUADAS-2 checklist. QUADAS-2 assesses the risk of bias in the study and the applicability of the study for the review question.

The quality of each piece of evidence, whether from meta-analysis or from individual studies was then assessed using adapted GRADE methodology. See Appendix G – GRADE tables for full GRADE tables.

The quality of the study providing evidence on the patient acceptability and satisfaction of ultrasound and hysteroscopy was assessed using the Cochrane Collaboration's tool for assessing risk of bias. The tool assesses the risk of selection bias, performance bias, detection bias, attrition bias, reporting bias, and other bias. The overall quality of the evidence was based on the assessment of all of these.

For more details on the quality assessment of clinical studies in the evidence review, please see Methods chapter.

Economic evidence

Included studies

One cost-effectiveness analysis (Cooper 2014) was found for inclusion in this review, see Table 4. A narrative review of this study is provided in the Health cconomics chapter.

Excluded studies

There were no excluded economic studies relating to the diagnosis of heavy menstrual bleeding.

Summary of studies included in the economic evidence review

Table 4: Summary of studies included in the economic evidence review

Study	Limitations	Applicability	Other comments	Costs	Effects	Inc. cost-effectiveness	Uncertainty
Cooper 2014 (UK) Population: Women with HMB, a mean age of 45 years and referred from primary care Interventions: 1. LNG-IUS alone 2. hysterectomy alone 3. OPH 4. TVS 5. EBx 6. SIS 7. OPH and EBx 8. TVS and EBx 9. SIS and EBx 10. OPH and SIS 11. OPH and TVS 12. SIS, OPH and EBx	Potentially serious [a, b]	Directly [c]	Decision trees developed for 13 investigation strategies.. Estimated that a woman's QoL is reduced by 0.5 for the 1 week per month of heavy menses ~ HMB is associated with a reduction of 0.125 QALYs in any year. Analyses on PMB also undertaken. Parameter inputs were derived from systematic reviews, individual patient data and focused searches and in the absence of estimates, the consensus view of an expert clinical panel was obtained.	Total cost per patient over 1 year: LNG-IUS alone £1067 OPH alone £1078 SIS alone £1083 TVUS alone £1085 TVUS+OPH £1139 OPH+EBx £1149 SIS+OPH £1170 EBx alone £1209 SIS+EBx £1223 TVUS+OPH+EBx £1227 TVUS+EBx £1231 SIS+OPH+EBx £1256 Hysterectomy alone £3182	Satisfaction (0 to 100% scale) at 1 year: LNG-IUS alone 0.933327 OPH alone 0.964122 SIS alone 0.962914 TVUS alone 0.955106 TVUS+OPH 0.964382 OPH+EBx 0.967421 SIS+OPH 0.96445 EBx alone 0.945963 SIS+EBx 0.964271 TVUS+OPH+EBx 0.964933 TVUS+EBx 0.953851 SIS+OPH+EBx 0.965028 Hysterectomy alone 0.9335	OPH alone and OPH+EBx remain non-dominated by alternative empirical treatment or diagnostic testing strategies. OPH alone dominates the testing strategies SIS alone and TVUS alone. The combination testing strategy TVUS + OPH is excluded by extended dominance between OPH alone and OPH + EBx. The remaining 7 alternative strategies are dominated by OPH + EBx. ICER (to gain an extra woman satisfied following treatment for HMB): The most effective strategy is the OPH+EBx, generating an ICER of £21,500 compared with OPH alone and an ICER of £21,859 when compared with LNG-IUS alone. OPH alone is less effective than OPH+EBx, but is less costly with an ICER of £360. Estimate that an ICER of £21,000 per case satisfied is approximately equivalent to an ICER of £26,500 per QALY, which falls within the £20,000–30,000 per QALY threshold range. Hence, OPH+EBx maybe cost-effective compared to OPH alone.	In PSA OPH remained more cost-effective than the LNG-IUS even at relatively low WTP thresholds. OPH+EBx also remained stable, to be at least 70% certain that it was a more cost-effective alternative than OPH alone, the WTP threshold would need to be £40,000 per patient satisfied (~£50,000 per QALY). Deterministic SA and additional PSA also performed [d].

Study	Limitations	Applicability	Other comments	Costs	Effects	Inc. cost-effectiveness	Uncertainty
13. TVS, OPH and EBx.							

EBx: endometrial biopsy; OPH: outpatient hysterectomy; SIS: saline infusion sonography; TVUS: transvaginal ultrasound scan; PMB: postmenopausal bleeding; PSA: probabilistic sensitivity analysis; QALY: quality-adjusted life year; WTP: willingness-to-pay

(a) 1-year time horizon may not be sufficiently long to reflect all important differences in costs and outcomes

(b) Unclear if all important outcomes have been captured using 'satisfaction' as the measure of effect

(c) This study does not use the preferred measure of effects (QALYs) as the primary measure of effects, but is still thought to be useful for decision making given that all other criteria are applicable and an approximation to QALYs is provided

(d) Sensitivity analysis (non-dominated strategies): women being managed during multiple clinic appointments (LNG-IUS alone, SIS alone, OPH+EBx), prior treatment with the levonorgestrel-releasing intrauterine system (OPH alone, TVUS+EBx), women wishing to preserve their fertility (LNG-IUS alone, SIS alone, OPH alone, SIS+OPH)

Economic model

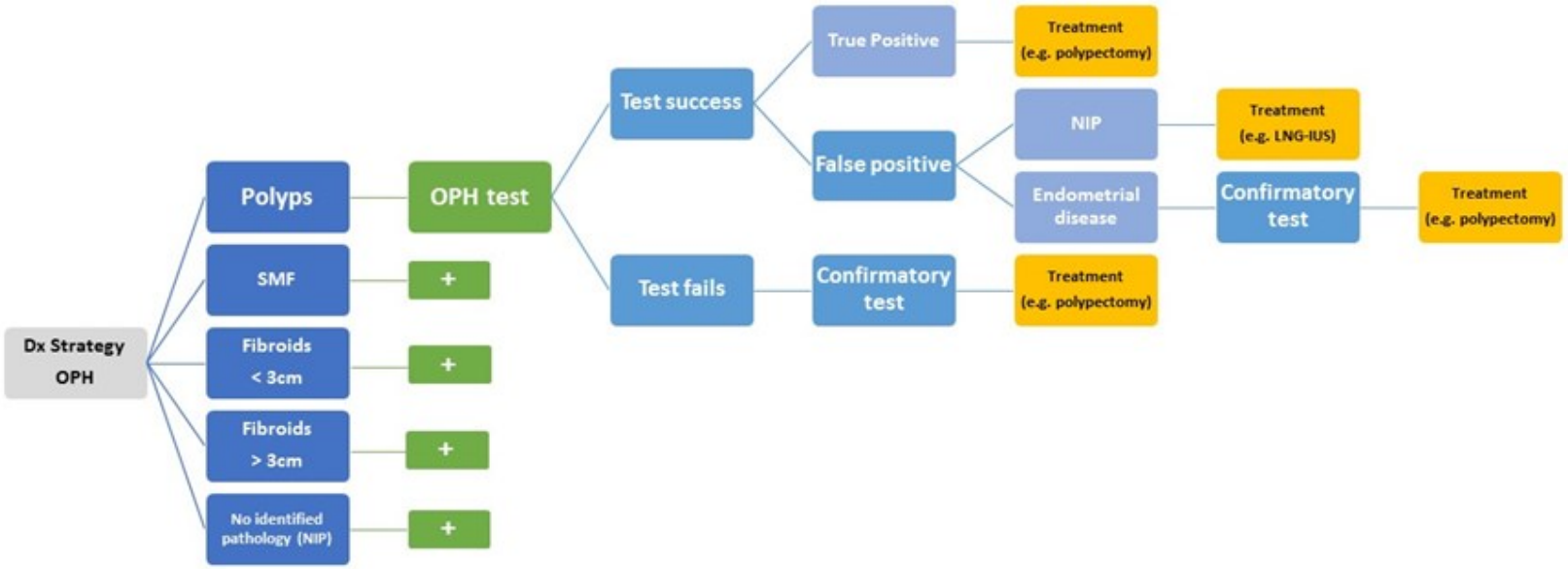
A cost utility analysis was developed to compare the cost-effectiveness of a range of diagnostic and management strategies. A Markov (state transition) model was developed to evaluate the costs and health related quality of life (HRQoL), measured in Quality Adjusted Life Years (QALYs), over a 5-year time frame.

The entry point for the model was women aged 42 years, presenting with heavy menstrual bleeding in an NHS primary care setting. The model could be run for up to 5 diagnostic strategies, including strategies involving treatment without any prior investigation:

- levonorgestrel-releasing intrauterine (LNG-IUS) alone
- hysterectomy alone
- outpatient hysteroscopy
- TVUS
- endometrial biopsy.

The decision analytic framework follows the approach used in a previous health economic evaluation (Cooper 2014) and the structure for one of the diagnostic strategies is illustrated in Figure 1.

Figure 1: An example decision tree to illustrate model structure for the diagnosis of underlying pathology in women with heavy menstrual bleeding



Dx: diagnosis; HMB: heavy menstrual bleeding; LNG-IUS: levonorgestrel-releasing intrauterine system; NIP: no identified pathology; OPH: outpatient hysteroscopy; SMF: submucosal fibroids

The outcome of the diagnostic tests was used to direct treatment according to the woman's underlying pathology. In the model, diagnostic test accuracy was used to estimate the proportion of women who would be correctly identified and receive the appropriate first line treatment. False positive rates were estimated to determine the proportion of women who would receive 'incorrect' treatment based on a diagnosis that differed from their true underlying pathology. The model comprised 5 underlying pathologies (polyps, submucosal fibroids, fibroids less than 3 cm in diameter, fibroids 3 cm or more in diameter and no identifiable pathology) and a simplifying assumption was made that the woman could only have a single underlying pathology.

The following surgical and pharmacological interventions were included as possible treatment alternatives in the model:

- LNG-IUS
- tranexamic acid
- combined oral contraceptives
- hysterectomy
- first generation endometrial ablation techniques
- second generation endometrial ablation techniques
- transcervical resection of fibroids
- nonsteroidal anti-inflammatory drugs
- medroxyprogesterone acetate
- polypectomy.

One important limitation of the model relates to how treatment gain was ascertained. Diagnosis of the underlying pathology is considered important to direct the woman to the most appropriate treatment. However, there are treatments which can be considered appropriate across a range of pathologies which is why empiric treatment with LNG-IUS, for example, can be considered as a plausible clinical option and is commonly used in UK current practice. This means that a woman incorrectly diagnosed can potentially still receive appropriate treatment. Furthermore, the network meta-analysis (NMA) EuroQol five dimensions questionnaire (EQ-5D) data allows a QALY gain to be estimated according to the treatment received but not according to the appropriateness of that treatment for the underlying pathology. To address this the committee were asked to dichotomise treatments according to whether they thought they were effective or not for a given pathology. Again this was a simplifying assumption as the committee acknowleged that a treatment may still provide some benefit for a given pathology even if it was considered sub-optimal. To reflect the uncertainty about the effectiveness of treatments the committee classified the treatments by pathology as always effective, partially effective and never effective. In the base case scenario it was assumed that both effective and partially effective treatments for a particular pathology would give the QALY gain derived from the NMA when used for a woman with that pathology. However, as a sensitivity analysis, a more conservative assumption was made where QALY gain was limited to only those treatments deemed effective for a given pathology. No QALY gain would result if the woman received a treatment assessed as only partially effective for her underlying uterine pathology in this sensitivity anaysis.

The analysis used probabilistic sensitivity analysis (PSA) to generate results. This involves Monte Carlo simulation where, instead of using a deterministic model input value, those values subject to uncertainty around their point estimate are sampled from a probability distribution over multiple iterations. The mean costs and QALYs are then calculated over these simulations and used to generate a mean net monetary benefit (NMB) which is calculated as follows:

NMB = QALY x cost-effectiveness threshold – cost

In line with NICE social value judgements (https://www.nice.org.uk/media/default/about/what-we-do/research-and-development/social-value-judgements-principles-for-the-development-of-nice-guidance.pdf) described in the methods chapter, a cost-effectiveness threshold of £20,000 per QALY was used. The strategy with the highest mean NMB was the most cost-effective strategy although to help assess any uncertainty an estimate was also made of the probability that a particular strategy was the most cost-effective. This was based on the proportion of simulations that the strategy is assessed as the most cost-effective.

The model results suggested that empiric LNG-IUS was a cost-effective strategy under the base case assumptions about treatment gain with a NMB of £30,131 and a 96.8% probability of being cost-effective. Under conservative assumption where LNG-IUS was deemed to be an appropriate treatment for a smaller subset of underlying pathology then that was no longer the case with the NMB falling to £19,039.

When treatment alternatives were held constant the model did not find large differences in the mean NMB of the 3 diagnostic tests compared. However, endometrial biopsy was dominated (the most expensive test strategy and the strategy generating the least QALY gain) by both TVUS and outpatient hysteroscopy. The comparison of the relative cost-effectiveness of outpatient hysteroscopy and TVUS is more complicated. Most of the analyses seemed to suggest that TVUS was more cost-effective than outpatient hysteroscopy despite being more costly. However, sensitivity analysis also indicated that this result was sensitive to prevalence of fibroids less than 3 cm in diameter as the underlying pathology in women presenting with HMB.

Pharmacological treatments generate relatively high mean NMB values when using base case assumptions about treatment gain across the underlying uterine pathologies.This may support their use in primary care, especially in women who are not refractory to treatment. Whilst hysterectomy is the most expensive intervention, analyses showed scenarios where it would be more cost-effective than other surgical intervention, such as second generation endometrial ablation.

This model is described in more detail in the Health economics chapter.

Resource impact

The guideline recommendations are likely to lead to a change in current practice with outpatient hysteroscopy being the recommended investigation in preference to pelvic ultrasound, for women with HMB when polyps, submucous fibroids or endometrial pathology are suspected. This change in practice will have a resource impact on service organisation and training.

Ultrasound is widely available through direct booking in primary care, whereas hysteroscopy is not. Changes to services will be needed to allow direct access booking into one-stop hysteroscopy services and to increase delivery in community-based clinics. GPs or nurses may need training to perform hysteroscopy in primary care. However, the commitee anticipate that there should be ongoing savings as the number of unnecessary investigations is reduced and women are offered effective treatment as a result of more accurate diagnosis.

Although diagnostic hysteroscopy is more expensive than transvaginal ultrasound the economic analysis undertaken for this guideline suggested that it may be a cheaper strategy overall. This is largely because diagnostic hysteroscopy can facilitate a ‘see and treat’ approach for women with polyps or submucosal fibroids which obviates the need for treatment at a later date.

Women who decline hysteroscopy can receive the investigation under anaesthesia. This was not formally considered in the analysis as the numbers of women was considered to be a

relatively small sub-group of the population having diagnostic hysteroscopy, approximately 10% in the oinion of the committee. These women in turn are a sub-group of those with HMB, many of whom will commence empiric treatment without diagnostic investigation. Therfore, the recommendation with respect to hysterectomy under anaesthesia is not expected to have a important resource impact on the NHS.

The committee also believe that the recommendation to not offer 'blind' endometrial biopsy will lead to a reduction in unnecessary investigations with resulting savings to the NHS.

Evidence statements

Diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting polyps

Very low quality evidence from 9 studies (n=955) found 2D-TVUS to have low sensitivity and high specificity in detecting polyps. LR+ indicated that 2D-TVUS is moderately useful in 'ruling in' polyps but LR- indicated that 2D-TVUS is not useful in 'ruling out' polyps. The studies came from India, Iran, Italy, Pakistan and Turkey and were published between 1995 and 2012.

This meta-analysis showed high heterogeneity especially for sensitivity, therefore, possible reasons for heterogeneity were explored and subgroup analysis was conducted when possible according to pre-specified groups in the protocol. Studies published before 2007 were excluded because the committee considered that technology has developed rapidly over the last years and older studies might not be comparable to newer studies. The subgroup analysis with 7 studies (n=853) including studies published in 2007 or later showed no considerable change in the results or heterogeneity of sensitivity. The quality of the evidence was very low.

Diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting fibroids

Very low quality evidence from 8 studies (n=870) found 2D-TVUS to have moderate sensitivity and high specificity in detecting fibroids. LR+ indicated that 2D-TVUS is very useful in 'ruling in' fibroids whereas LR- indicated that 2D-TVUS is not useful in 'ruling out' fibroids. The studies came from India, Iran, Italy, Pakistan and Turkey and were published between 1995 and 2012.

This meta-analysis showed moderate heterogeneity, therefore, possible reasons for heterogeneity were explored and subgroup analysis was conducted when possible according to pre-specified groups in the protocol. Studies published before 2007 were excluded. The subgroup analysis with 6 studies (n=768) including studies published in 2007 or later showed no considerable change in the results or heterogeneity. The quality of the evidence was very low.

Additionally, because hysteroscopy-guided biopsy or D&C are considered inadequate to histologically confirm fibroids, evidence on the diagnostic accuracy of 2D-TVUS compared to histopathology from hysterectomy alone was considered separately. Very low quality evidence from 2 studies showed mixed findings. One study (n=52) from Italy from 1995 showed 2D-TVUS to have high sensitivity in detecting fibroids whereas the other study (n=50) from India from 2002 showed 2D-TVUS to have low sensitivity in detecting fibroids. LR+ evidence from both studies indicated that 2D-TVUS is a very useful test in 'ruling in' fibroids. LR- evidence from one study indicated 2D-TVUS to be very useful in 'ruling out' fibroids whereas LR- evidence from the other study indicated 2D-TVUS to not be useful in 'ruling out' fibroids.

Diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting polyps or fibroids

Very low quality evidence from 1 study (n=189) from Denmark from 2001 showed 2D-TVUS to have high sensitivity but low specificity in detecting either polyps or fibroids. On the contrary, very low quality evidence from 1 study (n=88) from Norway from 2001 showed 2D-TVUS to have low sensitivity and high specificity in detecting polyps or fibroids. LR+ evidence from both studies indicated that 2D-TVUS is not useful in 'ruling in' polyps or fibroids. LR- evidence from the first study indicated that 2D-TVUS is moderately useful in 'ruling out' polyps or fibroids whereas LR- evidence from the second study indicated that it is not useful.

Diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting endometrial hyperplasia

Very low quality evidence from 3 studies (n=122; n=85; n=89) showed 2D-TVUS to have low to borderline moderate sensitivity and moderate to high specificity in detecting endometrial hyperplasia. LR+ indicated the test to be moderately to very useful in 'ruling in' hyperplasia whereas LR- indicated that the test is not useful in 'ruling out' endometrial hyperplasia. The studies were published in 2007 and 2012 and two came from Turkey and one from India.

Diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting endometrial carcinoma

Very low quality evidence from 1 study (n=89) from Turkey from 2012 showed 2D-TVUS to have low sensitivity and high specificity in detecting endometrial carcinoma. LR+ indicated the test to be moderately useful in 'ruling in' carcinoma and LR- indicated the test to be not useful in 'ruling out' carcinoma.

Diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting abnormal endometrial thickness (double-layer endometrium 12 mm or thicker and single-layer endometrium 6 mm or thicker)

Very low quality evidence from 1 study (n=88) from Norway from 2001 showed 2D-TVUS to have low sensitivity and moderate specificity in detecting abnormal endometrial thickness (double-layer endometrium 12 mm or thicker and single-layer endometrium 6 mm or thicker). LR+ and LR- indicated that the test is not useful in 'ruling in' or 'ruling out' abnormal endometrial thickness.

Diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting any abnormal finding

Very low to low quality evidence from 4 studies (n=213; n=83; n=89; n=39) showed mixed findings relating to any abnormal finding (for example abnormal thickness, hyperplasia, carcinoma, polyp or fibroid). Two studies showed 2D-TVUS to have moderate to high sensitivity whereas two studies showed low sensitivity. Two studies showed 2D-TVUS to have moderate to high specificity, whereas two studies showed low specificity. LR+ in the studies indicated 2D-TVUS to be either moderately useful or not useful in 'ruling in' any abnormal finding. LR- in the studies varied. In one study LR- indicated that 2D-TVUS is very useful in 'ruling out' any abnormal finding, in another study LR- indicated that 2D-TVUS is moderately useful, whereas in the two other studies LR- indicated that 2D-TVUS is not useful in 'ruling out' any abnormal finding.

Diagnostic accuracy of hysteroscopy under general anaesthetic (versus histopathology) in detecting polyps

Very low quality evidence from 2 studies (n=89; n=85) from Turkey and India from 2012 and 2007 show mixed results. The evidence from both studies showed hysteroscopy under general anaesthetic to have high specificity in detecting polyps, however, sensitivity was high in one study and low in the other study. LR+ in both studies indicated that hysteroscopy under general anaesthetic is very useful in 'ruling in' polyps. However, LR- in one study indicated that hysteroscopy under general anaesthetic is very useful in 'ruling out' polyps but

LR- in the other study indicated that hysteroscopy under general anaesthetic is not useful in 'ruling out' polyps.

Diagnostic accuracy of outpatient hysteroscopy (versus histopathology) in detecting polyps

Very low quality evidence from 2 studies (n=52; n=269) from Italy and Pakistan from 1995 and 2010 showed outpatient hysteroscopy to have moderate to high sensitivity and high specificity in detecting polyps. LR+ and LR- indicated that outpatient hysteroscopy is very useful in 'ruling in' and 'ruling out' polyps.

Diagnostic accuracy of hysteroscopy under general anaesthetic (versus histopathology) in detecting fibroids

Very low quality evidence from 1 study (n=89) from Turkey from 2012 showed hysteroscopy under general anaesthetic to have high sensitivity and specificity in detecting fibroids. LR+ and LR- indicated that hysteroscopy under general anaesthetic is very useful in 'ruling in' and 'ruling out' fibroids.

Diagnostic accuracy of outpatient hysteroscopy (versus histopathology) in detecting fibroids

Very low quality evidence from 1 study (n=51) from Italy from 1995 showed outpatient hysteroscopy to have high sensitivity and specificity in detecting fibroids. LR+ and LR- indicted outpatient hysteroscopy to be very useful in 'ruling in' and 'ruling out' fibroids.

Diagnostic accuracy of hysteroscopy under spinal or general anaesthetic (versus histopathology) in detecting polyps or fibroids

Very low quality evidence from 1 study (n=88) from Norway from 2001 showed day case hysteroscopy to have high sensitivity and moderate specificity in detecting either polyps or fibroids. LR+ indicated day case hysteroscopy to be moderately useful in 'ruling in' polyps or fibroids and LR- indicated that day case hysteroscopy to be very useful in 'ruling out' polyps or fibroids.

Diagnostic accuracy of hysteroscopy under general anaesthetic (versus histopathology) following a 2D-TVUS in detecting polyps or fibroids

Very low quality evidence from 1 study (n=50) from Egypt from 2016 showed hysteroscopy under general anaesthetic following a 2D-TVUS to have moderate sensitivity and high specificity in detecting either polyps or fibroids. LR+ indicated hysteroscopy under general anaesthetic following 2D-TVUS to be very useful in 'ruling in' polyps and fibroids and LR- indicated hysteroscopy under general anaesthetic following 2D-TVUS to be moderately useful in 'ruling out' polyps or fibroids.

Diagnostic accuracy of hysteroscopy under general anaesthetic (versus histopathology) in detecting endometrial hyperplasia

Very low quality evidence from 2 studies (n=89; n=85) from Turkey and India from 2012 and 2007 showed mixed findings. Both studies showed hysteroscopy under general anaesthetic to have high specificity in detecting endometrial hyperplasia, however, one study showed low sensitivity and the other study moderate sensitivity. LR+ from both studies indicated that hysteroscopy under general anaesthetic is very useful in 'ruling in' hyperplasia, whereas LR- in one study indicated hysteroscopy under general anaesthetic is moderately useful in 'ruling out' hyperplasia and LR- in the other study indicated hysteroscopy under general anaesthetic not to be useful in 'ruling out' hyperplasia.

Diagnostic accuracy of outpatient hysteroscopy (versus histopathology) in detecting endometrial hyperplasia

Very low quality evidence from 1 study (n=269) from Pakistan from 2010 showed outpatient hysteroscopy to have low sensitivity and high specificity in detecting endometrial hyperplasia. LR+ indicated that outpatient hysteroscopy is moderately useful in 'ruling in' endometrial hyperplasia whereas LR- indicated the test to not be useful in 'ruling out' endometrial hyperplasia.

Diagnostic accuracy of hysteroscopy under general anaesthetic (versus histopathology) in detecting endometrial carcinoma

Very low quality evidence from 1 study (n=89) from Turkey from 2012 showed hysteroscopy under general anaesthetic to have high sensitivity and specificity in detecting endometrial carcinoma. LR+ and LR- indicated that hysteroscopy under general anaesthetic is very useful in 'ruling in' and 'ruling out' endometrial carcinoma.

Diagnostic accuracy of outpatient hysteroscopy (versus histopathology) in detecting endometrial carcinoma

Very low quality evidence from 1 study (n=269) from Pakistan from 2010 showed outpatient hysteroscopy to have high sensitivity and specificity in detecting endometrial carcinoma. LR+ and LR- indicated that outpatient hysteroscopy is very useful in 'ruling in' and 'ruling out' endometrial carcinoma.

Diagnostic accuracy of hysteroscopy under spinal or general anaesthetic (versus histopathology) in detecting abnormal endometrial thickness (double-layer endometrium 12 mm or thicker and single-layer endometrium 6 mm or thicker) Very low quality evidence from 1 study (n=88) from Norway from 2001 showed day case hysteroscopy to have low sensitivity and moderately high specificity in detecting abnormal endometrial thickness (double-layer endometrium 12 mm or thicker and single-layer endometrium 6 mm or thicker). LR+ and LR- indicates that day case hysteroscopy is not useful in 'ruling in' or 'ruling out' abnormal endometrial thickness.

Diagnostic accuracy of outpatient hysteroscopy (versus histopathology) in detecting endometritis

Very low quality evidence from 1 study (n=269) from Pakistan from 2010 showed outpatient hysteroscopy to have low sensitivity but high specificity in detecting endometritis. LR+ indicated that outpatient hysteroscopy is very useful in 'ruling in' endometritis whereas LR- indicated that outpatient hysteroscopy is not useful in 'ruling out' endometritis.

Diagnostic accuracy of outpatient hysteroscopy (versus histopathology) in detecting retained products of conception

Very low quality evidence from 1 study (n=269) from Pakistan from 2010 showed outpatient hysteroscopy to have high sensitivity and specificity in detecting retained products of conception. LR+ and LR- indicated that outpatient hysteroscopy is very useful in 'ruling in' and 'ruling out' retained products of conception.

Diagnostic accuracy of hysteroscopy under general anaesthetic (versus histopathology) in detecting any abnormal finding

Very low quality evidence from 1 study (n=89) from Turkey from 2012 showed hysteroscopy under general anaesthetic to have high sensitivity and specificity in detecting any abnormal finding (defined as polyp, fibroid, hyperplasia or carcinoma). LR+ and LR- indicated that hysteroscopy under general anaesthetic is very useful in 'ruling in' and ruling out' any abnormal finding.

Diagnostic accuracy of hysteroscopy with or without general or local anaesthetic (versus histopathology) in detecting any abnormal finding

Very low quality evidence from 1 study (n=86) from Turkey from 2009 showed hysteroscopy (with general anaesthesia, with spinal anaesthesia or without any anaesthesia) to have high sensitivity and specificity in detecting any abnormal finding (defined as presence of adhesion, polyp, submucosal fibroid, pressure effect or any other abnormality in the uterine cavity). LR+ and LR- indicated that hysteroscopy is very useful in 'ruling in' or 'ruling out' abnormal endometrial finding.

Diagnostic accuracy of 2D-TVUS (versus hysteroscopy) in detecting polyps

Very low quality evidence from 1 study (n=196) from the UK from 2001 showed that 2D-TVUS and hysteroscopy compare poorly in terms of sensitivity but compare well in terms of specificity in detecting polyps.

Diagnostic accuracy of 2D-TVUS (versus hysteroscopy) in detecting suspicious focal thickening

Very low quality evidence from 1 study (n=196) from the UK from 2001 showed that 2D-TVUS and hysteroscopy compare poorly in terms of sensitivity but compare well in terms of specificity in detecting suspicious focal thickening.

Diagnostic accuracy of 2D-TVUS (versus hysteroscopy) in detecting any abnormal finding

Moderate quality evidence from 1 study (n=770) from Italy from 1997 showed that 2D-TVUS and hysteroscopy compare well in terms of sensitivity and moderately well in terms of specificity in detecting abnormal endometrial finding (defined as a lesion inside the cavity or when the maximum endometrial thickness measured in the sagittal plane was more than 14 mm).

Patient acceptability and satisfaction

Very low quality evidence was obtained from 1 study (n=326) from the UK from 2004 among premenopausal women either aged 40 years or more, or aged less than 40 years but with specific risk factors for endometrial cancer, who were undergoing investigations for abnormal uterine bleeding (68% had heavy menstrual bleeding). This study showed that more women who underwent hysteroscopy (with biopsy) and blind endometrial biopsy alone reported the investigation to be 'markedly unpleasant' than women who underwent ultrasound investigation. No difference were observed in the reporting of abdominal discomfort at home one day after investigation between women who underwent hysteroscopy and women who did not undergo hysteroscopy or between women who underwent ultrasound and women who did not undergo ultrasound.

The study also reported on feelings about the investigation one day after the clinic visit and the results did not show marked differences between women who underwent hysteroscopy, who did not undergo hysteroscopy, who underwent ultrasound, and who did not undergo ultrasound. Similarly, no marked differences were observed in how worthwhile the women in these different groups considered the clinic visit.

At 10 months post-investigation, only minimal differences were observed across intervention groups on self-report for persistence of symptoms and failure to cure the problem. The differences in responses for being satisfied with care, being reassured by clinic attendance, being glad to have attended the clinic, and feeling it was worthwhile attending the clinic were small, however, they seemed to favour either hysteroscopy or ultrasound rather than blind endometrial biopsy alone. Similar patterns were observed at 24 months post-investigation.

Health economic evidence statements

One cost-utility analysis (Cooper 2014) undertaken in the UK included 13 different strategies: transvaginal ultrasound scan, saline infusion sonography, global endometrial biopsy and

outpatient hysteroscopy used alone, or in combination. Using a 1-year time horizon, they found outpatient hysteroscopy alone and outpatient hysteroscopy plus endometrial biopsy to remain non-dominated by alternative empirical treatment or diagnostic testing strategies. The most effective strategy is outpatient hysteroscopy plus endometrial biopsy, generating an ICER of £21,500 (cost to gain an extra woman satisfied following treatment for HMB) compared with outpatient hysteroscopy alone and an ICER of £21,859 when compared with LNG-IUS alone. Outpatient hysteroscopy alone is less effective than outpatient hysteroscopy plus endometrial biopsy, but is less costly with an ICER of £360. If HMB is associated with a reduction of 0.125 QALYs in any year, they estimate that an ICER of £21,000 per case satisfied is approximately equivalent to an ICER of £26,500 per QALY. This analysis is directly applicable with potentially serious limitations.

An economic model developed for the guideline found that empiric treatment with LNG-IUS was cost-effective when compared to strategies where diagnosis was used to determine treatment under the assumption that LNG-IUS was effective for all underlying pathology (see Health Economics Chapter, Analyses 1, 3, and 10).

An economic model developed for the guideline found that TVUS was cost-effective when compared to empiric LNG-IUS, EBx and OPH under conservative assumptions about the treatment effectiveness of LNG-IUS ((see Health Economics Chapter, Analyses 2 and 11).

An economic model developed for the guideline found that TVUS was cost-effective compared to EBx and OPH when empiric treatment was not alternative for given set of commonly used first and second line treatments (Analyses 5 and 6).

An economic model developed for the guideline found that TVUS and OPH were of equivalent cost-effectiveness and dominated EBx when the prevalence of fibroids less than 3 cm in diameter was assumed to be 5% (see Health Economics Chapter, Analysis 7).

An economic model developed for the guideline found that EBx was dominated by TVUS and OPH when for commonly used first and second line treatments (see Health Economics Chapter, Analyses 5, 6 and 7).

Review question 2. What is the most clinically effective imaging strategy for diagnosing adenomyosis in women with heavy menstrual bleeding?

Introduction

Adenomyosis is a benign gynaecological disorder in which the inner lining of the uterus breaks through into the muscle wall of the uterus. The most common symptoms are pelvic pain and heavy menstrual bleeding. Traditionally, adenomyosis has been diagnosed from histopathology following a hysterectomy. While improvements in diagnostic imaging in recent years have shown that adenomyosis could be diagnosed through diagnostic radiology, there are currently no standard diagnostic imaging criteria for adenomyosis.

The aim of this review was to determine the most clinically effective imaging strategy for detecting adenomyosis in women presenting with heavy menstrual bleeding.

Summary of the protocol

Please see Table 5 for a summary of the characteristics of this review including population, index tests, reference standard and outcomes.

Table 5: Summary of the protocol

Population	Premenopausal women with heavy menstrual bleeding (HMB) At least 66% of the population should be women with HMB.

Index test(s)	• Magnetic resonance imaging (MRI)
	• Transvaginal ultrasound scan (TVUS)
	  ○ Two-dimensional (2D)
	  ○ Three-dimensional (3D)
	• Transabdominal ultrasound scan (TAUS)
	  ○ 2D
	  ○ 3D
	• Combination of two or more of the above

Reference standard	Histopathology
Outcomes	• sensitivity • specificity • positive likelihood ratio (LR+) • negative likelihood ratio (LR-) • area under the curve (AUC) if meta-analysis can be conducted
Target condition	Adenomyosis

2D: two-dimensional; 3D: three-dimensional; AUC: area under the curve; HMB: heavy menstrual bleeding; LR+: positive likelihood ratio; LR-: negative likelihood ratio; MRI: magnetic resonance imaging; sens: sensitivity; spec: specificity; TAUS: transabdominal ultrasound scan; TVUS: transvaginal ultrasound scan

For full details see Appendix A – Review protocols.

Clinical evidence

Histopathology was considered as the reference standard to which diagnostic accuracy of MRI, ultrasound were compared. Histopathology was derived from surgery specimen from hysterectomy or deep hysteroscopy-guided biopsy.

A combined literature search was undertaken for this review question and the review question on the diagnostic accuracy of ultrasound and hysteroscopy in the investigation of women presenting with heavy menstrual bleeding.

Due to the limited evidence available when the population was restricted only to women presenting with heavy menstrual bleeding, the committee agreed that one third of the population could be women without heavy menstrual bleeding and these results could be extrapolated to women presenting with heavy menstrual bleeding as long as the indirectness of the population was accounted for when assessing the quality of the evidence.

Meta-analysis was conducted when appropriate with a minimum of 5 studies. In the presence of heterogeneity, potential reasons for heterogeneity were explored and subgroup analyses were conducted when possible according to pre-specified groups in the protocol.

Test sensitivity or specificity was considered high when sensitivity or specificity was 90% or higher, and moderate when sensitivity or specificity was between 75% and 89%. A positive likelihood ratio (LR+) of 10 or higher was considered to indicate that the test was very useful in 'ruling in' the condition, and a LR+ between 5 and 9.9 was considered moderately useful in 'ruling in' the condition. A negative likelihood ratio (LR-) of 0.1 or lower was considered to indicate that the test was very useful in 'ruling out' the condition, and a LR- between 0.1 and 0.2 was considered to indicate that the test was moderate useful in 'ruling out' the condition.

Included studies

Eight studies (n=1028) were included in the review (Abdel Hak 2010; Alborzi 2007; Bazot 2002; Botsis 1998; Dakhly 2016; Duelholm 2001; Exacoustos 2011; Vercellini 1998).

All included studies included two-dimensional transvaginal ultrasound scan (2D-TVUS) as the index test (Abdel Hak 2010; Alborzi 2007; Bazot 2002; Botsis 1998; Dakhly 2016; Duelholm 2001; Exacoustos 2011; Vercellini 1998). One of these studies also looked at three-dimensional (3D) TVUS (Exacoustos 2011). One study also looked at magnetic resonance imaging (MRI) and a combination of 2D-TVUS and MRI as the index tests (Dueholm 2001b). One study also included 2D transabdominal ultrasound scan (2D-TAUS) as the index test (Bazot 2002).

All studied included histopathology as the reference standard. Seven studies used surgical specimen from hysterectomy as the reference standard (Abdel Hak 2010; Bazot 2002; Botsis 1998; Dakhly 2016; Duelholm 2001; Exacoustos 2011; Vercellini 1998) and one study used a deep biopsy from hysteroscopy as the reference standard (Alborzi 2007).

The included studies were from Egypt (Abdel Hak 2010; Dakhly 2016); Italy (Exacoustos 2011; Vercellini 1998); Denmark (Dueholm 2001b); France (Bazot 2002); Greece (Botsis 1998); and Iran (Alborzi 2007).

Evidence from these are presented in Appendix G – GRADE tables.

See also Appendix D – Clinical evidence study selection, Appendix H – Forest plots and hierarchical summary receiver operating characteristic (HSROC) plots, and Appendix F – Clinical evidence tables.

Excluded studies

Studies not included in this review with reasons for their exclusions are provided in Appendix I – Excluded studies.

Summary of clinical studies included in the evidence review

Table 6 provides a brief summary of the included studies

Table 6: Summary of included studies

Study	Population	Index test	Reference standard	Diagnositc criteria of adenomyosis in index test
Abdel Hak 2010 (Egypt)	N=50 perimenopausal women undergoing hysterectomy due to HMB.	2D-TVUS The paper did not report who performed or interpreted the TVUS or the level of experience of the person(s).	Histopathology (hysterectomy)	A poorly defined area of abnormal echo texture within the myometrium. Abnormal myometrial echo texture was defined if the myometrium demonstrated heterogeneity, decreased or increased echogenicity, and/or the presence of cysts, presence of linear striation, globular configuration of the uterus.
Alborzi 2007 (Iran)	N=81 women with abnormal uterine bleeding. Proportion of women with HMB not reported.	2D-TVUS The paper did not report who performed or interpreted the	Histopathology (hysteroscopic biopsy) The paper reports that 'in all patients a	Diffuse uterine enlargement with no alteration in echotexture and contour. Focal adenomyosis was diagnosed when a

Study	Population	Index test	Reference standard	Diagnositc criteria of adenomyosis in index test
		TVUS or the level of experience of the person(s).	relatively deep specimen from the anterior and posterior wall of the uterus was resected and sent to a pathologist'.	poorly defined area of abnormal echotexture is present in the myometrium with increased or decreased echogenecity.
Bazot 2002 (France)	N=129 women undergoing hysterectomy due to various reasons divided into two groups: Group 1 (n=23) women with recurrent menometrorrhagia but no evidence of leiomyomata and endometrial diseases on transabdominal examination. 100% had menometrorrhagia. Group 2 (n=106) all other women. 67.9% had menometrorrhagia.	2D-TAUS; 2D-TVUS The TVUS were performed by two investigators with 8 and 3 years of experience in female pelvic ultrasound scan.	Histopathology (hysterectomy)	TAUS Enlarged regular uterus with no evidence of leiomyoma and/or presence of myometrial cysts. TVUS A globular and/or asymmetric uterus, a poorly defined focus of abnormal myometrial echotexture, distorted and heterogeneous myometrial echotexture, myometrial linear striations, and myometrial cysts. With the exception of diffuse heterogeneous myometrium that appeared non-specific for adenomyosis, the diagnosis was made when at least one of the above criteria was met.
Botsis 1998 (Greece)	N=206 women undergoing hysterectomy for various reasons. 83% had HMB and/or dysmenorrhea.	2D-TVUS The paper did not report who performed or interpreted the TVUS or the level of experience of the person(s).	Histopathology (hysterectomy)	Heterogeneous myometrial areas that were not encapsulated and that contained anechoic lacunae measuring 1–3 mm in diameter and an area characterized by irregular cystic spaces measuring 1–7 mm in diameter (honeycomb pattern) and disrupting the normal fine speckled echo pattern of the uterus. The sonographic examination was considered diagnostic of adenomyosis when at least 3 parameters were positive.

Study	Population	Index test	Reference standard	Diagnositc criteria of adenomyosis in index test
Dakhly 2016 (Eqypt)	N=404 original sample N=292 included in analysis premenopausal women with clinical signs and symptoms of adenomyosis. 64% had HMB and 36% had menometrorrhagia.	2D-TVUS The paper did not report who performed or interpreted TVUS or the level of experience of the person(s).	Histopathology (hysterectomy)	Presence of 2 or more of the following 5 criteria: heterogeneous myometrial echo-texture; myometrial cysts; subendometrial echogenic linear striations; asymmetry of the anterior and posterior myometrium; and a poorly defined endometrial–myometrial junction.
Dueholm 2001b (Denmark)	N=106 premenopausal women undergoing hysterectomy for various reasons. 77% had abnormal uterine bleeding. Proportion of women with HMB not reported.	2D-TVUS; MRI The TVUS was performed by one experienced gynaecologists. The MRI was evaluated by one MRI specialist.	Histopathology (hysterectomy)	Presence of heterogeneity, increased or decreased areas of echogenicity, or presence of myometrial cysts.
Exacoustos 2011 (Italy)	N=72 premenopausal women undergoing hysterectomy due to benign pelvic pathology. 76% had HMB.	2D-TVUS; 3D-TVUS The TVUS were performed by one of three expert sonographers.	Histopathology (hysterectomy)	2D-TVUS Presence of 2 or more of the following individual ultrasonographic features: myometrial cysts; asymmetrical myometrial cysts; hypoechoic striations; heterogenous myometrial cysts. 3D-TVUS Presence of 2 or more of the following ultrasonographic features: JZmax 8 mm or more; JZmax − JZmin 4 mm or more; JZ ratio 50% or more; JZ alteration; myometrial cysts; asymmetrical myometrial cysts; heterogeneous myometrial cysts.
Vercellini 1998 (Italy)	N=102 premenopausal women undergoing hysterectomy for HMB and/or	2D-TVUS All the TVUS were performed by one expert sonographer.	Histopathology (hysterectomy)	Indistinctly demarcated heterogeneous myometrial areas with distorted echotexture. Myometrial echotexture was defined as

Study	Population	Index test	Reference standard	Diagnositc criteria of adenomyosis in index test
	worsening dysmenorrhea. Proportion of women with HMB not reported.			distorted by the presence of abnormally decreased or increased echogenicity and/or round anechoic areas.

2D: two-dimensiona; 3D: three-dimensional; HMB: heavy menstrual bleeding; MRI: magnetic resonance imaging; TAUS: transabdominal ultrasound scan; TVUS: transvaginal ultrasound scan

See Appendix F – Clinical evidence tables for full evidence tables.

Quality assessment of clinical studies included in the evidence review

The included studies were individually assessed using QUADAS-2 checklist. QUADAS-2 assesses the risk of bias in the study and the applicability of the study for the review question.

The quality of each piece of evidence, whether from meta-analysis or from individual studies was then assessed using adapted GRADE methodology.

See Appendix G – GRADE tables for full GRADE tables.

For more details on the quality assessment of clinical studies in the evidence review, please see Methods chapter.

Economic evidence

Included studies

No economic evaluations of imaging technologies for the diagnosis of adenomyosis were identified in the literature search conducted for this guideline.

Excluded studies

There were no excluded economic studies relating to the diagnosis of adenomyosis.

Summary of studies included in the economic evidence review

There were no included economics studies relating to the diagnosis of adenomyosis.

Economic model

No economic modelling was done for the evaluation of imaging technologies for the diagnosis of adenomyosis.

Resource impact

HMB is not the main presenting symptom of women with adenomyosis and therefore women with possible adenomyosis represent a small subset of the population covered by this guideline. TVUS is the recommended diagnostic test for women with possible adenomyosis which costs £147 (NHS Reference Costs, 2015-16) when provided in an outpatient setting. Whilst adenomyosis was not considered in previous NICE guidance the committee did not think that this would represent a change in practice. Transabdominal ultrasound or MRI are

less accurate and do not offer any significant saving when compared to TVUS although both can be considered if TVUS is not acceptable to the woman.

Evidence statements

Diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting adenomyosis in women with heavy menstrual bleeding

Very low quality evidence from 7 studies (n=1078) found 2D-TVUS to have moderately high sensitivity (76%, 95% CI 69 to 82%) and moderately high specificity (83%, 95% CI 73 to 90%) in diagnosing adenomyosis. 2D-TVUS was not found to be very useful in 'ruling in' or 'ruling out' adenomyosis (LR+ 4.6, 95% CI 2.8 to 7.4; LR- 0.29, 95% CI 0.22 to 0.37). The area under the curve (AUC) was 0.83 (95% CI 0.80-0.86). Histopathology on surgical specimens from hysterectomy was the reference standard for all studies except in one study where deep biopsy specimens from hysteroscopy examination were used for the reference standard. The studies included in this meta-analysis were published between 1998 and 2016 and came from five different countries (Denmark, Egypt, Greece, Iran, and Italy). The populations in these studies varied to an extent, however, at least around two thirds of the women included in these studies had heavy menstrual bleeding as a symptom.

Low quality evidence from 1 study (n=23) from France from 2002 among women with recurrent menometrorrhagia but no evidence of fibroids or endometrial diseases on transabdominal examination showed 2D-TVUS to have moderately high sensitivity (81%, 95% CI 58 to 95%) and high specificity (100%, 95% CI 16 to 100%) in diagnosing adenomyosis in this population. 2D-TVUS was found to be a useful test in 'ruling in' (LR+ infinite, not calculable) and moderately useful test in 'ruling out' (LR- 0.19, 95% CI 0.08 to 0.46) adenomyosis in this population. Moderate quality evidence from the same study among all other women (n=106, presumably women with evidence of fibroids or endometrial diseases on transabdominal examination) showed 2D-TVUS to have low sensitivity (39%, 95% CI 20 to 59%) but high specificity (98%, 95% CI 91 to 100%) in diagnosing adenomyosis in this population. 2D-TVUS was found to be a useful test in 'ruling in' adenomyosis (LR+ 15.4, 95% CI 3.6 to 65.7) but not a useful test in 'ruling out' adenomyosis (LR- 0.63, 95% CI 0.46 to 0.86) in this population.

Diagnostic accuracy of 3D-TVUS (versus histopathology) in detecting adenomyosis in women with heavy menstrual bleeding

Very low quality evidence from 1 study (n=72) from Italy from 2011 found 3D-TVUS to have high sensitivity (91%, 95% CI 74 to 97%) and moderately high specificity (88%, 95% CI 72 to 95%) in diagnosing adenomyosis (surgical specimen from hysterectomy as the reference standard). Likelihood ratios showed that 3D-TVUS is a moderately useful test in 'ruling in' and 'ruling out' adenomyosis (LR+ 7.3, 95% CI 3.2 to 16.6; LR- 0.11, 95% CI 0.03 to 0.31).

Diagnostic accuracy of 2D-TAUS (versus histopathology) in detecting adenomyosis in women with heavy menstrual bleeding

Moderate quality evidence from 1 study (n=23) from France from 2002 among women with recurrent menometrorrhagia but no evidence of fibroids or endometrial diseases on transabdominal examination showed 2D-TAUS to have low sensitivity (54%, 95% CI 34 to 78%) and specificity (50%, 95% CI 27 to 99%) in diagnosing adenomyosis and not to be useful in 'ruling in' or 'ruling out' adenomyosis (LR+ 1.1, 95% CI 0.3 to 4.8; LR- 0.9, 95% CI 0.2 to 3.7). Moderate quality evidence from the same study among all other women (n=106) who we assume, based on the group explicitly defined, have evidence of fibroids or endometrial diseases on transabdominal examination) showed 2D-TAUS to have very low sensitivity (8%, 95% CI 1 to 25%) but high specificity (96%, 95% CI 89 to 99%) and not to be useful in 'ruling in' or 'ruling out' adenomyosis (LR+ 2.1, 95% CI 0.4 to 11.6; LR- 0.96, 95% CI 0.85 to 1.08). Surgical specimen from hysterectomy was the reference standard.

Diagnostic accuracy of MRI (versus histopathology) in detecting adenomyosis in women with heavy menstrual bleeding

Low quality evidence from 1 study (n=106) from Denmark from 2001 found MRI to have low sensitivity (64%, 95% CI 41 to 83%) and moderately high specificity (88%, 95% CI 79 to 94%) in diagnosing adenomyosis (surgical specimen from hysterectomy as the reference standard). MRI was found to be moderately useful test in 'ruling in' adenomyosis (LR+ 5.4, 95% CI 2.8 to 10.4) but not useful in 'ruling out' adenomyosis (LR- 0.4, 95% CI 0.2 to 0.7). The study reported 'indefinite' results from the index test for 13 cases which have been included as negative test results in the results above.

Diagnostic accuracy of 2D-TVUS combined with MRI (versus histopathology) in detecting adenomyosis in women with heavy menstrual bleeding

Low quality evidence from 1 study (n=106) from Denmark from 2001 found the combination of 2D-TVUS and MRI to have a relatively low sensitivity (73%, 95% CI 50 to 89%) and moderately high specificity (77%, 67 to 86%) in diagnosing adenomyosis (surgical specimen from hysterectomy as the reference standard). 2D-TVUS combined with MRI was not found to be a useful test in 'ruling in' or 'ruling out' adenomyosis (LR+ 3.2, 95% CI 2.0 to 5.2; LR- 0.35, 95% CI 0.18 to 0.70) in this one study. The study reported 'indefinite' results from the index test for 41 cases which have been included as negative test results in the results above.

Recommendations

Investigations for women with HMB

A1. Consider starting pharmacological treatment for HMB without investigating the cause if the woman's history and/or examination suggests a low risk of intracavitary or histological abnormality or adenomyosis.

Hysteroscopy for women with suspected polyps, submucosal fibroids or endometrial pathology

A2. Offer outpatient hysteroscopy for women with HMB where history suggests polyps, submucosal fibroids or endometrial pathology because:

- they have symptoms such as intermenstrual bleeding **or**
- they have risk factors for endometrial pathology (see recommendation 1.3.8 in the short guideline).

A3. Ensure that outpatient hysteroscopy services are organised and the procedure is performed according to best practice, including:

- advising women to take oral analgesia before the procedure
- vaginoscopy as the standard technique, using miniature hysteroscopes (smaller than 3.5 mm)
- service organisation that enables ‘see-and-treat’ in a single setting if feasible.

A4. Explain to women with HMB who are offered outpatient hysteroscopy what the procedure involves and discuss the possible alternatives.

A5. If a woman declines outpatient hysteroscopy, consider hysteroscopy under anaesthesia.

Endometrial biopsy at the time of hysteroscopy

A6. Do not offer ‘blind’ endometrial biopsy to women with HMB.

Ultrasound

Women with possible larger fibroids

A7. Offer 2-dimensional pelvic ultrasound for women with HMB if any of the following apply:

- their uterus is palpable abdominally
- history or examination suggests a pelvic mass
- examination is inconclusive or difficult, for example in women who are obese.

Women who decline hysteroscopy

A8. For women who decline hysteroscopy, consider 2-dimensional pelvic ultrasound, explaining the limitations of this technique for detecting intracavitary causes of HMB.

Women with possible adenomyosis

A9. Offer transvaginal ultrasound (in preference to transabdominal ultrasound or MRI) for women with HMB who have:

- significant dysmenorrhoea (period pain) **or**
- a bulky, tender uterus on examination that suggests adenomyosis.

A10. If a woman declines transvaginal ultrasound or it is not suitable for her, consider transabdominal ultrasound or MRI, explaining the limitations of these techniques.

A11. Be aware that pain associated with HMB may be caused by endometriosis rather than adenomyosis (NICE is developing a guideline on endometriosis; publication expected September 2017).

Rationale and impact

Why the committee made the recommendations

The committee agreed that investigation is not necessary before starting treatment when history and examination do not suggest structural abnormalities or endometrial pathology.

Hysteroscopy for women with suspected polyps, submucosal fibroids or endometrial pathology

Outpatient hysteroscopy is recommended for women with HMB if intracavitary abnormalities or endometrial pathology are suspected because:

- the evidence showed that it is more accurate (higher sensitivity and specificity) in identifying them than pelvic ultrasound
- it is safe and has a low risk of complications
- it is acceptable to women if done according to best practice guidelines
- women can have polyps and submucosal fibroids removed during the procedure, and targeted biopsy if needed
- it is cost-effective as part of a diagnosis and treatment strategy.

For women who decline hysteroscopy, the committee agreed that hysteroscopy under general anaesthetic can be considered, because the benefits of accurate identification outweigh the the risks of anaesthesia.

Endometrial biopsy at the time of hysteroscopy

'Blind' endometrial biopsy is not recommended because it may not identify treatable lesions and is painful for women.

Ultrasound

Women with possible larger fibroids

Hysteroscopy is not able to detect abnormalities outside the uterine cavity, such as subserous or intramural fibroids, or adenomyosis. When an examination suggests a large or several fibroids, pelvic ultrasound (transvaginal or transabdominal) is recommended instead of hysteroscopy and is likely to be particularly cost-effective in this context.

When abdominal or vaginal examination is difficult to perform or inconclusive (for example, because the woman is obese) pelvic ultrasound should be offered to identify any abnormalities that might have otherwise been suggested by examination.

Women who decline hysteroscopy

The committee agreed that 2-dimensional pelvic ultrasound can be considered for women who decline hysteroscopy, provided that they understand and accept that it is less accurate in detecting intracavitary abnormalities and endometrial pathology.

Women with possible adenomyosis

The evidence showed that transvaginal ultrasound is more accurate than transabdominal ultrasound or MRI for detecting adenomyosis. Although transvaginal ultrasound is more intrusive than the other investigations, the committee's experience

suggests that many women find it acceptable. It is also widely available in primary and secondary care.

Transvaginal ultrasound may not be acceptable to or suitable for some women, such as women who have not been sexually active or women with female genital mutilation. The committee agreed that transabdominal ultrasound or MRI can be considered for these women, provided that they understand and accept that they are less accurate for detecting adenomyosis.

Impact of the recommendations on practice

Hysteroscopy

Hysteroscopy, in preference to pelvic ultrasound, is now being recommended as the investigation for causes of HMB when polyps, submucosal fibroids or endometrial pathology are suspected. This change in practice will have a resource impact on service organisation and training.

Ultrasound is available through direct booking in primary care, whereas hysteroscopy is not. Changes to services will be needed to allow direct access booking into one-stop hysteroscopy services and to increase delivery in community-based clinics. This would entail that GPs or nurses may need training to perform hysteroscopy in primary care. However, there should be ongoing savings as the number of unnecessary investigations is reduced and women are offered effective treatment as a result of more accurate diagnosis.

To ensure that outpatient hysteroscopy is acceptable to women, it is essential that the procedure is done according to best practice guidelines, including techniques and equipment to minimise discomfort and pain in women; adequately sized, equipped, and staffed facilities; staff with necessary training, skills and expertise; and the need for audit and benchmarking of outcomes.

Ultrasound

Two-dimensional transvaginal and transabdominal ultrasound are already widely available in primary and secondary care.

The committee noted that clinicians might need additional training and experience in interpreting transvaginal ultrasound scans to identify signs of adenomyosis.

The committee's discussion of the evidence

Interpreting the evidence

The outcomes that matter most

The committee agreed that the correct identification of the cause of HMB is important as this can impact the treatment options offered to women. The committee considered sensitivity, LR+ and LR- to be critical outcomes and specificity and AUC to be important outcomes. If a test is sensitive, it may help the clinicians to choose the right initial treatment to be offered to women. The committee recognised that it was important to avoid false positives because unnecessary treatment, especially surgical treatment, can cause harm. In relation to adenomyosis, the committee identified that it was important to avoid false negatives as adenomyosis typically does not respond well to conventional pharmacological treatments or endometrial ablation. Likelihood ratios were considered to be clinically helpful indicators of a useful test.

Owing to concerns about the acceptability of hysteroscopy to some women, patient satisfaction and acceptability was considered as a critical outcome for the review on the diagnostic accuracy of hysteroscopy and ultrasound for investigation of women with HMB. The committee considered that the experience and preference of women undergoing investigations is crucial in assessing which diagnostic test is most suitable.

The quality of the evidence

The evidence on diagnostic accuracy was assessed using adapted GRADE methodology. The evidence on patient satisfaction or acceptability was assessed using Cochrane Collaboration's tool for assessing risk of bias.

The quality of evidence in these reviews ranged from very low to moderate with most evidence being of very low quality. The committee recognised that the evidence was fragmented and with several limitations. The committee agreed that the quality of evidence was most often downgraded because of unclear sampling, unclear inclusion and exclusion criteria, unclear diagnostic criteria, and at times, considerable number of drop-outs.

The committee discussed that the populations in the included studies were not always identical to the population of interest in this guideline (premenopausal women with heavy menstrual HMB available, the committee had agreed in the review protocol that one third of the women included in the studies could be other than women with HMB, for example, women with pelvic pain without HMB. Therefore, most studies in these reviews included a proportion of women that were not directly the population of interest. This was accounted for in the quality assessment of the evidence.

The committee also agreed that the population of women in the review on the most effective investigation to detect adenomyosis in women with HMB do not represent the general HMB population since most of the women in the studies underwent hysterectomy. Milder forms of adenomyosis that do not require hysterectomy are much more common. The prevalence of adenomyosis in the included studies varied from 11% to 55%. The committee discussed that since the populations in the studies

included women undergoing hysterectomy, i.e. women with severe symptoms, the prevalence rates are likely underestimated. HMB is not the only, nor necessarily the most common presenting symptom for adenomyosis. Pelvic pain is another common symptom reported. In addition, an estimated 30% of women with adenomyosis are asymptomatic. Therefore, the prevalence of adenomyosis in the premenopausal population could be even higher.

The level of experience of the clinician performing and interpreting the diagnostic test was considered an important determinant of the diagnostic accuracy of the investigation, particularly for the diagnosis of adenomyosis for which there are no established diagnostic criteria. The committee recognised that the level of experience of the investigator was often not reported in the studies included in the reviews.

Considerable heterogeneity was observed in the meta-analyses conducted in the reviews, especially in the sensitivity of 2D-TVUS in detecting polyps and specificity of 2D-TVUS in detecting adenomyosis. Pre-specified subgroup analyses were conducted to investigate whether or not the year of publication could influence heterogeneity since diagnostic techniques have evolved over time, however, this was not found to impact the level of heterogeneity. The committee also considered other factors that might cause heterogeneity, including varying populations across studies, variation in the diagnostic criteria of adenomyosis on ultrasound, varying or unreported level of experience of the investigators, and varying reference standards used (hysterectomy, hysteroscopic biopsy, D&C). Despite the high heterogeneity of the analysis, the results of the meta-analysis on the accuracy of 2D-TVUS in diagnosing adenomyosis corresponded well with the results of two previous systematic reviews and meta-analysis, even though the inclusion criteria in these reviews were not the same as in this review (Champaneria 2010; Meredith 2009).

Limited or no evidence was found on the accuracy of 3D-TVUS and 2D- or 3D-TAUS to detect different pathologies and of MRI to detect adenomyosis in women with HMB. Furthermore, no evidence was found on vaginoscopic hysteroscopy. However, the committee agreed that the diagnostic accuracy of vaginoscopic hysteroscopy is expected to be similar to other outpatient hysteroscopy approaches, however, patient acceptability is expected to be higher.

Evidence on patient acceptability and satisfaction comparing hysteroscopy and ultrasound was scarce. The evidence came from one study published over a decade ago with limited and unclear reporting in relation to patient acceptability and satisfaction. The committee considered the techniques used in the study to be outdated, therefore, not reflecting the current standards of best practice. As with diagnostic accuracy, evidence on patient acceptability and satisfaction of diagnostic tests was limited to women presenting with heavy menstrual bleeding (at least 2/3 of the participants presenting with heavy menstrual bleeding). The committee acknowledged that there might have been more evidence available on patient acceptability and satisfaction of the diagnostic tests of interest in wider population groups. Whereas the committee considered that evidence on postmenopausal women could not be extrapolated to premenopausal women with heavy menstrual bleeding, evidence on the diagnostic accuracy of the tests in premenopausal women presenting with other symptoms might have been relevant. However, this was taken into account by allowing some deviation in the study population (i.e. one third of the study population could be other than women with heavy menstrual bleeding).

Benefits and harms

The committee agreed that many women presenting to primary care with symptoms of HMB can be offered treatment without the need for further examination or

investigation. However, investigation via a diagnostic technique might be warranted for women who have history or examination suggesting a structural or endometrial pathology or for whom the initial treatment has failed. The evidence reviewed showed that hysteroscopy had superior sensitivity, specificity, LR+, and LR- over 2D-TVUS in detecting polyps, submucosal fibroids and endometrial pathology. The committee considered outpatient hysteroscopy to be an efficient and safe technique with a low risk of complications, and acceptable if done according to best practice guidance, offering women the option of see-and-treat by having polyps or submucosal fibroids identified and removed in one process when appropriate. Therefore, the committee decided that outpatient hysteroscopy should be used in the diagnosis of heavy menstrual bleeding when history suggests submucosal fibroids, polyps or endometrial pathology. In detecting adenomyosis TVUS showed better accuracy compared to TAUS or MRI, therefore, the committee recommended the first-line investigation for suspected adenomyosis to be TVUS.

Traditionally, 2D-TVUS has been the first-line investigation for women presenting with heavy menstrual bleeding. However, technology has advanced which is evident from the clinical evidence showing hysteroscopy to have better accuracy in detecting most uterine and endometrial pathology compared to pelvic ultrasound scan. Diagnostic accuracy was not, however, the only factor considered in making the recommendation. Patient preference and acceptability was also very important to the committee. The evidence on patient acceptability and satisfaction in the evidence review came from a study that used techniques that were considered outdated and not according to the current best practice, and therefore, possibly painful for the women (larger hysteroscopes, blind biopsies taken, and use of non-vaginoscopic methods). The experience of the committee was that if performed according to current best practice standards, outpatient hysteroscopy is acceptable to women. One of the committee members had collected patient reported outcomes from over 1500 patients that confirms acceptability using the best practice guidelines with mean pain scores from the procedure less than that experienced by the patients at menstruation (not published). The other committee members agreed that this concurs with their own experience and is in line with the literature review presented in the Royal College of Obstetricians and Gynaecologists (RCOG) and British Society for Gynaecological Endoscopy (BSGE) Green-top Guideline No. 59 Best Practice in Outpatient Hysteroscopy (RCOG and BSGE 2011).

However, the committee did have concern that patient groups report poor experiences, and that some patients do experience significant distress. The committee discussed the variability in hysteroscopic techniques in the UK, with previous surveys among BSGE members suggesting poor provision of best practice. Implementing hysteroscopy as the first-line investigation in the diagnosis of heavy menstrual bleeding would, therefore, require the use of equipment, training of clinicians, and patient information provision in line with RCOG best current practice (RCOG and BSGE 2011). The committee felt that it was essential that all best practice elements needed to be delivered by services offering hysteroscopy, but that there was potential for these services to be offered more widely in community settings. The committee identified the following as essential elements for an integrated outpatient hysteroscopy service:

- vaginoscopy as the standard technique
- use of miniature hysteroscopes (less than 3.5 mm)
- service organisation to enable 'see-and-treat' in a single setting where feasible
- system to advise patients to take oral analgesia prior to their procedure.

In addition, the committee considered it is important that the staff has necessary training, skills and expertise to perform outpatient hysteroscopy; facilities are

appropriately sized, equipped and staffed with a toilet and private changing facilities; a nurse is available to act as the patient's advocate; and regular audit of outcomes that include patient reported outcomes is performed and benchmarked against local and national standards.

The committee recognised that not every woman would agree to undergo an outpatient hysteroscopy and emphasised that patient choice was important in deciding the diagnostic pathway. In women who declined an outpatient hysteroscopy the risks and benefits of anaesthesia (either general or local) should be discussed and a hysteroscopy under anaesthetic considered if appropriate.

The committee considered a blind biopsy to be painful and emphasised that blind biopsy may miss treatable pathology, thus should not be conducted in women presenting with HMB. An endometrial biopsy was only deemed necessary in women at high risk of endometrial hyperplasia or malignancy. According to the committee's knowledge and expertise, these include obese women, women with persistent intermenstrual bleeding, irregular bleeding or infrequent bleeding, women taking tamoxifen, and women in whom treatment has failed or was ineffective. In these circumstances a biopsy should be taken via hysteroscopic techniques in line with RCOG best practice guidelines (RCOG and BSGE 2011).

In cases where women decline hysteroscopy or for women for whom it is not appropriate, other options should be considered.

The place of pelvic ultrasound scan in the diagnosis of structural causes of heavy menstrual bleeding was discussed by the committee. Given ultrasound scan's superiority over hysteroscopy in detecting non-cavitary structural abnormalities but inferiority in detecting pathology within the uterine cavity, pelvic ultrasound scan would be restricted to women where the uterus is palpable abdominally, or where the history or a vaginal examination suggests a pelvic mass. Furthermore, in women where examination is inconclusive or difficult, for example, in women with high BMI, pelvic ultrasound scan is indicated to rule out fibroids that would normally be identified on examination. Even though the evidence was on TVUS, the committee considered TAUS to be clinically as appropriate in detecting fibroids that are large enough to be palpable. Therefore, pelvic ultrasound, meaning either TVUS or TAUS, was recommended as an investigation for these women.

If pelvic ultrasound scan is considered as an alternative to hysteroscopy or is requested by the woman, it should be recognised and explained to the patient that an ultrasound scan might not reliably detect or exclude treatable pathology within the uterine cavity. A trial of a pharmacological agent was deemed a better option than undergoing an ultrasound scan unless there are specific indications such as suspected pelvic mass or adenomyosis.

The evidence on the most effective imaging strategy for suspected adenomyosis showed TVUS to have better accuracy in terms of sensitivity and specificity compared to MRI and TAUS. The committee agreed that, despite the relatively small amount of evidence, the evidence reflected their expectations from a clinical perspective. In practice, 2D-TVUS is widely available and even though it is intrusive as the transducer is inserted in the vagina, it is known to be a widely accepted method among women and has greater diagnostic accuracy than transabdominal scanning of pelvic organs. 3D-TVUS, which showed better accuracy albeit with limited evidence, is rarely available outside of tertiary care or research centres. MRI is less accurate, less widely available and more expensive. It can also be an unpleasant or claustrophobic experience for women. 2D-TAUS on the other hand is widely available but has low accuracy. It can also cause discomfort in women since it

requires a full bladder. The committee, therefore, agreed that MRI and 2D-TAUS should not be recommended as the first-line diagnostic investigation for women with suspected adenomyosis.

However, the committee recognised that TVUS is not appropriate for all women and some women may decline it, for example adolescents, women who have not been sexually active, women who have had a traumatic sexual experience, or women with history of female genital mutilation. Therefore, the committee agreed that TAUS or MRI could be offered as an alternative as long as the lower accuracy of these tests was taken into consideration. The committee also considered that some women might also prefer a female practitioner to perform the test.

In addition to HMB, pelvic pain is a common symptom of adenomyosis. The committee agreed that these symptoms could also be signs of endometriosis and NICE guideline on Endometriosis: diagnosis and management (in development) should be referred to.

The committee noted that while diagnostic accuracy parameters give useful information about the test's ability to detect a condition or an absence of a condition, no matter how accurate a test is, it does not automatically indicate a better overall outcome for the woman. The overall outcome for the woman depends on a myriad of factors, including the treatment decisions after the diagnosis is made. According to the committee, a correct diagnosis of the cause of heavy menstrual bleeding is key in optimising treatment outcomes. The committee discussed the benefits and harms of false positives and false negatives. False positives can lead to over treatment, especially unnecessary surgical treatments, including hysterectomy which could carry risk and ideally should be avoided. However, it was thought by the committee that some women (who did not wish to preserve fertility or uterus) would still find hysterectomy acceptable, although a more aggressive treatment, because it would take away the symptoms (HMB and other symptoms such as pain or pressure symptoms) and improve quality of life. On the other hand, false negatives can prevent women from receiving a treatment that would improve their quality of life or ease the symptoms.

The committee also discussed that it can be important for a woman to obtain a diagnosis in order to explain the cause for her symptoms, even if the treatment strategy would be similar regardless of the diagnosis The committee recognised that many women have to endure symptoms without an identified pathology, but felt that some women may be reassured from knowing that there is no significant or treatable pathology.

Cost-effectiveness and resource use

A high quality economic evaluation from the UK (Cooper 2014) compared a number of diagnostic strategies for HMB. This analysis took an NHS perspective and the setting was a 'one-stop' secondary care clinical setting. The study concluded that either outpatient hysteroscopy or outpatient hysteroscopy in combination with endometrial biopsy represented cost-effective strategies for HMB. Treatment effectiveness was estimated through patient satisfaction although the authors also derived a cost per QALY estimate based on this.

An original economic model developed for the guideline compared combined diagnostic and treatment strategies for HMB. The diagnostic aspect was modelled using the approach used in the above study (Cooper 2014). However, treatment effectiveness was derived from 2 NMAs on short and long term quality of life based on EQ-5D data. This analysis suggested that endometrial biopsy was not cost-

effective relative to TVUS and outpatient hysteroscopy. However, the findings with respect to the relative cost-effectiveness of TVUS and outpatient hysteroscopy were less clear cut. Outpatient hysteroscopy was generally the cheapest strategy but TVUS generated a higher QALY gain, which was due to the ability of TVUS to detect intramural and subserosal fibroids. In many of the analyses the model suggested that this additional QALY gain warranted the additional cost of TVUS relative to outpatient hysteroscopy at a cost-effectiveness threshold of £20,000 per QALY. However, that finding was sensitive to the prevalence of intramural and subserosal fibroids as an underlying cause of HMB.

Diagnostic accuracy is important as it can lead to better targeted treatment, however this depends in the context of HMB, on the extent to which less appropriate treatments for a given underlying benefit may, neverthess, confer some treatment benefit. Outpatient hysteroscopy was the most expensive diagnostic test but the least expensive diagnostic strategy. An important contributing factor to this is that hysteroscopy can facilitate a one stop 'see-and-treat' approach which reduces treatment cost. The superior diagnostic accuracy of outpatient hysteroscopy for polyps, submucosal fibroids and no identified pathology means that this is likely to be the most cost-effective test in women where history is suggestive of those pathologies. Conversely, TVUS would be cost-effective where the woman's uterus is palpable abdominally or where history or examination suggests a pelvic mass.

It is anticipated that the guideline recommendation on investigations using hysteroscopy may increase the burden on secondary care at least in the short run. Furthermore, some units may have to adapt facilities and equipment in order to deliver hysteroscopy in lines with RCOG best practice guidelines (RCOG and BSGE 2011). However, it should be noted that the guideline recommends empiric LNG-IUS as a first-line treatment for many women presenting with HMB and therefore the population of women who would be offered hysteroscopy is only a subset of all women with HMB. The committee also believed that hysteroscopy could lead to a reduction in unnecessary scans and biopsies by better targeting treatment at the underlying pathology.

The committee have also recommended that 'blind' biopsy should not be offered and the committee expect that to generate savings to the NHS by also reducing the number of unnecessary biopsies.

Whilst there are recommendations in the guideline that will have a resource impact on the NHS, especially in the short term, it is not anticipated that any will have a significant resource impact (£1 million). Outpatient hysteroscopy is a more expensive investigation than pelvic ultrasound but there are potentiall off-setting savings to treatment costs as the technique can allow a 'see and treat' approach. It is expected that the recommendation to not perform 'blind' biopsy will lead to savings to the NHS but in the absence of data outlining baseline current practice it is not possible to quantify the potential size of any saving.

Other factors the committee took into account

The committee agreed that no special consideration regarding diagnosis is needed for women who wish to preserve their fertility since hysteroscopy, TVUS, TAUS and MRI do not affect fertility.

During the scoping phase the following groups were identified as needing special consideration:

- women who have difficulties communicating in English

- women with learning difficulties
- women from some minority ethnic groups (because women from some minority ethnic group might find it difficult to talk about HMB with health care professionals)
- women with from disadvantaged socio-economic group.

The committee noted that these group might need special consideration in terms of information provision and communication, however, the committee agreed that this is not specific to HMB and is covered by the NICE guideline on Patient experience in adult NHS services: improving the experience of care for people using adult NHS services (CG138). The committee also recognised that some women might prefer a female practitioner when discussing HMB and undergoing diagnostic tests.

Updated recommendations have taken into account new guidelines on the identification and management of high risk groups for endometrial hyperplasia (RCOG and BSGE 2016) and cancer (NICE guideline on suspected cancer: recognition and referral (NG12)), as these groups may also present with heavy menstrual bleeding. The committee agreed that in the case of suspected cancer, the clinicians should refer to NICE guideline on suspected cancer: recognition and referral (NG12).

In the evidence review on the diagnostic accuracy of hysteroscopy an ultrasound in investigating women with HMB, some studies used hysteroscopy as the reference standard. Even though in general histopathology was considered as the gold standard for detecting different uterine pathology, the committee discussed that blind biopsy can easily miss focal pathology (polyps and fibroids distorting to the cavity) and therefore, hysteroscopy with a direct visualisation of the uterine cavity (rather than histopathology) can be considered as the reference standard for these pathologies.

No evidence from test and treat RCTs was identified in this review. The committee agreed that a research recommendation is needed because a test-plus-treatment study design would offer better evidence of the clinical and cost-effectiveness benefits of choosing one test over another. The committee agreed it would be important to include empiric treatment as one of the treatment arms to reflect current clinical practice, as many women presenting with HMB in primary care are often offered pharmacological treatment without prior investigations. This research is needed because although diagnostic accuracy parameters provide useful information about a test's ability to detect the presence or absence of a condition, no matter how accurate a test is, it does not automatically translate to better clinical outcomes for the woman. The value of a diagnostic test lies in whether it actually guides treatment decisions (see Appendix C – Research recommendations).

See Appendix J – Diagnostic care path for illustration of the diagnostic recommendations.

References

Abd Elkhalek 2016

Abd Elkhalek, Y. I., Kamel, O. F., El-Sabaa, H., Comparison of 3 dimensional sonohysterography and hysteroscopy in Premenopausal women with abnormal uterine bleeding, Egyptian Journal of Radiology and Nuclear Medicine, 47, 1117-22, 2016

Abdel Hak 2010

Abdel Hak, A. M., Accuracy of sonographic criteria for diagnosis of adenomyosis in perimenopausal women with menorrhagia, Middle East Fertility Society Journal, 15, 35-8, 2010

Abe 2008

Abe, M., Ogawa, H., Ayhan, A., The use of non-three-layer ultrasound in biopsy recommendation for premenopausal women, Acta Obstetricia et Gynecologica Scandinavica, 87, 2008

Alborzi 2007

Alborzi, S., Parsanezhad, M. E., Mahmoodian, N., Alborzi, S., Alborzi, M., Sonohysterography versus transvaginal sonography for screening of patients with abnormal uterine bleeding, International Journal of Gynaecology & Obstetrics, 96, 20-3, 2007

Bazot 2002

Bazot, M., Darai, E., Rouger, J., Detchev, R., Cortez, A., Uzan, S., Limitations of transvaginal sonography for the diagnosis of adenomyosis, with histopathological correlation, Ultrasound in Obstetrics and Gynecology, 20, 605-11, 2002

Botsis 1998

Botsis, D., Kassanos, D., Antoniou, G., Pyrgiotis, E., Karakitsos, P., Kalogirou, D., Adenomyoma and leiomyoma: differential diagnosis with transvaginal sonography, Journal of Clinical Ultrasound, 26, 21-5, 1998

Cooper 2014

Cooper, N. A., Barton, P. M., Breijer, M., Caffrey, O., Opmeer, B. C., Timmermans, A., Mol, B. W., Khan, K. S., Clark, T. J., Cost-effectiveness of diagnostic strategies for the management of abnormal uterine bleeding (heavy menstrual bleeding and post-menopausal bleeding): a decision analysis, Health Technology Assessment, 18, 1-201, 2014

Champaneria 2010

Champaneria, R., Abedin, P., Daniels, J., Balogun, M., Khan, K.S., Ultrasound scan and magnetic resonance imaging for the diagnosis of adenomyosis: systematic review comparing test accuracy, Acta Obstetricia et Gynecologica, 89, 1374–84, 2010

Cicinelli 1995

Cicinelli, E., Romano, F., Anastasio, P. S., Blasi, N., Parisi, C., Galantino, P., Transabdominal sonohysterography, transvaginal sonography, and hysteroscopy in

the evaluation of submucous myomas, Obstet GynecolObstetrics and gynecology, 85, 42-7, 1995

Critchley 2004

Critchley, H. O. D., Warner, P., Lee, A. J., Brechin, S., Guise, J., Graham, B., Evaluation of abnormal uterine bleeding: Comparison of three outpatient procedures within cohorts defined by age and menopausal status, Health Technology Assessment, 8, iii-77, 2004

Dakhly 2016

Dakhly, D. M. R., Abdel Moety, G. A. F., Saber, W., Gad Allah, S. H., Hashem, A. T., Abdel Salam, L. O. E., Accuracy of Hysteroscopic Endomyometrial Biopsy in Diagnosis of Adenomyosis, Journal of Minimally Invasive Gynecology, 23, 364-71, 2016

Dasgupta 2011a

Dasgupta, S., Chakraborty, B., Karim, R., Aich, R. K., Mitra, P. K., Ghosh, T. K., Abnormal uterine bleeding in peri-menopausal age: Diagnostic options and accuracy, Journal of Obstetrics and Gynecology of India, 61, 189-94, 2011a

Dasgupta 2011b

Dasgupta, S., Sharma, P. P., Mukherjee, A., Ghosh, T. K., Ultrasound assessment of endometrial cavity in perimenopausal women on oral progesterone for abnormal uterine bleeding: comparison of diagnostic accuracy of imaging with hysteroscopy-guided biopsy, The journal of obstetrics and gynaecology research, 37, 2011b

Dueholm 2001a

Dueholm, M., Forman, A., Jensen, M. L., Laursen, H., Kracht, P., Transvaginal sonography combined with saline contrast sonohysterography in evaluating the uterine cavity in premenopausal patients with abnormal uterine bleeding, Ultrasound in Obstetrics and Gynecology, 18, 54-61, 2001a

Dueholm 2001b

Dueholm, M., Lundorf, E., Hansen, E. S., Sorensen, J. S., Ledertoug, S., Olesen, F., Magnetic resonance imaging and transvaginal ultrasonography for the diagnosis of adenomyosis, Fertility and Sterility, 76, 588-94, 2001b

Erdem 2007

Erdem, M., Bilgin, U., Bozkurt, N., Erdem, A., Comparison of transvaginal ultrasonography and saline infusion sonohysterography in evaluating the endometrial cavity in pre- and postmenopausal women with abnormal uterine bleeding, Menopause, 14, 2007

Exacoustos 2011

Exacoustos, C., Brienza, L., Di Giovanni, A., Szabolcs, B., Romanini, M. E., Zupi, E., Arduini, D., Adenomyosis: three-dimensional sonographic findings of the junctional zone and correlation with histology, Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology, 37, 471-9, 2011

Fakhar and Mahmud 2010

Fakhar, S., Mahmud, G., Validity of hysteroscopy and histopathology in patients with menstrual irregularity, Journal of Ayub Medical College, Abbottabad: JAMC, 22, 129-32, 2010

Krampl 2001

Krampl, E., Bourne, T., Hurlen-Solbakken, H., Istre, O., Transvaginal ultrasonography sonohysterography and operative hysteroscopy for the evaluation of abnormal uterine bleeding, Acta Obstetricia et Gynecologica Scandinavica, 80, 616-622, 2001

Meredith 2009

Meredith, S. M., Sanchez-Ramos, L., Kaunitz, A. M., Diagnostic accuracy of transvaginal sonography for the diagnosis of adenomyosis: systematic review and metaanalysis. American Journal of Obstetrics and Gynecology, 201:107, e1-6, 2009

Mukhopadhayay 2007

Mukhopadhayay, S., Bhattacharyya, S. K., Ganguly, R. P., Patra, K. K., Bhattacharya, N., Barman, S. C., Comparative evaluation of perimenopausal abnormal uterine bleeding by transvaginal sonography, hysteroscopy and endometrial biopsy, Journal of the Indian Medical Association, 105, 2007

Najeeb 2010

Najeeb, R., Awan, A. S., Bakhtiar, U., Akhter, S., Role of transvaginal sonography in assessment of abnormal uterine bleeding in perimenopausal age group, Journal of Ayub Medical College, Abbottabad : JAMC, 22, 2010

Nanda 2002

Nanda, S., Chadha, N., Sen, J., Sangwan, K., Transvaginal sonography and saline infusion sonohysterography in the evaluation of abnormal uterine bleeding, Australian and New Zealand Journal of Obstetrics and Gynaecology, 42, 530-4, 2002

RCOG and BSGE 2016

Royal Coll Royal College of Obstetricians and Gynaecologists, British Society for Gynaecological Endoscopy, Management of Endometrial Hyperplasia, Green-top Guideline No. 67, London: RCOG, 2016

RCOG and BSGE 2011

Royal College of Obstetricians and Gynaecologists, British Society for Gynaecological Endoscopy, Best Practice in Outpatient Hysteroscopu, Green-top Guideline No. 59, London: RCOG, 2011

Soguktas 2012

Soguktas, S., Cogendez, E., Kayatas, S. E., Asoglu, M. R., Selcuk, S., Ertekin, A., Comparison of saline infusion sonohysterography and hysteroscopy in diagnosis of premenopausal women with abnormal uterine bleeding, European Journal of Obstetrics, Gynecology, & Reproductive BiologyEur J Obstet Gynecol Reprod Biol, 161, 2012

Taylor 2001

Taylor, S., Jones, S., Dixon, A. M., O'Donovan, P., Evaluation of ultrasound in an outpatient hysteroscopy clinic: Does it alter management in premenopausal women?, Gynaecological Endoscopy, 10, 173-8, 2001

Vercellini 1998

Vercellini, P., Cortesi, I., De Giorgi, O., Merlo, D., Carinelli, S. G., Crosignani, P. G., Transvaginal ultrasonography versus uterine needle biopsy in the diagnosis of diffuse adenomyosis, Human Reproduction, 13, 1998

Vercellini 1997

Vercellini, P., Cortesi, I., Oldani, S., Moschetta, M., De Giorgi, O., Crosignani, P. G., The role of transvaginal ultrasonography and outpatient diagnostic hysteroscopy in the evaluation of patients with menorrhagia, Human Reproduction, 12, 1768-71, 1997

Williams and Marshburn 1998

Williams, C. D., Marshburn, P. B., A prospective study of transvaginal hydrosonography in the evaluation of abnormal uterine bleeding, Am J Obstet GynecolAmerican journal of obstetrics and gynecology, 179, 292-8, 1998

Yildiz 2009

Yildiz, A., Koksal, A., Ates, P. F., Ivit, H., Keklik, A., Cukurova, K., Hysteroscopy in the evaluation of intrauterine cavity. Is it more valuable than dilatation and curettage?, Turkiye Klinikleri Journal of Medical Sciences, 29, 2009

Appendices

Appendix A – Review protocols

Protocol for review question 1. Diagnostic accuracy of ultrasound and hysteroscopy for investigation of women presenting with heavy menstrual bleeding

Component	Description
Review question	What is the diagnostic accuracy of ultrasound and hysteroscopy for investigation of women presenting with heavy menstrual bleeding?
Objective	The CG44 surveillance report identified 1 UK based study evaluating the cost-effectiveness of different diagnostic strategies for HMB (including transvaginal scanning, endometrial biopsy, saline infusion sonography and outpatient hysteroscopy). The study concluded that outpatient hysteroscopy was the most cost-effective option for women referred to secondary care for investigation of HMB. Endometrial biopsy, in addition to hysteroscopy, had an incremental cost-effectiveness ratio of £21,000 per additional satisfied patient. Current guidance recommends the use of ultrasound as the first-line diagnostic tool, with hysteroscopy only used when ultrasound results are inconclusive. Therefore, the GC44 surveillance report concluded that this new evidence could have a potential impact on current recommendations. It is important to note, however, that cost-effectiveness was not assessed in terms of cost per quality adjusted life year, as is the usual NICE approach. The CG44 surveillance report did not identify other evidence that could affect current recommendations. Based on this, the guideline committee agreed it is necessary to look at the evidence comparing the effectiveness and cost-effectiveness of hysteroscopy compared with ultrasound. The objective of this review is to determine the diagnostic accuracy of hysteroscopy compared to ultrasound in diagnosing structural causes for heavy menstrual bleeding in women presenting with heavy menstrual bleeding.
Population	Women of reproductive age (post-pubertal and premenopausal) who present with heavy menstrual bleeding. We will include women with abnormal uterine bleeding, and studies where more than 66% of the population have heavy menstrual bleeding.
Subgroups and sensitivity analyses	The following sub-groups will be assessed separately in the presence of heterogeneity: • women with suspected fibroids • women who have inter-menstrual bleeding in addition to heavy menstrual bleeding

	• women who had a previous index test (hysteroscopy or ultrasound scan) • time of publication (pre 2007, post 2007)

Index test:	• Transvaginal ultrasound scan (TVUS)
	  ○ two-dimensional (2D)
	  ○ three-dimensional (3D)
	• Transabdominal ultrasound scan (TAUS)
	  ○ 2D
	  ○ 3D
	• Hysteroscopy
	  ○ inpatient/day case (under general anaesthesia/sedation or spinal/epidural anaesthesia)
	  ○ outpatient vaginoscopy (alternative method for performing hysteroscopy that does not require inserting a vaginal speculum or other cervical instruments)
	  ○ outpatient all other cases
	• Transvaginal ultrasound scan followed by hysteroscopy if needed
	
	Note: saline infusion sonography was not included as an index test because the committee agreed with the old guideline recommendation that it should not be used as a first-line diagnostic investigation.

Reference standard	• Histopathologic diagnosis (biopsy or surgery specimen) • Ultrasound scan (when compared with hysteroscopy) • Inpatient hysteroscopy (when compared to different outpatient types)
Outcomes	• Sensitivity • Specificity • Positive likelihood ratio (LR+) • Negative likelihood ratio (LR-) • Area under the curve (AUC) if meta-analysis can be conducted • Patient satisfaction and acceptability of the test, including pain score Results will be presented per pathology, for example submucosal fibroids, polyps, hyperplasia, cancer, abnormal uterine thickness, synechiae, endometritis, retained products of conception (RPOC), fibroids smaller than 3 cm in diameter, fibroids larger than 3cm in diameter, congenital abnormalities.
Importance of outcomes	Critical outcomes: • sensitivity • LR+ • LR- • patient satisfaction
Study design	Studies in which the index test and the reference standard would be compared in the same individuals and 2x2 tables will be constructed: • test and treat randomised controlled trials (RCTs) or systematic reviews of test and treat RCTs • cross-sectional studies

	• prospective cohort studies (where cross-sectional data were reported therefore 2 x 2 table could be tabulated) Exclude: case-control studies
Setting	Any healthcare setting where NHS care is delivered (primary, secondary, tertiary or community)
Search strategy	Sources searched: Embase, Medline, PreMedline, CENTRAL, CDSR, DARE Limits: The search undertaken for Clinical guideline CG44 met all quality criteria and was updated for articles published between 2007 (the end of the search period for CG44) to 13 October 2016. As a diagnostic review, the search strategy included all study types. Supplementary search techniques: Checking reference lists of included and potentially relevant studies.
Review strategy	Appraisal of methodological quality • The methodological quality of each study will be assessed using the QUADAS-2 checklist • The quality of the evidence for an outcome (i.e. across studies) will be assessed using adapted GRADE. • Studies with 80-99% women with HMB will be downgraded once for indirectness. Studies with 66-80% women with HMB, or where the proportion of women with HMB is not specified, will be downgraded twice for indirectness. Synthesis of data • Meta-analysis will be conducted when appropriate The cut-offs for diagnostic accuracy measures: • sensitivity and specificity: ○ high: higher than 90% ○ moderate: 75-90% ○ low: lower than 75% • positive likelihood ratio: ○ very useful test: higher than 10 ○ moderately useful test: 5-10 ○ not a useful test: lower than 5 • negative likelihood ratio: ○ very useful test: lower than 0.1 ○ moderately useful test: 0.1 to 0.2 ○ not a useful test: higher than 0.2
Equalities	Potential equality issues were considered. The following groups were identified in the scoping phase: • women who have difficulties communicating in English • women with learning difficulties • women from minority ethnic groups • women from disadvantaged socio-economic groups

Notes/additional information	The level of experience of the investigators should be considered. Clinician's judgement should be independent of index test/reference standard results and be blinded from the index test/reference standard results. |If blinding is not done, the evidence will be downgraded according to QUADAS-2.
Key papers	• Surveillance review of CG44: Heavy menstrual bleeding. • Critchley, H. O., Warner, P., Lee, A. J., Brechin, S., Guise, J., Graham, B., Evaluation of abnormal uterine bleeding: comparison of three outpatient procedures within cohorts defined by age and menopausal status, Health Technology Assessment, 8, 1-139, 2004 • Cooper, N. A., Barton, P. M., Breijer, M. et al., Cost-effectiveness of diagnostic strategies for the management of abnormal uterine bleeding (heavy menstrual bleeding and post-menopausal bleeding): a decision analysis, Health Technology Assessment, 18, 1-201, 2014

2D: two-dimenstional; 3D: three-dimensional; AUC: area under the curve; LR-: negative likelihood ratio; LR+: positive likelihood ratio; RCT: randomised controlled trial; TAUS: transabdominal ultrasound scan; TVUS: transvaginal ultrasound scan

Protocol for review question 2. Clinically effective imaging strategy for detecting adenomyosis in women presenting with heavy menstrual bleeding

Component	Description
Review question	What is the most clinically effective imaging strategy for diagnosing adenomyosis in women with heavy menstrual bleeding?
Objective	Adenomyosis is a benign gynaecological disorder in which the inner lining of the uterus breaks through the muscle wall of the uterus. The most common symptoms are pelvic pain and heavy menstrual bleeding. The diagnosis of adenomyosis is challenging because the symptoms are similar to other uterine pathology and might coexist with other uterine abnormalities. Traditionally adenomyosis was only diagnosed through histopathology after hysterectomy. Improvements in diagnostic imaging in recent years have shown that adenomyosis can be diagnosed with non-invasive methods as well. However, currently there are no standard diagnostic imaging criteria for adenomyosis. The objective of this review is to assess the accuracy of imaging techniques in diagnosing adenomyosis in women presenting with heavy menstrual bleeding.
Population	Women of reproductive age (post-pubertal and premenopausal) who present with heavy menstrual bleeding. Studies with indirect population (women with pelvic pain or dysmenorrhea) will be considered but at least two thirds of the population in the study should be women with heavy menstrual bleeding.
Subgroups and sensitivity analyses	The following sub-groups will be assessed separately in the presence of heterogeneity: • women who have coexisting uterine abnormalities such as fibroids • women who have inter-menstrual bleeding in addition to heavy menstrual bleeding
Index test: Severity assessment	• Magnetic resonance imaging (MRI) • Transvaginal ultrasound scan (TVUS) ○ two-dimensional (2D)

Component	Description
tools/clinical markers	○ three-dimensional (3D) • Transabdominal ultrasound scan (TAUS) ○ 2D ○ 3D • Combination of two or more of the above
Reference standard or target condition/patient outcomes	• Histopathologic diagnosis (biopsy or surgery specimen) • Ultrasound (when compared with MRI)
Outcomes	• Sensitivity • Specificity • Positive likelihood ratio (LR+) • Negative likelihood ratio (LR-) • Area under the curve (AUC) if meta-analysis can be conducted
Importance of outcomes	Critical outcomes: • sensitivity • LR+ • LR-
Study design	Studies in which the index test and the reference standard would be compared in the same individuals and 2x2 tables will be constructed: • test and treat randomised controlled trials (RCTs) or systematic reviews of test and treat RCTs • cross-sectional studies • prospective cohort studies where cross-sectional data were reported therefore 2 x 2 table could be tabulated Exclude: case-control studies
Setting	Any healthcare setting where NHS care is delivered (primary, secondary, tertiary or community)
Search strategy	Sources searched: Embase, Medline, PreMedline, CENTRAL, CDSR, DARE Limits: The search undertaken for Clinical guideline CG44 met all quality criteria and was updated for articles published between 1996 and 13 October 2016. As a diagnostic review, the search strategy included all study types. Supplementary search techniques: Checking reference lists of included and potentially relevant studies.
Review strategy	Appraisal of methodological quality • The methodological quality of each study will be assessed using the QUADAS-2 checklist • The quality of the evidence for an outcome (i.e. across studies) will be assessed using adapted GRADE.

Component	Description
	• Studies with 80-99% women with HMB will be downgraded once for indirectness. Studies with 66-80% women with HMB, or where the proportion of women with HMB is not specified, will be downgraded twice for indirectness. Synthesis of data • Meta-analysis will be conducted when appropriate The cut-offs for diagnostic accuracy measures: • sensitivity and specificity: ○ high: higher than 90% ○ moderate: 75-90% ○ low: lower than 75% • positive likelihood ratio: ○ very useful test: higher than 10 ○ moderately useful test: 5-10 ○ not a useful test: lower than 5 • negative likelihood ratio: ○ very useful test: lower than 0.1 ○ moderately useful test: 0.1 to 0.2 ○ not a useful test: higher than 0.2
Equalities	Potential equality issues were considered. The following groups were identified in the scoping phase: women who have difficulties communicating in English women with learning difficulties women from minority ethnic groups women from disadvantaged socio-economic groups
Notes/additional information	The level of experience of the investigators should be considered. Clinician's judgement should be independent of index test/reference standard results and be blinded from the index test/reference standard results. |If blinding is not done, the evidence will be downgraded according to QUADAS-2.
Key papers	• Champaneria, R., Abedin, P., Daniels, J., Balogun, M., Khan, K.,S. Ultrasound scan and magnetic resonance imaging for the diagnosis of adenomyosis: systematic review comparing test accuracy, Acta Obstet Gynecol Scand, 89, 1374-84, 2010 • Cooper, N. A., Barton, P. M., Breijer, M., Caffrey, O., Opmeer, B. C., Timmermans, A., Mol, B. W., Khan, K. S., Clark, T. J., Cost-effectiveness of diagnostic strategies for the management of abnormal uterine bleeding (heavy menstrual bleeding and post-menopausal bleeding): a decision analysis, Health Technology Assessment, 18, 1-201, 2014 • Herman, M. C., Mol, B. W., Bongers, M. Y., Diagnosis of heavy menstrual bleeding, Women's Health, 12, 15-20, 2016 • Meredith, S. M., Sanchez-Ramos, L., Kaunitz, A. M., Diagnostic accuracy of transvaginal sonography for the diagnosis of adenomyosis: systematic review and metaanalysis, American Journal of Obstetrics and Gynecology, 201, 107.e1-107.e6, 2009

Component	Description
	• Senturk, L. M., Imamoglu, M., Adenomyosis: what is new?, Womens Health, 11, 717–724, 2015 • Struble, J., Reid, S., Bedaiwy, M. A., Adenomyosis: A Clinical Review of a Challenging Gynecologic Condition, Journal of Minimally Invasive Gynecology, 23, 164-185, 2016

2D: two-dimenstional; 3D: three-dimensional; AUC: area under the curve; LR-: negative likelihood ratio; LR+: positive likelihood ratio; MRI: magnetic resonance imaging; RCT: randomised controlled trial; TAUS: transabdominal ultrasound scan; TVUS: transvaginal ultrasound scan

Appendix B – Health economic quality assessment

Study identification **Cooper, N. A., Barton, P. M., Breijer, M., Caffrey, O., Opmeer, B. C., Timmermans, A., Mol, B. W., Khan, K. S., Clark, T. J., Cost-effectiveness of diagnostic strategies for the management of abnormal uterine bleeding (heavy menstrual bleeding and post-menopausal bleeding): a decision analysis, Health Techonology Assessment, 18, 2014**		
Guidance topic: HMB		**Question no: 1**
Section 1: Applicability (relevance to specific review questions and the NICE reference case as described in section 7.5)	Yes/partly/no/unclear/NA	Comments
1.1 Is the study population appropriate for the review question?	Yes	Women referred to secondary care by their GP presenting with HMB
1.2 Are the interventions appropriate for the review question?	Yes	13 different investigation scenarios. TVUS, SIS, global EBx and OPH used in alone, or in combination. No investigation was also considered where treatment was initiated with LNG-IUS or hysterectomy.
1.3 Is the system in which the study was conducted sufficiently similar to the current UK context?	Yes	UK
1.4 Are the perspectives clearly stated and are they appropriate for the review question?	Yes	NHS
1.5 Are all direct effects on individuals included, and are all other effects included where they are material?	Yes	
1.6 Are all future costs and outcomes discounted appropriately?	N/A	Time horizon 1 year
1.7 Is QALY used as an outcome, and was it derived using NICE's preferred methods? If not, describe rationale and outcomes used in line with analytical perspectives taken (item 1.4 above).	No	Satisfaction used as the measure of effectiveness. Assumptions to translate

Study identification **Cooper, N. A., Barton, P. M., Breijer, M., Caffrey, O., Opmeer, B. C., Timmermans, A., Mol, B. W., Khan, K. S., Clark, T. J., Cost-effectiveness of diagnostic strategies for the management of abnormal uterine bleeding (heavy menstrual bleeding and post-menopausal bleeding): a decision analysis, Health Techonology Assessment, 18, 2014**		
		satisfaction to QALYs was reported in places.
1.8 Are costs and outcomes from other sectors fully and appropriately measured and valued?	Yes	
1.9 Overall judgement: Directly applicable		
Other comments: This study does not use the preferred measure of effects (QALYs), but is still thought to be useful for decision making given that all other criteria are applicable and the alternative outcome measure reported is unlikely to change the conclusions about cost-effectiveness		
Section 2: Study limitations (the level of methodological quality)	**Yes/partly/ no/unclear /NA**	**Comments**
2.1 Does the model structure adequately reflect the nature of the topic under evaluation?	Yes	Decision tree informed by clinical expertise
2.2 Is the time horizon sufficiently long to reflect all important differences in costs and outcomes?	Partly	Time horizon 1 year. Chosen primarily as satisfaction data was not reported beyond this
2.3 Are all important and relevant outcomes included?	Unclear	Satisfaction is treatment specific rather than health state specific and may not be able to incorporate all changes in health-related quality of life that can occur along the pathway
2.4 Are the estimates of baseline outcomes from the best available source?	Yes	Parameter inputs were derived from systematic reviews, individual patient data and focused searches and in the absence of estimates, the consensus view of an expert clinical panel was obtained.
2.5 Are the estimates of relative intervention effects from the best available source?	Yes	See 2.4
2.6 Are all important and relevant costs included?	Yes	See 2.4
2.7 Are the estimates of resource use from the best available source?	Yes	See 2.4
2.8 Are the unit costs of resources from the best available source?	Yes	NHS Reference Costs, PSSRU, BNF
2.9 Is an appropriate incremental analysis presented or can it be calculated from the data?	Yes	
2.10 Are all important parameters whose values are uncertain subjected to appropriate sensitivity analysis?	Yes	Extensive SA and PSA (1,000 iterations)
2.11 Is there any potential conflict of interest?	No	

Study identification
Cooper, N. A., Barton, P. M., Breijer, M., Caffrey, O., Opmeer, B. C., Timmermans, A., Mol, B. W., Khan, K. S., Clark, T. J., Cost-effectiveness of diagnostic strategies for the management of abnormal uterine bleeding (heavy menstrual bleeding and post-menopausal bleeding): a decision analysis, Health Techonology Assessment, 18, 2014
2.12 Overall assessment: Potentially serious limitations

Appendix C – Research recommendations

Research recommendation 1. Hysteroscopy compared with ultrasound or empiric pharmacological treatment in the diagnosis and management of heavy menstrual bleeding

Is initial testing using hysteroscopy more effective than testing with pelvic ultrasound or empiric pharmacological treatment in the diagnosis and management of heavy menstrual bleeding?

Why is this important?

There is no consensus about the best test-and-treat strategy for women with HMB and empiric pharmacological treatment is often initiated as a first treatment without investigation. Parameters of diagnostic accuracy give useful information about a test's ability to detect a condition (or the absence of a condition). But accurate diagnosis does not automatically result in a better overall outcome for the woman, because this also depends on treatment decisions after the diagnosis is made. However, it is thought that optimal treatment depends on accurate diagnosis of the underlying pathology causing HMB. In the absence of clinical trials, decision analytical economic models evaluating all possible outpatient testing algorithms have indicated that using ultrasound or hysteroscopy for initial diagnostic testing for women with HMB are the most effective diagnostic strategies. Pelvic ultrasound has been most commonly used because it has been more widely available and is considered less intrusive than hysteroscopy. However, advances in technology mean that the hysteroscopy is well tolerated in the outpatient setting, and it can potentially be performed outside the traditional hospital environment in a community setting. Moreover, in contrast with ultrasound, hysteroscopy allows concomitant treatment of intrauterine pathologies such as endometrial polyps and submucosal fibroids. It also facilitates the fitting of levonorgesterol-releasing intrauterine systems (LNG-IUS). A test-and-treat randomised controlled trial with cost-effectiveness analysis could help to answer the crucial question of whether hysteroscopy improves outcomes for women and results in more effective use of NHS resources.

Table 7: Reseach recommendation rationale

Research question	Is initial testing with hysteroscopy more effective than with pelvic ultrasound or empiric pharmacological treatment in the diagnosis and management of women with heavy menstrual bleeding?
Why this is needed	
Importance to 'patients' or the population	Heavy menstrual bleeding (HMB) is common and associated with significant morbidity including adverse impacts upon all domains of health related quality of life, restriction of daily activities including absenteeism from work and other responsibilities including impairment of family roles.
Relevance to NICE guidance	Evidence from economic decision analytical modelling has shown that an initial outpatient diagnostic strategies of women with HMB based upon hysteroscopy or transvaginal ultrasound are the most cost-effective strategies. In practice, many women presenting with HMB in primary care are offered empiric pharmacological treatment without investivations. However, there is substantial uncertainty regarding which approach is the more cost-effective.
Relevance to the NHS	HMB is common and adversely affects women's quality of life and utilises significant health service resources. HMB is caused by several different pathologies. Correct diagnosis allows the most appropriate medical or surgical treatments to be selected, thereby optimising clinical outcomes and saving health care resources by avoiding the need for further diagnostic testing and additional treatment.
National priorities	Women's health.
Current evidence base	The NICE HMB guideline update 2017 did not identify evidence from diagnostic randomised controlled trials. The available diagnostic cohort studies inform us about the diagnostic accuracy of ultrasound and hysteroscopy compared to a reference standard, however it is not possible to assess whether the differences in accuracy are clinically important or the impact on patient outcomes. One RCT conducted in Scotland between 1999 and 2001 compared three outpatient diagnostic tests (outpatient endometrial biopsy, pelvic ultrasound and hysteroscopy) (Critchley 2004). The utility of the findings from this RCT are limited because a mixed population of women with abnormal uterine bleeding were evaluated (women with HMB and post-menopausal women with unexpected vaginal bleeding). This is because the aim of investigation of women with postmenopausal bleeding is to exclude the possibility of endometrial cancer, whereas in pre-menopausal women it is to optimise management of benign uterine pathologies associated with HMB (i.e. selection of appropriate treatment modalities). The trial did not evaluate clinical outcomes, rather feasibility and acceptability. Furthermore, technological advancement in the 20 years since this study was conceived has restricted the relevance of the findings especially because imaging has improved, instruments are miniaturised and less traumatic, hysteroscopy is more widely available and now conducted in an outpatient setting and failure rates are much lower than reported here. Several decision analytical economic models (Cooper 2014; De Vries 2000; NICE 2007) have been conducted to evaluate the optimal testing strategies for HMB from a cost-effectiveness standpoint. No consistent findings have been found because of limitations in the constructed models, namely differing assumptions and simplifications being made; not accounting for clinical examination findings, not accounting for the range of possible underlying diagnoses, limited accuracy and effectiveness data from sequential testing such that a heavy reliance has been necessary from 'expert opinion' and the

Research question	Is initial testing with hysteroscopy more effective than with pelvic ultrasound or empiric pharmacological treatment in the diagnosis and management of women with heavy menstrual bleeding?
	use of outmoded or restricted therapeutic options as well as eclectic outcome measures. Not surprisingly the findings of these analysis were sensitive to changes in the key assumptions limiting the robustness of clinical inferences. Overall, the findings from the RCT and economic models has shown that an initial outpatient diagnostic strategies of women with HMB based upon hysteroscopy or transvaginal ultrasound are potentially the most effective and cost-effective strategies. However, there are no published trials including empiric pharmacological treatment without prior investigation as a comparator. Therefore, there remains considerable uncertainty surrounding clinical effectiveness and patient acceptability / preferences. Only an RCT can resolve this uncertainty.
Equality	HMB affects women of all reproductive ages and races. Thus the optimising clinical outcomes through better diagnosis will benefit all women with this condition.
Feasibility	HMB is common. Pelvic ultrasound and outpatient hysteroscopy are now widely available tests in gynaecological practice. However, whilst there is general consensus regarding treatment options for HMB, according to the suspected underlying causes, there is variation in current practice for diagnostic work in HMB across the NHS. Thus, given the high prevalence of HMB, the widespread availability of tests and uncertainty amongst clinicians regarding optimal diagnostic work up, an RCT to evaluate the effectiveness and cost-effectiveness of initial testing strategies based upon pelvic ultrasound and outpatient hysteroscopy is highly feasible.
Other comments	References: • Cooper, N. A. M., Barton, P. M., Breijer, M., Caffrey, O., Opmeer, B. C., Timmermans, A., et al, Cost-effectiveness of diagnostic strategies for the management of abnormal uterine bleeding (heavy menstrual bleeding and post-menopausal bleeding): a decision analysis, Health Technology Assessment, 18, 2014 • Critchley, H. O. D., Warner, P., Lee, A. J., Brechin, S., Guise, J., Graham, B., Evaluation of abnormal uterine bleeding: comparison of three outpatient procedures within cohorts defined by age and menopausal status, Health Technology Assessment, 8, 2004 • De Vries, L. D., Dijkhuizen, F. P. H. L., Mol, B. W. J., Brolmann, H. A. M., Moret, E., Heintz, A. P. M., Comparison of transvaginal sonography, saline infusion sonography, and hysteroscopy in premenopausal women with abnormal uterine bleeding, Journal of Clinical Ultrasound, 28, 217-23, 2000 • National Institute for Health and Care Excellence (NICE), NICE guideline on Heavy menstrual bleeding: assessment and management (CG44), London: NICE, 2007

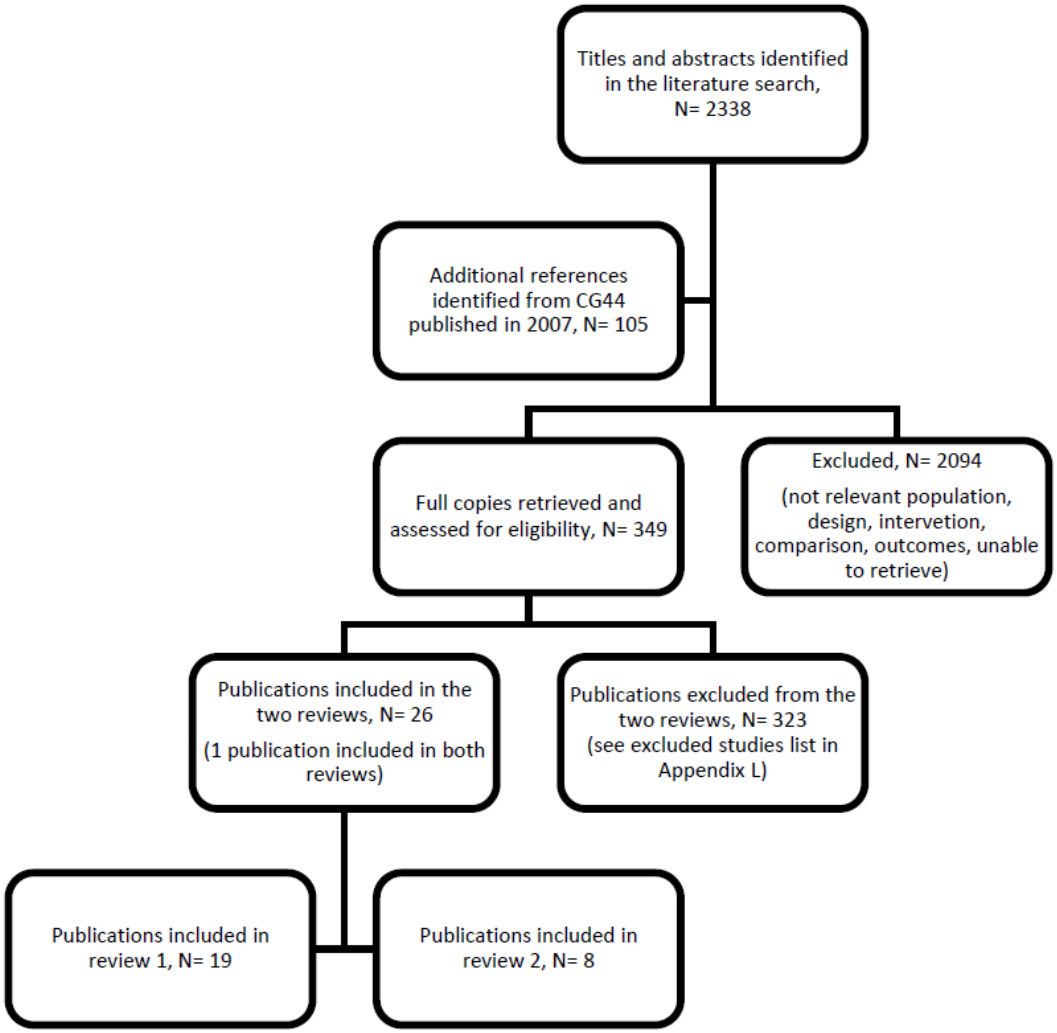
Table 8: Characteristics of the study design

Criterion	Explanation
Population	Women with heavy menstrual bleeding
Intervention	• Outpatient hysteroscopy followed by treatment • Pelvic ultrasound followed by treatment

Criterion	Explanation
	• Empiric pharmacological treatment only (no test)
Comparators	Any of the above
Outcomes	• Primary outcome at 12 months: Quality of life (condition specific); longer term effects measured at two years and 5 years • Secondary outcomes: Other clinical (e.g. satisfaction, generic quality of life, complications and feasibility of testing); qualitative (e.g. patient acceptability and preferences); economic (cost-utility (cost/QALY))
Study design	Test-and-treat randomised controlled trial
Timeframe	5 years

Appendix D – Clinical evidence study selection

Figure 2: Flow diagram of clinical article selection for reviews on diagnostic test accuracy for investigations of women presenting with heavy menstrual bleeding



Appendix E – Literature search strategies

Cochrane Library (CDSR, DARE, CENTRAL) – Wiley (last searched 13.10.2016)

ID	Search
#1	mesh descriptor: [menorrhagia] this term only
#2	(menorrhag* or hypermenorrh* or menometrorrhag* or metromenorrhag* or (menstru* near/3 (bleed* or blood loss)) or (heavy near/1 (period* or menses or menstruation)) or ((abnormal* or dysfunction*) near/3 (uterine or uterus) near/3 (bleed* or blood*))):ti
#3	(menorrhag* or hypermenorrh* or menometrorrhag* or metromenorrhag* or (menstru* near/3 (bleed* or blood loss)) or (heavy near/1 (period* or menses or menstruation)) or ((abnormal* or dysfunction*) near/3 (uterine or uterus) near/3 (bleed* or blood*))):ab
#4	#1 or #2 or #3
#5	mesh descriptor: [adenomyosis] this term only
#6	(adenomyo* or adenomyi*):ti
#7	(adenomyo* or adenomyi*):ab
#8	#5 or #6 or #7
#9	mesh descriptor: [area under curve] explode all trees
#10	mesh descriptor: [predictive value of tests] this term only
#11	mesh descriptor: [reproducibility of results] this term only
#12	mesh descriptor: [roc curve] this term only
#13	mesh descriptor: [sensitivity and specificity] this term only
#14	mesh descriptor: [validation studies] this term only
#15	(accurac* or accurat* or 'area under curve' or 'auc value*' or 'clinical utilit*' or (likelihood near/3 ratio*) or (diagnostic near/2 odds ratio*) or ((pretest or 'pre test' or posttest or 'post test') near/2 probabilit*) or (predict* near/3 value*) or 'receiver operating characteristic' or (roc near/2 curv*) or reliabil* or sensititiv* or ecificit* or valid*):ti
#16	(accurac* or accurat* or 'area under curve' or 'auc value*' or 'clinical utilit*' or (likelihood near/3 ratio*) or (diagnostic near/2 odds ratio*) or ((pretest or 'pre test' or posttest or 'post test') near/2 probabilit*) or (predict* near/3 value*) or 'receiver operating characteristic' or (roc near/2 curv*) or reliabil* or sensititiv* or specificit* or valid*):ab

ID	Search
#17	(satisf* or acceptab*):ti
#18	(satisf* or acceptab*):ab
#19	satisf*:kw
#20	#9 or #10 or #11 or #12 or #13 or #14 or #15 or #16
#21	#17 or #18 or #19 or #20
#22	mesh descriptor: [diagnostic imaging] this term only
#23	mesh descriptor: [imaging, three-dimensional] this term only
#24	mesh descriptor: [ultrasonography] explode all trees
#25	mesh descriptor: [ultrasonography] explode all trees
#26	mesh descriptor: [tomography] this term only
#27	mesh descriptor: [tomography, x-ray] explode all trees
#28	mesh descriptor: [magnetic resonance imaging] explode all trees
#29	mesh descriptor: [hysterosalpingography] this term only
#30	mesh descriptor: [hysteroscopy] this term only
#31	(ultrasonograph* or sonograph* or ultrasound* or ultrasonic* or echograph* or echotomograph* or endosonograph* or ((uterine or uterus or medical) near/3 imaging) or hysterosalpongograph* or hysterograph* or hysterosonograph* or sonohysterograph* or (magnetic near/3 (resonance or imaging)) or (mri or mr imaging or nmr) or hysteroscop* or taus or tvus):ti
#32	(ultrasonograph* or sonograph* or ultrasound* or ultrasonic* or echograph* or echotomograph* or endosonograph* or ((uterine or uterus or medical) near/3 imaging) or hysterosalpongograph* or hysterograph* or hysterosonograph* or sonohysterograph* or (magnetic near/3 (resonance or imaging)) or (mri or mr imaging or nmr) or hysteroscop* or taus or tvus):ab
#33	#22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32
#34	#4 and #33 publication year from 2007 to 2016, in cochrane reviews (reviews and protocols) and other reviews
#35	#4 and #21 and #33 publication year from 2007 to 2016, in trials, technology assessments and economic evaluations
#36	#8 and #33 publication year from 1996 to 2016, in cochrane reviews (reviews and protocols) and other reviews
#37	#8 and #20 and #33 publication year from 1996 to 2016, in trials, technology assessments and economic evaluations

ID	Search
#38	#34 or #35 or #36 or #37

OVID MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R); Embase 1980 to 2016 (last searched 13.10.2016)

#	Searches
1	menorrhagia/ or 'menorrhagia and metrorrhagia'/ (12537)
2	1 use emez (8425)
3	menorrhagia/ (12470)
4	3 use mesz, prem (4113)
5	(menorrhag* or hypermenorrh* or menometrorrhag* or metromenorrhag* or (menstru* adj3 (bleed* or blood loss)) or (heavy adj1 (period* or menses or menstruation)) or ((abnormal* or dysfunction*) adj3 (uterine or uterus) adj3 (bleed* or blood*))).ti,ab. (18679)
6	or/2,4-5 (22025)
7	adenomyosis/ (3927)
8	7 use emez (3549)
9	adenomyosis/ (3927)
10	9 use mesz, prem (378)
11	(adenomyo* or adenomyi*).ti,ab. (7190)
12	or/8,10-11 (8216)
13	'*area under the curve'/ or *predictive validity/ or *receiver operating characteristic/ or *reliability/ or '*sensitivity and specificity'/ or *test retest reliability/ or *validity/ (35178)
14	13 use emez (31257)
15	'area under curve'/ or 'predictive value of tests'/ or 'reproducibility of results'/ or roc curve/ or 'sensitivity and specificity'/ or validation studies/ (1358993)
16	15 use mesz, prem (806587)

#	Searches
17	(accurac* or accurat* or area under curve or auc value* or clinical utilit* or (likelihood adj3 ratio*) or (diagnostic adj2 odds ratio*) or ((pretest or pre test or posttest or post test) adj2 probabilit*) or (predict* adj3 value*) or receiver operating characteristic or (roc adj2 curv*) or reliabil* or sensititiv* or specificit* or valid*).tw. (3326518)
18	or/14,16-17 (3747277)
19	imaging/ (357914)
20	diagnostic imaging/ (174189)
21	exp echography/ or hysteroscopy/ (916865)
22	gynecological radiography/ (116)
23	hysterosalpingography/ (8923)
24	tomography/ (38627)
25	computer assisted tomography/ (637444)
26	exp nuclear magnetic resonance/ (1259850)
27	three dimensional imaging/ (134632)
28	or/19-27 (2894095)
29	28 use emez (2282821)
30	diagnostic imaging/ (174189)
31	exp imaging, three-dimensional/ (148746)
32	exp ultrasonography/ (905916)
33	tomography/ or exp tomography, x-ray/ (1272329)
34	exp magnetic resonance imaging/ (1121066)
35	hysterosalpingography/ (8923)
36	hysteroscopy/ (14119)
37	uterus/us (2297)
38	leiomyoma/us (961)
39	or/30-38 (2989587)

#	Searches
40	39 use mesz, prem (1028816)
41	(ultrasonograph* or sonograph* or ultrasound* or ultrasonic* or echograph* or echotomograph* or endosonograph* or ((uterine or uterus or medical) adj3 imaging) or hysterosalpongograph* or hysterograph* or hysterosonograph* or sonohysterograph* or (magnetic adj3 (resonance or imaging)) or (mri or mr imaging or nmr) or hysteroscop* or tvus or taus).ti,ab. (1853552)
42	or/29,40-41 (3881741)
43	meta analysis/ or 'meta analysis (topic)'/ or systematic review/ (332848)
44	43 use emez (252004)
45	meta analysis.sh,pt. or 'meta-analysis as topic'/ or 'review literature as topic'/ (310231)
46	45 use mesz, prem (102226)
47	(exp bibliographic database/ or (((electronic or computer$ or online) adj database$) or bids or cochrane or embase or index medicus or isi citation or medline or psyclit or psychlit or scisearch or science citation or (web adj2 science)).ti,ab.) and (review$.ti,ab,sh,pt. or systematic$.ti,ab.) (253845)
48	(exp databases, bibliographic/ or (((electronic or computer$ or online) adj database$) or bids or cochrane or embase or index medicus or isi citation or medline or psyclit or psychlit or scisearch or science citation or (web adj2 science)).ti,ab.) and (review$.ti,ab,sh,pt. or systematic$.ti,ab.) (253845)
49	((analy$ or assessment$ or evidence$ or methodol$ or quantativ$ or systematic$) adj2 (overview$ or review$)).tw. or ((analy$ or assessment$ or evidence$ or methodol$ or quantativ$ or systematic$).ti. and review$.ti,pt.) or (systematic$ adj2 search$).ti,ab. (486134)
50	(metaanal$ or meta anal$).ti,ab. (228536)
51	(research adj (review$ or integration)).ti,ab. (2657)
52	reference list$.ab. (28511)
53	bibliograph$.ab. (30896)
54	published studies.ab. (40368)
55	relevant journals.ab. (2105)
56	selection criteria.ab. (53262)
57	(data adj (extraction or synthesis)).ab. (36561)
58	(handsearch$ or ((hand or manual) adj search$)).ti,ab. (22026)
59	(mantel haenszel or peto or dersimonian or der simonian).ti,ab. (11716)

#	Searches
60	(fixed effect$ or random effect$).ti,ab. (73352)
61	((pool$ or combined or combining) adj2 (data or trials or studies or results)).ti,ab. (113198)
62	47 use emez (153761)
63	48 use mesz, prem (100084)
64	or/44,46,49-63 (961624)
65	64 or review$.ti,ab,pt. (6564182)
66	(12 and 18 and 42) or (12 and 42 and 65) (1321)
67	limit 66 to yr='1996 -current' (1178)
68	(satisf* or acceptab*).ti,ab. or 18 (4447557)
69	(6 and 68 and 42) or (6 and 42 and 65) (2548)
70	limit 69 to yr='2007 -current' (1405)
71	or/67,70 (2436)

Appendix F – Clinical evidence tables

The clinical evidence tables are presented in a separate document.

Appendix G – GRADE tables

Review question 1. Adapted GRADE tables for diagnostic accuracy of ultrasound and hysteroscopy for investigations of women presenting with heavy menstrual bleeding review

Table 9: Clinical evidence profile: Diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting polyps

No of studies	n	Risk of biasa	Inconsistency[b]	Indirectnessc	Imprecisiond	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Polyps - All eligible studies												
9	955	very serious risk of bias[1]	very serious inconsistency[2]	very serious indirectness[3]	serious imprecision[4]	none	57.4 (30.0-81.0)	94.1 (91.6-95.9)	9.73 (5.32-17.79)	0.45 (0.23-0.87)	0.93 (0.91-0.95)	VERY LOW
Polyps - Studies published in 2007 or later												
7	853	very serious risk of bias[5]	very serious inconsistency[2]	very serious indirectness[6]	serious imprecision[4]	none	60.2 (29.0-84.9)	93.1 (90.8-94.8)	8.7 (4.8-15.9)	0.43 (0.19-0.94)	0.94 (0.92-0.96)	VERY LOW

2D-TVUS: two-dimensional transvaginal ultrasound scan; AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 2 because 8 studies did not report if a consecutive or random sample of women were enrolled; 4 studies failed to report exclusion criteria or reported them unclearly; 3 studies did not reported diagnostic criteria or reported unclearly; 3 studies had drop-puts, however, the reasons were described; 2 studies did not report clearly if reference standard was interpreted without knowledge of the index test; 1 study had an interval of 1 month between the investigations; 1 study used varying methods of obtaining histological sample including D&C, hysteroscopy-guided biopsy and hysterectomy.
2 The evidence was downgraded by 2 because of high heterogeneity.
3 The evidence was downgraded by 2 because of high concern about applicability of population since 8 studies did not report the exact proportion of women with HMB and 1 study had less than 80% women with HMB; unclear concern about the interpretation of the index test since the level of experience of the investigators were not reported in most studies.
4 The evidence was downgraded by 1 because 95% CI for sensitivity crosses 75%.

5 The evidence was downgraded by 2 because 6 studies did not report if a consecutive or random sample of women were enrolled; 2 studies did not report diagnostic criteria or reported it unclearly; 2 studies did not report clearly if reference standard was interpreted without knowledge of the index test; 1 study made unclear exclusions; 1 study reported inclusion and exclusion criteria unclearly; 1 study had an interval of 1 month between the investigations; 1 study used varying methods of obtaining histological sample including D&C, hysteroscopy-guided biopsy and hysterectomy; 3 studies had drop-puts, however, the reasons were described.
6 The evidence was downgraded by 2 because of high concern about applicability of population since 7 studies did not report the exact proportion of women with HMB, unclear concern about the interpretation of the index test since the level of experience of the investigators were not reported in most studies.

Table 10: Clinical evidence profile: Diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting fibroids

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectness[c]	Imprecisiond	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Fibroids - All eligible studies												
8	870	very serious risk of bias[1]	serious inconsistency[2]	very serious indirectness[3]	serious indirectness[4]	none	77.0 (65.8-85.4)	96.5 (92.7-98.4)	22.3 (10.5-47.5)	0.23 (0.16-0.36)	0.92 (0.89-0.94)	VERY LOW
Fibroids - Studies published in 2007 or later												
6	768	serious risk of bias5	very serious inconsistency6	very serious indirectness[7]	very serious imprecision[8]	none	81.6 (65.3-91.2)	96.8 (91.4-98.9)	25.9 (9.1-73.8)	0.19 (0.09-0.38)	0.96 (0.94-0.97)	VERY LOW
Fibroids – Studies using hysterectomy as reference standard												
1	52	no serious risk of bias	no serious inconsistency	very serious indirectness[9]	very serious imprecision[8]	none	90 (55.5-99.8)	97.6 (87.4-99.9)	37.80 (3.59-265.0)	0.10 (0.02-0.66)	N/A	VERY LOW
1	50	serious risk of bias[10]	no serious inconsistency	very serious indirectness[11]	very serious imprecision[8]	none	70 (48.8-90.9)	96.7 (83.3-99.9)	21.2 (3.25-160.0)	0.30 (0.13-0.58)	N/A	VERY LOW

2D-TVUS: two-dimensional transvaginal ultrasound scan; AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted..
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 2 because 7 studies did not report if a consecutive or random sample of women were enrolled; 3 studies failed to report exclusion criteria or reported them unclearly; 2 studies did not reported diagnostic criteria or reported unclearly; 3 studies had drop-puts, however, the reasons were described; 1 study did not report clearly if reference standard was interpreted without knowledge of the index test; 1 study had an interval of 1 month between the investigations; 1 study used varying methods of

obtaining histological sample including D&C, hysteroscopy-guided biopsy and hysterectomy.
2 The evidence was downgraded by 1 because of moderately high heterogeneity.
3 The evidence was downgraded by 2 because of high concern about applicability of population since 7 studies did not report the exact proportion of women with HMB and 1 study had less than 80% women with HMB; unclear concern about the interpretation of the index test since the level of experience of the investigators were not reported in most studies.
4 The evidence was downgraded by 1 because 95% CI for sensitivity crosses 75%.
5 The evidence was downgraded by 1 because 5 studies did not report if a consecutive or random sample of women were enrolled; 1 study reported diagnostic criteria unclearly; 1 study did not report clearly if reference standard was interpreted without knowledge of the index test; 1 study reported inclusion and exclusion criteria unclearly; 1 study had an interval of 1 month between the investigations; 1 study used varying methods of obtaining histological sample including D&C, hysteroscopy-guided biopsy and hysterectomy; 3 studies had drop-puts, however, the reasons were described.
6 The evidence was downgraded by 2 because of high heterogeneity.
7 The evidence was downgraded by 2 because of high concern about applicability of population since 6 studies did not report the exact proportion of women with HMB; unclear concern about the interpretation of the index test since the level of experience of the investigators were not reported in most studies.
8 The evidence was downgraded by 2 because 95% CI for sensitivity crosses 75% and 90%.
9 The evidence was downgraded by 2 because of high concern about applicability since the study had less than 80% women with HMB; unclear concern about the interpretation of the index test since the level of experience of the investigator was not reported.
10 The evidence was downgraded by 1 because the study did not report if a consecutive or random sample of women were enrolled; the study did not report any exclusion criteria; the study did not define the diagnostic criteria.
11 The evidence was downgraded by 2 because of high concern about applicability since the study did not report the proportion of women with HMB; unclear concern about the interpretation of the index test since the level of experience of the investigator was not reported.

Table 11: Clinical evidence profile: Diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting polyps or fibroids

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectnessc	Imprecisiond	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Polyps or fibroids												
1	189	very serious risk of bias[1]	no serious inconsistency	very serious indirectness[2]	serious imprecision[3]	none	92 (85-96)	62 (50-73)	2.41 (1.78-3.26)	0.14 (0.07-0.25)	N/A	VERY LOW
1	88	serious risk of bias[4]	no serious inconsistency	very serious indirectness[2]	no serious imprecision	none	23.5 (6.8-49.9)	93.0 (78.0-97.7)	3.19 (1.02-9.96)	0.82 (0.64-1.06)	N/A	VERY LOW

2D-TVUS: two-dimensional transvaginal ultrasound scan; AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.

b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 2 because the study did not report if a consecutive or random sample of women were enrolled; study made unclear or inappropriate exclusions; unclear if reference standard was interpreted without knowledge of the index test; interval between investigations was 4 months; reference standard differed between women (hysteroscopy or hysterectomy).
2 The evidence was downgraded by 2 because the study did not report the proportion of women with HMB; unclear who interpreted the index test and what was his/her level of experience.
3 The evidence was downgraded by 1 because 95% CI for sensitivity crosses 90%.
4 The evidence was downgraded by 1 because the study did not report if a consecutive or random sample of women were enrolled; unclear if reference standard was interpreted without knowledge of the index test; 12% of enrolled women not included in analysis, however, reasons for dropping out were described.

Table 12: Clinical evidence profile: Diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting endometrial hyperplasia

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectness[c]	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Endometrial hyperplasia												
1	122	very serious risk of bias[1]	no serious inconsistency	very serious indirectness[2]	very serious imprecision[3]	none	75.0 (19.4-99.4)	92.4 (86.0-96.5)	9.83 (4.22-22.90)	0.27 (0.05-1.48)	N/A	VERY LOW
1	85	very serious risk of bias[4]	no serious inconsistency	very serious indirectness[2]	no serious imprecision	none	43.8 (19.8-70.1)	95.7 (87.8-99.1)	10 (2.92-34.72)	0.59 (0.38-0.91)	N/A	VERY LOW
1	89	no serious risk of bias	no serious inconsistency	very serious indirectness[2]	very serious imprecision3	none	71.4 (29.0-96.3)	85.4 (75.8-92.2)	4.9 (3.0-7.9)	0.3 (0.09-1.2)	N/A	VERY LOW

2D-TVUS: two-dimensional transvaginal ultrasound scan; AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 2 because the study did not report if a consecutive or random sample of women were enrolled; unclear if reference standard was interpreted

without knowledge of the index test; interval between investigations was 1 month; study used varying methods of obtaining histological sample including D&C, hysteroscopy-guided biopsy and hysterectomy; 11/133 of enrolled women were not included in analysis and the reasons for dropping out were not described.
2 The evidence was downgraded by 2 because the study did not report the proportion of women with HMB; unclear who interpreted the index test and what was his/her level of experience.
3 The evidence was downgraded by 2 because 95% CI for sensitivity crosses 75% and 90%.
4 The evidence was downgraded by 2 because the study did not report if a consecutive or random sample of women were enrolled; study made unclear exclusions; study did not report diagnostic criteria; unclear if reference standard was interpreted without knowledge of the index test.

Table 13: Clinical evidence profile: Diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting endometrial carcinoma

No of studies	n	Risk of biasa	Inconsistencyb	Indirectnessc	Imprecisiond	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Endometrial carcinoma												
1	89	no serious risk of bias	no serious inconsistency	very serious indirectness[1]	very serious imprecision[2]	none	50 (1.3-98.7)	93.1 (85.6-97.4)	7.25 (1.8-29)	0.54 (0.1-2.6)	N/A	VERY LOW

2D-TVUS: two-dimensional transvaginal ultrasound scan; AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 2 because the study did not report the proportion of women with HMB; unclear who interpreted the index test and what was his/her level of experience.
2 The evidence was downgraded by 2 because 95% CI for sensitivity crosses 75% and 90%.

Table 14: Clinical evidence profile: Diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting abnormal endometrium (double-layer endometrium 12 mm or thicker and single-layer endometrium 6 mm or thicker)

No of studies	n	Risk of biasa	Inconsistencyb	Indirectnessc	Imprecisiond	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Abnormal endometriume (abnormal thickness)												

No of studies	n	Risk of biasa	Inconsistencyb	Indirectnessc	Imprecisiond	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
1	88	serious risk of bias[1]	no serious inconsistency	very serious indirectness[2]	no serious imprecision	none	33.3 (7.5-70.1)	88.6 (79.5-94.7)	2.93 (0.96-8.88)	0.75 (0.47-1.20)	N/A	VERY LOW

2D-TVUS: two-dimensional transvaginal ultrasound scan; AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
e In premenopausal women, double-layer endometrium thickness of less than 12mm and single-layer endometrium thickness of less than 6mm were arbitrarily considered to be normal, and thicker endometrium was classified as abnormal. In postmenopausal women 4mm was used as a cut-off level to define normality. Irregularly thickened hyperechogenic endometrium was considered to be suggestive of endometrial carcinoma.
1 The evidence was downgraded by 1 because the study did not report if a consecutive or random sample of women were enrolled; unclear if reference standard was interpreted without knowledge of the index test; 12% of enrolled women were not included in analysis, however, the reasons for dropouts were described.
2 The evidence was downgraded by 2 because the study did not report the proportion of women with HMB; unclear who interpreted the index test and what was his/her level of experience.

Table 15: Clinical evidence profile: Diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting any abnormal finding

No of studies	n	Risk of biasa	Inconsistencyb	Indirectnessc	Imprecisiond	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Any abnormal finding												
1	213	no serious risk of bias	N/A	very serious indirectness[1]	no serious imprecision	none	94.6 (91-98)	77.2 (67-87)	4.16 (2.66-6.50)	0.07 (0.04-0.14)	N/A	LOW
1	83	serious risk of bias2	N/A	very serious indirectness[1]	serious imprecision[3]	none	74 (61-84)	55 (37-71)	1.65 (1.07-2.55)	0.47 (0.27-0.82)	N/A	VERY LOW
1	89	no serious risk of bias	N/A	very serious indirectness[1]	serious imprecision[4]	none	89.4 (76.9-96.5)	71.4 (55.4-84.3)	3.13 (2.5-3.9)	0.15 (0.06-0.40)	N/A	VERY LOW

No of studies	n	Risk of biasa	Inconsistencyb	Indirectnessc	Imprecisiond	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
1	39	very serious risk of bias5	N/A	very serious indirectness[1]	very serious imprecision[6]	none	67 (35-90)	93 (76-99)	9.0 (2.24-36.22)	0.36 (0.16-0.81	N/A	VERY LOW

2D-TVUS: two-dimensional transvaginal ultrasound scan; AUC: area under the curve; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
e Defined slightly differently across studies but in general either abnormal endometrial thickness, hyperplasia, carcinoma, or focal abnormalities such as fibroids or polyps.
1 The evidence was downgraded by 2 because the study did not report the proportion of women with HMB; unclear who interpreted the index test and what was his/her level of experience.
2 The evidence was downgraded by 1 because the study did not report if a consecutive or random sample of women were enrolled; 22/274 were not included in analysis, however, reasons for dropping out were described.
3 The evidence was downgraded by1 because 95% CI for sensitivity crosses 75%.
4 The evidence was downgraded by1 because 95% CI for sensitivity crosses 90%.
5 The evidence was downgraded by 2 because the study did not report if a consecutive or random sample of women were enrolled; diagnostic criteria not reported; unclear if reference standard was interpreted without knowledge of the index test; 8/47 women not included in analysis, however, reasons for dropping out were described; the study used varying methods of obtaining histological sample.
6 The evidence was downgraded by2 because 95% CI for sensitivity crosses 75% and 90%.

Table 16: Clinical evidence profile: Diagnostic accuracy of hysteroscopy under general anaesthetic (versus histopathology) in detecting polyps

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectnes[sc]	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Polyps												
1	89	no serious risk of bias	no serious inconsistency	very serious indirectness[1]	serious imprecision[2]	none	91.1 (76.3-98.1)	98.2 (90.3-100)	50.2 (44.9-56)	0.09 (0.01-0.8)	N/A	VERY LOW

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectnes[sc]	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
1	85	very serious risk of bias3	no serious inconsistency	very serious indirectness[1]	very serious imprecision4	none	71.4 (41.9-91.6)	100 (94.9-100)	Inf	0.29 (0.12-0.65)	N/A	VERY LOW

AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 2 because the study did not report the proportion of women with HMB; the level of experience of the index test investigator not reported.
2 The evidence was downgraded by 1 because 95% CI for sensitivity crosses 90%.
3 The evidence was downgraded by 2 because the study did not report if a consecutive or random sample of women were enrolled; unclear exclusions made; diagnostic criteria not reports; unclear if reference standard was interpreted without knowledge of the index test.
4 The evidence was downgraded by 2 because 95% CI for sensitivity crosses 75% and 90%.

Table 17: Clinical evidence profile: Diagnostic accuracy of outpatient hysteroscopy (versus histopathology) in detecting polyps

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectness[c]	Imprecisiond	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Polyps												
1	52	no serious risk of bias	no serious inconsistency	very serious indirectness[1]	very serious imprecision[2]	none	100 (2.5-100)	100 (93.02-100)	Inf	0.00	N/A	VERY LOW
1	223	serious risk of bias3	no serious inconsistency	very serious indirectness[1]	very serious imprecision[2]	none	88 (67.64-97.34)	93 (88.48-96.10)	12.44 (7.34-21.07)	0.13 (0.05-0.39	N/A	VERY LOW

AUC: area under the curve; CI: confidence interval; Inf: infinite; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.

d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 2 because the proportion of women with HMB was less than 80%; the level of experience of the index test investigator not reported.
2 The evidence was downgraded by 2 because 95% CI for sensitivity crosses 75% and 90%.
3 The evidence was downgraded by 1 because the study did not report if a consecutive or random sample of women were enrolled; diagnostic criteria unclear; 67/290 were not included in analysis, however, the reasons for dropping out were described.

Table 18: Clinical evidence profile: Diagnostic accuracy of hysteroscopy under general anaesthetic (versus histopathology) in detecting fibroids

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectnessc	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Fibroids												
1	89	no serious risk of bias	no serious inconsistency	very serious indirectness1	very serious imprecision[2]	none	100 (39.8-100)	100 (95.8-100)	Inf	0.00	N/A	VERY LOW

AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 2 because the study did not report the proportion of women with HMB; the level of experience of the index test investigator not reported.
2 The evidence was downgraded by 2 because 95% CI for sensitivity crosses 75% and 90%.

Table 19: Clinical evidence profile: Diagnostic accuracy of outpatient hysteroscopy (versus histopathology) in detecting fibroids

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectness[c]	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Fibroids												
1	51	no serious risk of bias	no serious inconsistency	very serious indirectness[1]	very serious imprecision[2]	none	100 (69.15-100)	100 (91.59-100)	Inf	0.00	N/A	VERY LOW

AUC: area under the curve; CI: confidence interval; Inf: infinite; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 2 because the proportion of women with HMB was less than 80%; level of experience of investigator not reported.
2 The evidence was downgraded by 2 because 95% CI for sensitivity crosses 75% and 90%.

Table 20: Clinical evidence profile: Diagnostic accuracy of hysteroscopy under spinal or general anaesthetic (versus histopathology) in detecting polyps or fibroids

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectness[c]	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95 % CI)	Quality
Polyps or fibroids												
1	88	serious risk of bias[1]	no serious inconsistency	very serious indirectness[2]	serious imprecision[3]	Not clearly reported how hysteroscopy was performed but likely as a day case according to reporting.	100 (80.5-100)	87.3 (77.3-94.0)	7.44 (4.05-13.67)	Inf	N/A	VERY LOW

AUC: area under the curve; CI: confidence interval; Inf: infinite; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 1 because the study did not report if a consecutive or random sample of women were enrolled; unclear if reference standard was interpreted without knowledge of the index test; 12% of enrolled women were not included in analysis, however, the reasons for dropouts were described.

2 The evidence was downgraded by 2 because the study did not report the proportion of women with HMB; level of experience of investigator not reported.
3 The evidence was downgraded by 1 because 95% CI for sensitivity crosses 90%.

Table 21: Clinical evidence profile: Diagnostic accuracy of hysteroscopy under general anaesthetic (versus histopathology) following a 2D transvaginal ultrasound in detecting polyps or fibroids

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectness[c]	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95 % CI)	Quality
Polyps or fibroids												
1	50	serious risk of bias[1]	no serious inconsistency	very serious indirectness[2]	very serious imprecision[3]	none	87.5 (71.01-96.49)	100 (81.47-100)	Inf	0.12 (0.05-0.31)	N/A	VERY LOW

AUC: area under the curve; CI: confidence interval; Inf: infinite; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 1 because the study did not report if a consecutive or random sample of women were enrolled; diagnostic criteria not reported; unclear if reference standard was interpreted without knowledge of the index test.
2 The evidence was downgraded by 2 because the study did not report the proportion of women with HMB; level of experience of investigator not reported.
3 The evidence was downgraded by 2 because 95% CI for sensitivity crosses 75% and 90%.

Table 22: Clinical evidence profile: Diagnostic accuracy of hysteroscopy under general anaesthetic (versus histopathology) in detecting endometrial hyperplasia

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectnessc	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Endometrial hyperplasia												
1	89	no serious risk of bias	no serious inconsistency	very serious indirectness[1]	very serious imprecision[2]	none	85.7 (42.1-99.6)	97.6 (91.5-99.7)	35.1 (25.9-47.6)	0.15 (0.02-1.4)	N/A	VERY LOW

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectnessc	Imprecision[d]	Other conside rations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
1	85	very serious risk of bias[3]	no serious inconsistency	very serious indirectness[1]	serious imprecision[4]	none	50 (23.0-77.0)	95.8 (88.1-99.1)	11.8 (3.48-40.29)	0.52 (0.31-0.88)	N/A	VERY LOW

AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 2 because the study did not report the proportion of women with HMB; level of experience of the investigator not reported.
2 The evidence was downgraded by 2 because 95% CI for sensitivity crosses 75% and 90%.
3 The evidence was downgraded by 2 because the study did not report if a consecutive or random sample of women were enrolled; unclear exclusions made; diagnostic criteria not reports; unclear if reference standard was interpreted without knowledge of the index test.
4 The evidence was downgraded by 1 because 95% CI for sensitivity crosses 75%.

Table 23: Clinical evidence profile: Diagnostic accuracy of outpatient hysteroscopy (versus histopathology) in detecting endometrial hyperplasia

No of studies	n	Risk of biasa	Inconsistency[b]	Indirectness[c]	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Endometrial hyperplasi[a]												
1	223	serious risk of bias[1]	no serious inconsistency	very serious indirectness[2]	serious imprecision[3]	none	63 (43.7-78.9)	92 (86.8-95.1)	7.46 (4.35-12.81)	0.41 (0.26-0.64)	N/A	VERY LOW

AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the

test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 1 because the study did not report if a consecutive or random sample of women were enrolled; diagnostic criteria unclear; 67/290 were not included in analysis, however, the reasons for dropping out were described.
2 The evidence was downgraded by 2 because the proportion of women with HMB was less than 80%; level of experience of investigator not reported.
3 The evidence was downgraded by 1 because 95% CI for sensitivity crosses 75%.

Table 24: Clinical evidence profile: Diagnostic accuracy of hysteroscopy under general anaesthetic (versus histopathology) in detecting endometrial carcinoma

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectness[c]	Imprecision[d]	Other conside rations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Endometrial carcinoma												
1	89	no serious risk of bias	no serious inconsistency	very serious indirectness[1]	very serious imprecision2	none	100 (15.8-100)	96.4 (90.3-99.3)	29.0 (27.9-30.2)	0.00	N/A	VERY LOW

AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 2 because the study did not report the proportion of women with HMB; the level of experience of the investigator not reported.
2 The evidence was downgraded by 2 because 95% CI for sensitivity crosses 75% and 90%.

Table 25: Clinical evidence profile: Diagnostic accuracy of outpatient hysteroscopy (versus histopathology) in detecting endometrial carcinoma

No of studies	n	Risk of biasa	Inconsistencyb	Indirectnessc	Imprecisiond	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Endometrial carcinoma												
1	223	serious risk of bias[1]	no serious inconsistency	very serious indirectness[2]	very serious imprecision3	none	100 (15.8-100)	98 (95.4-99.5)	55.2 (20.9-145.9)	0.00	N/A	VERY LOW

AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 1 because the study did not report if a consecutive or random sample of women were enrolled; diagnostic criteria unclear; 67/290 were not included in analysis, however, the reasons for dropping out were described.
2 The evidence was downgraded by 2 because the proportion of women with HMB was less than 80%; level of experience of investigator not reported.
3 The evidence was downgraded by 2 because 95% CI for sensitivity crosses 75% and 90%.

Table 26: Clinical evidence profile: Diagnostic accuracy of hysteroscopy under spinal or general anaesthetic (versus histopathology) in detecting abnormal endometrium (double-layer endometrium 12 mm or thicker and single-layer endometrium 6 mm or thicker)

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectness[c]	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Abnormal endometriume (abnormal thickness)												
1	88	serious risk of bias[1]	no serious inconsistency	very serious indirectness2	no serious imprecision	Not clearly reported how hysteroscopy was performed but likely as a day case according to reporting.	22.2 (2.8-60.6)	87.3 (78.0-93.8)	1.76 (0.45-6.79)	0.89 (0.062-1.28)	N/A	VERY LOW

AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.

e In premenopausal women, double-layer endometrium thickness of less than 12mm and single-layer endometrium thickness of less than 6mm were arbitrarily considered to be normal, and thicker endometrium was classified as abnormal. In postmenopausal women 4mm was used as a cut-off level to define normality. Irregularly thickened hyperechogenic endometrium was considered to be suggestive of endometrial carcinoma.
1 The evidence was downgraded by 1 because the study did not report if a consecutive or random sample of women were enrolled; unclear if reference standard was interpreted without knowledge of the index test; 12% of enrolled women were not included in analysis, however, the reasons for dropouts were described.
2 The evidence was downgraded by 2 because the study did not report the proportion of women with HMB; level of experience of investigator not reported.

Table 27: Clinical evidence profile: Diagnostic accuracy of outpatient hysteroscopy (versus histopathology) in detecting endometritis

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectness[c]	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Endometritis												
1	223	serious risk of bias[1]	no serious inconsistency	very serious indirectness[2]	no serious imprecision	none	41 (27.00-56.77)	99 (95.98-99.86)	36.55 (8.83-1451.30)	0.59 (0.47-0.76)	N/A	VERY LOW

AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 1 because the study did not report if a consecutive or random sample of women were enrolled; diagnostic criteria unclear; 67/290 were not included in analysis, however, the reasons for dropping out were described.
2 The evidence was downgraded by 2 because the proportion of women with HMB was less than 80%; level of experience of investigator not reported.

Table 28: Clinical evidence profile: Diagnostic accuracy of outpatient hysteroscopy (versus histopathology) in detecting retained products of conception

No of studies	n	Risk of biasa	Inconsistency[b]	Indirectness[c]	Imprecisiond	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Retained products of conception												

No of studies	n	Risk of biasa	Inconsistency[b]	Indirectness[c]	Imprecisiond	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
1	223	serious risk of bias[1]	no serious inconsistency	very serious indirectness[2]	very serious imprecision[3]	none	100 (47.8-100)	100 (97.5-100)	218 (30.85-1540)	0.00	N/A	VERY LOW

AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 1 because the study did not report if a consecutive or random sample of women were enrolled; diagnostic criteria unclear; 67/290 were not included in analysis, however, the reasons for dropping out were described.
2 The evidence was downgraded by 2 because the proportion of women with HMB was less than 80%; level of experience of investigator not reported.
3 The evidence was downgraded by 2 because 95% CI for sensitivity crosses 75% and 90%.

Table 29: Clinical evidence profile: Diagnostic accuracy of hysteroscopy under general anaesthetic (versus histopathology) in detecting any abnormal finding

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectness[c]	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Any abnormal findinge												
1	89	no serious risk of bias	no serious inconsistency	very serious indirectness[1]	serious imprecision[2]	none	97.9 (88.7-99.9)	92.9 (80.5-98.5)	13.7 (12.5-15.1)	0.02 (0.002-0.2)	N/A	VERY LOW

AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the

test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
e Defined as polyp, fibroid, hyperplasia, or carcinoma.
1 The evidence was downgraded by 2 because the study did not report the proportion of women with HMB; the level of experience of the investigator not reported.
2 The evidence was downgraded by 1 because 95% CI for sensitivity crosses 90%.

Table 30: Clinical evidence profile: Diagnostic accuracy of hysteroscopy with or without general or local anasthesia (versus histopathology) in detecting any abnormal finding

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectness[c]	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Any abnormal findinge												
1	86	very serious risk of bias[1]	no serious inconsistency	very serious indirectness[2]	serious imprecision[3]	none	94 (86.01-98.42)	100 (79.41-100)	Inf	0.06 (0.02-0.15)	N/A	VERY LOW

AUC: area under the curve; CI: confidence interval; Inf: infinite; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
e Defined as presence of adhesion, polyp, submucosal fibroid, pressure effect or any other abnormality in the uterine cavity.
1 The evidence was downgraded by 2 because the study did not report if a consecutive or random sample of women were enrolled; diagnostic criteria not defined; unclear if reference standard was interpreted without knowledge of the index test.
2 The evidence was downgraded by 2 because the proportion of women with HMB was less than 80%; level of experience of investigator not reported.
3 The evidence was downgraded by 1 because 95% CI for sensitivity crosses 90%.

Table 31: Clinical evidence profile: Diagnostic accuracy of 2D-TVUS (versus hysteroscopy) in detecting polyps

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectness[c]	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Polyps												

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectness[c]	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
1	196	very serious risk of bias[1]	no serious inconsistency	very serious indirectness[2]	no serious imprecision	none	32.35 (17.39-50.53)	95.06 (90.50-97.84)	6.55 (2.85-15.06)	0.71 (0.56-0.90)	N/A	VERY LOW

AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 2 because the study did not report if a consecutive or random sample of women were enrolled; exclusion criteria not reported; diagnostic criteria not defined; unclear if reference standard was interpreted without knowledge of the index test; 23/219 women were not included in analysis, however, the reasons for dropping out were described.
2 The evidence was downgraded by 2 because the study did not report the proportion of women with HMB.

Table 32: Clinical evidence profile: Diagnostic accuracy of 2D-TVUS (versus hysteroscopy) in detecting suspicious focal thickening

No of studies	n	Risk of biasa	Inconsistencyb	Indirectnessc	Imprecisiond	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Suspicious focal thickening												
1	196	very serious risk of bias[1]	no serious inconsistency	very serious indirectness[2]	no serious imprecision	none	0 (0-45.93)	93.68 (89.23-96.69)	0.00	1.07 (1.03-1.11)	N/A	VERY LOW

AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.

1 The evidence was downgraded by 2 because the study did not report if a consecutive or random sample of women were enrolled; exclusion criteria not reported; diagnostic criteria not defined; unclear if reference standard was interpreted without knowledge of the index test; 23/219 women were not included in analysis, however, the reasons for dropping out were described.
2 The evidence was downgraded by 2 because the study did not report the proportion of women with HMB.

Table 33: Clinical evidence profile: Diagnostic accuracy of 2D-TVUS (versus hysteroscopy) in detecting any abnormal finding

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectnessc	Imprecisiond	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
Any abnormal finding												
1	770	serious risk of bias[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	96 (93.41-97.41)	86 (82.25-89.99)	7.07 (5.37-9.31)	0.05 (0.03-0.08)	N/A	MODERATE

AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
e Defined as a lesion inside the cavity or when the maximum endometrial thickness measured in the sagittal plane was more than 14 mm.
1 The evidence was downgraded by 1 because the study did not report if a consecutive or random sample of women were enrolled; exclusion criteria not reported; inappropriate/unclear exclusions; unclear if reference standard was interpreted without knowledge of the index test.

Review question 2. Adapted GRADE tables for the most clinically effective imaging strategy for diagnosing adenomyosis in women presenting with heavy menstrual bleeding review

Table 34: Clinical evidence profile: Diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting adenomyosis

No of studies	n	Risk of biasa	Inconsistency[b]	Indirectness[c]	Imprecisiond	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
7	1078	serious risk of bias[2]	very serious inconsistency[3]	very serious indirectness[4]	serious imprecision[5]	none	76 (69-82)	83 (73-90)	4.59 (2.84-7.42)	0.29 (0.22-0.37)	0.83 (0.80-0.86)	VERY LOW
1	23	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious imprecision[6]	Population: women with recurrent menometrorrhagia but no evidence of leiomyomata and endometrial diseases on transabdominal examination	81.0 (58.1-94.6)	100.0 (15.8-100.0)	Infinite	0.19 (0.08-0.46)	N/A	LOW
1	106	no serious risk of bias	no serious inconsistency	very serious indirectness[7]	no serious imprecision	Population: women with recurrent menometrorrhagia and evidence of leiomyomata and endometrial diseases on transabdominal examination	38.5 (20.2-59.4)	97.5 (91.3-99.7)	15.4 (3.6-65.7)	0.63 (0.46-0.86)	N/A	LOW

2D-TVUS: two-dimensional transvaginal ultrasound scan; AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.

d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The study reported 'indefinite' results from the index test for 36 cases which have been included as negative test results in the results.
2 The evidence was downgraded by 1 because three studies did not report if consecutive or random sample of patients were enrolled; three studied did not report inclusion and exclusion criteria clearly; in one study it was unclear if the index test results were interpreted without the knowledge of the results of the reference standard; in one study 28% of the original sample were not included in the analysis
3 The evidence was downgraded by 2 because of high heterogeneity.
4 The evidence was downgraded by 2 because the population in 6 studies also included a varying proportion of women that did not have heavy menstrual or the exact proportion was unclearly reported; four papers do not report who performed or interpreted the index test or what was the level of experience of the person(s)
5 The evidence was downgraded by 1 because the 95% CI of test sensitivity crosses 75%.
6 The evidence was downgraded by 2 because the 95% CI of test sensitivity crosses 75% and 90%.
7 The evidence was downgraded by 2 because less than 80% of included women did not had HMB; the paper does not report who interpreted the index test or what was the level experience of the person(s).

Table 35: Clinical evidence profile: Diagnostic accuracy of 3D-TVUS (versus histopathology) in detecting adenomyosis

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectness[c]	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
1	72	no serious risk of bias	no serious inconsistency	very serious indirectness[1]	very serious imprecision[2]	none	91 (74-97)	88 (72-95)	7.3 (3.2-16.6)	0.11 (0.03-0.31)	N/A	VERY LOW

3D-TVUS: three-dimensional transvaginal ultrasound scan; AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The evidence was downgraded by 2 because less than 80% of included women had HMB.
2 The evidence was downgraded by 2 because the 95% CI of test sensitivity crosses 75% and 90%.

Table 36: Clinical evidence profile: Diagnostic accuracy of 2D-TAUS (versus histopathology) in detecting adenomyosis

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectness[c]	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
1	23	no serious risk of bias	no serious inconsistency	no serious indirectness	serious imprecision[1]	Population: women with recurrent menometrorrhagia but no evidence of leiomyomata and endometrial diseases on transabdominal examination	54.1 (34.0-78.2)	50.0 (27.0-98.7)	1.1 (0.3-4.8)	0.9 (0.2-3.7)	N/A	MODERATE
1	106	no serious risk of bias	no serious inconsistency	very serious indirectness[2]	no serious imprecision	Population: women with recurrent menometrorrhagia and evidence of leiomyomata and endometrial diseases on transabdominal examination	7.7 (1.0-25.1)	96.3 (89.4-99.2)	2.1 (0.4-11.6)	0.96 (0.85-1.08)	N/A	LOW

2D-TAUS: two-dimensional transabdominal ultrasound scan; AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; N/A: not applicable; sens: sensitivity; spec: specificity

The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.

a Risk of bias was assessed using the QUADAS-2 checklist.

b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.

c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.

d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.

1 The evidence was downgraded by 1 because the 95% CI of test sensitivity crosses 75%.

2 The evidence was downgraded by 2 because less than 80% of included women had HMB.

Table 37: Clinical evidence profile: Diagnostic accuracy of MRI (versus histopathology) in detecting adenomyosis

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectness[c]	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
11	106	no serious risk of bias	no serious inconsistency	very serious indirectness[2]	serious imprecision[3]	none	63.6 (40.7-82.8)	88.1 (79.2-94.1)	5.4 (2.8-10.4)	0.4 (0.2-0.7)	N/A	VERY LOW

AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; MRI: magnetic resonance imaging; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.
c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The study reported 'indefinite' results from the index test for 13 cases which have been included as negative test results in the results.
2 The evidence was downgraded by 2 because less than 80% of included women had HMB.
3 The evidence was downgraded by 1 because the 95% CI of test sensitivity crosses 75%.

Table 38: Clinical evidence profile: Diagnostic accuracy of 2D-TVUS combined with MRI (versus histopathology) in detecting adenomyosis

No of studies	n	Risk of bias[a]	Inconsistency[b]	Indirectness[c]	Imprecision[d]	Other considerations	Sens % (95% CI)	Spec % (95% CI)	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)	Quality
11	106	no serious risk of bias	no serious inconsistency	very serious indirectness2	serious imprecision3	none	72.7 (49.8-89.3)	77.4 (67.0-85.8)	3.2 (2.0-5.2)	0.35 (0.18-0.70)	N/A	VERY LOW

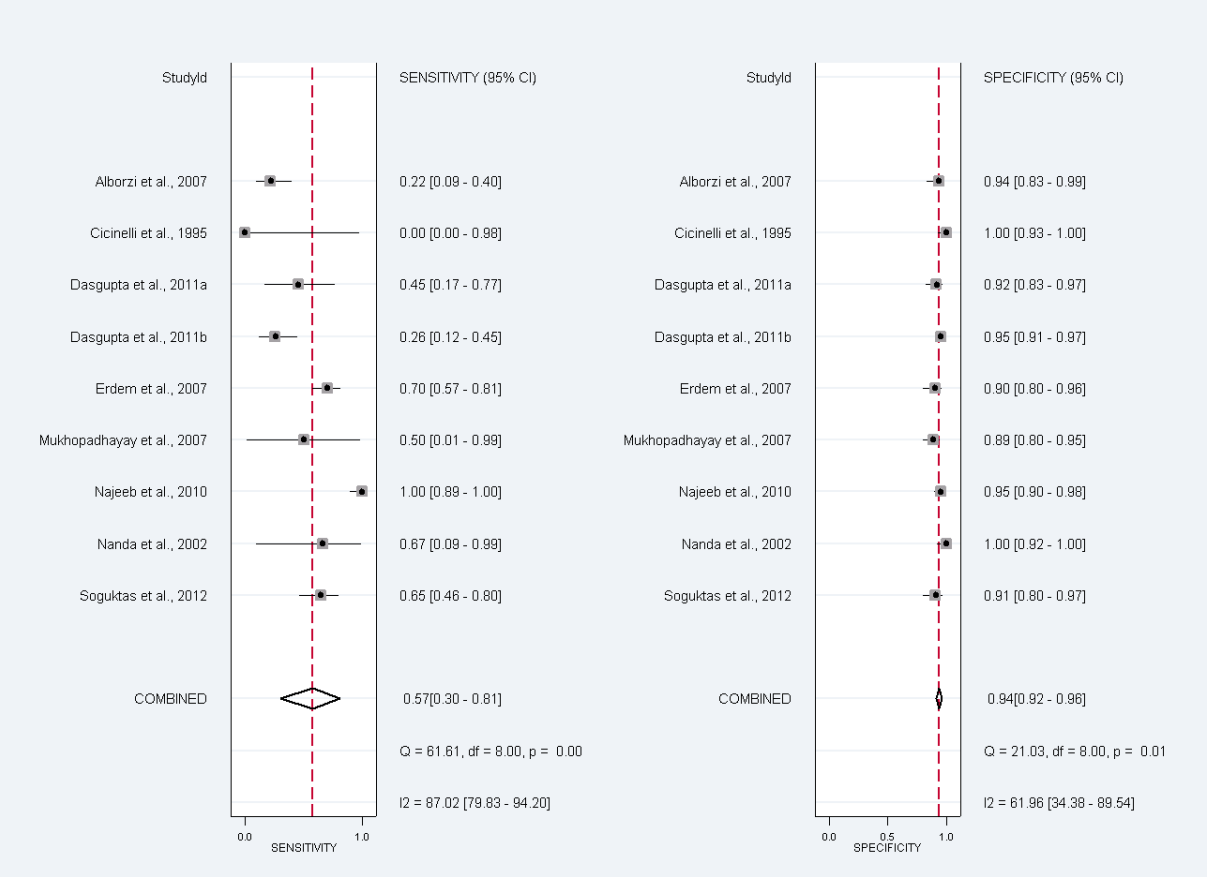
2D-TVUS: two-dimensional transvaginal ultrasound scan; AUC: area under the curve; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; MRI: magnetic resonance imaging; N/A: not applicable; sens: sensitivity; spec: specificity
The assessment of the evidence quality was conducted with emphasis on test sensitivity as this was the primary measure discussed in decision-making.
a Risk of bias was assessed using the QUADAS-2 checklist.
b Inconsistency was assessed by inspection of the 95% prediction region in a HSROC plot if a diagnostic meta-analysis was conducted.

c Indirectness was assessed using the QUADAS-2 checklist items referring to applicability.
d The judgement of precision was based on the confidence interval of test sensitivity as this was considered to be the primary measure of interest. If the 95% CI crosses either 75% or 90%, the result was judged to be seriously imprecise (90% was considered to be the cut-off for the test to be highly sensitive and if the sensitivity was less than 75% the test was considered to be of low sensitivity). If the 95% CI crosses both 75% and 90%, the results was judged to be very seriously imprecise.
1 The study reported 'indefinite' results from the index test for 41 cases which have been included as negative test results in the results.
2 The evidence was downgraded by 2 because less than 80% of included women had HMB.
3 The evidence was downgraded by 1 because the 95% CI of test sensitivity crosses 75%.

Appendix H – Forest plots and hierarchical summary receiver operating characteristic (HSROC) plots

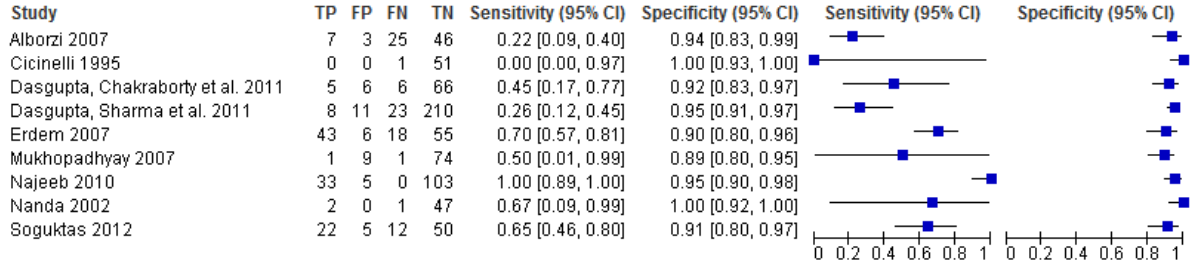
Review question 1. Forest plots and HSROC plots for diagnostic accuracy of ultrasound and hysteroscopy for investigations of women presenting with heavy menstrual bleeding review

Figure 3: Forest plot for sensitivity and specificity of 2D-TVUS (versus histopathology) in detecting polyps (meta-analysis with all eligible studies)



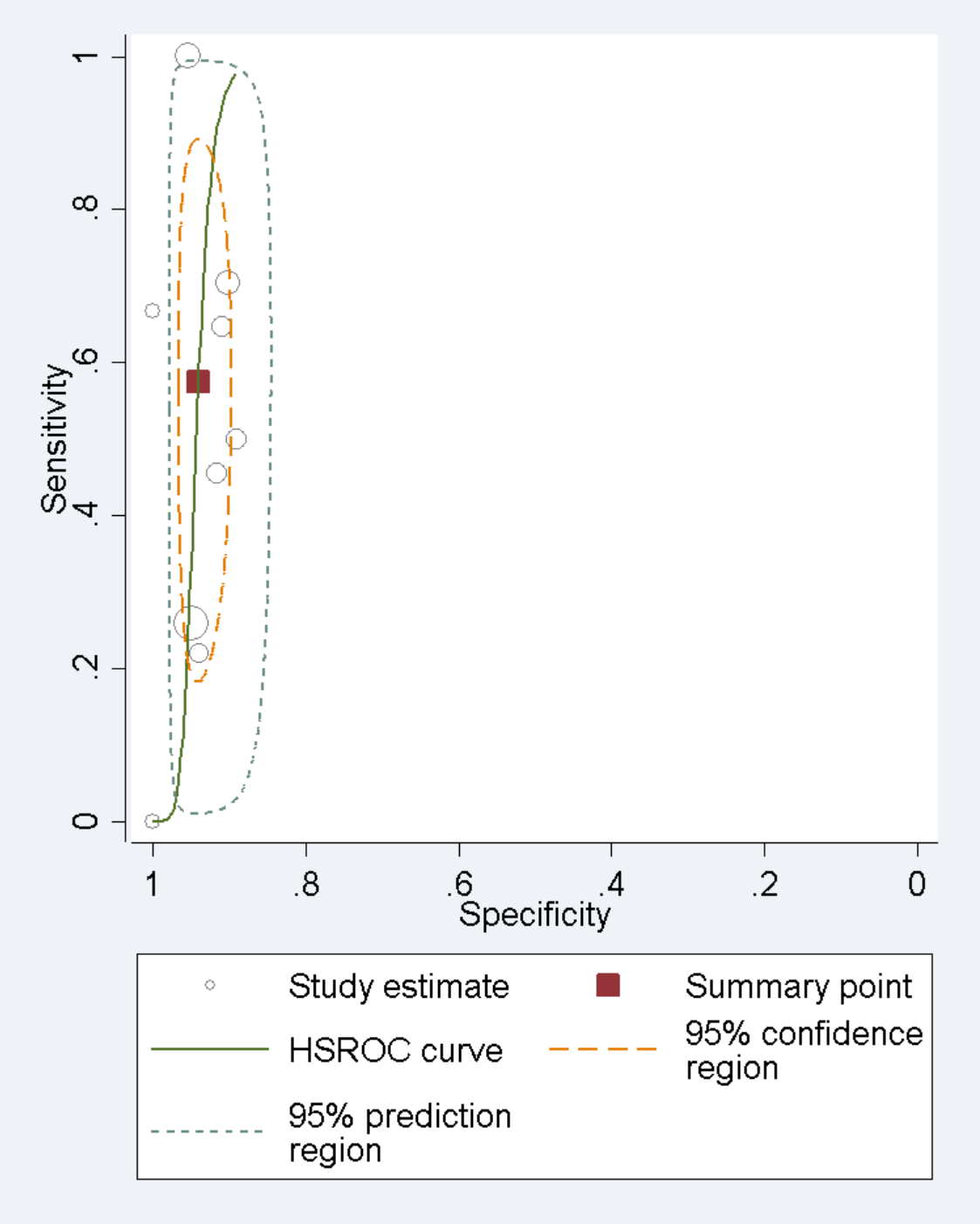
2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval

Figure 4: Forest plot for sensitivity and specificity of 2D-TVUS (versus histopathology) in detecting polyps (all eligible studies without meta-analysis)



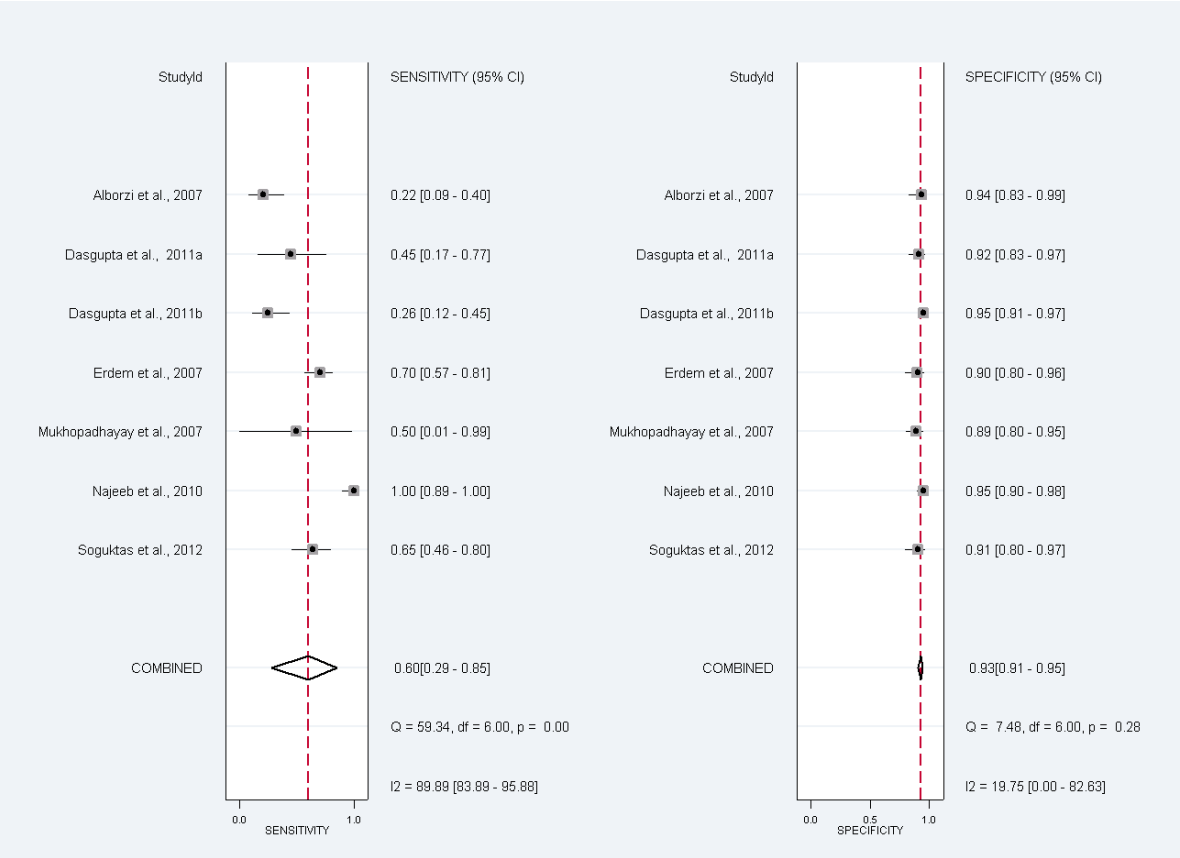
2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 5: HSROC plot for diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting polyps (meta-analysis with all eligible studies)



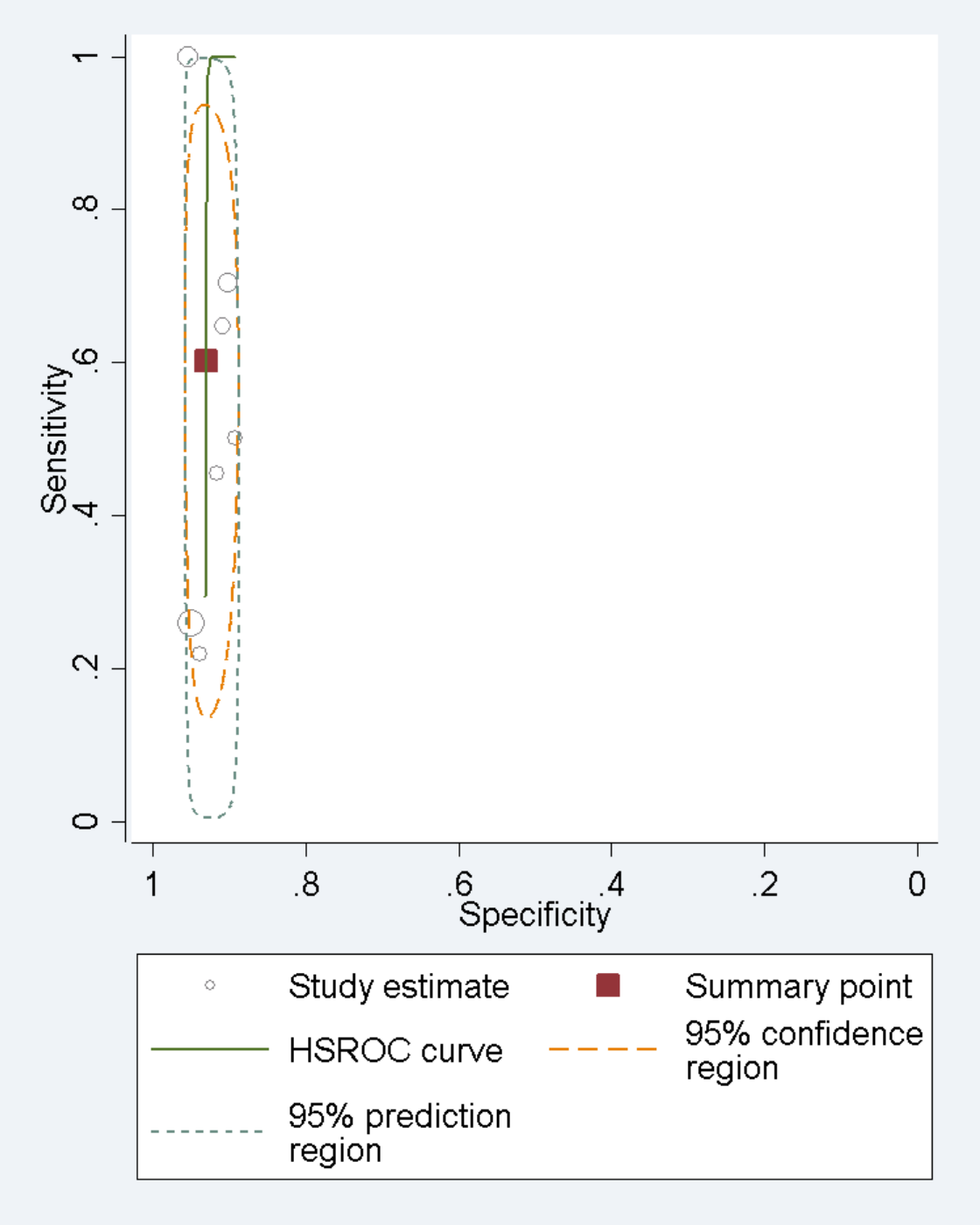
2D-TVUS: two-dimensional transvaginal ultrasound scan; HSROC: hierarchical summary receiver operating characteristic

Figure 6: Forest plot for sensitivity and specificity of 2D-TVUS (versus histopathology) in detecting polyps (meta-analysis with studies published in 2007 or later)



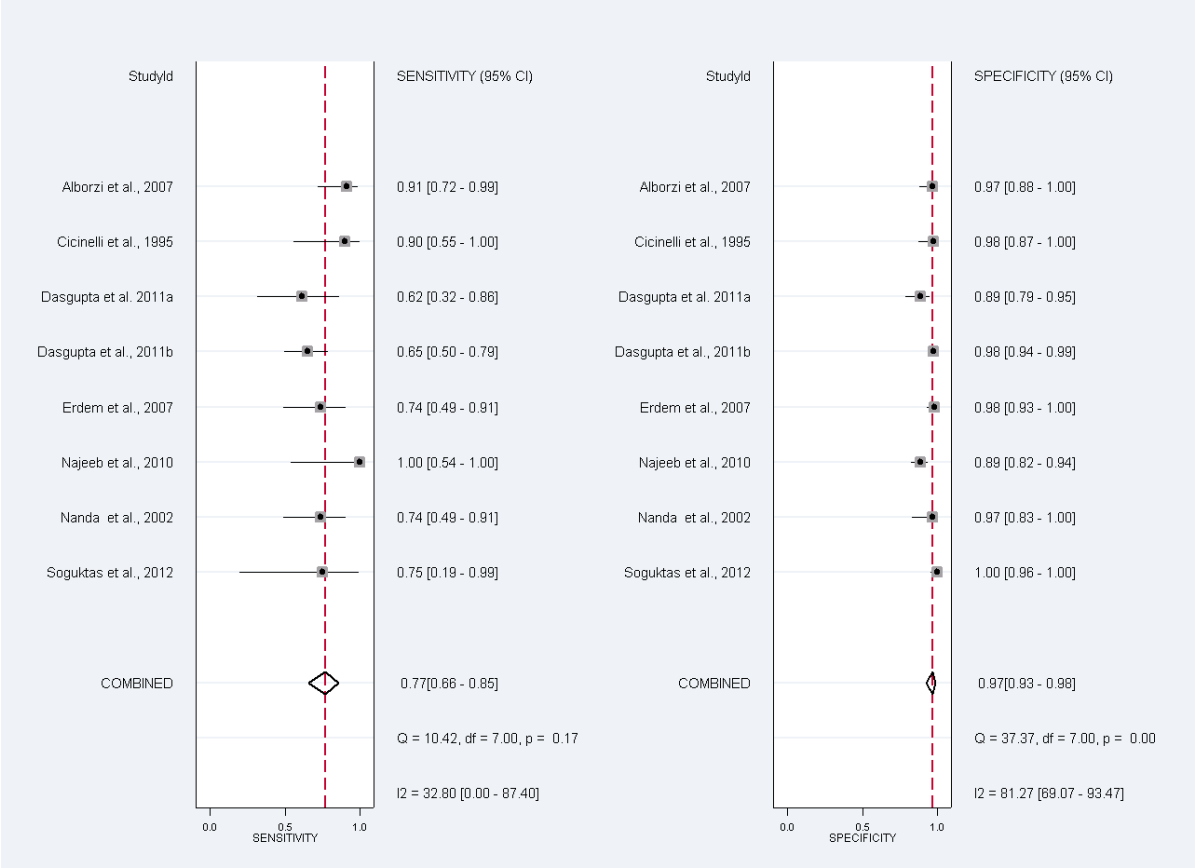
2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval

Figure 7: HSROC plot for diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting polyps (meta-analysis of studies published in 2007 or later)



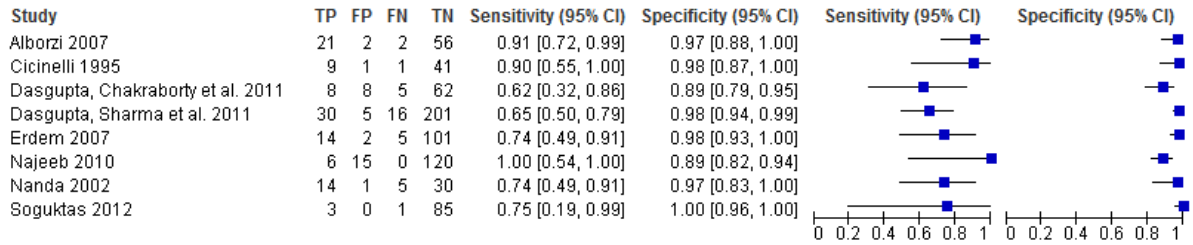
2D-TVUS: two-dimensional transvaginal ultrasound scan; HSROC: hierarchical summary receiver operating characteristic

Figure 8: Forest plot for sensitivity and specificity of 2D-TVUS (versus histopathology) in detecting fibroids (meta-analysis with all eligible studies)



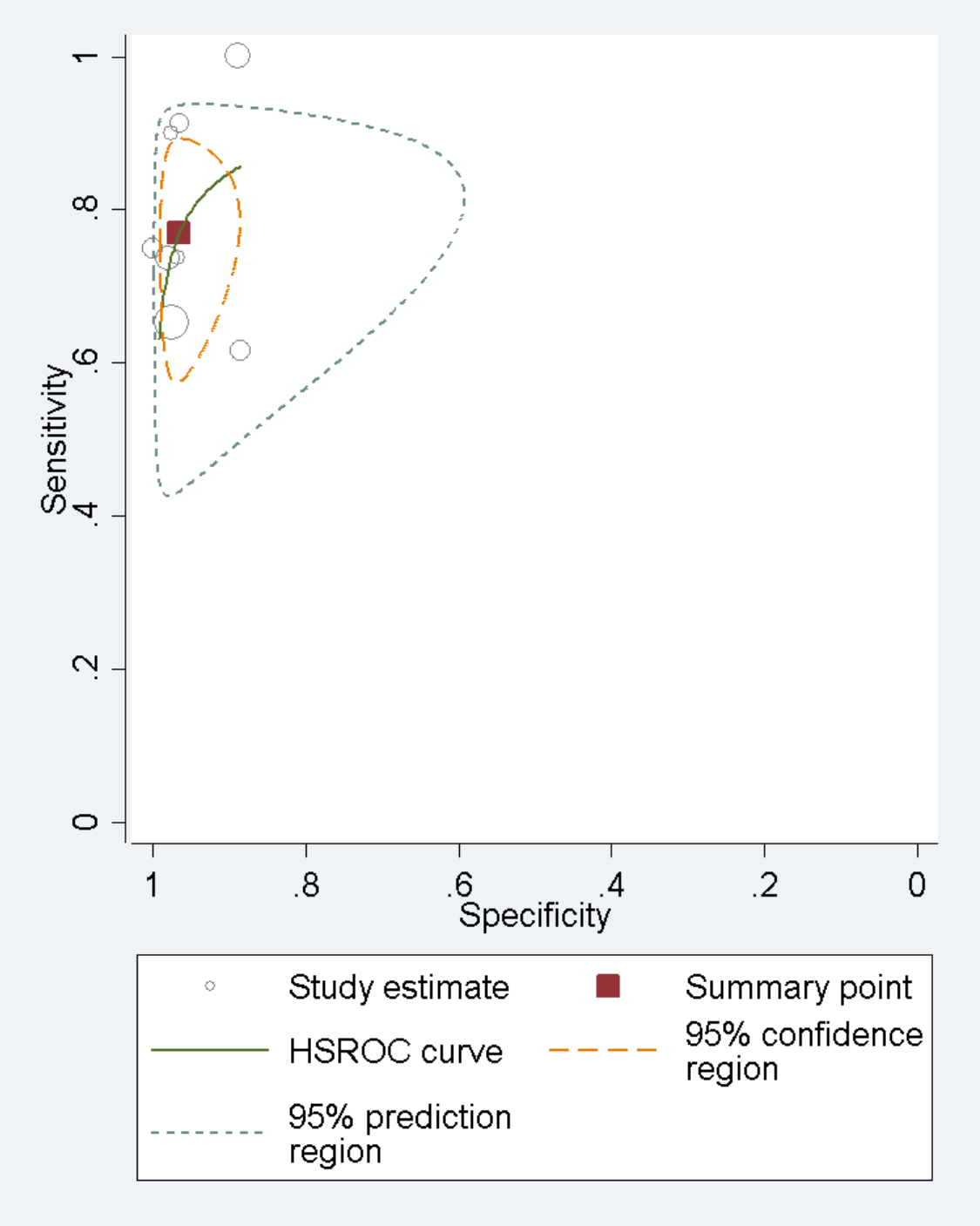
2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval

Figure 9: Forest plot for sensitivity and specificity of 2D-TVUS (versus histopathology) in detecting fibroids (all eligible studies without meta-analysis)



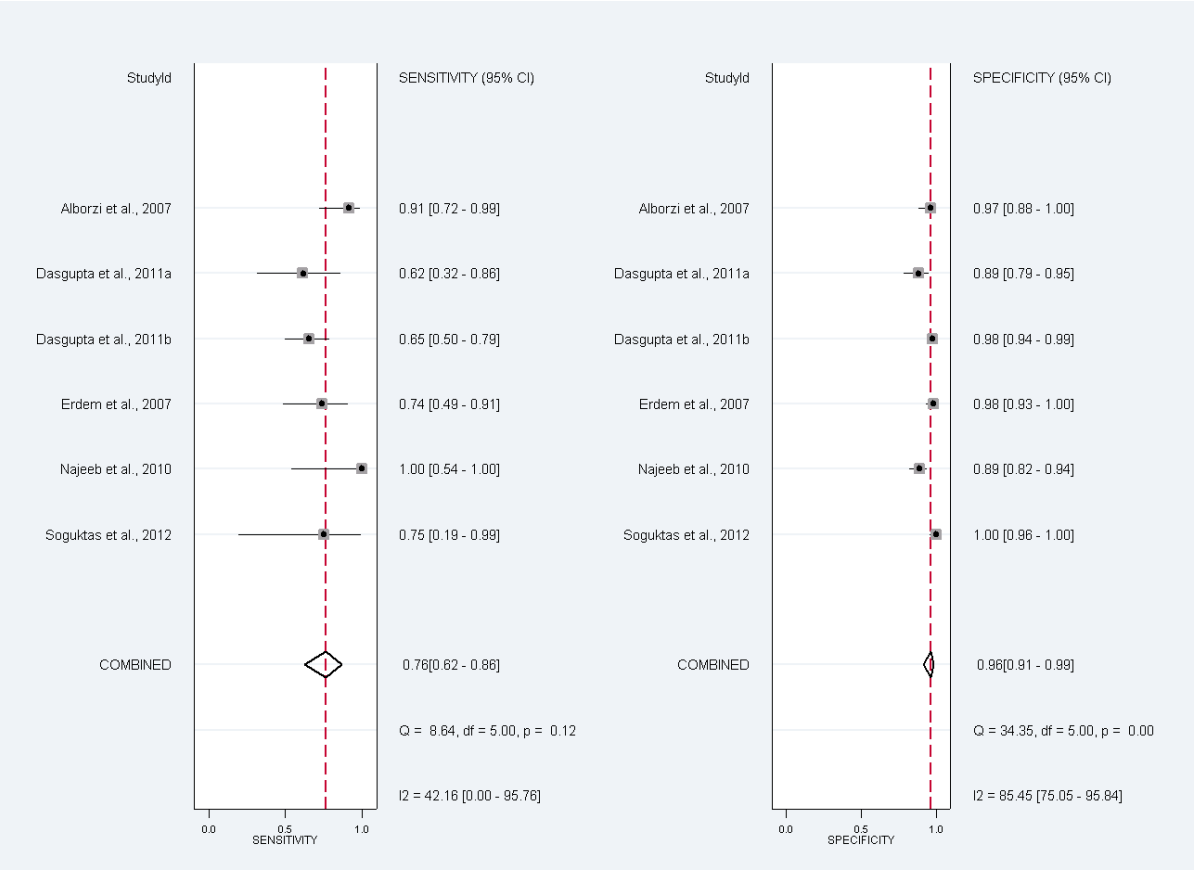
2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 10: HSROC plot for diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting fibroids (meta-analysis with all eligible studies)



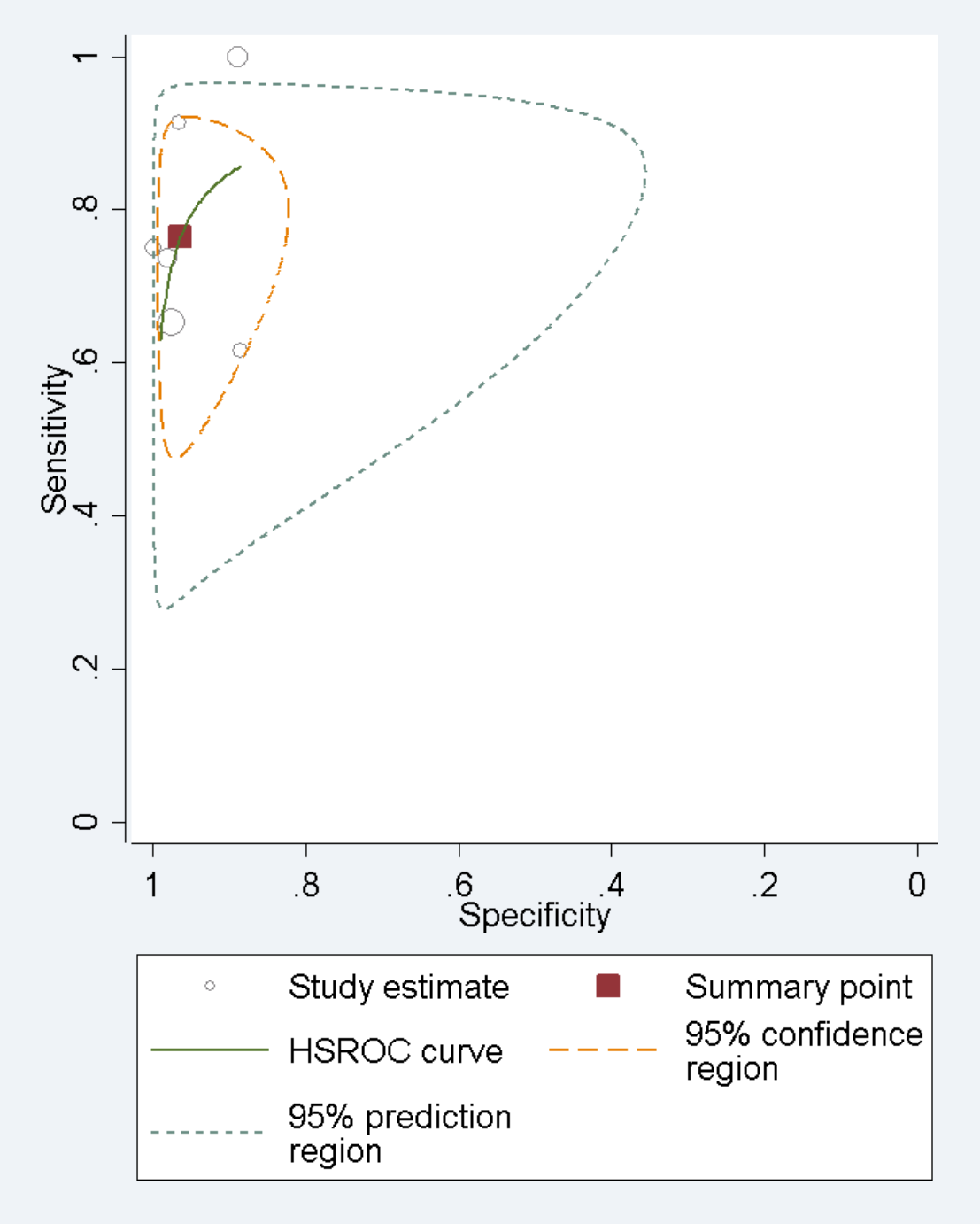
2D-TVUS: two-dimensional transvaginal ultrasound scan; HSROC: hierarchical summary receiver operating characteristic

Figure 11: Forest plot for sensitivity and specificity of 2D-TVUS (versus histopathology) in detecting fibroids (meta-analysis with studies published in 2007 or later)



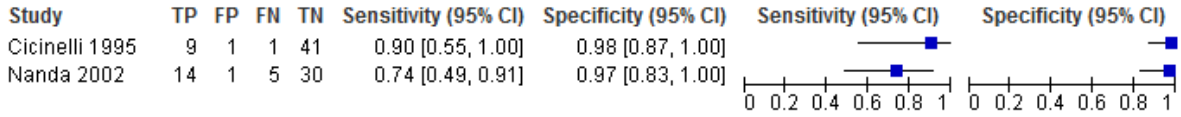
2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval

Figure 12: HSROC plot for diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting fibroids (meta-analysis with studies published in 2007 or later)



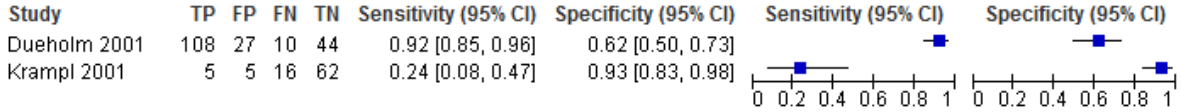
2D-TVUS: two-dimensional transvaginal ultrasound scan; HSROC: hierarchical summary receiver operating characteristic

Figure 13: Forest plot for sensitivity and specificity of 2D-TVUS (versus histopathology) in detecting fibroids (studies using hysterectomy as reference standard)



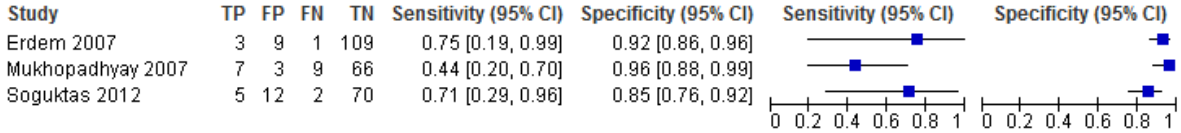
2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 14: Forest plot for sensitivity and specificity of 2D-TVUS (versus histopathology) in detecting polyps or fibroids



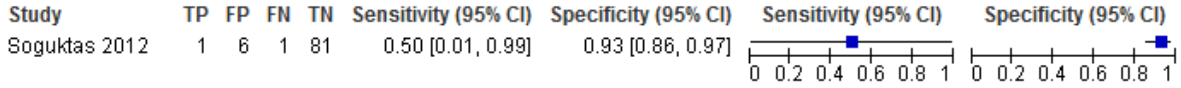
2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 15: Forest plot for sensitivity and specificity of 2D-TVUS (versus histopathology) in detecting endometrial hyperplasia



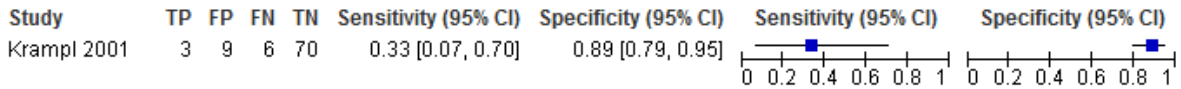
2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 16: Forest plot for sensitivity and specificity of 2D-TVUS (versus histopathology) in detecting endometrial carcinoma



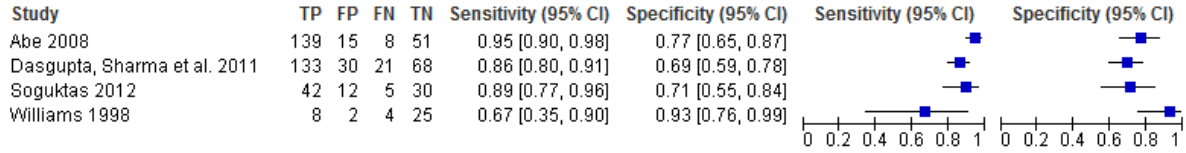
2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 17: Forest plot for sensitivity and specificity of 2D-TVUS (versus histopathology) in detecting abnormal endometrium (double-layer endometrium 12 mm or thicker and single-layer endometrium 6 mm or thicker)



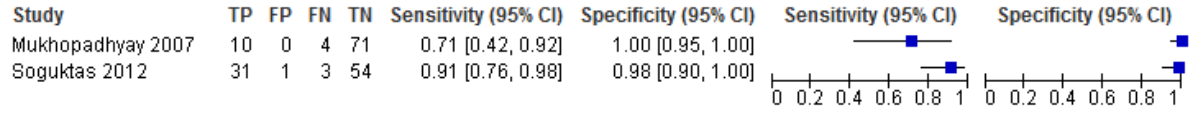
2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 18: Forest plot for sensitivity and specificity of 2D-TVUS (versus histopathology) in detecting any abnormal finding



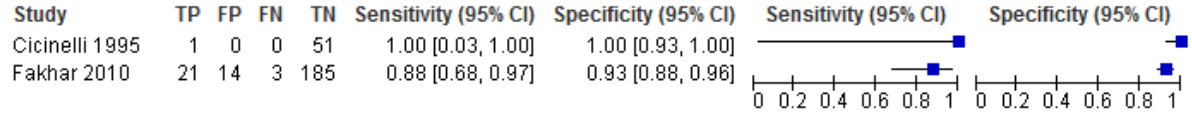
2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 19: Forest plot for sensitivity and specificity of hysteroscopy under general anaesthetic (versus histopathology) in detecting polyps



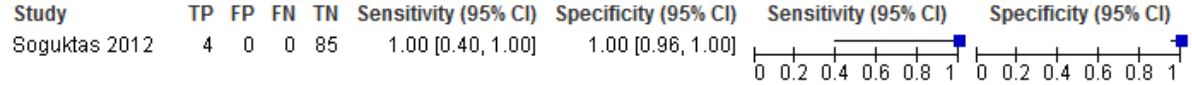
CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 20: Forest plot for sensitivity and specificity of outpatient hysteroscopy (versus histopathology) in detecting polyps



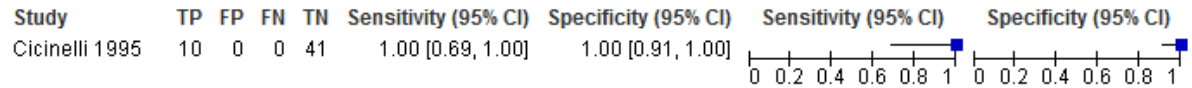
CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 21: Forest plot for sensitivity and specificity of hysteroscopy under general anaesthetic (versus histopathology) in detecting fibroids



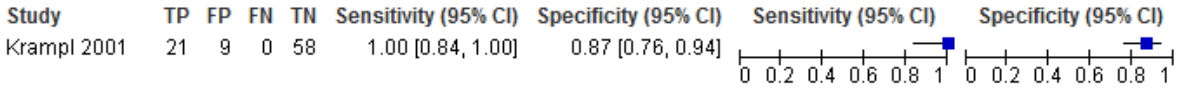
CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 22: Forest plot for sensitivity and specificity of outpatient hysteroscopy (versus histopathology) in detecting fibroids



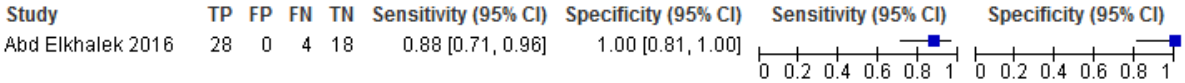
CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 23: Forest plot for sensitivity and specificity of hysteroscopy under spinal or general anaesthetic (versus histopathology) in detecting polyps or fibroids



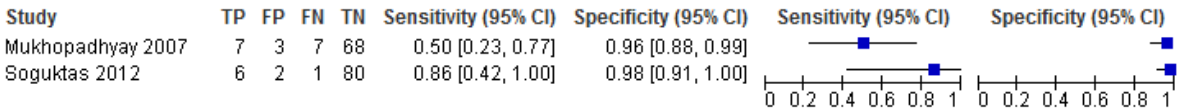
CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 24: Forest plot for sensitivity and specificity of hysteroscopy under general anaesthetic following 2D-TVUS (versus histopathology) in detecting polyps or fibroids



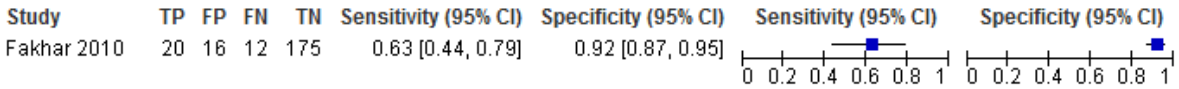
CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 25: Forest plot for sensitivity and specificity of hysteroscopy under general anaesthetic (versus histopathology) in detecting endometrial hyperplasia



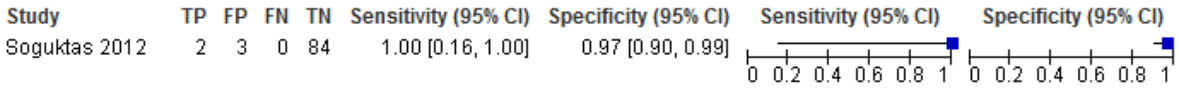
CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 26: Forest plot for sensitivity and specificity of outpatient hysteroscopy (versus histopathology) in detecting endometrial hyperplasia



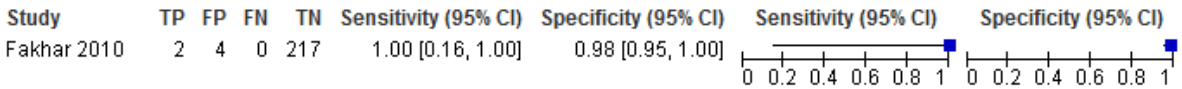
CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 27: Forest plot for sensitivity and specificity of hysteroscopy under general anaesthetic (versus histopathology) in detecting endometrial carcinoma



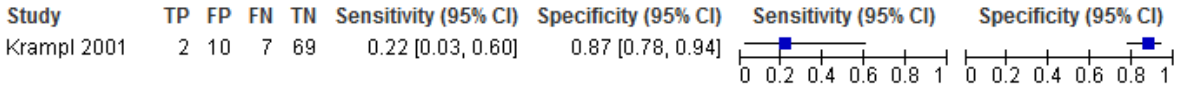
CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 28: Forest plot for sensitivity and specificity of outpatient hysteroscopy (versus histopathology) in detecting endometrial carcinoma



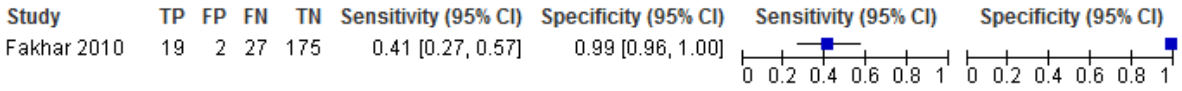
CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 29: Forest plot for sensitivity and specificity of hysteroscopy under spinal or general anaesthetic (versus histopathology) in detecting abnormal endometrium (double-layer endometrium 12 mm or thicker and single-layer endometrium 6 mm or thicker)



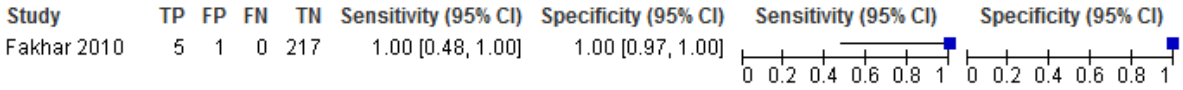
CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 30: Forest plot for sensitivity and specificity of outpatient hysteroscopy (versus histopathology) in detecting endometritis



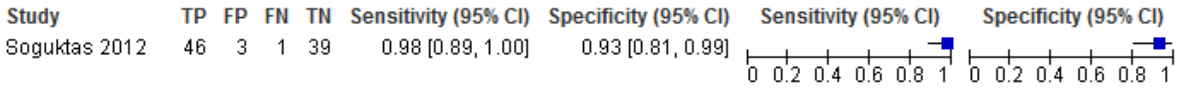
CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 31: Forest plot for sensitivity and specificity of outpatient hysteroscopy (versus histopathology) in detecting retained products of conception



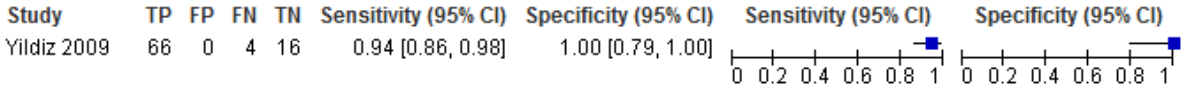
CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 32: Forest plot for sensitivity and specificity of hysteroscopy under general anaesthetic (versus histopathology) in detecting any abnormal finding



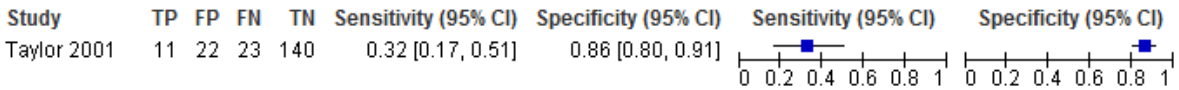
CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 33: Forest plot for sensitivity and specificity of hysteroscopy with or without general or local anaesthesia (versus histopathology) in detecting any abnormal finding



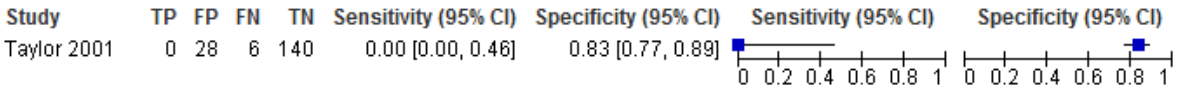
CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 34: Forest plot for sensitivity and specificity of 2D-TVUS (versus hysteroscopy) in detecting polyps



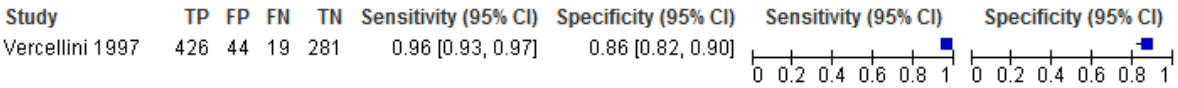
2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 35: Forest plot for sensitivity and specificity of 2D-TVUS (versus hysteroscopy) in detecting suspicious focal thickening



2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

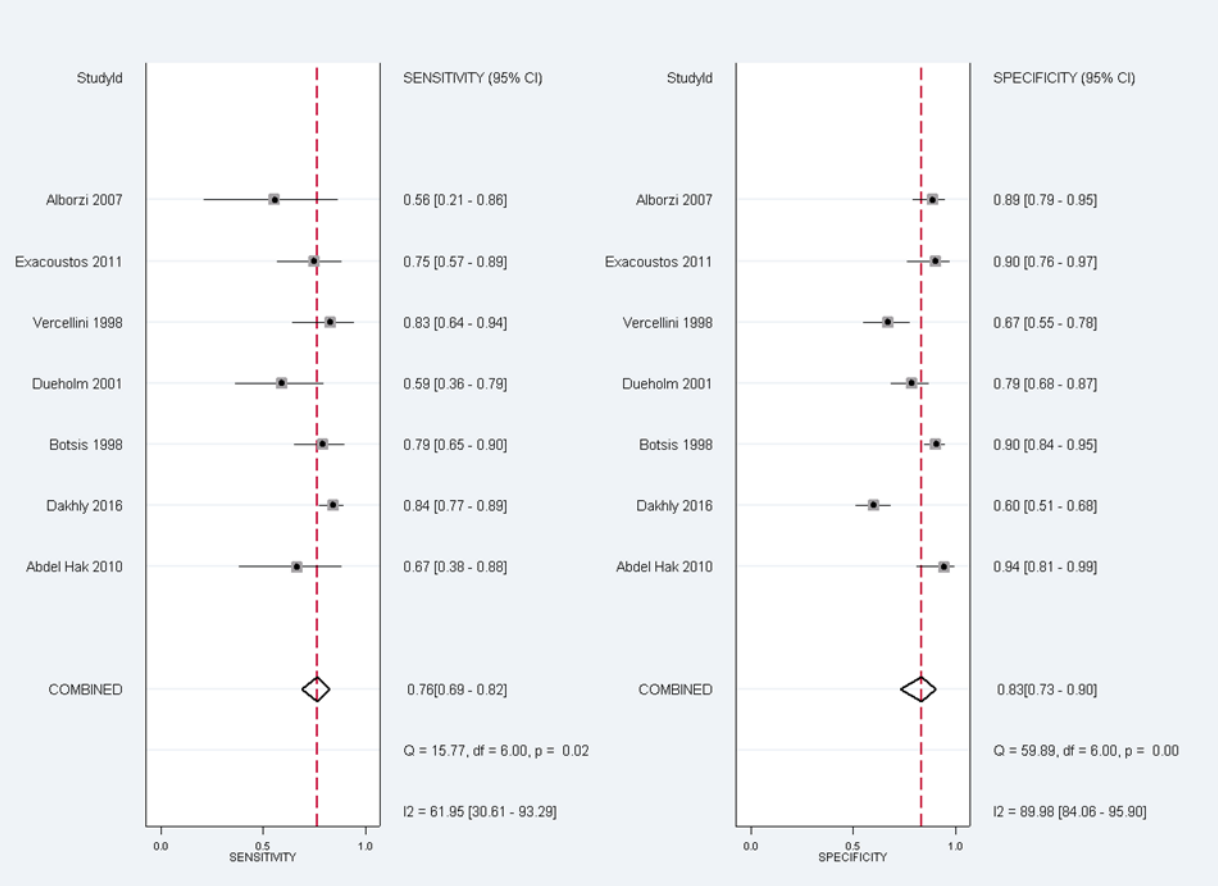
Figure 36: Forest plot for sensitivity and specificity of 2D-TVUS (versus hysteroscopy) in detecting any abnormal finding



2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

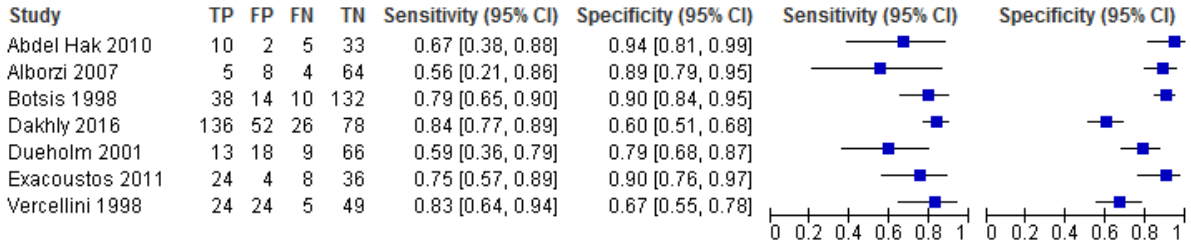
Review question 2. Forest plots and HSROC plots for the most clinically effective imaging strategy for diagnosing adenomyosis in women presenting with heavy menstrual bleeding review

Figure 37: Forest plot for sensitivity and specificity of 2D-TVUS (versus histopathology) in detecting adenomyosis (meta-analysis)



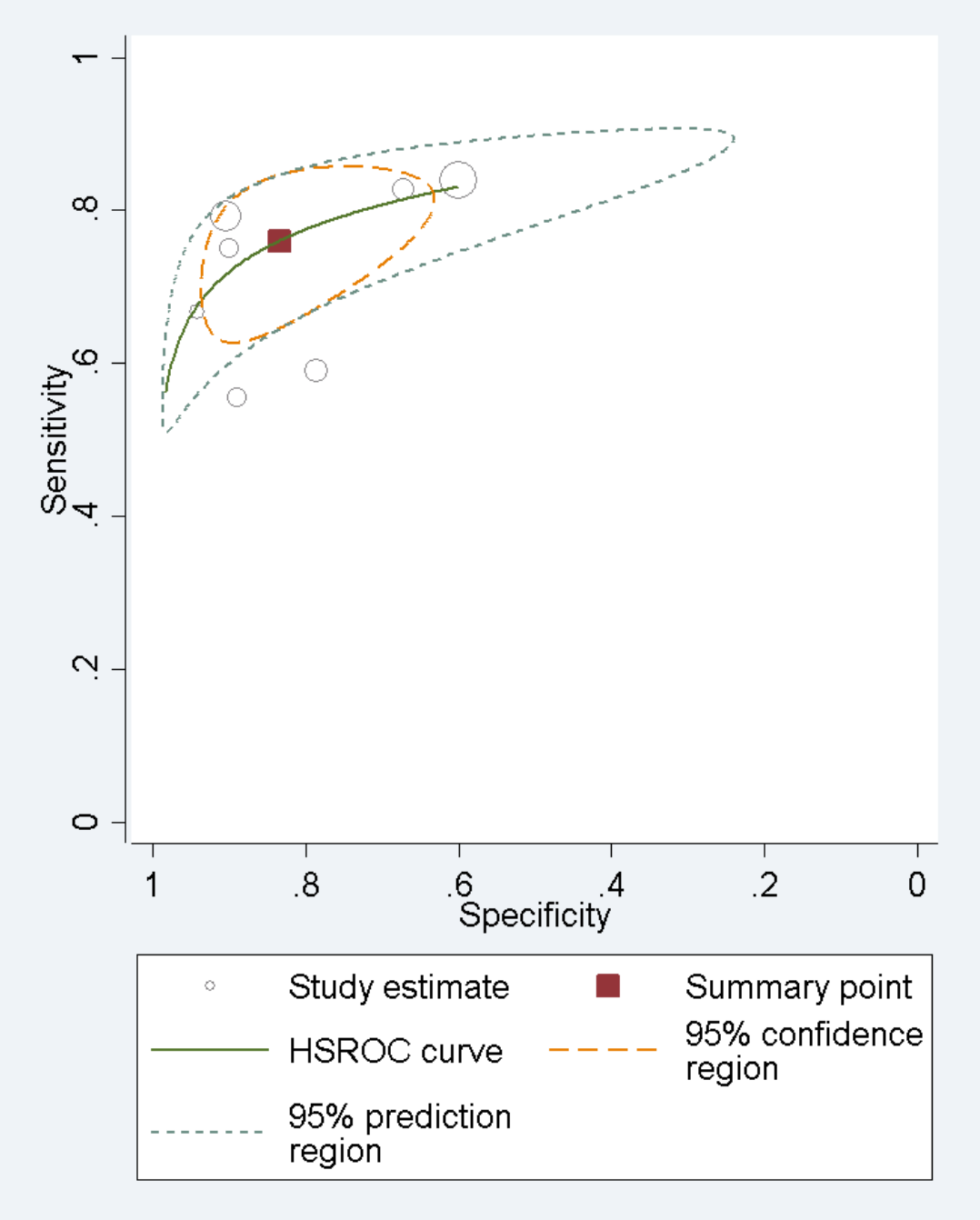
2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval

Figure 38: Forest plot for sensitivity and specificity of 2D-TVUS (versus histopathology) in detecting adenomyosis (studies included in the meta-analysis)



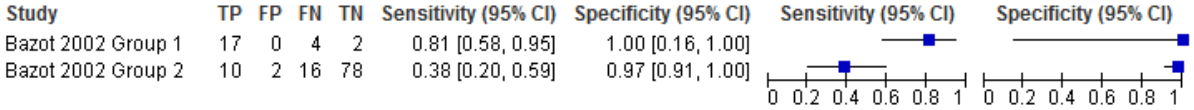
Dueholm 2001 reported 'indefinite' results from the index test for 36 cases which have been included as negative test results in the results. 2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 39: HSROC plot for diagnostic accuracy of 2D-TVUS (versus histopathology) in detecting adenomyosis (meta-analysis)



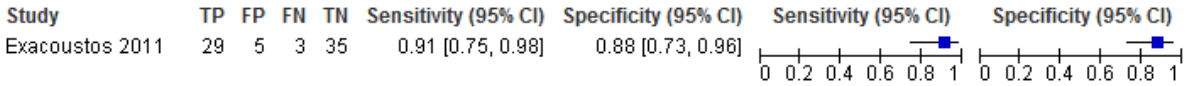
2D-TVUS: two-dimensional transvaginal ultrasound scan; HSROC: hierarchical summary receiver operating characteristic

Figure 40: Forest plot for sensitivity and specificity of 2D-TVUS (versus histopathology) in detecting adenomyosis in Bazot 2002



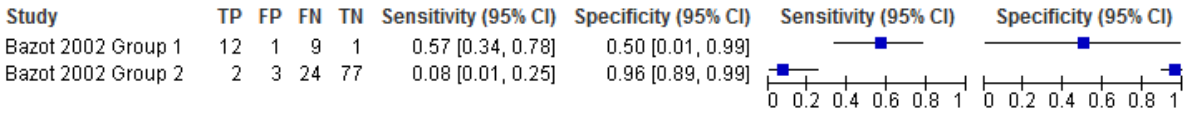
Group 1, women with recurrent menometrorrhagia but no evidence of leiomyomata and endometrial diseases on transabdominal examination; Group 2, all other women. 2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 41: Forest plot for sensitivity and specificity of 3D-TVUS (versus histopathology) in detecting adenomyosis



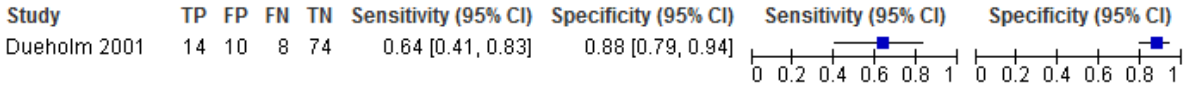
3D-TVUS: three-dimensional transvaginal ultrasound scan; CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 42: Forest plot for sensitivity and specificity of 2D-TAUS (versus histopathology) in detecting adenomyosis



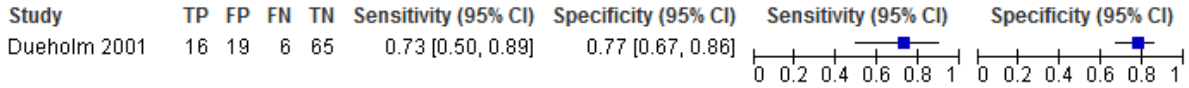
Group 1, women with recurrent menometrorrhagia but no evidence of leiomyomata and endometrial diseases on transabdominal examination; Group 2, all other women. 2D-TAUS: two-dimensional transabdominal ultrasound scan; CI: confidence interval; FN: false negative; FP: false positive; TN: true negative; TP: true positive

Figure 43: Forest plot for sensitivity and specificity of MRI (versus histopathology) in detecting adenomyosis



The study reported 'indefinite' results from the index test for 13 cases which have been included as negative test results in the results. CI: confidence interval; FN: false negative; FP: false positive; MRI: magnetic resonance imaging; TN: true negative; TP: true positive

Figure 44: Forest plot for sensitivity and specificity of 2D-TVUS combined with MRI (versus histopathology) in detecting adenomyosis



The study reported 'indefinite' results from the index test for 41 cases which have been included as negative test results in the results. 2D-TVUS: two-dimensional transvaginal ultrasound scan; CI: confidence interval; FN: false negative; FP: false positive; MRI: magnetic resonance imaging; TN: true negative; TP: true positive

Appendix I – Excluded studies

Clinical studies

Table 39: Studies excluded from the clinical evidence reviews for diagnostic test accuracy for investigations of women presenting with heavy menstrual bleeding

Reference	Reason for exclusion
Abd El Aal, D. E., Ali, M., Shaaban, O., Sabra, A., Color doppler ultrasound to improve the diagnostic accuracy of the transvaginal ultrasound in diagnosis of adenomyosis: A cross sectional study, International Journal of Gynecology and Obstetrics, 131, E203, 2015	Conference abstract.
Abou-Salem, N., Elmazny, A., El-Sherbiny, W., Value of 3-Dimensional Sonohysterography for Detection of Intrauterine Lesions in Women with Abnormal Uterine Bleeding, Journal of Minimally Invasive Gynecology, 17, 2010	Wrong population: 57% of participants with heavy menstrual bleeding.
Acog Committee on Gynecologic Practice, Committee Opinion: number 263, December 20001. von Willebrand's disease in gynecologic practice, Obstetrics & GynecologyObstet Gynecol, 98, 1185-6, 2001	No relevant data.
Adishesh, M., Subramanian, M., Patient satisfaction survey of outpatient hysteroscopy service, BJOG: An International Journal of Obstetrics and Gynaecology, Conference, 2013	Conference abstract.
Agarwal, A., Lee, L., The feasibility and yield of outpatient hysteroscopy, International Journal of Gynecology and Obstetrics, Conference, 2012	Conference abstract.
Aggarwal, A., Lilley, A., Sproston, T., Outpatient hysteroscopy audit, BJOG: An International Journal of Obstetrics and Gynaecology, Conference, 2013	Conference abstract.
Aghajanyan, H., The improvement of the outpatient diagnostics of intrauterine pathology, Gynecological Surgery, Conference, 2013	Conference abstract.
Aghajanyan, H., Outpatient diagnostics of endometrial polyps, Gynecological Surgery, Conference, 2013	Conference abstract.

Reference	Reason for exclusion
Ahmadi,F., Zafarani,F., Niknejadi,M., Vosough,A., Uterine leiomyoma: Hysterosalpingographic appearances, International Journal of Fertility and Sterility, 1, 137-144, 2008	Pictorial review, references checked for relevant studies.
Ajao, M. O., Laughlin-Tommaso, S. K., Hopkins, M. R., Breitkopf, D. M., Famuyide, A. O., Years of hysteroscopic experience in correctly predicting pathologic diagnosis, Journal of Minimally Invasive Gynecology, Conference, 41st Global Congress of Minimally Invasive Gynecology, AAGL 2012. Las Vegas, NV United States. Conference Start: 20121105. Conference End: 20121109. Conference Publication: (var.pagings). 19 (6 SUPPL. 1) (pp S111), 2012	Conference abstract.
Akhavan Tafti, M., Davar, R., Nafisi Moghadam, R., Lotfi, M. H., Panahande Ali, M., Comparing findings of transvaginal ultrasonography and endometrial histopathology in patients with abnormal uterine bleeding, Iranian Journal of Obstetrics, Gynecology and Infertility, 15, 7-13, 2012	Wrong population: 48% of women with HMB.
Alcazar, J. L., Auba, M., Olartecoechea, B., Three-dimensional ultrasound in gynecological clinical practice, Reports in Medical Imaging, 5, 2012	Expert review. Relevant references checked.
Alcazar, J. L., Galvan, R., Garcia-Manero, M., Ajossa, S., Guerriero, S., Melis, G. B., Transvaginal ultrasound in the diagnosis of uterine pathology, Expert Review of Obstetrics and Gynecology, 3, 753-760, 2008	Expert review. Relevant references checked.
Anastasiadis, P. G., Koutlaki, N. G., Skaphida, P. G., Galazios, G. Ch, Tsikouras, P. N., Liberis, V. A., Endometrial polyps: Prevalence, detection, and malignant potential in women with abnormal uterine bleeding, European Journal of Gynaecological Oncology, 21, 180-183, 2000	Unclear/wrong population. HMB status or premenopausal status of the women are not reported for the whole sample. For the 126 women with polyps and who were studied more, more than half were postmenopausal.
Andrade, A. T. L., Souza, J. P., Shaw Jr, S. T., Belsey, E. M., Rowe, P. J., Menstrual blood loss and body iron stores in Brazilian women, Contraception, 43, 241-249, 1991	No relevant data.
Andreotti, R. F., Fleischer, A. C., The sonographic diagnosis of adenomyosis, Ultrasound Quarterly, 21, 167-70, 2005	Narrative review, references checked.
Anonymous,, AAGL Practice Report: Practice Guidelines for the Diagnosis and Management of Submucous Leiomyomas, Journal of Minimally Invasive Gynecology, 19, 152-171, 2012	Practice guideline. Relevant references checked.
Antunes, A., The efficacy of hysteroscopy in diagnosis and treatment of endometrial pathology: See and treat, Gynecological Surgery, Conference, 2009	Conference abstract.

Reference	Reason for exclusion
Arslan, M., Erdem, A., Erdem, M., Yazici, G., Himmetoglu, O., Gursoy, R., Transvaginal color Doppler ultrasonography for prediction of pre-cancerous endometrial lesions, International Journal of Gynecology and Obstetrics, 80, 299-306, 2003	Wrong population, majority of the women are postmenopausal.
Ascher, S. M., Jha, R. C., Reinhold, C., Benign Myometrial Conditions: Leiomyomas and Adenomyosis, Topics in Magnetic Resonance Imaging, 14, 281-304, 2003	A narrative/pictorial review, references checked.
Ash, S. J., Farrell, S. A., Flowerdew, G., Endometrial biopsy in DUB, Journal of Reproductive Medicine for the Obstetrician and Gynecologist, 41, 892-896, 1996	Index test not of interest.
Atri, M., Reinhold, C., Mehio, A. R., Chapman, W. B., Bret, P. M., Adenomyosis: US features with histologic correlation in an in vitro study, Radiology, 215, 783-790, 2000	Wrong population.
Atzori, E., Tronci, C., Sionis, L., Transvaginal ultrasound in the diagnosis of diffuse adenomyosis, Gynecologic and Obstetric Investigation, 42, 39-41, 1996	Wrong/unclear population.
Babacan, A., Gun, I., Kizilaslan, C., Ozden, O., Muhcu, M., Mungen, E., Atay, V., Comparison of transvaginal ultrasonography and hysteroscopy in the diagnosis of uterine pathologies, International Journal of Clinical and Experimental Medicine, 7, 764-769, 2014	Population not well defined: % of women with HMB unclear, age of patients not reported.
Badawy, A., Ash, A., Nagele, F., Mohamed, H., O'Connor, H., Magos, A., Ultrasonography, hysteroscopy or both?, Journal of Obstetrics and Gynaecology, 16, 551-555, 1996	No sufficient data to form 2x2 tables and calculate diagnostic accuracy.
Bain, C., Parkin, D. E., Cooper, K. G., Is outpatient diagnostic hysteroscopy more useful than endometrial biopsy alone for the investigation of abnormal uterine bleeding in unselected premenopausal women? A randomised comparison, BJOG: An International Journal of Obstetrics and Gynaecology, 109, 805-811, 2002	No data on diagnostic accuracy but some data on patient acceptability/satisfaction, however, wrong test comparisons (hysteroscopy compared to endometrial biopsy alone).
Balogun, M., Imaging diagnosis of adenomyosis, Reviews in Gynaecological and Perinatal Practice, 6, 63-69, 2006	A narrative review, references checked.
Banerjee, R., Ofuasia, E., Study to assess correlation between diagnostic imaging and histology of adenomyosis at hysterectomy, Gynecological Surgery, 1), S61, 2012	Conference abstract.
Banu, L. F., Hysteroscopy a diagnostic and therapeutic tool-500 case series, BJOG: An International Journal of Obstetrics and Gynaecology, Conference, 2014	Conference abstract.

Reference	Reason for exclusion
Basu, A., Lewis, P., Transvaginal scan, saline infusion sonography and hysteroscopy in abnormal uterine bleeding, International Journal of Gynecology and Obstetrics, Conference, 2009	Conference abstract.
Batistatou, A., Ntoulia, A., Gkrozou, F., Paschopoulos, M., Malamou-Mitsi, V., Correlation of hyperspectral hysteroscopy with histopathological features in endometrial pathology, Virchows Archiv, Conference, 2013	Conference abstract.
Baxter, A. J., Beck, B., Phillips, K., A randomized prospective trial of rigid and flexible hysteroscopy in an outpatient setting, Gynaecological Endoscopy, 11, 357-364, 2002	Wrong population, less than two thirds with HMB.
Bazot, M., Cortez, A., Darai, E., Rouger, J., Chopier, J., Antoine, J. M., Uzan, S., Ultrasonography compared with magnetic resonance imaging for the diagnosis of adenomyosis: Correlation with histopathology, Human Reproduction, 16, 2427	Wrong population.
Bazot, M., Darai, E., De Givry, S. C., Boudghene, F., Uzan, S., Le Blanche, A. F., Fast breath-hold T2-weighted MR imaging reduces interobserver variability in the diagnosis of adenomyosis, American Journal of Roentgenology, 180, 1291	Wrong population, less than two thirds with HMB.
Bazot, M., Robert, Y., Diagnostic imaging of menometrorrhagia, Journal de Gynecologie Obstetrique et Biologie de la Reproduction, 37, 2008	Full article in French language.
Bazot, M., Salem, C., Frey, I., Darai, E., Imaging of myomas: Is MRI useful before operating?, Gynecologie Obstetrique Fertilite, 30, 2002	Full article in French language.
Becker Jr, E., Lev-Toaff, A. S., Kaufman, E. P., Halpern, E. J., Edelweiss, M. I., Kurtz, A. B., The added value of transvaginal sonohysterography over transvaginal sonography alone in women with known or suspected leiomyoma, Journal of Ultrasound in Medicine, 21, 237-247, 2002	Population not defined.
Belli, P., De Gaetano, A. M., Mirk, P., Speca, S., Valentini, A. L., Uterine adenomyosis and tubal endometriosis: diagnostic imaging, Rays, 23, 693-701, 1998	Not a study, not relevant.
Benagiano, G., Habiba, M., Brosens, I., The pathophysiology of uterine adenomyosis: an update, Fertility & Sterility, 98, 572-9, 2012	Narrative review. Possibly relevant references checked.
Ben-Baruch, G., Seidman, D. S., Schiff, E., Moran, O., Menczer, J., Outpatient endometrial sampling with the pipelle curette, Gynecologic and Obstetric Investigation, 37, 260-262, 1994	Index test not of interest.

Reference	Reason for exclusion
Ben-Yehuda, O. M., Kim, Y. B., Leuchter, R. S., Does hysteroscopy improve upon the sensitivity of dilatation and curettage in the diagnosis of endometrial hyperplasia or carcinoma?, Gynecologic Oncology, 68, 4-7, 1998	Wrong population: 46% of women postmenopausal.
Bernard, J. P., Lecuru, F., Darles, C., Robin, F., De Bievre, P., Taurelle, R., Saline contrast sonohysterography as first-line investigation for women with uterine bleeding, Ultrasound in Obstetrics and Gynecology, 10, 121-125, 1997	Wrong population: 67.3% premenopausal women with abnormal uterine bleeding (rest postmenopausal), proportion with HMB not reported.
Bettocchi, S., Ceci, O., Vicino, M., Marello, F., Impedovo, L., Selvaggi, L., Diagnostic inadequacy of dilatation and curettage, Fertility and Sterility, 75, 803-805, 2001	Index test not of interest.
Bignardi, T., Van den Bosch, T., Condous, G., Abnormal uterine and post-menopausal bleeding in the acute gynaecology unit, Best Practice and Research: Clinical Obstetrics and Gynaecology, 23, 595-607, 2009	Expert review. Possible relevant references checked.
Bingol,B., Gunenc,Z., Gedikbasi,A., Guner,H., Tasdemir,S., Tiras,B., Comparison of diagnostic accuracy of saline infusion sonohysterography, transvaginal sonography and hysteroscopy, Journal of Obstetrics and Gynaecology, 31, 54-58, 2011	Wrong population: women with infertility, 43.4% women without abnormal bleeding or hypomenorrhea.
Bocca, S. M., Oehninger, S., Stadtmauer, L., Agard, J., Duran, E. H., Sarhan, A., Horton, S., Abuhamad, A. Z., A study of the cost, accuracy, and benefits of 3-dimensional sonography compared with hysterosalpingography in women with uterine abnormalities, Journal of Ultrasound in Medicine, 31, 81-5, 2012	Wrong/unclear population: % of women with HMB unclear.
Bodur, S., Dundar, O., Pektas, M. K., Babayigit, M. A., Ozden, O., Kucukodaci, Z., The clinical significance of classical and new emerging determinants of adenomyosis, International Journal of Clinical and Experimental Medicine, 8, 7958-7964, 7958	Wrong index test.
Bougie, O., Wang, V., Lortie, K., Shenassa, H., Singh, S. S., High patient satisfaction with outpatient hysteroscopy with use of tailored analgesia protocols, Journal of Minimally Invasive Gynecology, Conference, 2013	Conference abstract.
Bourdel, N., Modaffari, P., Tognazza, E., Pertile, R., Chauvet, P., Botchorishivili, R., Savary, D., Pouly, J. L., Rabischong, B., Canis, M., Does experience in hysteroscopy improve accuracy and inter-observer agreement in the management of abnormal uterine bleeding?, Surgical Endoscopy and Other Interventional Techniques., 2016	Wrong population: 48.9% of women post-menopausal.

Reference	Reason for exclusion
Bouzid, A., Ayachi, A., Ben Khedija, M., Menjli, S., Mkaouar, L., Mourali, M., Journal de Gynecologie Obstetrique et Biologie de la Reproduction. (no pagination), 2016, Date of Publication, February 08, 2016	Full text in French language.
Bradley, L. D., Falcone, T., Magen, A. B., Radiographic imaging techniques for the diagnosis of abnormal uterine bleeding, Obstetrics and Gynecology Clinics of North America, 27, 245-276, 2000	Narrative review, not relevant.
Bradley,L.D., Diagnosis of abnormal uterine bleeding with biopsy or hysteroscopy, Menopause, 18, 425-433, 2011	Expert review. Relevant references checked.
Brasic, N., Feldstein, V. A., Dysfunctional uterine bleeding: Diagnostic approach and therapeutic options, Ultrasound Clinics, 5, 2010	Expert review. Relevant references checked.
Breitkopf, D. M., Frederickson, R. A., Snyder, R. R., Detection of benign endometrial masses by endometrial stripe measurement in premenopausal women, Obstetrics and Gynecology, 104, 120-125, 2004	Index test not of interest.
Bromley, B., Shipp, T. D., Benacerraf, B., Adenomyosis: Sonographic findings and diagnostic accuracy, Journal of Ultrasound in Medicine, 19, 529-534, 2000	Wrong population.
Bronz, L., Suter, T., Rusca, T., The value of transvaginal sonography with and without saline instillation in the diagnosis of uterine pathology in pre- and postmenopausal women with abnormal bleeding or suspect sonographic findings, Ultrasound in Obstetrics and Gynecology, 9, 53-58, 1997	Wrong population: only a small proportion of the women had HMB.
Brown, M. A., MR Imaging of Benign Uterine Disease, Magnetic Resonance Imaging Clinics of North America, 14, 2006	Expert review on MRI on benign uterine disease. Relevant references checked.
Byun, J. Y., Kim, S. E., Choi, B. G., Ko, G. Y., Jung, S. E., Choi, K. H., Diffuse and focal adenomyosis: MR imaging findings, Radiographics : a review publication of the Radiological Society of North America, Inc, 19 Spec No, S161-170, 1999	Wrong population.
Cepni, I., Ocal, P., Erkan, S., Saricali, F. S., Akbas, H., Demirkiran, F., Idil, M., Bese, T., Comparison of transvaginal sonography, saline infusion sonography and hysteroscopy in the evaluation of uterine cavity pathologies, Australian and New Zealand Journal of Obstetrics and Gynaecology, 45, 30-35, 2005	Wrong population: only 33% had abnormal uterine bleeding (% of HMB not reported).

Reference	Reason for exclusion
Champaneria, R., Abedin, P., Daniels, J., Balogun, M., Khan, K. S., Ultrasound scan and magnetic resonance imaging for the diagnosis of adenomyosis: Systematic review comparing test accuracy, Acta Obstetricia et Gynecologica Scandinavica, 89, 1374	A systematic review with wrong population. Included studies assessed individually.
Chandra, C., Shekhar, S., Vyas, J., The role of transvaginal ultrasonography in initial evaluation of abnormal uterine bleeding (AUB) in premenopausal women, BJOG: An International Journal of Obstetrics and Gynaecology, Conference, 2014	Conference abstract.
Chauhan, M., Dahiya, P., Nanda, S., Role of hysteroscopy in evaluation of abnormal uterine bleeding in perimenopausal women, Maturitas, Conference, 2009	Conference abstract.
Chawla, I., Tripathi, S., Vohra, P., Singh, P., To evaluate the accuracy of saline infusion sonohysterography (SIS) for evaluation of uterine cavity abnormalities in patients with abnormal uterine bleeding, Journal of Obstetrics and Gynecology of India, 64, 2014	Wrong population: 51.7% of women postmenopausal.
Chittacharoen, A., Theppisai, U., Linasmita, V., Manonai, J., Sonohysterography in the diagnosis of abnormal uterine bleeding, Journal of Obstetrics and Gynaecology Research, 26, 277-281, 2000	Wrong index test.
Chopra, S., Lev-Toaff, A. S., Ors, F., Bergin, D., Adenomyosis: Common and uncommon manifestations on sonography and magnetic resonance imaging, Journal of Ultrasound in Medicine, 25, 617-627, 2006	Wrong/unclear population.
Chu, L. C., Coquia, S. F., Hamper, U. M., Ultrasonography evaluation of pelvic masses, Radiologic Clinics of North America, 52, 1237	Expert review. Possible relevant references checked.
Claessens, E. A., Cowell, C. A., Acute adolescent menorrhagia, American Journal of Obstetrics and Gynecology, 139, 277-280, 1981	No relevant data.
Clark, T. J., Voit, D., Gupta, J. K., Hyde, C., Song, F., Khan, K. S., Accuracy of hysteroscopy in the diagnosis of endometrial cancer and hyperplasia: a systematic quantitative review, JAMAJama, 288, 1610-21, 2002	Systematic review with different inclusion criteria. Included studies checked individually.
Cockerham, A. Z., Adenomyosis: a challenge in clinical gynecology, Journal of Midwifery & Women's Health, 57, 212-20, 2012	A narrative review. References checked.
Cogendez, E., Eken, M. K., Bakal, N., Gun, I., Kaygusuz, E. I., Karateke, A., The role of transvaginal power Doppler ultrasound in the differential diagnosis of benign intrauterine focal lesions, Journal of Medical Ultrasonics, 42, 2015	Wrong index test (power Doppler).

Reference	Reason for exclusion
Cooper, N. A. M., Barton, P. M., Breijer, M. C., Opmeer, B. C., Mol, B. W. J., Khan, K. S., Clark, T. J., Cost-effectiveness of diagnostic strategies for the management of heavy menstrual bleeding, BJOG: An International Journal of Obstetrics and Gynaecology, Conference, 2013	Conference abstract.
Cooper, N., Barton, P., Breijer, M., Khan, K., Mol, B., Clark, J., Outpatient hysteroscopy is the most costeffective investigation for heavy menstrual bleeding, Gynecological Surgery, Conference, 2012	Conference abstract.
Cooper,N.A.M., Barton,P.M., Breijer,M., Caffrey,O., Opmeer,B.C., Timmermans,A., Mol,B.W.J., Khan,K.S., Clark,T.J., Cost-effectiveness of diagnostic strategies for the management of abnormal uterine bleeding (heavy menstrual bleeding and post-menopausal bleeding): A decision analysis, Health Technology Assessment, 18, 1-201, 2014	Clinical and cots-effectiveness review, wrong inclusion criteria, relevant references checked individually.
Cowan, F., Shirol, V., Burnham, A., Patient feedback on outpatient hysteroscopy in a busy UK district general hospital-the importance of improving patient care in ambulatory gynaecology, Gynecological Surgery, Conference, 2015	Conference abstract.
Crispi Jr, C. P., Soares, L. C., Adenomyosis, Revista Brasileira de Medicina, 70, 409-414, 2013	Full text in Portuguese.
Daniele,A., Ferrero,A., Maggiorotto,F., Perrini,G., Volpi,E., Sismondi,P., Suspecting malignancy in endometrial polyps: Value of hysteroscopy, Tumori, 99, 204-209, 2013	Wrong population: 61.5% postmenopausal women.
Daniilidis, A., Pantelis, A., Dinas, K., Tantanasis, T., Loufopoulos, P. D., Angioni, S., Carcea, F., Indications of diagnostic hysteroscopy, a brief review of the literature, Gynecological Surgery., 2011	Narrative review, references checked for relevant studies.
Dartmouth, K., A systematic review with meta-analysis: the common sonographic characteristics of adenomyosis, Ultrasound, 22, 148-57, 2014	Systematic review and meta-analysis on the ultrasound characteristics of adenomyosis.
Darwish, A. M. M., Sayed, E. H., Reliability of out-patient hysteroscopy in one-stop clinic for abnormal uterine bleeding atef M. Darwish MD PHD, Ezzat H. Sayed MD, Safwat A. Mohammad MD, Ibraheem I. Mohammad MSC, Hoida I Hassan , PHD, Gynecological Surgery, Conference, 2011	Conference abstract.
De Crespigny, L., Kuhn, R., McGinnes, D., Saline infusion sonohysterosalpingography, an underutilized technique, Australian and New Zealand Journal of Obstetrics and Gynaecology, 37, 206-209, 1997	Wrong comparison of tests, no relevant data.

Reference	Reason for exclusion
De Kroon, C. D., De Bock, G. H., Dieben, S. W. M., Jansen, F. W., Saline contrast hysterosonography in abnormal uterine bleeding: A systematic review and meta-analysis, BJOG: An International Journal of Obstetrics and Gynaecology, 110, 938-947, 2003	Systematic review, wrong index test.
De Sousa Filogonio, I. D., De Avila, I., Gouvea, P. S., Carneiro, M. M., Accuracy of hysteroscopic view in the diagnosis of intrauterine pathology: A Brazilian experience, Journal of Gynecologic Surgery, 26, 23-30, 2010	Wrong population: 29% of women with abnormal uterine bleeding.
De Vries, L. D., Dijkhuizen, F. P. H. L. J., Mol, B. W. J., Brolmann, H. A. M., Moret, E., Heintz, A. P. M., Comparison of transvaginal sonography, saline infusion sonography, and hysteroscopy in premenopausal women with abnormal uterine bleeding, Journal of Clinical Ultrasound, 28, 217-223, 2000	Wrong population: only 53% with HMB.
Decloedt, J. F., Fenton, D. W., Outpatient hysteroscopy: Indications and hysteroscopic findings in pre- and postmenopausal patients, Gynaecological Endoscopy, 8, 137-141, 1999	No relevant data.
Deffieux, X., Fernandez, H., Adenomyosis: New concepts on pathophysiology, diagnosis and treatment, Journal de Gynecologie Obstetrique et Biologie de la Reproduction, 33, 703-712, 2004	Full text in French language.
Dijkhuizen, F. P. H. L. J., Brolmann, H. A. M., Potters, A. E., Bongers, M. Y., Heintz, A. P. M., The accuracy of transvaginal ultrasonography in the diagnosis of endometrial abnormalities, Obstetrics and Gynecology, 87, 345-349, 1996	Wrong population: women with metrorrhagia or postmenopausal bleeding included.
Dijkhuizen, F. P. H. L. J., De Vries, L. D., Mol, B. W. J., Brolmann, H. A. M., Peters, H. M., Moret, E., Heintz, A. P. M., Comparison of transvaginal ultrasonography and saline infusion sonography for the detection of intracavitary abnormalities in premenopausal women, Ultrasound in Obstetrics and Gynecology, 15, 372-376, 2000	Wrong population: only around half of the population with HMB.
Dijkhuizen, F. P., Mol, B. W., Brolmann, H. A., Heintz, A. P., The accuracy of endometrial sampling in the diagnosis of patients with endometrial carcinoma and hyperplasia: a meta-analysis, Cancer, 89, 1765-72, 2000	Index test not of interest.
Dinic, S. P. T., Kopitovic, V., Antic, V., Stamenovic, S., Mitic, D., Milosevic, J., Role of hysteroscopy in evaluation of patients with abnormal uterine bleeding, Acta Facultatis Medicae Naissensis, 28, 2011	Number of post-menopausal women not reported.

Reference	Reason for exclusion
Dipi, R. M., Amin, M. S., Islam, M. N., Khan, N. A., Chaiti, M. M., Hossain, M. M., Comparison of transabdominal and transvaginal sonography in the evaluation of uterine mass with histopathological correlation, Mymensingh medical journal : MMJ, 22, 69-74, 2013	Only 60% of the women present with HMB.
Donnez, J., Menometrorrhagia during the premenopause: An overview, Gynecological Endocrinology, 27, 1114	Expert review. Relevant references checked.
Doubilet,P.M., Diagnosis of abnormal uterine bleeding with imaging, Menopause, 18, 421-424, 2011	Wrong population: post-menopausal women.
Dueholm, M., Transvaginal ultrasound for diagnosis of adenomyosis: a review, Best Practice and Research: Clinical Obstetrics and Gynaecology, 20, 569-582, 2006	A review, different inclusion criteria. Included studies assessed individually.
Dueholm, M., Jensen, M. L., Laursen, H., Kracht, P., Can the endometrial thickness as measured by trans-vaginal sonography be used to exclude polyps or hyperplasia in pre-menopausal patients with abnormal uterine bleeding?, Acta Obstetricia et Gynecologica Scandinavica, 80, 645-651, 2001	No relevant diagnostic accuracy data. Another publication with the same cohort already included in the review.
Dueholm, M., Lundorf, E., Transvaginal ultrasound or MRI for diagnosis of adenomyosis, Current Opinion in Obstetrics & Gynecology, 19, 505-12, 2007	A review, different inclusion criteria. Included studies assessed individually.
Dueholm, M., Lundorf, E., Olesen, F., Imaging techniques for evaluation of the uterine cavity and endometrium in premenopausal patients before minimally invasive surgery, Obstetrical and Gynecological Survey, 57, 388-403, 2002	Literature review, included studies checked individually.
Eldred, J. M., Thomas, E. J., Pituitary and ovarian hormone levels in unexplained menorrhagia, Obstetrics and Gynecology, 84, 775-778, 1994	No relevant data.
Elgantri, R., Khalifa, N., Rhoma, N., Almsalati, S., Hysteroscopy in abnormal uterine bleeding in Libyan women. Is it helpful?, International Journal of Gynecology and Obstetrics, Conference, 2009	Conference abstract.
ElKattan, E. A., Omran, E. F., Al Inany, H. G., The accuracy of transvaginal ultrasound and uterine artery Doppler in the prediction of adenomyosis, Middle East Fertility Society Journal, 15, 73-78, 2010	Wrong population.
El-Khayat,W., Sleet,M.E., Mahdi,E.Y., Comparative study of transvaginal sonography and ysteroscopy for the detection of pathological endometrial esions in women with perimenopausal bleeding, Middle East Fertility Society Journal, 16, 77-82, 2011	The paper does not report confidence intervals or enough data to calculate them.

Reference	Reason for exclusion
El-Sharkawy, M., El-Mazny, A., Ramadan, W., Hatem, D., Abdel-Hafiz, A., Hammam, M., Nada, A., Three-dimensional ultrasonography and power Doppler for discrimination between benign and malignant endometrium in premenopausal women with abnormal uterine bleeding, BMC Women's Health. 16 (1) (no pagination), 2016, Article Number, 2016	Wrong population: women with polyps, fibroids, adenomyosis etc. excluded. A cut-off of endometrial thickness used in the study (19mm) is not used normally, thus, is not useful.
El-Sherbiny, W., El-Mazny, A., Abou-Salem, N., Mostafa, W. S., The diagnostic accuracy of two- vs three-dimensional sonohysterography for evaluation of the uterine cavity in the reproductive age, Journal of Minimally Invasive Gynecology, 22, 2015	Wrong population: only 35% women with abnormal uterine bleeding.
Emanuel, M. H., Is dilatation and curettage obsolete for diagnosing intrauterine disorders in premenopausal patients with persistent abnormal uterine bleeding?, Acta Obstetricia et Gynecologica Scandinavica, 76, 65-68, 1997	Index test not of interest.
Emanuel, M. H., Verdel, M. J., Wamsteker, K., Lammes, B., A prospective comparison of transvaginal ultrasonography and diagnostic hysteroscopy in the evaluation of patients with abnormal uterine bleeding: Clinical implications, American Journal of Obstetrics and Gynecology, 172, 547-552, 1995	Wrong population: less than half of the population with HMB.
Evans, P., Brunsell, S., Uterine fibroid tumors: Diagnosis and treatment, American Family Physician, 75, 1503	Expert review. Possible relevant references checked.
Farquhar, C., Ekeroma, A., Furness, S., Arroll, B., A systematic review of transvaginal ultrasonography, sonohysterography and hysteroscopy for the investigation of abnormal uterine bleeding in premenopausal women, Acta Obstetricia et Gynecologica Scandinavica, 82, 493-504, 2003	A systematic review, not the same inclusion criteria. Included studies checked individually.
Fascilla, F. D., Cramarossa, P., Cannone, R., Olivieri, C., Vimercati, A., Exacoustos, C., Ultrasound diagnosis of uterine myomas, Minerva Ginecologica, 68, 297-312, 2016	Expert review. Relevant references checked.
Fedele, L., Bianchi, S., Dorta, M., Arcaini, L., Zanotti, F., Carinelli, S., Transvaginal ultrasonography in the diagnosis of diffuse adenomyosis, Fertility and Sterility, 58, 94-97, 1992	Wrong/unclear population, old study published in 1992.
Fedele, L., Bianchi, S., Dorta, M., Brioschi, D., Zanotti, F., Vercellini, P., Transvaginal ultrasonography versus hysteroscopy in the diagnosis of uterine submucous myomas, Obstetrics and Gynecology, 77, 745-748, 1991	Wrong/unclear population: women with menometrorrhagia and/or pelvic pain included, proportions not reported.

Reference	Reason for exclusion
Ferry, J., Farnsworth, A., Webster, M., Wren, B., The efficacy of the Pipelle endometrial biopsy in detecting endometrial carcinoma, Australian and New Zealand Journal of Obstetrics and Gynaecology, 33, 76-78, 1993	Index test not of interest.
Filiz, T., Doger, E., Corakci, A., Ozeren, S., Caliskan, E., The efficacy, cost and patient satisfaction of classic versus office hysteroscopy in cases with suspected intrauterine space occupying lesions with 3-dimension ultrasound and abnormal uterine bleeding, Journal of the Turkish German Gynecology Association, 10, 2009	Wrong population: only 45% of women with menorrhagia or menometrorrhagia.
Fothergill, D. J., Brown, V. A., Hill, A. S., Histological sampling of the endometrium - A comparison between formal curettage and the Pipelle sampler, British Journal of Obstetrics and Gynaecology, 99, 779-780, 1992	Index test not of interest.
Fukuda, M., Shimizu, T., Fukuda, K., Yomura, W., Shimizu, S., Transvaginal hysterosonography for differential diagnosis between submucous and intramural myoma, Gynecologic and Obstetric Investigation, 35, 236-239, 1993	Unclear population: a total of 36 patients with hypermenorrhea, dysmenorrhea or anemia, with patients with distinct intramural myomas in TVUS were excluded. Proportion with HMB not reported.
Gao, J., Zeng, S., Sun, B. L., Fan, H. M., Han, L. H., Menstrual blood loss and hematologic indices in healthy Chinese women, Journal of Reproductive MedicineJ Reprod Med, 32, 822-6, 1987	No relevant data.
Garcia, L. E., Morris, S., Isaacson, K., Adenomyosis: Risk factors and diagnosis, Journal of Minimally Invasive Gynecology, 1), S17-S18, 2011	Conference abstract.
Garcia,L., Isaacson,K., Adenomyosis: Review of the Literature, Journal of Minimally Invasive Gynecology, 18, 428-437, 2011	Narrative review on adenomyosis, references checked.
Garuti, G., Sambruni, I., Colonnelli, M., Luerti, M., Accuracy of hysteroscopy in predicting histopathology of endometrium in 1500 women, Journal of the American Association of Gynecologic Laparoscopists, 8, 207-213, 2001	Wrong population: only 310 out of 1500 with HMB.
Georgantopoulou, C., Simm, A., Roberts, M., Transvaginal saline hysterosonography: A comparison with local anaesthetic hysteroscopy for the diagnosis of benign lesions associated with menorrhagia, Gynecological Surgery, 5, 2008	Wrong index test: saline hysterosonography.

Reference	Reason for exclusion
Gimpelson, R. J., Rappold, H. O., A comparative study between panoramic hysteroscopy with directed biopsies and dilatation and curettage: A review of 276 cases, American Journal of Obstetrics and Gynecology, 158, 489-492, 1988	Wrong comparison: compares biopsy and curettage; unclear population: only 76% have abnormal uterine bleeding of which proportion of HMB not reported.
Gkrozou, F., Dimakopoulos, G., Vrekoussis, T., Lavasidis, L., Koutlas, A., Navrozoglou, I., Stefos, T., Paschopoulos, M., Hysteroscopy in women with abnormal uterine bleeding: a meta-analysis on four major endometrial pathologies, Archives of Gynecology and Obstetrics., 19, 2014	Systematic review; no additional relevant studies.
Goldschmit, R., Katz, Z., Blickstein, I., Caspi, B., Dgani, R., The accuracy of endometrial pipelle sampling with and without sonographic measurement of endometrial thickness, Obstetrics and Gynecology, 82, 727-730, 1993	Wrong index test; unclear population.
Goldstein,S.R., Zeltser,I., Horan,C.K., Snyder,J.R., Schwartz,L.B., Ultrasonography-based triage for perimenopausal patients with abnormal uterine bleeding, American Journal of Obstetrics and Gynecology, 177, 102-108, 1997	No sufficient data to calculate diagnostic accuracy of TVUS. Not all who underwent TVUS had a reference standard test.
Gordts, S., Brosens, J. J., Fusi, L., Benagiano, G., Brosens, I., Uterine adenomyosis: a need for uniform terminology and consensus classification, Reproductive Biomedicine Online, 17, 244-8, 2008	A narrative review. References checked.
Goyal, B. K., Gaur, C. I., Sharma, G. C. S., Saha, M. A., Das, N. K., Transvaginal sonography versus hysteroscopy in evaluation of abnormal uterine bleeding, Medical Journal Armed Forces India, 71, 2015	Wrong comparison; wrong population: less than two thirds of the population with HMB.
Graziano, A., Lo Monte, G., Piva, I., Caserta, D., Karner, M., Engl, B., Marci, R., Diagnostic findings in adenomyosis: A pictorial review on the major concerns, European Review for Medical and Pharmacological Sciences, 19, 1146	Pictorial review on adenomyosis. References checked.
Griffin, Y., Sudigali, V., Jacques, A., Radiology of benign disorders of menstruation, Seminars in Ultrasound, CT & MR, 31, 414-32, 2010	Expert review. Relevant references checked.
Grimbizis, G. F., Tsolakidis, D., Mikos, T., Anagnostou, E., Asimakopoulos, E., Stamatopoulos, P., Tarlatzis, B. C., A prospective comparison of transvaginal ultrasound, saline infusion sonohysterography, and diagnostic hysteroscopy in the evaluation of endometrial pathology, Fertility and Sterility, 94, 2720	Wrong population: 52% pre-menopausal women with abnormal uterine bleeding.

Reference	Reason for exclusion
Guido, R. S., Ultrasound and magnetic resonance imaging (MRI) screening of patients with heavy menstrual bleeding and suspected uterine fibroids, Journal of Minimally Invasive Gynecology, Conference, 2013	Conference abstract.
Gungoren, A., Dolapcioglu, K., Hakverdi, A. U., Ciftci, S. C., Yetim, C., How effective TV-USG, SIS and hysteroscopy are in evaluating uterine cavity?, Turkiye Klinikleri Jinekoloji Obstetrik, 21, 2011	Full text in Turkish.
Guven, M. A., Bese, T., Demirkiran, F., Comparison of hydrosonography and transvaginal ultrasonography in the detection of intracavitary pathologies in women with abnormal uterine bleeding, International Journal of Gynecological Cancer, 14, 57-63, 2004	Wrong population: only 66% premenopausal with abnormal uterine bleeding, proportion of these with HMB not reported.
Guyatt, G. H., Oxman, A. D., Ali, M., Willan, A., McIlroy, W., Patterson, C., Laboratory diagnosis of iron-deficiency anemia: An overview, Journal of General Internal Medicine, 7, 145-153, 1992	No relevant data.
Hallberg, L., Hogdahl, A. M., Nilsson, L., Rybo, G., Menstrual blood loss--a population study. Variation at different ages and attempts to define normality, Acta Obstetricia et Gynecologica Scandinavica, 45, 320-351, 1966	No relevant data.
Hanafi, M., Ultrasound diagnosis of adenomyosis, leiomyoma or combined with histopathological correlation, Gynecological Surgery, Conference, 2012	Unclear population.
Harmanli, O. H., Bevilacqua, S. A., Dandolu, V., Chatwani, A. J., Hernandez, E., Adenomyosis interferes with accurate ultrasonographic detection of uterine leiomyomas, Archives of Gynecology and Obstetrics, 273, 146-149, 2005	Unclear population.
Haynes, P J; Anderson ABM., Patterns of menstrual blood loss in menorrhagia., Research and Clinical Forums, 1, 73-8, 1979	No relevant data.
Herman, M. C., Mol, B. W., Bongers, M. Y., Diagnosis of heavy menstrual bleeding, Women's Health, 12, 2016	Expert review. Relevant references checked.
Herman, P., Gaspard, U., Meno-metrorrhagia, Revue Medicale de Liege, 54, 1999	Full text in French.
Hong, S. C., Khoo, C. K., An update on adenomyosis uteri, Gynecology and Minimally Invasive Therapy, 5, 132-133, 2016	Review on adenomyosis. References checked.
Hunter, D. C., McClure, N., Abnormal uterine bleeding: an evaluation endometrial biopsy, vaginal ultrasound and outpatient hysteroscopy, The Ulster medical journal, 70, 25-30, 2001	Wrong population: 42% postmenopausal women.

Reference	Reason for exclusion
Hurskainen, R., Grenman, S., Komi, I., Kujansuu, E., Luoto, R., Orrainen, M., Patja, K., Penttinen, J., Silventoinen, S., Tapanainen, J., Toivonen, J., Diagnosis and treatment of menorrhagia, Acta Obstetricia et Gynecologica Scandinavica, 86, 2007	An article about Finnish guideline on diagnosis and treatment of heavy menstrual bleeding from 2007. Relevant references checked.
Indman, P. D., Abnormal uterine bleeding: Accuracy of vaginal probe ultrasound in predicting abnormal hysteroscopic findings, Journal of Reproductive Medicine for the Obstetrician and Gynecologist, 40, 545-548, 1995	Wrong/unclear population: 40% of women 50 years, menopausal status or proportion with HMB not reported.
Issat, T., Beta, J., Nowicka, M. A., Jakimiuk, A. J., Accuracy and diagnostic value of outpatient hysteroscopy for malign and benign disease, European Journal of Gynaecological Oncology, 35, 52-55, 2014	Wrong population: 27% abnormal uterine bleeding, 56% postmenopausal.
Jain, N., Verma, R., Imaging diagnosis of accessory and cavitated uterine mass, a rare mullerian anomaly, Indian Journal of Radiology and Imaging, 24, 2014	A case report.
Jakimiuk, A. J., Grzybowski, W., Beta, J., Dysfunctional uterine bleeding--diagnostics and treatment, Ginekologia Polska, 79, 2008	Full text in Polish.
James, A., Matchar, D. B., Myers, E. R., Testing for von Willebrand disease in women with menorrhagia: A systematic review, Obstetrics and Gynecology, 104, 381-388, 2004	No relevant data.
Janssen, C. A. H., Scholten, P. C., Heintz, A. P. M., Reconsidering menorrhagia in gynecological practice. Is a 30-year-old definition still valid?, European Journal of Obstetrics Gynecology and Reproductive Biology, 78, 69-72, 1998	No relevant data.
Jason, M., Dechenne, V., Chantraine, F., Nisolle, M., Adenomyosis, Revue Medicale de Liege, 68, 160-162, 2013	Full text in French.
Jeng, C. J., Huang, S. H., Shen, J., Chou, C. S., Tzeng, C. R., Laparoscopy-guided myometrial biopsy in the definite diagnosis of diffuse adenomyosis, Human Reproduction, 22, 2016	Wrong/unclear population.
Kadir, R. A., Economides, D. L., Sabin, C. A., Owens, D., Lee, C. A., Frequency of inherited bleeding disorders in women with menorrhagia, Lancet, 351, 485-489, 1998	No relevant data.
Katsetos, C., Radhakrishnan, S., Koumousidis, A., Kontoyannis, M., Boutas, I., Kalampokas, E., Kalampokas, T., Sofoudis, C., Comparison of transvaginal 3D sonohysterography with outpatient hysteroscopy in the evaluation of abnormal uterine bleeding, International Journal of Gynecological Cancer, Conference, 2013	Conference abstract.

Reference	Reason for exclusion
Kavak, Z., Ceyhan, N., Pekin, S., Combination of vaginal ultrasonography and Pipelle sampling in the diagnosis of endometrial disease, Australian and New Zealand Journal of Obstetrics and Gynaecology, 36, 63-66, 1996	Wrong population: more than half postmenopausal women.
Keckstein, J., Hysteroscopy and adenomyosis, Contributions to Gynecology & Obstetrics, 20, 41-50, 2000	Not a study. Hysteroscopy is not of interest for the diagnosis of adenomyosis in the review.
Keeling, A. N., Reidy, J. F., Imaging and treatment of uterine fibroids, including the role of uterine artery embolization, Imaging, 19, 374-384, 2007	Expert review. Possible relevant references checked.
Kelekci, S., Kaya, E., Alan, M., Alan, Y., Bilge, U., Mollamahmutoglu, L., Comparison of transvaginal sonography, saline infusion sonography, and office hysteroscopy in reproductive-aged women with or without abnormal uterine bleeding, Fertility and Sterility, 84, 682-686, 2005	Wrong population: less than two thirds with HMB.
Kent, A. S. H., Haines, P., Manners, B. T. B., Coats, P. M., Blind endometrial biopsies: Insufficient for diagnosis in women with intrauterine pathology, Gynaecological Endoscopy, 7, 273-278, 1998	Index test not of interest.
Kepkep, K., Tuncay, Y. A., Goynumer, G., Tutal, E., Transvaginal sonography in the diagnosis of adenomyosis: Which findings are most accurate?, Ultrasound in Obstetrics and Gynecology, 30, 341-345, 2007	Wrong population.
Khan, F., Jamaat, S., Al-Jaroudi, D., Saline infusion sonohysterography versus hysteroscopy for uterine cavity evaluation, Annals of Saudi Medicine, 31, 2011	Wrong population: only 43% with HMB.
Khan, Z., Famuyide, A. O., Hopkins, M. R., Breitkopf, D. M., Laughlin-Tommaso, S. K., Sensitivity of flexible hysteroscopy for endometrial polyps and the role of endometrial biopsies, Journal of Minimally Invasive Gynecology, Conference, 2012	Conference abstract.
Khan,A.T., Shehmar,M., Gupta,J.K., Uterine fibroids: Current perspectives, International Journal of Women's Health, 6, 95-114, 2014	Expert review. Relevant references checked.
Khanna, A., Gupta, M., Shukla, R. C., Saline perfusion sonography and transvaginal sonography in abnormal uterine bleeding, Ultrasound International, 7, 31-36, 2001	Wrong population: 9/70 women had HMB.
Kinkel, K., Vincent, B., Balleyguier, C., Helenon, O., Moreau, J., Value of MR imaging in the diagnosis of benign uterine conditions, Journal de Radiologie, 81, 2000	Full text in French language.

Reference	Reason for exclusion
Knol, H. M., Bogchelman, D. H., Meijer, K., Van Der Zee, A. G. J., Van Der Meer, J., Unexplained menorrhagia and hemostatic evaluation in gynaecological practice, a retrospective study, Journal of Thrombosis and Haemostasis, Conference, 2009	Conference abstract.
Koc, H., Noyan, V., Yucel, A., Sagsoz, N., Comparison of uterine artery Doppler sonography, saline infusion sonography and endometrial biopsy in premenopausal patients with abnormal uterine bleeding who were found to have endometrial polyps, Turkiye Klinikleri Jinekoloji Obstetrik, 20, 2010	Full text in Turkish.
Kolhe, S., Prospective study of one-stop diagnostic and see-and-treat outpatient hysteroscopy service at royal derby hospital, BJOG: An International Journal of Obstetrics and Gynaecology, Conference, 2014	Wrong population: 42% of women with postmenopausal bleeding.
Koonings, P. P., Moyer, D. L., Grimes, D. A., A randomized clinical trial comparing Pipelle and Tis-u-trap for endometrial biopsy, Obstetrics and Gynecology, 75, 293-295, 1990	Index test not of interest.
Korczynski, J., Sobkiewicz, S., Adenomyosis. Diagnostic technique and treatment, Ginekologia Polska, 72, 317-321, 2001	Full text in Polish language.
Koss, L. G., Schreiber, K., Oberlander, S. G., Moussouris, H. F., Lesser, M., Detection of endometrial carcinoma and hyperplasia in asymptomatic women, Obstetrics and Gynecology, 64, 1-11, 1984	Index test not of interest.
Krampl, E., Soby, B., Istre, O., How representative are Pipelle endometrial biopsies? A retrospective analysis of 324 biopsies followed by transcervical resection of the endometrium or hysterectomy, Gynaecological Endoscopy, 6, 277-281, 1997	Index test not of interest.
Krassas, G. E., Pontikides, N., Kaltsas, Th, Papadopoulu, Ph, Batrinos, M., Menstrual disturbances in thyrotoxicosis, Clinical Endocrinology, 40, 641-644, 1994	No relevant data.
Kroencke, T. J., MRI for diagnosis of adenomyosis: Unsung and underutilized, Gynecologic and Obstetric Investigation, 60, 2005	A letter to editor.
Kunz, G., Herbertz, M., Beil, D., Huppert, G., Leyendecker, G., Adenomyosis as a disorder of the early and late human reproductive period, Reproductive BioMedicine Online, 15, 681-685, 2007	Wrong population.
Kuruba, N., Thilina, P., Medha, S., Sonal, G., An audit on outpatient hysteroscopy (OPH) in the evaluation of abnormal uterine bleeding, Gynecological Surgery, Conference, 22nd Annual	Conference abstract.

Reference	Reason for exclusion
Congress of the European Society of Gynaecological Endoscopy, ESGE 2013. Berlin Germany. Conference Start: 20131016. Conference End: 20131019. Conference Publication: (var.pagings). 10 (pp S114), 2013	
La Sala, G. B., Blasi, I., Gallinelli, A., Debbi, C., Lopopolo, G., Vinci, V., Villani, M. T., Iannotti, F., Diagnostic accuracy of sonohysterography and transvaginal sonography as compared with hysteroscopy and endometrial biopsy: A prospective study, Minerva Ginecologica, 63, 2011	Wrong population: 56% with abnormal bleeding, post versus pre-menopausal status not reported.
Lane, B. F., Wong-You-Cheong, J. J., Imaging of endometrial pathology, Clinical Obstetrics and Gynecology, 52, 2009	Expert review. No relevant information.
Laughead, M. K., Stones, L. M., Burry, K. A., Holbert, T. R., Gamberdella, F. R., Powers, T. W., Steinke, R. G. F., Margolin, M. L., Jennett, R. J., Clinical utility of saline solution infusion sonohysterography in a primary care obstetric-gynecologic practice, American Journal of Obstetrics and Gynecology, 176, 1313-1318, 1997	Wrong index test, no relevant data.
Law, J., Fothergill, D., Histological sampling of the endometrium - A comparison between formal curettage and the Pipelle sampler [7], British Journal of Obstetrics and Gynaecology, 100, 503-504, 1993	Index test not of interest.
Leah, M., Hunter, D., Outpatient hysteroscopy: Outcome study, Gynecological Surgery, Conference, 2011	Wrong population: abnormal bleeding in 45% of participants.
Lee, A. Y., Poder, L., Qayyum, A., Wang, Z. J., Yeh, B. M., Coakley, F. V., Imaging malignant and apparent malignant transformation of benign gynaecological disease, Clinical Radiology, 65, 1031	Expert review. Possible relevant references checked.
Lee, L., Agarwal, A., Fai, F. Y., The feasibility and yield of outpatient hysteroscopy, Annals of the Academy of Medicine Singapore, Conference, 2011	Conference abstract.
Lee, S. I., Imaging for gynecological cancers, International Journal of Gynecology and Obstetrics, Conference, 2015	Conference abstract.
Lemercier, E., Genevois, A., Dacher, J. N., Benozio, M., Descargues, G., Marpeau, L., Endometrial imaging, Journal de Radiologie, 81, 1845	Full text in French.
Levgur, M., Diagnosis of adenomyosis: a review, Journal of Reproductive Medicine, 52, 177-93, 2007	Expert review on adenomyosis. Relevant references checked.

Reference	Reason for exclusion
Levine, D. J., Berman, J. M., Harris, M., Chudnoff, S. G., Whaley, F. S., Palmer, S. L., Sensitivity of myoma imaging using laparoscopic ultrasound compared with magnetic resonance imaging and transvaginal ultrasound, Journal of Minimally Invasive Gynecology, 20, 2013	Unclear population: criteria for inclusion includes presence of fibroids in TVUS; proportion of patients with HMB not reported.
Levy, G., Dehaene, A., Laurent, N., Lernout, M., Collinet, P., Lucot, J. P., Lions, C., Poncelet, E., An update on adenomyosis, Diagnostic and Interventional Imaging, 94, 3-25, 2013	Narrative review on adenomyosis. References checked.
Lin, P., Sun, Y. L., Wang, C. B., Lee, C. Y., Wun, T. H., Lin, Y. H., Tseng, C. C., Chen, C. H., Tseng, C. J., Transvaginal Sonographic Criteria for the Diagnosis of Adenomyosis Based on Histopathologic Correlation, Taiwanese Journal of Obstetrics and Gynecology, 49, 2010	Wrong population.
Lipscomb, G. H., Lopatine, S. M., Stovall, T. G., Ling, F. W., A randomized comparison of the Pipelle, Accurette, and Explora endometrial sampling devices, American Journal of Obstetrics and Gynecology, 170, 591-594, 1994	Index test not of interest.
Litta, P., Vasile, C., Quintieri, F., Blandamura, S., Correlation between hysteroscopy and histology in abnormal uterine bleeding, Italian Journal of Gynaecology and Obstetrics, 8, 22-24, 1996	Wrong population: 40% postmenopausal.
Lobo Abascal, P., Alvarez Bernardi, J., Rubio Valtuena, J., Heras Aznar, J., Garcia de Santiago, C., Operative office hysteroscopy, Gynecological Surgery, Conference, 2010	Conference abstract.
Loffer, F. D., Hysteroscopy with selective endometrial sampling compared with D&C for abnormal uterine bleeding: The value of a negative hysteroscopic view, Obstetrics and Gynecology, 73, 16-20, 1989	Wrong population: not women with HMB.
Lone, F. W., Balogun, M., Khan, K. S., Adenomyosis: Not such an elusive diagnosis any longer, Journal of Obstetrics and Gynaecology, 26, 225-228, 2006	A commentary. Relevant references checked.
Looker, A. C., Dallman, P. R., Carroll, M. D., Gunter, E. W., Johnson, C. L., Prevalence of iron deficiency in the United States, Journal of the American Medical Association, 277, 973-976, 1997	No relevant data.
Luciano, D. E., Albrecht, L., LaMonica, R., Luciano, A. A., Exacoustos, C., Two and three-dimensional ultrasound evaluation of adenomyosis and histological correlation on ultrasound targeted biopsies of the myometrium at hysterectomy, Journal of Minimally Invasive Gynecology, 1), S7, 2011	Conference abstract.

Reference	Reason for exclusion
Luciano, D. E., Exacoustos, C., Albrecht, L., LaMonica, R., Proffer, A., Zupi, E., Luciano, A. A., Three-dimensional ultrasound in diagnosis of adenomyosis: Histologic correlation with ultrasound targeted biopsies of the uterus, Journal of Minimally Invasive Gynecology, 20, 803-810, 2013	Wrong population.
Luterek, K., Szymusik, I., Bartkowiak, R., Wielgos, M., Sonohysterography in peri- and postmenopausal women with abnormal uterine bleeding or abnormal endometrial appearance, Neuroendocrinology Letters, 35, 2014	Wrong/unclear population: peri- and postmenopausal women with AUB or abnormal endometrial appearence on TVUS. Proportion with HMB not reported.
MacKenzie, I. Z., Bibby, J. G., Critical assessment of dilatation and curettage in 1029 women, Lancet, 2, 566-8, 1978	Index test not of interest.
Maiti, S., Naidoo, K., Patients' satisfaction survey at the outpatient hysteroscopy service at St. Mary's hospital, Manchester, UK, Menopause International, Conference, 2008	Conference abstract.
Majmudar, T., Abdel-Rahman, H., Pelvic mass - diagnosis and management, Obstetrics, Gynaecology and Reproductive Medicine, 18, 193-198, 2008	Expert article, no relevant references.
Majos, A., Bakalarz, M., Stachowiak, G., Stefanczyk, L., Jedrzejczyk, S., Adenomyosis in magnetic resonance imaging - The value of SE T2 sequence using body arrey, Przeglad Menopauzalny, 4, 43-46, 2005	Full text in Polish.
Makris, N., Kalmantis, K., Skartados, N., Papadimitriou, A., Mantzaris, G., Antsaklis, A., Three-dimensional hysterosonography versus hysteroscopy for the detection of intracavitary uterine abnormalities, International Journal of Gynaecology & ObstetricsInt J Gynaecol Obstet, 97, 2007	Wrong population: women with infertility, menorrhagia in 15%.
Makris, N., Skartados, N., Kalmantis, K., Mantzaris, G., Papadimitriou, A., Antsaklis, A., Evaluation of abnormal uterine bleeding by transvaginal 3-D hysterosonography and diagnostic hysteroscopy, European Journal of Gynaecological Oncology, 28, 2007	Wrong/unclear population: 30% postmenopausal women, rest premenopausal but with AUB, % of HMB not reported.
Malartic, C., Morel, O., Rivain, A. L., Place, V., Le Dref, O., Dohan, A., Gayat, E., Barranger, E., Soyer, P., Evaluation of symptomatic uterine fibroids in candidates for uterine artery embolization: Comparison between ultrasonographic and MR imaging findings in 68 consecutive patients, Clinical Imaging, 37, 83-90, 2013	All participants with confirmed fibroids. Not a study on diagnosis but on therapeutic decisions based on MRI versus ultrasound.

Reference	Reason for exclusion
Mancini, F., Regnani, G., Persico, N., de Aloysio, D., Battaglia, C., Sonohysterography in the evaluation of endometrial abnormalities, Italian Journal of Gynaecology and Obstetrics, 14, 69-72, 2002	Wrong index test (sonohysterography).
Mangano, U., Garofalo, A., Santonocito, V., Salemi, I., Role of ultrasound in evaluation of endometrial disease, Giornale Italiano di Ostetricia e Ginecologia, 31, 2009	Full text in Italian.
Markov, D., Markov, P., Pavlova, E., Atanaova, D., Three-dimensional (3D) transvaginal ultrasound--clinical implementation in benign gynecological disorders, Akusherstvo i ginekologiia, 49, 2010	Full text in Bulgarian.
Martin, B., Levy, L., Owczarczak, W., Rapoport, C., Adenomyosis: Imaging features, Imagerie de la Femme, 16, 85-94, 2006	Full text in French.
Mathew, M., Gowri, V., Rizvi, S. G., Saline infusion sonohysterography - an effective tool for evaluation of the endometrial cavity in women with abnormal uterine bleeding, Acta Obstetricia et Gynecologica Scandinavica, 89, 2010	Wrong population: less than two thirds with HMB.
Mathew, M., Gupta, R., Krolikowski, A., Role of transvaginal ultrasonography and diagnostic hysteroscopy in the evaluation of patients with abnormal uterine bleeding, International Journal of Gynecology and Obstetrics, 71, 251-253, 2000	Wrong population: less than two thirds with HMB.
McIlwaine, K., Readman, E., Cameron, M., Maher, P., Outpatient hysteroscopy: Factors influencing post-procedure acceptability in patients attending a tertiary referral centre, Australian and New Zealand Journal of Obstetrics and Gynaecology, 49, 2009	Unclear population: not reported the reason for hysteroscopy or the proportion with HMB.
McLucas, B., Diagnosis, imaging and anatomical classification of uterine fibroids, Best Practice & Research in Clinical Obstetrics & Gynaecology, 22, 627-42, 2008	Expert review, relevant references checked.
Megh, M., Katke, R., Use of diagnostic hysteroscopy in abnormal uterine bleeding in perimenopausal age group and its clinicopathological co-relation with ultrasound and histopathology findings: A study in teritary care teaching institute, Mumbai, International Journal of Gynecology and Obstetrics, Conference, 2015	Conference abstract.
Meredith, S. M., Sanchez-Ramos, L., Kaunitz, A. M., Diagnostic accuracy of transvaginal sonography for the diagnosis of adenomyosis: systematic review and metaanalysis, American Journal of Obstetrics & Gynecology, 201, 107.e1-6, 2009	A systematic review and meta-analysis with wrong population and wrong years of publication. Included studies assessed individually.

Reference	Reason for exclusion
Mihm, L. M., Quick, V. A., Brumfield, J. A., Connors Jr, A. F., Finnerty, J. J., The accuracy of endometrial biopsy and saline sonohysterography in the determination of the cause of abnormal uterine bleeding, American Journal of Obstetrics and Gynecology, 186, 858-860, 2002	Index test not of interest.
Miller, C. H., Dilley, A., Richardson, L., Hooper, W. C., Evatt, B. L., Population differences in von Willebrand factor levels affect the diagnosis of von Willebrand disease in African-American women, American Journal of Hematology, 67, 125-9, 2001	No relevant data.
Moawad, N. S., Santamaria, E., Johnson, M., Shuster, J., Cost-effectiveness of office hysteroscopy for abnormal uterine bleeding, Journal of the Society of Laparoendoscopic Surgeons, 18, 2014	No relevant data; no comparison test.
Modaffari, P., Tognazza, E., Panuccio, E., Rabischong, B., Canis, M., Bourdel, N., Accuracy and reproducibility of diagnostic hysteroscopy in abnormal uterine bleeding, Gynecological Surgery, Conference, 23rd Annual Congress of the European Society of Gynaecological Endoscopy, ESGE 2014. Brussels Belgium. Conference Start: 20140924. Conference End: 20140927. Conference Publication: (var.pagings). 11 (1 SUPPL. 1) (pp 146-147), 2014	Conference abstract.
Moghadam, R., Lathi, R. B., Shahmohamady, B., Saberi, N. S., Nezhat, C. H., Nezhat, F., Nezhat, C., Predictive value of magnetic resonance imaging in differentiating between leiomyoma and adenomyosis, JSLS : Journal of the Society of Laparoendoscopic Surgeons / Society of Laparoendoscopic Surgeons, 10, 216-219, 2006	Index test and population not of interest.
Molinas, C. R., Campo, R., Office hysteroscopy and adenomyosis, Best Practice and Research: Clinical Obstetrics and Gynaecology, 20, 557-567, 2006	Hysteroscopy not of interest for review on diagnosis of adenomyosis.
Moschos,E., Ashfaq,R., McIntire,D.D., Liriano,B., Twickler,D.M., Saline-infusion sonography endometrial sampling compared with endometrial biopsy in diagnosing endometrial pathology, Obstetrics and Gynecology, 113, 881-887, 2009	Wrong population: 49% postmenopausal women.
Murase, E., Siegelman, E. S., Outwater, E. K., Perez-Jaffe, L. A., Tureck, R. W., Uterine leiomyomas: histopathologic features, MR imaging findings, differential diagnosis, and treatment, Radiographics : a review publication of the Radiological Society of North America, Inc, 19, 1179	Expert review, relevant references checked.

Reference	Reason for exclusion
Nagele, F., Bournas, N., O'Connor, H., Broadbent, M., Richardson, R., Magos, A., Comparison of carbon dioxide and normal saline for uterine distension in outpatient hysteroscopy, Fertility and Sterility, 65, 305-309, 1996	Wrong comparison; wrong population.
Nagele, F., O'Connor, H., Davies, A., Badawy, A., Mohamed, H., Magos, A., 2500 Outpatient diagnostic hysteroscopies, Obstetrics and Gynecology, 88, 87-92, 1996	Wrong population: less than two thirds with HMB.
Nandan, N., Manjeera, L., Rai, S., Gowri, M., Diagnostic hysteroscopy in abnormal uterine bleeding & it's histopathologic correlation: Our experience, Nitte University Journal of Health Science, 3, 2013	Only 50% of women with heavy menstrual bleeding.
Nazim,F., Hayat,Z., Hannan,A., Ikram,U., Nazim,K., Role of transvaginal ultrasound in identifying endometrial hyperplasia, Journal of Ayub Medical College, Abbottabad: JAMC, 25, 100-102, 2013	Wrong population: less than two thirds with HMB.
Nicholson, Y., Chan, J., Patient satisfaction in outpatient hysteroscopy, BJOG: An International Journal of Obstetrics and Gynaecology, Conference, 2015	Conference abstract.
Niinimaki, M., Paakko, E., Kyllonen, A. P., Santala, M., The improved diagnostics of adenomyosis, Duodecim, 119, 2003	Full text in Finnish.
Novellas, S., Chassang, M., Delotte, J., Toullalan, O., Chevallier, A., Bouaziz, J., Chevallier, P., MRI characteristics of the uterine junctional zone: From normal to the diagnosis of adenomyosis, American Journal of Roentgenology, 196, 1206	Narrative review on adenomyosis, references checked.
Ogutcuoglu, B., Karadag, C., Inan, C., Dolgun, Z. N., Yoldemir, A. T., Aslanova, L., Diagnostic utility of saline infusion doppler sonohysterography in endometrial mass lesions, Pakistan Journal of Medical Sciences, 32, 2016	Unclear population: 85% premenopausal, proportion with HMB not reported; inclusion criteria was women with suspected masses in TVUS, index test of interest in this study was SIS.
Ossola, M. W., Bertulessi, C., Iasi, L., Bianchini, B., Hanozet, F., Grassini, E., Capetta, P., Comparison of saline infusion sonography to transvaginal echography and hysteroscopy in the diagnostic evaluation of abnormal uterine bleeding, Italian Journal of Gynaecology and Obstetrics, 11, 145-150, 1999	Wrong population: less than two thirds premenopausal women.
Outwater, E. K., Siegelman, E. S., Van Deerlin, V., Adenomyosis: current concepts and imaging considerations, AJR. American Journal of Roentgenology, 170, 437-41, 1998	A pictorial essay. References checked.

Reference	Reason for exclusion
Ozer, A., Ozer, S., Kanat-Pektas, M., Correlation between transvaginal ultrasound measured endometrial thickness and histopathological findings in Turkish women with abnormal uterine bleeding, Journal of Obstetrics and Gynaecology Research, 42, 2016	Unclear population; unclear exclusions.
Pallavi, P., Krishna, L., Havaldar, N. A., Shailaja, N., Saravana, A., Pradeep, S., A comparative diagnostic evaluation of abnormal uterine bleeding by hysteroscopy and transvaginal sonography, Biomedicine, 33, 2013	Wrong/unclear population: number of women postmenopausal versus premenopausal not reported.
Parveen, S., Azad, S. A., Mohammad, J., Haque, S., Shanta, R. S., Rahman, S. M., Role of sonohysterography in evaluation of abnormally thickened endometrium causing abnormal uterine bleeding with histopathological correlation, Bangladesh Medical Research Council Bulletin, 40, 2014	Unclear population: 85% pre/perimenopausal but proportion of women with HMB not reported; reporting of results of index test and reference standard unclear, the text contradicts itself and figure explaining the results is also unclear/missing some observations. Therefore, difficult to form a definite 2x2 table.
Paschopoulos, M., Lolis, E. D., Alamanos, Y., Koliopoulos, G., Paraskevaidis, E., Vaginoscopic hysteroscopy and transvaginal sonography in the evaluation of patients with abnormal uterine bleeding, Journal of the American Association of Gynecologic Laparoscopists, 8, 506-510, 2001	Wrong population: less than half with HMB.
Pascual, A., Graupera, B., Tresserra, F., Ubeda, A., Hereter, L., Rodriguez, I., Grases, P. J., Color Doppler transvaginal ultrasound for detecting intrauterine disorders in patients with abnormal uterine bleeding, Gynaecologia et Perinatologia, 14, 157-160, 2005	Wrong index test; wrong/unclear population.
Pasqualotto,E.B., Margossian,H., Price,L.L., Bradley,L.D., Accuracy of preoperative diagnostic tools and outcome of hysteroscopic management of menstrual dysfunction, Journal of the American Association of Gynecologic Laparoscopists, 7, 201-209, 2000	Wrong population: 44% postmenopausal women.
Pasrija, S., Trivedi, S. S., Narula, M. K., Prospective study of saline infusion sonohysterography in evaluation of perimenopausal and postmenopausal women with abnormal uterine bleeding, Journal of Obstetrics and Gynaecology Research, 30, 27-33, 2004	Wrong population: less than two thirds with HMB.
Patil, S. G., Bhute, S. B., Inamdar, S. A., Acharya, N. S., Shrivastava, D. S., Role of diagnostic hysteroscopy in abnormal uterine bleeding and its histopathologic correlation, Journal of Gynecological Endoscopy and Surgery, 1, 2009	Wrong population: 25% with HMB.
Pehlivan, H., Keskin, U., Yenen, M. C., Guler, A. E., Dede, M., Ergun, A., The comparison of the effectiveness of biopsy combined with hysteroscopy and dilatation and curettage for the	Conference abstract.

Reference	Reason for exclusion
detection of endometrial pathologies, International Journal of Gynecological Cancer, Conference, 2012	
Perrot, N., Mergui, J. L., Frey, I., Uzan, M., Ultrasound exams in the diagnosis of menorrhagia, Gynecologie Obstetrique Fertilite, 30, 2002	Full text in French.
Philipp, C. S., Dilley, A., Miller, C. H., Evatt, B., Baranwal, A., Schwartz, R., Bachmann, G., Saidi, P., Platelet functional defects in women with unexplained menorrhagia, Journal of thrombosis and haemostasis : JTH, 1, 477-484, 2003	No relevant data.
Philipp, C. S., Miller, C. H., Faiz, A., Dilley, A., Michaels, L. A., Ayers, C., Bachmann, G., Dowling, N., Saidi, P., Screening women with menorrhagia for underlying bleeding disorders: The utility of the platelet function analyser and bleeding time, Haemophilia, 11, 497-503, 2005	No relevant data.
Phillips, D. R., Nathanson, H. G., Milim, S. J., Haselkorn, J. S., Magnetic resonance imaging for diagnosing adenomyomata, Journal of the American Association of Gynecologic Laparoscopists, 3, 245-250, 1996	Wrong/unclear population.
Pinheiro, W., Pereira, A. K. C., Ejzenberg, D., Ricci, M. D., Motta, E. V., Baracat, M. C. P., Simoes, R. S., Anzai, M., Soares Jr, J. M., Baracat, E. C., How is the accucary of ultrasound diagnosis of polyp after hysteroscopy?, International Journal of Gynecology and Obstetrics, Conference, 2015	Conference abstract.
Rajan, D. K., Margau, R., Kroll, R., Simons, M. E., Tan, K., Beecroft, R., Kachura, J., Sniderman, K. W., Comparison of ultrasound versus MRI prior to uterine artery embolization in symptomatic patients with presumed uterine fibroids, CardioVascular and Interventional Radiology, Conference, 2009	Conference abstract.
Rashid, S. Q., Chou, Y. H., Tiu, C. M., Ultrasonography of Uterine Leiomyomas, Journal of Medical Ultrasound, 24, 3-12, 2016	Expert review. Relevant references checked.
Reinhold, C., McCarthy, S., Bret, P. M., Mehio, A., Atri, M., Zakarian, R., Glaude, Y., Liang, L., Seymour, R. J., Diffuse adenomyosis: Comparison of endovaginal US and MR imaging with histopathologic correlation, Radiology, 199, 151-158, 1996	Wrong population.
Reinhold, C., Tafazoli, F., Mehio, A., Wang, L., Atri, M., Siegelman, E. S., Rohoman, L., Uterine adenomyosis: endovaginal US and MR imaging features with histopathologic correlation, Radiographics, 19 Spec No, S147-60, 1999	A review. References checked.

Reference	Reason for exclusion
Reinhold, C., Tafazoli, F., Wang, L., Imaging features of adenomyosis, Human Reproduction Update, 4, 337-49, 1998	A review. References checked.
Riaz, S., Ibrar, F., Dawood, N. S., Jabeen, A., Endometrial pathology by endometrial curettage in menorrhagia in premenopausal age group, Journal of Ayub Medical College, Abbottabad : JAMC, 22, 2010	No comparison of diagnostic tests.
Robert, Y., Bazot, M., Meno-metrorrhagia imaging, Journal de Radiologie, 89, 2008	Full article in French.
Rodeghiero, F., Castaman, G., Dini, E., Epidemiological investigation of the prevalence of von Willebrand's disease, BloodBlood, 69, 454-9, 1987	No relevant data.
Rodriguez, M. A., Gualtieri, B. M., Mouhayar, Y., Mejia, A., Brookfield, K. F., Cruz-Pachano, F., Comparison of 3D/4D ultrasound to 3D saline infused sonohysterographyfor evaluation of endometrial polyps in pre and postmenopausal women, Fertility and Sterility, Conference, 2012	Conference abstract.
Rodriquez, M., Gualtieri, M., Mouhayar, Y., Mejia, A., Cruz, F., Brookfield, K., 3D/4D ultrasoundtransvaginal ultrasound can beauseful alternative to saline infusion sonohysterography for evaluation of polyps, Fertility and Sterility, Conference, 60th Annual Meeting of The Pacific Coast Reproductive Society, PCRS 2012. Rancho Mirage, CA United States. Conference Start: 20120418. Conference End: 20120422. Conference Publication: (var.pagings). 97 (3 SUPPL. 1) (pp S21-S22), 2012	Conference abstract.
Rokita, W., Stanislawska, M., Nowak-Markwitz, E., Kedzia, W., Spaczynski, M., Karowicz-Bilinska, A., Bednarek, W., Diagnostic hysteroscopy in women with endometrial hyperplasia and endometrial cancer - 9 years of experience, analysis of 142 cases, International Journal of Gynecology and Obstetrics, Conference, 2009	Conference abstract.
Ryu, J. A., Kim, B., Lee, J., Kim, S., Lee, S. H., Comparison of transvaginal ultrasonography with hysterosonography as a screening method in patients with abnormal uterine bleeding, Korean Journal of Radiology, 5, 39-46, 2004	Wrong population: only 55% premenopausal women.
Saadati, N., Mohammad Jafary, R., Shojaie, K., Comparison of sensitivity and specificity of transvaginal ultra sonography, saline infusion sonohysterography and hysteroscopy in evaluation of the women with abnormal uterine bleeding in Ahvaz Imam Khomeini and Razi Hospitals during 2008, Iranian Journal of Reproductive Medicine, Conference, 2013	Conference abstract.

Reference	Reason for exclusion
Sadecky, A. M., Guido, R., Lee, T., Mansuria, S., Rindos, N., Donnellan, N. M., Clinical diagnosis of adenomyosis: Use of predictive uterine characteristics to improve accuracy and reliability, American Journal of Obstetrics and Gynecology, 1), S485, 2016	Conference abstract.
Saidi, M. H., Sadler, R. K., Theis, V. D., Akright, B. D., Farhart, S. A., Villanueva, G. R., Comparison of sonography, sonohysterography, and hysteroscopy for evaluation of abnormal uterine bleeding, Journal of Ultrasound in Medicine, 16, 587-91, 1997	Unclear population: proportion of women with HMB or who are premenopausal not reported. Age range 40-89 years, mean age around 50.
Sakhel, K., Abuhamad, A., Sonography of adenomyosis, Journal of Ultrasound in Medicine, 31, 805-8, 2012	A narrative review. References checked.
Saleh, S. S., Fram, K., Histopathology diagnosis in women who underwent a hysterectomy for a benign condition, Archives of Gynecology and Obstetrics, 285, 1339	Wrong population: 39% with HMB.
Salim, S., Won, H., Nesbitt-Hawes, E., Campbell, N., Abbott, J., Diagnosis and Management of Endometrial Polyps: A Critical Review of the Literature, Journal of Minimally Invasive Gynecology, 18, 2011	Expert review. Possible relevant references checked.
Salim,R., Lee,C., Davies,A., Jolaoso,B., Ofuasia,E., Jurkovic,D., A comparative study of three-dimensional saline infusion sonohysterography and diagnostic hysteroscopy for the classification of submucous fibroids, Human Reproduction, 20, 253-257, 2005	Wrong index test.
Sanna, S., Stochino Loi, E., Pirarba, S., Arena, I., Maricosu, G., Melis, G. B., Angioni, S., Office operative hysteroscopy. Results on pain, diagnosis and treatment of uterine pathologies, Gynecological Surgery, Conference, 2010	Conference abstract.
Scarpellini, F., Curto, C., Caracussi, U., Letta, C., Scarpellini, L., Transvaginal ultrasound versus histology in endometrial hyperplasia, Clinical and Experimental Obstetrics and Gynecology, 21, 266-269, 1994	Wrong population: most women postmenopausal.
Schneider, S. L., Craig, M., Branning, P., Improved sonographic accuracy in the presurgical diagnosis of diffuse adenomyosis: A case series and review, Journal of Diagnostic Medical Sonography, 18, 2002	A case series and review. No relevant data or references.
Schwarzler, P., Concin, H., Bosch, H., Berlinger, A., Wohlgenannt, K., Collins, W. P., Bourne, T. H., An evaluation of sonohysterography and diagnostic hysteroscopy for the assessment of intrauterine pathology, Ultrasound in Obstetrics and Gynecology, 11, 337-342, 1998	Wrong population: 70.4% of women were premenopausal with abnormal uterine bleeding including HMB, or irregular bleeding or postcoital bleeding. Proportion of HMB not

Reference	Reason for exclusion
	reported but presumably lower than two thirds of the whole population since only 70% were premenopausal.
Sconfienza, L. M., Lacelli, F., Caldiera, V., Perrone, N., Piscopo, F., Gandolfo, N., Serafini, G., Three-dimensional sonohysterography for examination of the uterine cavity in women with abnormal uterine bleeding: Preliminary findings, Journal of Ultrasound, 13, 2010	Unclear population: menstrual/menopausal status not reported.
Senturk,L.M., Adenomyosis: What is new?, Journal of Endometriosis, 4, 142-143, 2012	Narrative review. Possibly relevant references checked.
Seto, M. T. Y., Ip, P. P. C., Ngu, S. F., Cheung, A. N. Y., Pun, T. C., Positive predictive value of endometrial polyps in Pipelle aspiration sampling: A histopathological study of 195 cases, European Journal of Obstetrics Gynecology and Reproductive Biology, 203, 2016	Unclear population; wrong index test.
Shankar, M., Lee, C. A., Sabin, C. A., Economides, D. L., Kadir, R. A., Von Willebrand disease in women with menorrhagia: A systematic review, BJOG: An International Journal of Obstetrics and Gynaecology, 111, 734-740, 2004	No relevant data.
Sharma, K., Bora, M. K., Venkatesh, B. P., Barman, P., Roy, S. K., Jayagurunathan, U., Sellamuthu, E., Moidu, F., Role of 3D Ultrasound and Doppler in Differentiating Clinically Suspected Cases of Leiomyoma and Adenomyosis of Uterus, Journal of Clinical and Diagnostic Research JCDR, 9, QC08-12, 2015	Wrong population.
Sheth, S., Macura, K., Sonography of the Uterine Myometrium: Myomas and Beyond, Ultrasound Clinics, 2, 267-295, 2007	Expert review. Relevant references checked.
Shinar, S., Bibi, G., Barzilay, L., Rubens, P., Almog, B., Levin, I., The value of diagnostic hysteroscopy before operative hysteroscopy for suspected abnormal intrauterine findings, Journal of Minimally Invasive Gynecology, 21, 2014	Only 46% women with abnormal uterine bleeding.
Shwayder, J., Sakhel, K., Imaging for Uterine Myomas and Adenomyosis, Journal of Minimally Invasive Gynecology, 21, 362-376, 2014	Expert review. No relevant data.
Sincan, S., Keskin, H. L., Ustuner, I., Avsar, A. F., Correlation of preoperative diagnosis and postoperative histopathologic findings in patients operated for presumed leiomyoma and the increased accuracy rate in preoperative diagnose, Turkiye Klinikleri Jinekoloji Obstetrik, 21, 2011	Full text in Turkish.

Reference	Reason for exclusion
Singh, N., Agarwal, G., Singh, U., Jaiswar, S. P., Qureshi, S., Role of chromohysteroscopy in detection of endometrial pathology in perimenopausal & post menopausal women, International Journal of Gynecological Cancer, Conference, 2014	Conference abstract.
Smith, A., O'Brien, K., Ficquet, J., Can the number of hysteroscopies for the investigation of abnormal uterine bleeding be reduced?, International Journal of Gynecology and Obstetrics, Conference, 20th FIGO World Congress of Gynecology and Obstetrics. Rome Italy. Conference Start: 20121007. Conference End: 20121012. Conference Publication: (var.pagings). 119 (pp S743), 2012	Wrong population: 64% of women postmenopausal.
Smith,P., Bakos,O., Heimer,G., Ulmsten,U., Transvaginal ultrasound for identifying endometrial abnormality, Acta Obstetricia et Gynecologica Scandinavica, 70, 591-594, 1991	Wrong population: almost half of the women were postmenopausal.
Song, Y., Shen, L. C., Huang, W., Lei, H. K., Wang, Q. S., Zhu, H. L., Diagnostic value of endometrial thickness determined by transvaginal sonography in infertile women with endometrial polyps, Chinese Medical Journal, 125, 2279	Wrong population: infertile women.
Stamatopoulos, C. P., Mikos, T., Grimbizis, G. F., Dimitriadis, A. S., Efstratiou, I., Stamatopoulos, P., Tarlatzis, B. C., Value of Magnetic Resonance Imaging in Diagnosis of Adenomyosis and Myomas of the Uterus, Journal of Minimally Invasive Gynecology, 19, 620-626, 2012	HMB in 58% of participants.
Steiner, R. A., Fink, D., Abnormal menstrual bleeding, Praxis, 91, 1967-1974, 1967	Full text in German.
Stoelinga, B., Hehenkamp, W., Bromann, H., Huirne, J., Accuracy and reproducibility of ultrasound elastography for the assessment fibroids and adenomyosis, with MRI as reference standard, Gynecological Surgery, 1), S15, 2015	Conference abstract.
Stovall, T. G., Ling, F. W., Morgan, P. L., A prospective, randomized comparison of the Pipelle endometrial sampling device with the Novak curette, American Journal of Obstetrics and Gynecology, 165, 1287-1290, 1991	Index test not of interest.
Struble, J., Reid, S., Bedaiwy, M. A., Adenomyosis: A Clinical Review of a Challenging Gynecologic Condition, Journal of Minimally Invasive Gynecology, 23, 164-85, 2016	A review. Possibly relevant references assessed individually.
Sudderuddin, S., Helbren, E., Telesca, M., Williamson, R., Rockall, A., MRI appearances of benign uterine disease, Clinical Radiology, 69, 1095	Expert review on uterine MRI. Relevant references checked.

Reference	Reason for exclusion
Sydow, B. D., Seigelman, E. S., Uterine MRI: A review of technique and diagnosis, Applied Radiology, 37, 18-29, 2008	Expert review on uterine MRI. Relevant references checked.
Tafazoli, F., Reinhold, C., Uterine adenomyosis: current concepts in imaging, Seminars in Ultrasound, CT & MR, 20, 267-77, 1999	Narrative review on adenomyosis. References checked.
Tahir, M. M., Bigrigg, M. A., Browning, J. J., Brookes, S. T., Smith, P. A., A randomised controlled trial comparing transvaginal ultrasound, outpatient hysteroscopy and endometrial biopsy with inpatient hysteroscopy and curettage, British Journal of Obstetrics and Gynaecology, 106, 1259-1264, 1999	Wrong population: less than half of women had HMB.
Takeuchi, M., Matsuzaki, K., Adenomyosis: usual and unusual imaging manifestations, pitfalls, and problem-solving MR imaging techniques, Radiographics, 31, 99-115, 2011	Narrative review on adenomyosis. References checked.
Tamai, K., Koyama, T., Umeoka, S., Saga, T., Fujii, S., Togashi, K., Spectrum of MR features in adenomyosis, Best Practice & Research in Clinical Obstetrics & Gynaecology, 20, 583-602, 2006	A narrative review. References checked
Tamai, K., Togashi, K., Ito, T., Morisawa, N., Fujiwara, T., Koyama, T., MR imaging findings of adenomyosis: correlation with histopathologic features and diagnostic pitfalls, Radiographics, 25, 21-40, 2005	A narrative review. References checked
Teale, G. R., Dunster, G. D., The Pipelle endometrial suction curette: How useful is it in clinical practice?, Journal of Obstetrics and Gynaecology, 18, 53-55, 1998	Index test not of interest.
Thubert, T., Demoulin, G., Lamazou, F., Rivain, A. L., Trichot, C., Faivre, E., Deffieux, X., Menometrorrhagia, Revue du Praticien, 64, 2014	Full text in French.
To, J. K., Levie, M., Chudnoff, S., Endometrial pipelle biopsy identifying a polyp versus office hysteroscopy, Journal of Minimally Invasive Gynecology, Conference, 2015	Conference abstract.
Torrejon, R., Fernandez-Alba, J. J., Carnicer, I., Martin, A., Castro, C., Garcia-Cabanillas, J., Rodriguez-Cornejo, J., Moreno, L. J., Comino, R., The value of hysteroscopic exploration for abnormal uterine bleeding, Journal of the American Association of Gynecologic Laparoscopists, 4, 453-456, 1997	Wrong population: 43% postmenopausal women.
Touqmatchi, D., Sharma, M., Joash, K., Outpatient hysteroscopy (OPH) service; patient satisfaction and adequacy of service, BJOG: An International Journal of Obstetrics and Gynaecology, Conference, 2013	Conference abstract.

Reference	Reason for exclusion
Towbin,N.A., Gviazda,I.M., March,C.M., Office hysteroscopy versus transvaginal ultrasonography in the evaluation of patients with excessive uterine bleeding, American Journal of Obstetrics and Gynecology, 174, 1678-1682, 1996	Wrong population: only 61.7% of women had HMB.
Tsai, M. C., Goldstein, S. R., Office diagnosis and management of abnormal uterine bleeding, Clinical Obstetrics and Gynecology, 55, 635-650, 2012	Expert review. Relevant references checked.
Turkgeldi, E., Urman, B., Ata, B., Role of Three-Dimensional Ultrasound in Gynecology, Journal of Obstetrics and Gynecology of India, 65, 146-154, 2015	Expert review. Relevant references checked.
Valentini, A. L., Speca, S., Gui, B., Soglia, G., Micco, M., Bonomo, L., Adenomyosis: from the sign to the diagnosis. Imaging, diagnostic pitfalls and differential diagnosis: a pictorial review, La Radiologia medica, 116, 1267-1287, 1267	A pictorial review. References checked.
Valle, R. F., Hysteroscopic evaluation of patients with abnormal uterine bleeding, Surgery Gynecology and Obstetrics, 153, 521-526, 1981	No sufficient data to form 2x2 table and calculate diagnostic accuracy.
Van den Bosch, T., Valentin, L., Van Schoubroeck, D., Luts, J., Bignardi, T., Condous, G., Epstein, E., Leone, F. P., Testa, A. C., Van Huffel, S., Bourne, T., Timmerman, D., Detection of intracavitary uterine pathology using offline analysis of three-dimensional ultrasound volumes: interobserver agreement and diagnostic accuracy, Ultrasound in Obstetrics & Gynecology, 40, 459-63, 2012	Wrong/unclear population: 30% postmenopausal women, rest premenopausal but with AUB, % with HMB not reported.
Van den Bosch, T., Verguts, J., Daemen, A., Gevaert, O., Domali, E., Claerhout, F., Vandenbroucke, V., De Moor, B., Deprest, J., Timmerman, D., Pain experienced during transvaginal ultrasound, saline contrast sonohysterography, hysteroscopy and office sampling: a comparative study, Ultrasound in Obstetrics & GynecologyUltrasound Obstet Gynecol, 31, 2008	Wrong population: 39% were postmenopausal women.
Van Dongen, H., De Kroon, C. D., Jacobi, C. E., Trimbos, J. B., Jansen, F. W., Diagnostic hysteroscopy in abnormal uterine bleeding: A systematic review and meta-analysis, BJOG: An International Journal of Obstetrics and Gynaecology, 114, 2007	Systematic review with different inclusion criteria; included studies checked individually.
Veena, B. T., Shivalingaiah, N., Role of transvaginal sonography and diagnostic hysteroscopy in abnormal uterine bleeding, Journal of Clinical and Diagnostic Research, 8, OC06-OC08, 2014	Wrong population: less than two thirds with HMB.

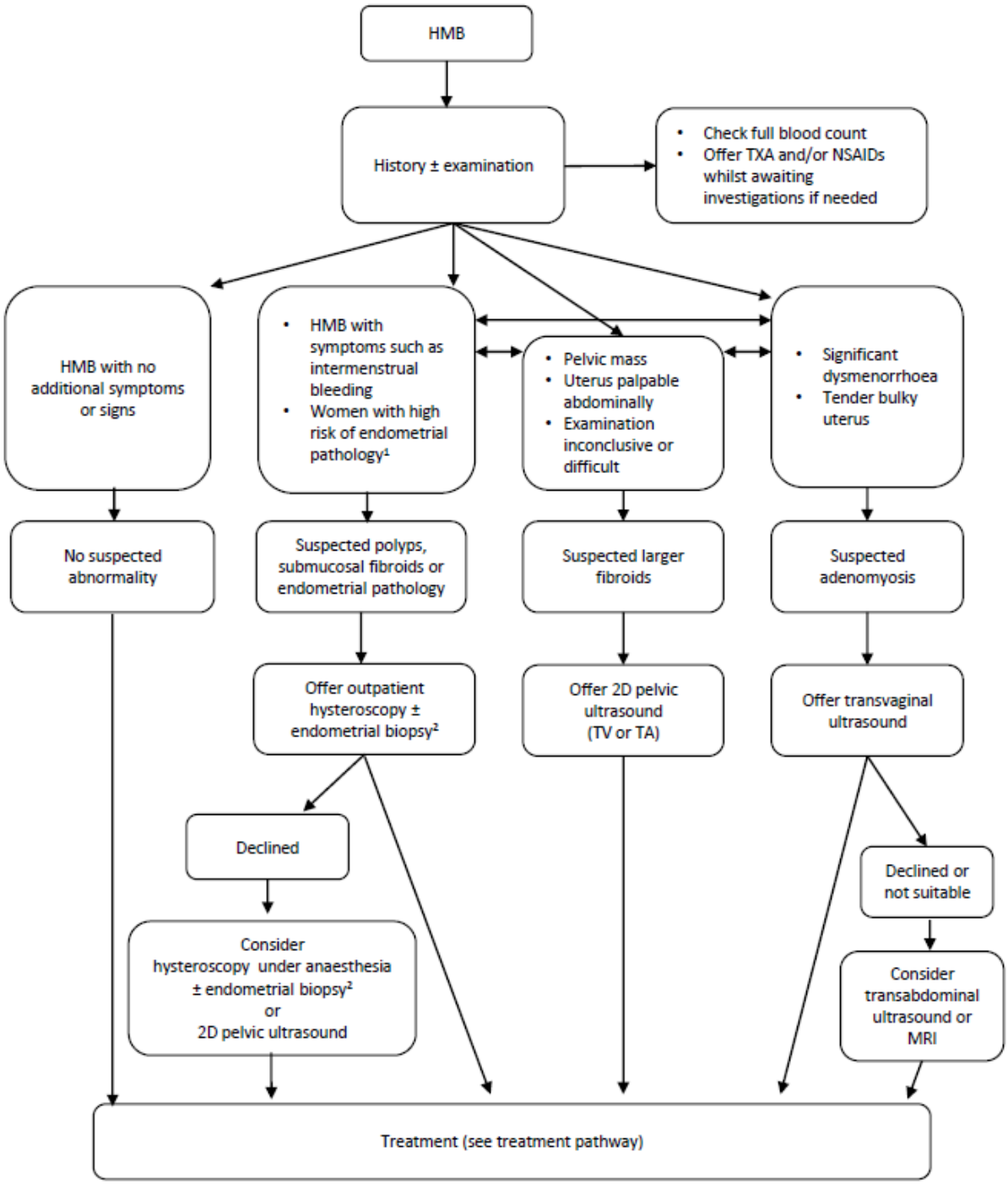
Reference	Reason for exclusion
Venugopalan, S. K., Pandian, N. S., Pavani, M., Srinivasa Rao, T., Rajini, Y., Khadeer, S. K., Ravichandiran, V., Abnormal uterine bleeding in reproductive women: Diagnosis, management and treatment, Asian Journal of Pharmaceutical and Clinical Research, 8, 42-45, 2015	No relevant data.
Vitner, D., Filmer, S., Goldstein, I., Khatib, N., Weiner, Z., A comparison between ultrasonography and hysteroscopy in the diagnosis of uterine pathology, European Journal of Obstetrics Gynecology and Reproductive Biology, 171, 2013	Unclear population: referring symptoms and menopausal status not reported.
Walker, K., Jayaprakasan, K., Raine-Fenning, N. J., Ultrasound in benign gynaecology, Obstetrics, Gynaecology and Reproductive Medicine, 17, 2007	Expert review. Possible relevant references checked.
Walker, W. J., Jones, K., Transvaginal ultrasound guided biopsies in the diagnosis of pelvic lesions, Minimally Invasive Therapy and Allied Technologies, 12, 2003	Unclear/wrong population.
Wery, O., Thille, A., Gaspard, U., Van Den Brule, F., Adenomyosis: Update on a frequent but difficult diagnosis, Journal de Gynecologie Obstetrique et Biologie de la Reproduction, 34, 633-648, 2005	Full text in French.
Whitaker, L., Critchley, H. O. D., Abnormal uterine bleeding, Best Practice and Research: Clinical Obstetrics and Gynaecology., 2016	Expert review. Relevant references checked.
Widrich, T., Bradley, L. D., Mitchinson, A. R., Collins, R. L., Comparison of saline infusion sonography with office hysteroscopy for the evaluation of the endometrium, American Journal of Obstetrics and Gynecology, 174, 1327-1334, 1996	Wrong population: 73% premenopausal, 68% had abnormal uterine bleeding, proportion with HMB not reported.
Wolfman, D. J., Allison, S. J., Ascher, S. M., Imaging of benign uterine conditions, Applied Radiology, 40, 2011	Narrative review, possible relevant references checked.
Wong, I. K. C., Visvanathan, D., Do women having an outpatient hysteroscopy have a higher anxiety level compared to a day case hysteroscopy? A prospective questionnaire research study, BJOG: An International Journal of Obstetrics and Gynaecology, Conference, 2016	Conference abstract.
Wood, C., Hurley, V. A., Leoni, M., The value of vaginal ultrasound in the management of menorrhagia, Australian and New Zealand Journal of Obstetrics and Gynaecology, 33, 198-200, 1993	Not data to calculate diagnostic accuracy measures.
Wu, K. K., Wang, Y. Y., Qian, B., Pan, Y., Chang, C., The classification of uterine myoma and uterine adenomyosis based on ultrasound image features, Chinese Journal of Biomedical Engineering, 26, 2007	Full text in Chinese.

Reference	Reason for exclusion
Xyda, A., Moyle, P., Addley, H., Freeman, S., Imaging of the female pelvis, Obstetrics, Gynaecology and Reproductive Medicine, 25, 283-294, 2015	Narrative expert review. Possibly relevant references checked.
Yang, P. Y., Wu, J. L., Wu, P. W., Wu, C. H., Lin, C. J., Fan, L. R., Yang, Y. H., Yeh, G. P., Accuracy of transvaginal ultrasonography for detecting intrauterine lesions at a Taiwan medical center: A correlation with ultrasound and hysteroscopic histopathology, Journal of Medical Ultrasound, 22, 37-42, 2014	Wrong/unclear population: 66.8% with abnormal uterine bleeding, % of HMB not reported and presumably less than 66.6%.
Yildizhan, B., Yildizhan, R., Ozkesici, B., Suer, N., Transvaginal ultrasonography and saline infusion sonohysterography for the detection of intra-uterine lesions in pre- and post-menopausal women with abnormal uterine bleeding, Journal of International Medical Research, 36, 1205	Wrong population: less than two thirds with HMB.
Yumru, A. E., Bozkurt, M., Coskun, E. I., Erkilincoglu, M., Ayanoglu, Y. T., Evaluation of diagnostic methods in abnormal uterine bleeding and endometrial carcinoma: Sectional prospective study, Nobel Medicus, 5, 2009	Wrong population: less than two thirds with HMB.
Zegura, B., Office operative hysteroscopy, Gynecological Surgery, Conference, 2015	Wrong population: 45% women post-menopausal.
Zhao, F., Zhang, H., Ren, Y., Kong, F., Transvaginal sonographic characteristics of paraovarian borderline tumor, International Journal of Clinical and Experimental Medicine, 8, 2684	Wrong population: women with diagnosis of parovarian tumour; small sample (n=4).
Zhu, H. L., Liang, X. D., Wang, J. L., Cui, H., Wei, L. H., Hysteroscopy and directed biopsy in the diagnosis of endometrial carcinoma, Chinese Medical Journal, 123, 3524	Wrong population: women with diagnosed endometrial carcinoma.

Economic studies

There were no excluded economic studies.

Appendix J – Diagnostic care pathway



2D: two-dimensional; HMB: heavy menstrual bleeding; LNG-IUS: levonorgestrel-releasing intrauterine system; MRI: magnetic resonance imaging; NSAIDs: nonsteroidal anti-inflammatory drugs; TA: transabdominal; TV: transvaginal; TXA: tranexamic acid

[1]women who are obese, women with persistent intermenstrual bleeding, or irregular or infrequent bleeding (for example, women with polycystic ovary syndrome), women taking tamoxifen, women for whom treatment for HMB has been unsuccessful

[2]if high risk for endometrial pathology (see footnote 1)