National Institute for Health and Care Excellence

Final

Early and locally advanced breast cancer: diagnosis and management

[C] Evidence reviews for adjuvant systemic therapy planning

NICE guideline NG101

Evidence reviews

July 2018

Final

These evidence reviews were developed by the National Guideline Alliance hosted by the Royal College of Obstetricians and Gynaecologists

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 All NICE guidance is subject to regular review and may be updated or withdrawn.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: 978-1-4731-3008-1

Contents

Adjuvant systemic therapy planning

This evidence report contains information on 2 reviews relating to adjuvant systemic therapy planning.

- Review question 3.1. Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?
- Review question 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?

Review question 3.1. Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?

Introduction

Current UK recommendations in the previous guideline CG80 (NICE 2009), and from the Royal College of Pathologists (RCPath, 2016), state that oestrogen receptor (ER) and human epidermal growth factor receptor 2 (HER2) biomarkers should be assessed in all invasive breast cancers. This biomarker analysis can provide important prognostic and predictive information to help direct further adjuvant management breast cancer after surgery.

ER positivity in breast cancers can predict a potential response to endocrine-based treatments and these cancers are known to have an overall better prognosis than ER-negative cancers. Progesterone receptor (PR) is from the same family of molecules as ER, but CG80 recommended not to routinely test all breast cancers for PR as, at the time, there was no strong evidence to support PR being predictive of a response to endocrine therapy (despite being independently prognostic for relapse-free survival and overall survival).

The co-expression of ER and PR does vary between breast cancers. Whilst the majority of breast cancers which are ER positive are also PR positive, many are PR negative, and studies have now shown these to have a worse prognosis and to be less responsive to endocrine therapies. Some people have breast cancers that are negative for each of ER, PR and HER2. As none of the 3 biomarkers are expressed in these cancers, they are conventionally referred to as 'triple negative' and are associated with a poor prognosis without treatment, but the cancer may respond well to certain forms of chemotherapy.

The purpose of this review question is to determine if establishing PR status affects planning for adjuvant chemotherapy.

PICO table

See Table 1 for a summary of the population, intervention, comparison and outcome (PICO) characteristics of this review.

Table 1: Summary of the protocol (PICO table)

Population	Adults (18 or over) with invasive breast cancer (M0)
Intervention	• ER and HER2 plus PR test followed by chemotherapy as indicated based on test results
Comparison	• ER and HER2 test followed by chemotherapy as indicated based on test results
Outcome	**Critical** • Disease-free survival • Overall survival **Important** • Treatment-related morbidity

ER, oestrogen receptor; HER2, human epidermal growth factor receptor-2; M0, no distant metastases; PR, progesterone receptor

For full details see review protocol in appendix A.

Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual; see the methods chapter for further information. Methods specific to this review question are described in the review protocol in appendix A.

Declarations of interest were recorded according to NICE's 2014 conflicts of interest policy.

Clinical evidence

Included studies

There were no test and treat randomised controlled trials (RCTs) identified by the literature search. In test and treat studies, only participants who get discrepant test results (for example, chemotherapy indicated versus chemotherapy not indicated) would receive different treatment and only a proportion of those may benefit from differences in treatment. Therefore, the committee deemed it inappropriate to drop down the evidence hierarchy to include non-randomised studies as it is likely that bias inherent in such studies would dominate any treatment effect.

The study selection flow chart is in appendix C.

Excluded studies

Studies not included in this review with reasons for their exclusions are provided in appendix K.

Quality assessment of clinical studies included in the evidence review

No studies were included in this review question.

Economic evidence

A systematic review of the economic literature was conducted but no relevant studies were identified which were applicable to this review question. Economic modelling was not undertaken for this question because other topics were agreed as higher priorities for economic evaluation.

Formal consensus

The committee decided that a modified form of the nominal group technique would be the most appropriate method for producing recommendations regarding the appropriateness of PR testing. The method used for the nominal group technique is described in full within the methods chapter.

Key issues related to progesterone receptor testing were identified from relevant papers identified by the current search results (Duffy 2005; Hammond 2010, Harris 2007; Henry 2016) the previous guideline CG80 (NICE 2009), key papers and guidelines identified by the guideline committee (Early Breast Cancer Trialists' Collaborative Group [EBCTCG] 2011), and from protocol discussions with the committee. These were used to generate statements covering the following areas: prognosis based on progesterone receptor status, impact of progesterone receptor status on endocrine therapy and chemotherapy, and assessment of progesterone receptor status. These statements were placed into a questionnaire and distributed to the guideline committee to be rated.

The first round of rating was completed by 11 of 16 committee members. Percentage agreement values were calculated and comments collated for each statement; the rankings

and comments were then presented to the committee members to facilitate a structured discussion. Two statements were redrafted based on the comments from the committee members and re-distributed for rating as a second questionnaire; this round was completed by 10 of 16 committee members. A blank copy of the questionnaire (including re-rated statements) can be found in appendix M and consensus ratings can be found in appendix N.

A brief summary of level of consensus is depicted in Table 2 below.

Table 2: Summary of nominal group technique process followed for the development of recommendations on progesterone receptor testing

Round 1		Round 2		Number of recommendations generated
Level of consensus	Statements N (total = 16)	Level of consensus	Statements N (total = 2)	1
High (≥80%)	4	High (≥80%)	2	
Moderate (60-80%)	1	Moderate (60-80%)	0	
Low (<60%)	11	Low (<60%)	0	

Evidence statements

The committee agreed that:

- positive progesterone receptor status is associated with favourable prognosis and negative progesterone receptor status is associated with worse prognosis
- progesterone receptor status provides additional information to oestrogen receptor status that may be beneficial when considering the benefit of adjuvant hormone treatment
- negative progesterone receptor status is one factor that may increase benefit from chemotherapy, and likelihood that it is offered in borderline cases
- progesterone receptor status is relevant when making decisions regarding adjuvant therapy and should be assessed in all newly diagnosed invasive breast cancers.

The committee's discussion of the evidence

Interpreting the evidence

The outcomes that matter most

The committee identified disease-free survival and overall survival as critical outcomes. Treatment-related morbidity was selected as an important outcome to examine the impact of any additional treatment required as a result of progesterone receptor testing. These outcomes are valued by service users as increased survival is prioritised; however, treatment-related morbidities can have a significant impact on health-related quality of life and adherence to treatment.

No test and treat studies were identified from the literature search, therefore there was no evidence for any of the outcomes reported in the PICO.

The impact of PR status on prognosis, benefit from endocrine therapy and benefit from chemotherapy were identified through discussions with the committee as key areas related to the need for PR testing. These areas were used as guides for generating statements to be ranked by the committee using a modified form of the nominal group technique.

The quality of the evidence

No published evidence was identified for this review. Although there were high levels of agreement in the nominal group technique for statements which informed and supported recommendations, this formal consensus method constitutes low quality evidence.

Benefits and harms

The addition of PR testing to ER testing will provide further information on which to base decisions regarding adjuvant hormone therapy and chemotherapy resulting in better tailored treatment. Specifically, tumours that are negative for PR have a worse prognosis and therefore may receive greater benefit from chemotherapy. Determining negative PR status may therefore increase the likelihood of chemotherapy being offered in borderline cases.

Assessing PR status upfront in all newly diagnosed invasive cancers reduces delays in decision making that may occur if determining progesterone receptor status is carried out at a later stage, and not at the same time as ER testing. This may allow earlier commencement of treatment. Earlier treatment may lead to a reduction in recurrence and mortality. The committee made an additional recommendation therefore that the 3 tests (ER, PR and HER2) should all be requested simultaneously at the time of initial histopathological diagnosis, to prevent delays in treatment.

No harms were identified by the committee as no additional procedures are required for progesterone receptor testing. Treatment-related morbidities were discussed, but the committee thought that there would not be significant increases in morbidities as currently the majority of patients do receive progesterone receptor testing and corresponding treatment when indicated, but the testing is not routinely done upfront.

The benefits identified combined with the lack of harms led the committee to make a strong recommendation in this area despite the low quality of the evidence.

Cost effectiveness and resource use

A systematic review of the economic literature was conducted but no relevant studies were identified which were applicable to this review question.

There is the potential for cost increases associated with testing PR status. However the committee did not think the increase would be large as some centres are already performing upfront PR testing and it would be more expensive if a second test was required to determine PR status.

There is also a potential cost increase associated with additional adjuvant endocrine therapy and chemotherapy that may be indicated as a result of the PR test. This is unlikely to be large as the recommendation will have a greater impact on the timing of PR testing rather than whether the test, and indicated treatment, occurs.

In contrast, performing the PR test upfront will produce cost savings as pathology results will not need to be discussed at multiple multidisciplinary team meetings, and fewer second appointments will be required for decision making and adjuvant treatment planning; currently, an additional multidisciplinary team meeting may be required to discuss the impact of PR status if this information is not available at the initial meeting. There is also a potential for cost savings if treatment improvements reduce recurrence and/or mortality as there will be a decreased need for future procedures, treatments and hospice care.

Other factors the committee took into account

Ethical considerations

There are elevated rates of triple-negative breast cancer among some ethnic groups, for example Afro-Caribbean people, and they are therefore more likely to be affected by delays to optimal treatment if progesterone receptor status is not known. The current recommendation will reduce this inequality as progesterone receptor testing will be performed upfront allowing for earlier determination of triple-negative status.

Methods for assessing and reporting progesterone-receptor status

The committee recommended that PR status be assessed using immunohistochemical techniques. This is standard clinical practice and all UK laboratories using hormone receptor assays are subject to national quality assurance. Furthermore, the committee recommended that results are reported quantitatively (as opposed to dichotomously) as the degree of positivity is directly correlated with a better prognosis.

References

Duffy 2005

Duffy, M. J. (2005). Predictive markers in breast and other cancers: a review. Clinical chemistry, 51, 494-503.

EBCTCG 2011

Early Breast Cancer Trialists' Collaborative Group (EBCTCG). (2011). Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. Lancet, 378, 771-784.

Hammond 2010

Hammond, M. E. H., Hayes, D. F., Dowsett, M., Allred, D. C., Hagerty, K. L., Badve, S., Fitzgibbons, P. L., Francis, G., Goldstein, N. S., Hayes, M., Hicks, D. G., Lester, S., Love, R., Mangu, P. B., McShane, L., Miller, K., Osborne, C. K., Paik, S., Perlmutter, J., Rhodes, A., Sasano, H., Schwartz, J. N., Sweep, F. C. G., Taube, S., Torlakovic, E. E., Valenstein, P., Viale, G., Visscher, D., Wheeler, T., Williams, R. B., Wittliff, J. L., Wolff, A. C. (2010). American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). Archives of Pathology & Laboratory Medicine, 134, e48-e72.

Harris 2007

Harris, L., Fritsche, H., Mennel, R., Norton, L., Ravdin, P., Taube, S., Somerfield, M. R., Hayes, D. F., Bast Jr, R. C. (2007). American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. Journal of Clinical Oncology, 25, 5287-5312.

Henry 2016

Henry, N. L., Somerfield, M. R., Abramson, V. G., Allison, K. H., Anders, C. K., Chingos, D. T., Hurria, A., Openshaw, T. H., Krop, I. E. (2016). Role of patient and disease factors in adjuvant systemic therapy decision making for early-stage, operable breast cancer: American Society of Clinical Oncology endorsement of Cancer Care Ontario guideline recommendations. Journal of Clinical Oncology, 34, 2303-2311.

NICE 2009

National Institute for Health and Clinical Excellence. (2009). Early and locally advanced breast cancer: diagnosis and treatment. NICE guideline (CG80).

RCPath 2016

Ellis, I. O., Carder, P., Hales, S., Lee, A. H. S., Pinder, S. E., Rakha, E., Stephenson, T., Al-Sam, S., Deb, R., Hanby, A., Liebmann, R., Provenzano, E., Rowlands, D., Wells, A., Anderson, N., Girling, A., Ibrahim, M., Mallon, E., Quinn, C. (2016). Pathology reporting of breast disease in surgical excision specimens incorporating the dataset for histological reporting of breast cancer. The Royal College of Pathologists: London, UK.

Review question 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?

Introduction

Planning adjuvant treatment is complex and incorporates a variety of prognostic and predictive factors. In order to identify which people would benefit from adjuvant therapy, a number of prognostic tools have been developed. These take into account a number of factors such as age, comorbidities, tumour staging and biomarkers, and assess the risk of an individual person developing recurrent disease and/or dying within 10 years when receiving a specific treatment. These prognostic tools can be used jointly by the person and their doctor to determine the most appropriate adjuvant treatment (chemotherapy, endocrine therapy, or no therapy).

The aim of this review is to determine which of the currently available prognostic tools is most reliable at correctly predicting survival and the benefits of adjuvant treatment.

See Table 3 for a description of the prognostic tools included in this review.

Table 3: Description of the prognostic tools

Prognostic tool	Description	Factors included in the model
Adjuvant! Online	Adjuvant! is an online tool that aims to help healthcare professionals and people with early cancer discuss the benefits and risks of receiving additional adjuvant therapy after surgery. For further details please see https://www.adjuvantonline.com/	Adjuvant therapy Age Comorbidity ER status Menopausal status Number of positive lymph nodes Tumour size
PREDICT	The PREDICT tool is a free online computer programme developed by the NHS and the University of Cambridge, and it aims to help patients and healthcare professionals decide on the ideal course of treatment following surgery for breast cancer. There are different versions of PREDICT (personal communication,): v1.0 (2011) v1.1 (also known as PREDICT Plus) – modified version of PREDICT v1.0 + HER2 v1.2 – modified version of PREDICT v1.1 + KI67 v2.0 (2017) – updated version with substantial modifications to the underlying model For further details please see http://www.predict.nhs.uk/index.html	Age at diagnosis ER status Gen chemo regimen HER2 status KI67 status Mode of detection Number of positive nodes Tumour grade Tumour size in mm
Nottingham Prognostic Index (NPI)	The Nottingham prognostic index (NPI) is a tool used to determine prognosis following breast cancer surgery. For further details please see http://www.pmidcalc.org/?sid=3689666&newtest=Y	Grade of the tumour Number of involved lymph nodes Size of the lesion
FinProg	FinProg is an online-based system for individualised survival estimation in breast cancer. For further details please see http://www.finprog.org/	Adjuvant therapy Age ER

Prognostic tool	Description	Factors included in the model
		HER2 Histologic grade Histologic type Lymph node status Method of detection PR Tumour size
CancerMath	Cancer-Math is an online tool aimed to provide healthcare professionals with web-based calculators for: 1) accurately predicting the clinical outcome for people with cancer (including breast cancer), and 2) accurately estimating the impact of various treatment choices on that outcome. For further details please see http://www.lifemath.net/cancer/?cancer	Age Chemotherapy ER status Grade HER2 status Histological type Hormonal therapy Number of positive nodes PR status Tumour diameter in mm
Oxford Prognostic Index (OPI)	The Oxford Prognostic Index (OPI) is a tool aimed to predict the long-term risk of a recurrent event in women diagnosed with early breast cancer.	Age ER status Nodal status Tumour grade Tumour size

ER: oestrogen receptor; HER2: human epidermal growth factor receptor 2; NPI: Nottingham Prognostic Index; OPI: Oxford Prognostic Index; PR: progesterone receptor

PICOTS table

See Table 4 for a summary of the population, intervention (predictive prognostic tool), comparison, outcome, timing and setting (PICOTS) characteristics of this review.

Table 4: Summary of the protocol (PICOTS table)

Population	Adults (18 or over) with invasive breast cancer (M0) who have undergone surgery and who are candidates for adjuvant systemic therapy. Only studies conducted with UK population will be considered for inclusion.
Intervention (Predictive prognostic tools)	Any appropriate predictive prognostic tools, for example, • Adjuvant! Online • PREDICT • Nottingham Prognostic Index (NPI) • FinProg • CancerMath • Only studies assessing validated tools will be considered for inclusion.
Outcome	**Critical** • Calibration • Discrimination (AUROC) **Important** • Accuracy of prediction (sensitivity, specificity)

Timing	• 5 years • 10 years
Setting	• UK

AUROC, area under receiver operating characteristic curve; M0, no distant metastases; NPI, Nottingham prognostic index

For full details see the review protocol in appendix A.

Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual; see the methods chapter for further information. Methods specific to this review question are described in the review protocol in appendix A.

Declarations of interest were recorded according to NICE's 2014 conflicts of interest policy.

Clinical evidence

We included validation studies that reported sensitivity, specificity, calibration and discrimination of validated predictive prognostic tools.

It was agreed with the committee that sensitivity or specificity would be considered high when sensitivity or specificity was 90% or higher, and moderate when sensitivity or specificity was between 75% and 89%. However none of the studies reported these outcomes.

The mortality ratio is defined as the ratio of observed number of deaths in a study population and the expected number of deaths. In this review, a tool was judged to have good calibration if the ratio ranged from 0.8 to 1.2 (as suggested by Debray 2017).

Discrimination is a measure to assess how well a tool identifies people with worse survival, and it is often reported by the concordance c-statistic (also known as AUC). In this review a tool was judged to have good discrimination if c-statistic was above 0.75 (as suggested by Debray 2017).

Included studies

Seven studies (number of participants, N=27,287) were included in this review (Blamey 2007, Campbell 2009, Campbell 2010, Candido dos Reis 2017, Maishman 2015, Wishart 2010 and Wishart 2014).

One study looked at the Nottingham Prognostic Index (NPI) (Blamey 2007), 1 study looked at Adjuvant! Online (Campbell 2009), 1 study looked at the Oxford Prognostic Index (OPI) (Campbell 2010), and 4 studies looked at PREDICT (Candido dos Reis 2017, Maishman 2015, Wishart 2010 and Wishart 2014). Because a number of versions of PREDICT exist, the authors were contacted to seek clarification.

All studies were conducted with a UK population.

The clinical evidence from these studies is summarised in Table 5. Please note that GRADE profiles are not applicable to this review question. See also the study selection flow chart in appendix C, forest plots in appendix E, study evidence tables in appendix D and exclusion list in appendix K.

This review updates a question from the previous guideline CG80 (NICE 2009). However no studies were identified in the previous guideline.

Excluded studies

Studies not included in this review with reasons for their exclusions are provided in appendix K.

Summary of clinical studies included in the evidence review

Table 5: Summary of included studies

Study	Population	Predictive prognostic tool	Outcomes	Comments
Blamey 2007	Women diagnosed with or treated for primary operable invasive breast cancer at Nottingham city hospital Dates: • 1980 to 1986 (n=892) • 1990 to 1999 (n=2238)	Nottingham Prognostic Index (NPI)	• Observed and predicted 10-year breast cancer survival	
Campbell 2009	Data from 1,065 women with early breast cancer diagnosed at the Churchill hospital Oxford Dates: 1986 to 1996 UK population	Adjuvant! Online	• Observed and predicted 10-year overall survival (%) • Observed and predicted 10-year breast cancer specific survival (%) • Observed and predicted 10-year event free survival (%)	
Campbell 2010	N=1787 women with invasive ductal carcinoma, a sub-set obtained from the Adjuvant Breast Cancer trial from 70 UK centres Dates: 1992 to 2000 UK population	Oxford Prognostic Index (OPI)	• Observed and predicted 5-year recurrence free survival	
Candido dos Reis 2017	Tool development Data from 5738 people from the ECRIC database Dates: 1999 to 2003 Validations study Data from the following databases: • BCOS: n=981 (dates: 1990 to 2000)	PREDICT v2.0*	• Observed and predicted 10-year breast-cancer mortality • Observed and predicted 10-year all-cause mortality	• This validation study reports data for PREDICT v2.0 and v1. Data for v1 was not used in the analysis as for many of the cases in the validation data the authors did not have either HER2 status or KI67 status*

Study	Population	Predictive prognostic tool	Outcomes	Comments
	• NTBCS: n=1726 (dates: 1989 to 1998) • POSH: n=2609 (dates: 2000 to 2008) • UK population			
Maishman 2015	Data from 3000 women aged ≤40 years at diagnosis (POSH cohort) UK population	PREDICT v1.2*	• Observed and predicted 5-year all-cause mortality • Observed and predicted 10-year all-cause mortality	
Wishart 2010	Data from 5468 people with breast cancer from the West Midlands Cancer Intelligence Unit (WMCIU) Dates: 1999 to 2003 UK population	PREDICT v1.1*	• Observed and predicted 5-year all-cause mortality • Observed and predicted 8-year all-cause mortality	• Validation study (data from the primary analysis has not been reported)
Wishart 2014	Data for 2232 cases of invasive breast cancer treated in Nottingham - 506 node-negative cases were excluded, so data from n=1726 people was included in the study Dates: 1989 to 1998 UK population	PREDICT v1.1* and v1.2*	• Observed and predicted 10-year all-breast cancer mortality	

**This information was provided by PREDICT (info@predict.nhs.uk)*
BCOS, Breast Cancer Outcomes Simulator; ECRIC, East Anglia cancer registration and information centre; NPI, Nottingham Prognostic Index; NTBC, Nottingham Tenovus Breast Cancer; OPI, Oxford Prognostic Index; POSH, Prospective study of Outcomes in Sporadic versus Hereditary breast cancer; UK, United Kingdom; WMCIU, West Midlands Cancer Intelligence Unit

See appendix D for full evidence tables.

Quality assessment of clinical studies included in the evidence review

The included studies were individually assessed using the Critical Appraisal Skills Programme (CASP) tool for clinical prediction rule. Studies were rated as of moderate or high quality. The reasons for rating down the quality of the studies were that the tool had not been validated in a different population, or that the tool did not include all the relevant prognostic factors.

The clinical evidence profiles for this review question (prognostic tools) are presented in Table 6 to Table 9.

Predictive prognostic tool 1: Adjuvant! Online

Table 6: Summary of included studies and results for Adjuvant! Online

Study	Study details	Population	Findings	Comments	Quality
10-year overall survival (OS)					
Campbell (2009)	Validation study Study period: 1986 to 1996	Data from 1,065 women with early breast cancer diagnosed at the Churchill hospital I Oxford	Prognostic accuracy (sensitivity, specificity) Not reported Tool calibration All population (N=1065): • OS ratio O:E = 0.93 • Difference O-E = -5.54 (p<0.01) *Subgroup: age* • 20 to 35 (n=34) ○ OS ratio O:E = 0.97 ○ Difference O-E = -2.27% (ns) • 36 to 50 (n=363) ○ OS ratio O:E = 0.95 ○ Difference O-E = -4.33% (p<0.05) • 51 to 65 (n=458) ○ OS ratio O:E = 0.95 ○ Difference O-E = -4.02% (p<0.05) • 66 to 75 (n=194) ○ OS ratio O:E = 0.82 ○ Difference O-E = -12.17% (p<0.01) • ≥76 (n=16) ○ OS ratio O:E = 0.94 ○ Difference O-E = -3.11% (ns) *Subgroup: grade:* • Grade 1 (n=152)	• Unknown confounders • OS ratio O:E was calculated by the NGA technical team based on available data on the paper • Other factors included in the model: histology, local therapy, systemic therapy	• High quality, as assessed by CASP Clinical Prediction Rule checklist.

Study	Study details	Population	Findings	Comments	Quality
			○ OS ratio O:E: 0.96 ○ Difference O-E: - 3.65% (ns) • Grade 2 (n=421) ○ OS ratio O:E: 0.91 ○ Difference O-E: -7.05% (p<0.01) • Grade 3 (n=248) ○ OS ratio O:E: 0.86 ○ Difference O-E: -9.82% (p<0.01) • Unknown grade (n=244) ○ OS ratio O:E: 1.00 ○ Difference O-E: 0.26% (ns) • *Subgroup: tumour size:* • 0.1 to 1 cm (n=150) ○ OS ratio O:E: 0.93 ○ Difference O-E: -6.10% (ns) • 1.1 to 2 cm (n=471) ○ OS ratio O:E: 0.92 ○ Difference O-E: -6.57% (p<0.01) • 2.1 to 5 cm (n=444) ○ OS ratio O:E: 0.94 ○ Difference O-E: -4.26% (ns) • *Subgroup: nodal status* • Negative (n=733) ○ OS ratio O:E = 0.94 ○ Difference O-E = -4.70% (p<0.01) • Positive (n=332) ○ OS ratio O:E = 0.89 ○ Difference O-E = -7.38% (p<0.01)		

Study	Study details	Population	Findings	Comments	Quality
			• *Subgroup: ER status* • Negative (n=261) ○ OS ratio O:E = 0.97 ○ Difference O-E = -1.93% (ns) • Positive (n=495) ○ OS ratio O:E = 0.89 ○ Difference O-E = -9.00% (p<0.01) • Unknown (n=309) ○ OS ratio O:E = 0.96 ○ Difference O-E = -3.04% (ns) • Tool discrimination • Not reported		
10-year breast cancer specific survival (BCSS)					
Campbell (2009)	Validation study Study period: 1986 to 1996	Data from 1,065 women with early breast cancer diagnosed at the Churchill hospital I Oxford	Prognostic accuracy (sensitivity, specificity) Not reported Tool calibration All population (N=1058): • BCSS ratio O:E: 0.95 • Difference O-E: -4.53% (p<0.01) *Subgroup: age* • 20 to 35 (n=34) ○ BCSS ratio O:E = 0.99 ○ Difference O-E = -0.67% (ns) • 36 to 50 (n=361) ○ BCSS ratio O:E = 0.94 ○ Difference O-E = -4.62% (p<0.05) • 51 to 65 (n=454) ○ BCSS ratio O:E = 0.96 ○ Difference O-E = -3.51% (ns)	• Unknown confounders • BCSS ratio O:E was calculated by the NGA technical team based on available data on the paper • Other factors included in the model: histology, local therapy, systemic therapy	• High quality, as assessed by CASP Clinical Prediction Rule checklist.

Study	Study details	Population	Findings	Comments	Quality
			• 66 to 75 (n=193) ○ BCSS ratio O:E = 0.89 ○ Difference O-E = -9.28% (p<0.05) • ≥76 (n=16) ○ BCSS ratio O:E = 1.08 ○ Difference O-E = 7.04% (ns) *Subgroup: grade:* • Grade 1 (n=152) ○ BCSS ratio O:E: 0.99 ○ Difference O-E: -1.29% (ns) • Grade 2 (n=420) ○ BCSS ratio O:E: 0.93 ○ Difference O-E: -5.89% (p<0.01) • Grade 3 (n=243) ○ BCSS ratio O:E: 0.92 ○ Difference O-E: -6.10 (p<0.05) • Unknown grade (n=243) ○ BCSS ratio O:E: 0.96 ○ Difference O-E: -2.78 (ns) • *Subgroup: tumour size:* • 0.1 to 1 cm (n=148) ○ BCSS ratio O:E: 0.92 ○ Difference O-E: -7.95% (p<0.01) • 1.1 to 2 cm (n=470) ○ BCSS ratio O:E: 0.95 ○ Difference O-E: -4.54% (p<0.01) • 2.1 to 5 cm (n=440) ○ BCSS ratio O:E: 0.95 ○ Difference O-E: -3.53% (ns		

Study	Study details	Population	Findings	Comments	Quality
			• *Subgroup: nodal status* • Negative (n=729) ○ BCSS ratio O:E = 0.96 ○ Difference O-E = -3.53% (p<0.01) • Positive (n=329) ○ BCSS ratio O:E = 0.91 ○ Difference O-E = -6.73% (p<0.01) • *Subgroup: ER status* • Negative (n=259) ○ BCSS ratio O:E = 0.96 ○ Difference O-E = -2.76% (ns • Positive (n=491) ○ BCSS ratio O:E = 0.92 ○ Difference O-E = -6.62% (p<0.01) • Unknown (n=308) ○ BCSS ratio O:E = 0.96 ○ Difference O-E = -2.74% (ns • Tool discrimination • Not reported		

BCSS, breast cancer specific survival; CASP, Critical Appraisal Skills Programme; ER, oestrogen receptor; NGA, National Guideline Alliance; ns, not significant; O-E, observed minus expected; O:E, observed/ expected; OS, overall survival; UK, United Kingdom

Predictive prognostic tool 2: PREDICT

Table 7: Summary of included studies and results for PREDICT

Study	Study details	Population	Findings	Comments	Quality
5-year all-cause mortality [PREDICT v1.0]					
Wishart (2010)	Validation study	Data from 5468 people with breast cancer	• Prognostic accuracy (sensitivity, specificity)	• PREDICT v1.0	Moderate quality, as assessed by

Study	Study details	Population	Findings	Comments	Quality
	Study period: 1999 to 2003	from the West Midlands Cancer Intelligence Unit (WMCIU)	• Not reported • Tool calibration • Total cohort (N=5468) • Mortality ratio O:E = 0.91 • Difference O-E = -1.61% • *Subgroup: age* • <35 (n=108) ○ Mortality ratio O:E = 0.88 ○ Difference O-E = -2.78% • 35 to 49 (n=1195) ○ Mortality ratio O:E = 0.83 ○ Difference O-E = -2.68% • 50 to 67 (n=2393) ○ Mortality ratio O:E = 0.90 ○ Difference O-E = -1.34% • 65 to 74 (n=1101) ○ Mortality ratio O:E = 0.98 ○ Difference O-E = -0.45% • 75+ (n=671) ○ Mortality ratio O:E = 0.98 ○ Difference O-E = -0.75% • *Subgroup: grade* • Grade 1 (n=1017) ○ Mortality ratio O:E = 0.98 ○ Difference O-E = -0.1% • Grade 2 (n=2442) ○ Mortality ratio O:E = 0.98 ○ Difference O-E = -0.16% • Grade 3 (n=2009) ○ Mortality ratio O:E = 0.87 ○ Difference O-E = -3.58% • *Subgroup: tumour size*	• Validation study (data from the primary analysis has not been reported)	CASP Clinical Prediction Rule checklist

Study	Study details	Population	Findings	Comments	Quality

			• <10 mm (n=485)
			  ○ Mortality ratio O:E = 0.84
			  ○ Difference O-E = -1.03%
			• 10 to 19 mm (n=2136)
			  ○ Mortality ratio O:E = 0.88
			  ○ Difference O-E = -2.01%
			• 20 to 29 mm (n=1566)
			  ○ Mortality ratio O:E = 0.94
			  ○ Difference O-E = -0.96%
			• 30 to 49 mm (n=923)
			  ○ Mortality ratio O:E = 0.99
			  ○ Difference O-E = -0.11%
			• 50+ mm (n=358)
			  ○ Mortality ratio O:E = 0.91
			  ○ Difference O-E = -3.91%
			• *Subgroup: nodal status*
			• Negative (n=3184)
			  ○ Mortality ratio O:E = 0.80
			  ○ Difference O-E = -2.14%
			• Positive (n=2284)
			  ○ Mortality ratio O:E = 0.98
			  ○ Difference O-E = -0.39%
			• *Subgroup: ER status*
			• Negative (n=1116)
			  ○ Mortality ratio O:E = 0.87
			  ○ Difference O-E = -4.21%
			• Positive (n=4352)
			  ○ Mortality ratio O:E = 0.95
			  ○ Difference O-E = -0.69%
			• **Tool discrimination**
			• ER+: AUC=0.81; SE=0.0111
			• ER-: AUC=0.75; SE=0.0169		

Study	Study details	Population	Findings	Comments	Quality
5-year all-cause mortality [PREDICT v1.2]					

Maishman (2015)	Validation study
	Study period: 2000 to 2008
		Data from 3000 women aged ≤40 years at diagnosis (POSH cohort)	• **Prognostic accuracy (sensitivity, specificity)**
			• Not reported
			• **Tool calibration**
			• Total cohort (N=2827)
			• Mortality ratio O:E = 1.33
			• Difference O-E = 25% (n=152)
			• *Subgroup: age at diagnosis*
			• 18 to 25 (n=40)
			  ○ Mortality ratio O:E = 1.4
			  ○ Difference O-E = 28.6% (n=2)
			• 26 to 30 (n=258)
			  ○ Mortality ratio O:E = 1.35
			  ○ Difference O-E = 25.8% (n=16)
			• 31 to 35 (n=864)
			  ○ Mortality ratio O:E = 1.38
			  ○ Difference O-E = 27.6% (n=58)
			• 36 to 40 (n=1665)
			  ○ Mortality ratio O:E = 1.30
			  ○ Difference O-E = 23.2% (n=76)
			• *Subgroup: grade*
			• Grade 1 (n=156)
			  ○ Mortality ratio O:E = 1.25
			  ○ Difference O-E = 20% (n=1)
			• Grade 2 (n=929)
			  ○ Mortality ratio O:E = 2.40
			  ○ Difference O-E = 58.4% (n=94)
			• Grade 3 (n=1676)
			  ○ Mortality ratio O:E = 1.13
			  ○ Difference O-E = 11.9% (n=51)
			• Unknown (n=66)
				• PREDICT v1.2
				• Other factors in the tool: menopausal status, morphology, LV invasion, ER status, local treatment, systemic treatment, HER2 status, ethnicity
				• Other outcomes reported: 8-year all-cause mortality
					• High quality, as assessed by CASP Clinical Prediction Rule checklist

Study	Study details	Population	Findings	Comments	Quality
			○ Mortality ratio O:E = 1.71 ○ Difference O-E = 41.7% (n=5) • *Subgroup: tumour size* • 0 to 10 mm (n=265) ○ Mortality ratio O:E = 2.1 ○ Difference O-E = 52.4% (n=22) • 11 to 20 mm (n=930) ○ Mortality ratio O:E = 1.25 ○ Difference O-E = 20% (n=25) • 21 to 50 mm (n=1229) ○ Mortality ratio O:E = 1.26 ○ Difference O-E = 22.8% (n=69) • >50 mm (n=244) ○ Mortality ratio O:E = 1.16 ○ Difference O-E = 14% (n=85) • Unknown (n=159) ○ Mortality ratio O:E = 2.44 ○ Difference O-E = 59% (n=23) • *Subgroup: node status* • Negative (n=1370) ○ Mortality ratio O:E = 1.26 ○ Difference O-E = 20.5% (n=33) • Positive (n=1431) ○ Mortality ratio O:E = 1.35 ○ Difference O-E = 26.2% (n=115) • Unknown (n=26) ○ Mortality ratio O:E = 1.75 ○ Difference O-E = 42.9% (n=3) • *Subgroup: ER status* • Negative (n=965) ○ Mortality ratio O:E = 0.82		

Study	Study details	Population	Findings	Comments	Quality
			○ Difference O-E = -21.2% (n=-52) • Positive (n=1862) ○ Mortality ratio O:E = 2.29 ○ Difference O-E = 56.4% (n=204) • *Subgroup: HER2 status* • Negative (n=1773) ○ Mortality ratio O:E = 1.50 ○ Difference O-E = 33.4% (n=128) • Positive (n=679) ○ Mortality ratio O:E = 1.15 ○ Difference O-E = 13.1% (n=24) • Borderline (n=40) ○ Mortality ratio O:E = 1.67 ○ Difference O-E = 40% (n=4) • Unknown (n=335) ○ Mortality ratio O:E = 0.88 ○ Difference O-E = -12.9% (n=-4) • **Tool discrimination** • AUC ER- vs ER+ = 0.718 vs 0.730		
8-year all-cause mortality [PREDICT 1.0] (proxy for long-term OS)					
Wishart (2010)	Validation study Study period: 1999 to 2003	Data from 5468 people with breast cancer from the West Midlands Cancer Intelligence Unit (WMCIU)	• **Prognostic accuracy (sensitivity, specificity)** • Not reported • **Tool calibration and discrimination** • Total cohort (N=5468): • Mortality ratio O:E = 0.95 • Difference O-E = -0.93% • *Subgroup: age*	• PREDICT v1.0 • Validation study (data from the primary analysis has not been reported) • 10-year all-cause mortality was not reported in the paper. 8-year all-cause mortality was taken as a proxy outcome instead	• Moderate quality, as assessed by CASP Clinical Prediction Rule checklist.

Study	Study details	Population	Findings	Comments	Quality

			• <35 (n=108)
			  ○ Mortality ratio O:E = 1.08
			  ○ Difference O-E = 1.85%
			• 35 to 49 (n=1195)
			  ○ Mortality ratio O:E = 0.87
			  ○ Difference O-E = -2.18%
			• 50 to 67 (n=2393
			  ○ Mortality ratio O:E = 0.92
			  ○ Difference O-E = -1%
			• 65 to 74 (n=1101)
			  ○ Mortality ratio O:E = 1.00
			  ○ Difference O-E = -0.09%
			• 75+ (n=671)
			  ○ Mortality ratio O:E = 0.98
			  ○ Difference O-E = -0.6%
			• *Subgroup: grade*
			• Grade 1 (n=1017)
			  ○ Mortality ratio O:E = 1.04
			  ○ Difference O-E = 0.29%
			• Grade 2 (n=2442)
			  ○ Mortality ratio O:E = 1.04
			  ○ Difference O-E = 0.61%
			• Grade 3 (n=2009)
			  ○ Mortality ratio O:E = 0.88
			  ○ Difference O-E = -3.38%
			• *Subgroup: tumour size*
			• <10 mm (n=485)
			  ○ Mortality ratio O:E = 0.85
			  ○ Difference O-E = -1.03%
			• 10 to 19 mm (n=2136)
			  ○ Mortality ratio O:E = 0.84
			  ○ Difference O-E = -1.73%		

Study	Study details	Population	Findings	Comments	Quality

			• 20 to 29 mm (n=1566)
			  ○ Mortality ratio O:E = 0.97
			  ○ Difference O-E = -0.57%
			• 30 to 49 mm (n=923)
			  ○ Mortality ratio O:E = 0.98
			  ○ Difference O-E = -0.43%
			• 50+ mm (n=358)
			  ○ Mortality ratio O:E = 0.56
			  ○ Difference O-E = -3.35%
			• *Subgroup: nodal status*
			• Negative (n=3184)
			  ○ Mortality ratio O:E = 0.84
			  ○ Difference O-E = -1.76%
			• Positive (n=2284)
			  ○ Mortality ratio O:E = 1.01
			  ○ Difference O-E = 0.26%
			• *Subgroup: ER status*
			• Negative (n=1116)
			  ○ Mortality ratio O:E = 0.90
			  ○ Difference O-E = -3.49%
			• Positive (n=4352)
			  ○ Mortality ratio O:E = 0.98
			  ○ Difference O-E = -0.25%
			• **Tool discrimination (AUC)**
			• Total cohort (N=5468): AUC (SE) = 0.79 (0.008)
			• *Subgroup: age*
			• <35 (n=108); AUC (SE) = 0.70 (0.057)
			• 35 to 49 (n=1195); AUC (SE) = 0.79 (0.018)
			• 50 to 67 (n=2393); AUC (SE) = 0.80 (0.013)		

Study	Study details	Population	Findings	Comments	Quality
			• 65 to 74 (n=1101); AUC (SE) =0.76 (0.018) • 75+ (n=671); AUC (SE) = 0.72 (0.021) • *Subgroup: grade* • Grade 1 (n=1017): AUC (SE) = 0.79 (0.029) • Grade 2 (n=2442): AUC (SE) = 0.77 (0.013) • Grade 3 (n=2009): AUC (SE) = 0.75 (0.012) • *Subgroup: tumour size* • <10 mm (n=485): AUC (SE) = 0.82 (0.040) • 10 to 19 mm (n=2136): AUC (SE) = 0.76 (0.018) • 20 to 29 mm (n=1566): AUC (SE) = 0.71 (0.017) • 30 to 49 mm (n=923): AUC (SE) = 0.72 (0.018) • 50+ mm (n=358): AUC (SE) = 0.72 (0.027) • *Subgroup: nodal status* • Negative (n=3184): AUC (SE) = 0.74 (0.015) • Positive (n=2284): AUC (SE) = 0.75 (0.011) • *Subgroup: ER status* • Negative (n=1116): AUC (SE) = 0.76 (0.016) • Positive (n=4352): AUC (SE) = 0.78 (0.010)		
10-year all-cause mortality [PREDICT v1.2]					

Study	Study details	Population	Findings	Comments	Quality

Maishman (2015)	Validation study
	Study period: 2000 to 2008
		Data from 3000 women aged ≤40 years at diagnosis (POSH cohort)	• **Prognostic accuracy (sensitivity, specificity)**
			• Not reported
			• **Tool calibration**
			• Total cohort (N=597)
			• Mortality ratio O:E = 0.93
			• Difference O-E = -7.9% (n=-12)
			• *Subgroup: age at diagnosis*
			• 18 to 25 (n=8)
			  ○ Mortality ratio O:E = 1
			  ○ Difference O-E = 0% (n=0)
			• 26 to 30 (n=55)
			  ○ Mortality ratio O:E = 0.94
			  ○ Difference O-E = -6.7% (n=-1)
			• 31 to 35 (n=203)
			  ○ Mortality ratio O:E = 1.05
			  ○ Difference O-E = 5% (n=3)
			• 36 to 40 (n=331)
			  ○ Mortality ratio O:E = 0.84
			  ○ Difference O-E = -18.4% (n=-14)
			• *Subgroup: grade*
			• Grade 1 (n=31)
			  ○ Mortality ratio O:E = 1.5
			  ○ Difference O-E = 33% (n=1)
			• Grade 2 (n=200)
			  ○ Mortality ratio O:E = 1.42
			  ○ Difference O-E = 30% (n=13)
			• Grade 3 (n=351)
			  ○ Mortality ratio O:E = 0.80
			  ○ Difference O-E = -25.5% (n=-26)
				• PREDICT v1.2
				• Other factors in the tool: menopausal status, morphology, LV invasion, local treatment, systemic treatment, ethnicity
				• Tool discrimination: very limited data reported (based on ER status and HER2 status)
					• High quality, as assessed by CASP Clinical Prediction Rule checklist

Study	Study details	Population	Findings	Comments	Quality

			• Unknown (n=15)
			  ○ Mortality ratio O:E = 1
			  ○ Difference O-E = 0% (n=0)
			• *Subgroup: tumour size*
			• 0 to 10 (n=48)
			  ○ Mortality ratio O:E = 2
			  ○ Difference O-E = 50% (n=7)
			• 11 to 20 (n=221)
			  ○ Mortality ratio O:E = 0.91
			  ○ Difference O-E = -9.8% (n=-4)
			• 21 to 50 (n=244)
			  ○ Mortality ratio O:E = 0.99
			  ○ Difference O-E = -1.3% (n=-1)
			• >50 (n=54)
			  ○ Mortality ratio O:E = 0.46
			  ○ Difference O-E = -115.4% (n=-15)
			• Unknown (n=30)
			  ○ Mortality ratio O:E = 1.2
			  ○ Difference O-E = 16.7% (n=1)
			• *Subgroup: node status*
			• Negative (n=266)
			  ○ Mortality ratio O:E = 0.93
			  ○ Difference O-E = -7.7% (n=-3)
			• Positive (n=327)
			  ○ Mortality ratio O:E = 0.92
			  ○ Difference O-E = 8% (n=9)
			• Unknown (n=4)
			  ○ Mortality ratio O:E = 1
			  ○ Difference O-E = 0% (n=0)
			• *Subgroup: ER status*
			• Negative (n=231)		

Study	Study details	Population	Findings	Comments	Quality
			○ Mortality ratio O:E = 0.68 ○ Difference O-E = -46.9% (n=-30) • Positive (n=366) ○ Mortality ratio O:E = 1.26 ○ Difference O-E = 20.5% (n=18) • *Subgroup: HER2 status* • Negative (n=327) ○ Mortality ratio O:E = 0.99 ○ Difference O-E = -1.2% (n=-1) • Positive (n=140) ○ Mortality ratio O:E = 0.94 ○ Difference O-E = -6% (n=-3) • Borderline (n=14) ○ Mortality ratio O:E = 1.25 ○ Difference O-E = 20% (n=1) • Unknown (n=116) ○ Mortality ratio O:E = 0.62 ○ Difference O-E = -60% (n=-9) • **Tool discrimination** • AUC ER- vs ER+ = 0.694 vs 0.724 (discrimination was better for ER+ tumours, compared to ER- tumours) • AUC HER2- vs HER2+ =0.724 vs 0.592 (discrimination was better for HER2- tumours, compared to HER2+ tumours)		
10-year breast cancer mortality [PREDICT v1.1]					
Wishart (2014)	Validation study Study period: 1989 to 1998	Data for 2232 cases of invasive breast cancer treated in Nottingham - 506 node-negative cases were excluded,	• **Prognostic accuracy (sensitivity, specificity)** • Not reported • **Tool calibration**	• PREDICT v1.1	• Moderate quality, as assessed by CASP Clinical Prediction Rule checklist

Study	Study details	Population	Findings	Comments	Quality

		so data from n=1726 people was included in the study	• Total cohort (N=1726)
			• BC mortality ratio O:E = 1.13
			• *Subgroup: age*
			• <40 (n=67)
			  ○ BC mortality ratio O:E = 1.15
			• 40 to 49 (n=274)
			  ○ BC mortality ratio O:E = 1.18
			• 50 to 59 (n=436)
			  ○ BC mortality ratio O:E = 1.18
			• 60+ (n=497)
			  ○ BC mortality ratio O:E = 1.06
			• *Subgroup: tumour size*
			• <10 (n=144)
			  ○ BC mortality ratio O:E = 0.78
			• 10 to 19 (n=574)
			  ○ BC mortality ratio O:E = 1.09
			• 20 to 29 (n=404)
			  ○ BC mortality ratio O:E = 1.32
			• 30 to 49 (n=140)
			  ○ BC mortality ratio O:E = 0.95
			• 50+ (n=11)
			  ○ BC mortality ratio O:E = 0.5
			• Missing (n=1)
			  ○ BC mortality ratio O:E = 1
			• *Subgroup: node status*
			• Negative (n=709)
			  ○ BC mortality ratio O:E = 1.19
			• 1+ (n=241)
			  ○ BC mortality ratio O:E = 1.23
			• 2 to 4+ (n=184)
			  ○ BC mortality ratio O:E = 1.05
			• 5 to 9+ (n=37)		

Study	Study details	Population	Findings	Comments	Quality
			○ BC mortality ratio O:E = 1.10 • 10+ (n=6) ○ BC mortality ratio O:E = 0.8 • Missing (n=97) ○ BC mortality ratio O:E = 1.07 • *Subgroup: grade* • Grade 1 (n=235) ○ BC mortality ratio O:E = 1.8 • Grade 2 (n=528) ○ BC mortality ratio O:E = 1.16 • Grade 3 (n=395) ○ BC mortality ratio O:E = 1.14 • Missing grade (n=116) ○ BC mortality ratio O:E = 0.31 • *Subgroup: HER2 status* • Negative (n=792) ○ BC mortality ratio O:E = 1.35 • Positive (n=77) ○ BC mortality ratio O:E = 1.35 • Missing (n=405) ○ BC mortality ratio O:E = 0.44 • **Tool discrimination** • AUC = 0.7611 (CI not reported)		
10-year breast cancer mortality [PREDICT v1.2]					
Wishart (2014)	Validation study Study period: 1989 to 1998	Data for 2232 cases of invasive breast cancer treated in Nottingham - 506 node-negative cases were excluded, so data from n=1726 people was included in the study	• **Prognostic accuracy (sensitivity, specificity)** • Not reported • **Tool calibration** • Total cohort (N=1726) • BC mortality ratio O:E = 1.08 • *Subgroup: age* • <40 (n=67)	• PREDICT v1.2	• High quality, as assessed by CASP Clinical Prediction Rule checklist

Study	Study details	Population	Findings	Comments	Quality
			○ BC mortality ratio O:E = 1.07 • 40 to 49 (n=274) ○ BC mortality ratio O:E = 1.13 • 50 to 59 (n=436) ○ BC mortality ratio O:E = 1.15 • 60+ (n=497) ○ BC mortality ratio O:E = 1.01 • *Subgroup: tumour size* • <10 (n=144) ○ BC mortality ratio O:E = 0.78 • 10 to 19 (n=574) ○ BC mortality ratio O:E = 1.05 • 20 to 29 (n=404) ○ BC mortality ratio O:E = 1.26 • 30 to 49 (n=140) ○ BC mortality ratio O:E = 0.91 • 50+ (n=11) ○ BC mortality ratio O:E = 0.5 • Missing (n=1) ○ BC mortality ratio O:E = 1 • *Subgroup: grade* • Grade 1 (n=235) ○ BC mortality ratio O:E = 1.8 • Grade 2 (n=528) ○ BC mortality ratio O:E = 1.14 • Grade 3 (n=395) ○ BC mortality ratio O:E = 1.07 • Missing grade (n=116) ○ BC mortality ratio O:E = 0.31 • *Subgroup: node status* • Negative (n=709) ○ BC mortality ratio O:E = 1.15		

Study	Study details	Population	Findings	Comments	Quality

			• 1+ (n=241)
			  ○ BC mortality ratio O:E = 1.17
			• 2 to 4+ (n=184)
			  ○ BC mortality ratio O:E = 1
			• 5 to 9+ (n=37)
			  ○ BC mortality ratio O:E = 1.05
			• 10+ (n=6)
			  ○ BC mortality ratio O:E = 0.8
			• Missing (n=97)
			  ○ BC mortality ratio O:E = 1.07
			• *Subgroup: HER2 status*
			• Negative (n=792)
			  ○ BC mortality ratio O:E = 1.29
			• Positive (n=77)
			  ○ BC mortality ratio O:E = 1.24
			• Missing (n=405)
			  ○ BC mortality ratio O:E = 0.44
			• **Tool discrimination**
			• AUC = 0.7676 (CI not reported) (significant improvement compared to v1.1 (p-value = 0.0008) (see Wishart 2014, data for v1.1.)		

10-year breast cancer mortality [PREDICT v2.0]					
Candido dos Reis (2017)	Validation study Validations study Study period: • BCOS: 1990 to 2000 • NTBCS: 1989 to 1998 • POSH: 2000 to 2008	Validations study Study period: • BCOS: n=981 • NTBCS: n=1726 • POSH: n=2609	• **Prognostic accuracy (sensitivity, specificity)** • Not reported • **Tool calibration** • Total cohort: not reported • *Subgroup: age at diagnosis* • *ER-* • 20 to 29 (n=92):	• Validation study (data from the primary analysis has not been reported)	• High quality, as assessed by CASP Clinical Prediction Rule checklist

Study	Study details	Population	Findings	Comments	Quality
			○ BC mortality ratio = 0.94; ○ difference O:E = -6% (p-value = 0.76) • 30 to 39 (n=855): ○ BC mortality ratio = 0.92; ○ difference O:E = -9% (p-value = 0.18) • 40 to 49 (n=414): ○ BC mortality ratio = 0.98; ○ difference O:E = -2% (p-value = 0.83) • 50 to 59 (n=165): ○ BC mortality ratio = 0.97; ○ difference O:E = -3% (p-value = 0.85) • 60 to 69 (n=117): ○ BC mortality ratio = 0.82; ○ difference O:E = -21% (p-value = 0.32) • 70 to 79 (n=11): ○ BC mortality ratio = 0.36; ○ difference O:E = -180% (p-value = 0.28) • *ER+* • 20 to 29 (n=140): ○ BC mortality ratio = 0.71; ○ difference O:E = -40% (p-value = 0.047) • 30 to 39 (n=1633): ○ BC mortality ratio = 0.96; ○ difference O:E = -4% (p-value = 0.48) • 40 to 49 (n=1063): ○ BC mortality ratio = 0.90;		

Study	Study details	Population	Findings	Comments	Quality
			○ difference O:E = -11% (p-value = 0.16) • 50 to 59 (n=467): ○ BC mortality ratio = 0.96; ○ difference O:E = -4% (p-value = 0.77) • 60 to 69 (n=517): ○ BC mortality ratio = 1.08; ○ difference O:E = 7% (p-value = 0.53) • 70 to 79 (n=55): ○ BC mortality ratio = 0.38; ○ difference O:E = -26% (p-value = 0.54) • *Subgroup: tumour size* • *ER-* • 0 to 9 mm (n=96): ○ BC mortality ratio = 0.90; ○ difference O:E = -10% (p-value = 0.73) • 10 to 19 mm (n=559): ○ BC mortality ratio = 0.92; ○ difference O:E = -8% (p-value = 0.41) • 20 to 29 mm (n=524): ○ BC mortality ratio = 0.97; ○ difference O:E = -3% (p-value = 0.72) • 30 to 49 mm (n=354): ○ BC mortality ratio = 0.99; ○ difference O:E = -1% (p-value = 0.91) • 50+ mm (n=121): ○ BC mortality ratio = 0.75;		

Study	Study details	Population	Findings	Comments	Quality
			○ difference O:E = -33% (p-value = 0.04) • *ER+* • 0 to 9 mm (n=352): ○ BC mortality ratio = 1.54; ○ difference O:E = 35% (p-value = 0.024) • 10 to 19 mm (n=1428): ○ BC mortality ratio = 1.06; ○ difference O:E = 6% (p-value = 0.46) • 20 to 29 mm (n=1111): ○ BC mortality ratio = 0.98; ○ difference O:E = -2% (p-value = 0.80) • 30 to 49 mm (n=695): ○ BC mortality ratio = 0.87; ○ difference O:E = -15% (p-value = 0.07) • 50+ mm (n=289): ○ BC mortality ratio = 0.74; ○ difference O:E = -35% (p-value = 0.00) • *Subgroup: tumour grade* • *ER-* • 1 (n=44): ○ BC mortality ratio = 0.96; ○ difference O:E = -4% (p-value = 0.91) • 2 (n=183): ○ BC mortality ratio = 0.86; ○ difference O:E = -17% (p-value = 0.33) • 3 (n=1427):		

Study	Study details	Population	Findings	Comments	Quality
			○ BC mortality ratio = 0.94; ○ difference O:E = -7% (p-value = 0.19) • *ER+* • 1 (n=658): ○ BC mortality ratio = 0.86; ○ difference O:E = -16% (p-value = 0.43) • 2 (n=1730): ○ BC mortality ratio = 0.95; ○ difference O:E = -5% (p-value = 0.44) • 3 (n=1487): ○ BC mortality ratio = 0.93; ○ difference O:E = -7% (p-value = 0.17) • *Subgroup: nodes positive* • *ER-* • 0 (n=937): ○ BC mortality ratio = 1.01; ○ difference O:E = 1% (p-value = 0.89) • 1 (n=232): ○ BC mortality ratio = 0.86; ○ difference O:E = -17% (p-value = 0.23) • 2 to 4 (n=300): ○ BC mortality ratio = 0.88; ○ difference O:E = -13% (p-value = 0.19) • 5 to 9 (n=101): ○ BC mortality ratio = 0.96; ○ difference O:E = -4% (p-value = 0.77)		

Study	Study details	Population	Findings	Comments	Quality

			• 10+ (n=84):
			  ○ BC mortality ratio = 0.85;
			  ○ difference O:E = -17% (p-value = 0.28)
			• *ER+*
			• 0 (n=2085):
			  ○ BC mortality ratio = 0.99;
			  ○ difference O:E = -1% (p-value = 0.85)
			• 1 (n=675):
			  ○ BC mortality ratio = 0.92;
			  ○ difference O:E = -9% (p-value = 0.39)
			• 2 to 4 (n=734):
			  ○ BC mortality ratio = 0.96;
			  ○ difference O:E = -4% (p-value = 0.63)
			• 5 to 9 (n=245):
			  ○ BC mortality ratio = 0.86;
			  ○ difference O:E = -17% (p-value = 0.14)
			• 10+ (n=136):
			  ○ BC mortality ratio = 0.87;
			  ○ difference O:E = -15% (p-value = 0.25)
			• **Tool discrimination**
			• ER-: AUC = 0.696
			• ER+: AUC = 0.760
			• All population: AUC = 0.752		

AUC, area under the curve; BC, breast cancer; BCOS, Breast Cancer Outcomes Simulator BCS, breast cancer survival; CASP, Critical Appraisal Skills Programme; ER, oestrogen receptor; HER2, Human epidermal growth factor receptor 2; LV; lymphovascular; NGA, National Guideline Alliance; NTBC, Nottingham Tenovus Breast Cancer; O:E, observed/expected; POSH, Prospective study of Outcomes in Sporadic versus Hereditary breast cancer; SE, standard error; UK, United Kingdom; WMCIU, West Midlands Cancer Intelligence Unit

Predictive prognostic tool 3: Nottingham Prognostic Index (NPI)

Table 8: Summary of included studies and results for Nottingham Prognostic Index (NPI)

Study	Study details	Population	Findings	Comments	Quality
10 year breast cancer survival (BCS)					
Blamey (2007)	Validation study Study period: • 1980 to 1986 (n=892) • 1990 to 1999 (n=2238)	Women diagnosed with or treated for primary operable invasive breast cancer at Nottingham city hospital Based on the NPI score, women were allocated to the following 6 categories: • EPG: excellent prognostic group • GPG: good prognostic group • MPG I: moderate prognostic group I • MPG II: moderate prognostic group II • PPG: poor prognostic group • VPG: very poor prognostic group	**Prognostic accuracy (sensitivity, specificity)** Not reported **Tool calibration** Results only available for the 2000 to 2009 cohort Total cohort = not reported EPG (n=320) • BCS ratio O:E = 0.98 GPG (n=475) • BCS ratio O:E = 0.99 MPG I (n=634) • BCS ratio O:E = 1.03 MPG II (n=489) • BCS ratio O:E = 1.00 PPG (n=233) • BCS ratio O:E = 1.02 VPG (n=86) • BCS ratio O:E = 0.89 • **Tool discrimination** Not reported	• Observed survival was adjusted by subtracting expected number of deaths for all causes was subtracted. Data was obtained from the Office of National Statistics for England and Wales • BCS ratio O:E was calculated by the NGA technical team based on available data on the paper	• High quality, as assessed by CASP Clinical Prediction Rule checklist.

BCS: breast cancer survival; CASP: Critical Appraisal Skills Programme; NGA: National Guideline Alliance; NPI: Nottingham Prognostic Index; O:E: observed: expected; UK: United Kingdom

Predictive prognostic tool 4: FinProg

No studies were identified for this prognostic tool.

Predictive prognostic tool 5: CancerMath

No studies were identified for this prognostic tool.

Predictive prognostic tool 6: Oxford Prognostic Index (OPI)

Table 9: Summary of included studies and results for Oxford Prognostic Index (OPI)

Study	Study details	Population	Findings	Comments	Quality
5 year recurrence free survival (RFS)					
Campbell (2010)	Validation study Study period: 1992 to 2000	N=1787 women with invasive ductal carcinoma, a sub-set obtained from the Adjuvant Breast Cancer trial from 70 UK centres	**Prognostic accuracy (sensitivity, specificity)** Not reported **Tool calibration** Total cohort (N=1789) ○ RFS ratio O:E:1.01 ○ Difference O-E: 0.7% *Subgroup: age* • ≤50 years (n=1097) ○ RFS ratio O:E: 1.03 ○ Difference O-E: 1.92% • > 50 years (n=690) ○ RFS ratio O:E: 1.00 ○ Difference O-E: -0.10% *Subgroup: tumour grade* • Grade 1 (n=196) ○ RFS ratio O:E: 1.06 ○ Difference O-E: 5.15% • Grade 2 (n=772) ○ RFS ratio O:E: 1.03 ○ Difference O-E: 2.44% • Grade 3 (n=819) ○ RFS ratio O:E: 0.98 ○ Difference O-E: -1.04% *Subgroup: tumour size*	• Study provided data as E:O. The NGA technical team has calculated the RFS ratio O:E and the difference O-E • No other factors were in the tool	• Moderate quality, as assessed by CASP Clinical Prediction Rule checklist.

Study	Study details	Population	Findings	Comments	Quality
			• ≤2 cm (n=954) ○ RFS ratio O:E: 1.06 ○ Difference O-E: 4.6% • >2 cm to ≤5 cm (n=772) ○ RFS ratio O:E:0.95 ○ Difference O-E: -3.16% • >5 cm (n=61) ○ RFS ratio O:E: 1.04 ○ Difference O-E: -2.47% *Subgroup: nodal status* • Negative (n=674) ○ RFS ratio O:E: 1.02 ○ Difference O-E: 1.82 • Positive (n=1113) ○ RFS ratio O:E: 1.01 ○ Difference O-E: 0.71% *Subgroup: ER status* • Negative (n=1097) ○ RFS ratio O:E: 1.03 ○ Difference O-E: 1.92% • Positive (n=690) ○ RFS ratio O:E: 1.00 ○ Difference O-E: -0.10% • **Tool discrimination** • Overall C = 0.720 (95%CI 0.693 to 0.746)		

BCS, breast cancer survival; NGA, National Guideline Alliance; O:E, observed/ expected; OPI, Oxford Prognostic Index RFS, recurrence free survival; UK, United Kingdom

Full GRADE tables are not available as GRADE is not appropriate to assess the quality of evidence for prediction model performance reviews.

Economic evidence

A systematic review of the economic literature was conducted but no relevant studies were identified which were applicable to this review question. Economic modelling was not undertaken for this question because other topics were agreed as higher priorities for economic evaluation.

Evidence statements

Predictive prognostic tool 1: Adjuvant! Online

Critical outcomes

Tool calibration - 10-year breast cancer specific survival

There was good quality evidence from 1 validation study conducted in the UK with 1065 (data available for n=1058) women with early breast cancer that reported the following:

- For the whole cohort (N=1058), Adjuvant! Online showed good calibration (O:E = 0.95). The tool overestimated 10-year breast cancer specific survival by 4.53% (p-value <0.01).

The study also reported the calibration according to different factors:

Subgroup: age

- for women aged 20 to 35 (n=34), Adjuvant! Online showed good calibration (O:E = 0.99). This tool overestimated 10-year breast cancer specific survival by 0.67% (p-value: ns);
- for women aged 36 to 50 (n=361), Adjuvant! Online showed good calibration (O:E = 0.94). This tool overestimated 10-year breast cancer specific survival by 4.62% (p-value <0.05);
- for women aged 51 to 65 (n=454), Adjuvant! Online showed good calibration (O:E = 0.96). This tool overestimated 10-year breast cancer specific survival by 3.51% (p-value: ns);
- for women aged 66 to 75 (n=193), Adjuvant! Online showed good calibration (O:E = 0.89). This tool underestimated 10-year breast cancer specific survival by 7.04% (p-value <0.05);
- for women aged ≥76 (n=16), Adjuvant! Online showed good calibration (O:E = 0.94). This tool overestimated 10-year breast cancer specific survival by 3.11% (p-value: n.s).

Subgroup: tumour grade

- for women with grade 1 breast cancer (n=152), Adjuvant! Online showed good calibration (O:E = 0.99). This tool overestimated 10-year breast cancer specific survival by 1.29% (p-value: ns);
- for women with grade 2 breast cancer (n=420), Adjuvant! Online showed good calibration (O:E = 0.93). This tool overestimated 10-year breast cancer specific survival by 5.89% (p-value <0.01);
- for women with grade 3 breast cancer (n=243), Adjuvant! Online showed good calibration (O:E = 0.92). This tool overestimated 10-year breast cancer specific survival by 6.10% (p-value <0.05);
- for women with unknown grade (n=243), Adjuvant! Online showed good calibration (O:E = 0.96). This tool overestimated 10-year breast cancer specific survival by 2.78% (p-value: ns).

Subgroup: tumour size

- for women with tumour size 0.1 to 1 cm (n=148), Adjuvant! Online showed good calibration (O:E = 0.92). This tool overestimated 10-year breast cancer specific survival by 7.95% (p-value <0.01);

- for women with tumour size 1.1 to 2 cm (n=470), Adjuvant! Online showed good calibration (O:E = 0.95). This tool overestimated 10-year breast cancer specific survival by 4.54% (p-value <0.01);
- for women with tumour size 2.1 to 5 cm (n=440), Adjuvant! Online showed good calibration (O:E = 0.95). This tool overestimated 10-year breast cancer specific survival by 3.53% (p-value: n.s).

Subgroup: nodal status

- for women with negative nodal status (n=729), Adjuvant! Online showed good calibration (O:E = 0.96). This tool overestimated 10-year breast cancer specific survival by 3.53% (p-value <0.01);
- for women with positive nodal status (n=329), Adjuvant! Online showed good calibration (O:E = 0.91). This tool overestimated 10-year breast cancer specific survival by 6.73% (p-value <0.01).

Subgroup: ER status

- for women with negative ER status (n=259), Adjuvant! Online showed good calibration (O:E = 0.97). This tool overestimated 10-year breast cancer specific survival by 2.76% (p-value: n.s);
- for women with positive ER status (n=491), Adjuvant! Online showed good calibration (O:E = 0.89). This tool overestimated 10-year breast cancer specific survival by 6.62% (p-value <0.01);
- for women with unknown ER status (n=308), Adjuvant! Online showed good calibration (O:E = 0.96). This tool overestimated 10-year breast cancer specific survival by 2.74% (p-value: n.s).

Tool calibration - 10-year overall survival

There was good quality evidence from 1 validation study conducted in the UK with 1065 women with early breast cancer that reported the following:

- For the whole cohort (N=1065), Adjuvant! Online showed good calibration (O:E = 0.93). The tool overestimated 10-year overall survival by 5.54% (p-value <0.01).

The study also reported the tool calibration according to different factors:

Subgroup: age

- for women aged 20 to 35 (n=34), Adjuvant! Online showed good calibration (O:E = 0.97). This tool overestimated 10-year overall survival by 2.27% (p-value: n.s);
- for women aged 36 to 50 (n=363), Adjuvant! Online showed good calibration (O:E = 0.95). This tool overestimated 10-year overall survival by 4.33% (p-value <0.05);
- for women aged 51 to 65 (n=458), Adjuvant! Online showed good calibration (O:E = 0.95). This tool overestimated 10-year overall survival by 4.02% (p-value <0.05);
- for women aged 66 to 75 (n=194), Adjuvant! Online showed good calibration (O:E = 0.82). This tool overestimated 10-year overall survival by 11.17% (p-value <0.01);
- for women aged ≥76 (n=16), Adjuvant! Online showed good calibration (O:E = 0.94). This tool underestimated 10-year overall survival by 3.11% (p-value: ns).

Subgroup: tumour grade

- for women with grade 1 breast cancer (n=152), Adjuvant! Online showed good calibration (O:E = 0.96). This tool overestimated 10-year overall survival by 3.6% (p-value: ns);
- for women with grade 2 breast cancer (n=421), Adjuvant! Online showed good calibration (O:E = 0.91). This tool overestimated 10-year overall survival by 7.0% (p-value <0.01);
- for women with grade 3 breast cancer (n=248), Adjuvant! Online showed good calibration (O:E = 0.86). This tool overestimated 10-year overall survival by 9.8% (p-value <0.01);

- for women with unknown grade (n=244), Adjuvant! Online showed perfect calibration (O:E = 1.00). This tool overestimated 10-year overall survival by 0.2% (p-value: ns).

Subgroup: tumour size

- for women with tumour size 0.1 to 1 cm (n=150), Adjuvant! Online showed good calibration (O:E = 0.93). This tool overestimated 10-year overall survival by 6.10% (p-value: ns);
- for women with tumour size 1.1 to 2 cm (n=471), Adjuvant! Online showed good calibration (O:E = 0.92). This tool overestimated 10-year overall survival by 6.57% (p-value <0.01);
- for women with tumour size 2.1 to 5 cm (n=444), Adjuvant! Online showed good calibration (O:E = 0.94). This tool overestimated 10-year overall survival survival by 4.26% (p-value: ns).

Subgroup: nodal status

- for women with negative nodal status (n=733), Adjuvant! Online showed good calibration (O:E = 0.94). This tool overestimated 10-year overall survival by 4.70% (p-value <0.01);
- for women with positive nodal status (n=332), Adjuvant! Online showed good calibration (O:E = 0.89). This tool overestimated 10-year overall survival by 7.38% (p-value <0.01).

Subgroup: ER status

- for women with negative ER status (n=261), Adjuvant! Online showed good calibration (O:E = 0.97). This tool overestimated 10-year overall survival by 6.57% (p-value: ns);
- for women with positive ER status (n=495), Adjuvant! Online showed good calibration (O:E = 0.89). This tool overestimated 10-year overall survival by 4.26% (p-value <0.01);
- for women with unknown ER status (n=309), Adjuvant! Online showed good calibration (O:E = 0.96). This tool overestimated 10-year overall survival by 4.26% (p-value: ns).

Tool discrimination

- No evidence was found for this outcome.

Important outcomes

Prognostic accuracy (sensitivity, specificity)

- No evidence was found for this outcome.

Predictive prognostic tool 2: PREDICT

Critical outcomes

Tool calibration and discrimination - 5 year all-cause mortality [PREDICT v1.0]

There was moderate quality evidence from 1 validation study conducted in the UK with 5648 women diagnosed with invasive breast cancer that reported the following:

- for the whole cohort (N=5468), PREDICT v1.0 showed good calibration (O:E = 0.91). The tool overestimated the number of deaths at 5 years by 1.61%.

The study also reported the tool calibration and discrimination according to different factors:

Subgroup: age at diagnosis

- for the women aged <35 (n=108), PREDICT v1.0 showed good calibration (O:E = 0.88). The tool overestimated the number of deaths at 5 years by 2.78%;
- for the women aged 35 to 49 (n=1195), PREDICT v1.0 showed good calibration (O:E = 0.83). The tool overestimated the number of deaths at 5 years by 2.68%;

- for the women aged 50 to 67 (n=2393), PREDICT v1.0 showed good calibration (O:E = 0.90). The tool overestimated the number of deaths at 5 years by 1.34%;
- for the women aged 65 to 74 (n=1101), PREDICT v1.0 showed good calibration (O:E = 0.98). The tool overestimated the number of deaths at 5 years by 0.45%;
- for the women aged 75+ (n=671), PREDICT v1.0 showed good calibration (O:E = 0.98). The tool overestimated the number of deaths at 5 years by 0.75%.

Subgroup: tumour grade

- for women with grade 1 tumour (n=1017), PREDICT v1.0 showed good calibration (O:E = 0.98). The tool overestimated the number of deaths at 5 years by 0.1%;
- for women with grade 2 tumour (n=2442), PREDICT v1.0 showed good calibration (O:E = 0.98). The tool overestimated the number of deaths at 5 years by 0.16%;
- for women with grade 3 tumour (n=2009), PREDICT v1.0 showed good calibration (O:E = 0.87). The tool overestimated the number of deaths at 5 years by 3.58%.

Subgroup: tumour size

- for women with tumours <10 mm. (n=485), PREDICT v1.0 showed good calibration (O:E = 0.84). The tool overestimated the number of deaths at 5 years by 1.03%;
- for women with tumours 11 to 19 mm. (n=2136), PREDICT v1.0 showed good calibration (O:E = 0.88). The tool overestimated the number of deaths at 5 years by 2.01%;
- for women with tumours 20 to 19 mm. (n=1566), PREDICT v1.0 showed good calibration (O:E = 0.94). The tool overestimated the number of deaths at 5 years by 0.96%;
- for women with tumours 30 to 49 mm. (n=923), PREDICT v1.0 showed good calibration (O:E = 0.99). The tool overestimated the number of deaths at 5 years by 0.11%;
- for women with tumours 50+ mm. (n=358), PREDICT v1.0 showed good calibration (O:E = 0.91). The tool overestimated the number of deaths at 5 years by 3.91%.

Subgroup: nodal status

- for women with negative nodal status (n=3184), PREDICT v1.0 showed good calibration (O:E = 0.80). The tool overestimated the number of deaths at 5 years by 2.14%;
- for women with positive nodal status (n=2284), PREDICT v1.0 showed good calibration (O:E = 0.98). The tool overestimated the number of deaths at 5 years by 0.39%.

Subgroup: ER status

- for women with negative ER status (n=1116), PREDICT v1.0 showed good calibration (O:E = 0.87). The tool overestimated the number of deaths at 5 years by 4.21%. The tool also showed good discrimination (AUC = 0.81);
- for women with positive ER status (n=4352), PREDICT v1.0 showed good calibration (O:E = 0.95). The tool overestimated the number of deaths at 5 years by 0.69%. The tool also showed good discrimination (AUC = 0.75).

Tool calibration - 5 year all-cause mortality [PREDICT v1.2]

There was good quality evidence from 1 validation study conducted in the UK with 3000 women aged ≤40 years at diagnosis that reported the following:

- for the whole cohort (N=2827), PREDICT v1.2 showed poor calibration (O:E = 1.33). The tool overestimated the number of deaths at 5 years by 25%. The tool also showed poor discrimination for both ER- (AUC=0.718) and ER+ and (AUC=0.730) groups.

The study also reported the calibration according to different factors:

Subgroup: age at diagnosis

- for the women aged 18 to 25 (n=40), PREDICT v1.2 showed poor calibration (O:E = 1.4). The tool underestimated the number of deaths at 5 years by 28.6%;
- for the women aged 26 to 30 (n=258), PREDICT v1.2 showed poor calibration (O:E = 1.35). The tool underestimated the number of deaths at 5 years by 25.8%;
- for the women aged 31 to 35 (n=864), PREDICT v1.2 showed poor calibration (O:E = 1.38). The tool underestimated the number of deaths at 5 years by 27.6%;
- for the women aged 36 to 40 (n=1665), PREDICT v1.2 showed poor calibration (O:E = 1.30). The tool underestimated the number of deaths at 5 years by 23.2%.

Subgroup: tumour grade

- for women with grade 1 tumour (n=156), PREDICT v1.2 showed poor calibration (O:E = 1.25). The tool underestimated the number of deaths at 5 years by 20%;
- for women with grade 2 tumour (n=929), PREDICT v1.2 showed poor calibration (O:E = 2.40). The tool underestimated the number of deaths at 5 years by 58.4%;
- for women with grade 3 tumour (n=1676), PREDICT v1.2 showed good calibration (O:E = 1.13). The tool underestimated the number of deaths at 5 years by 11.9%;
- for women with unknown grade tumour (n=66), PREDICT v1.2 showed poor calibration (O:E = 1.71). The tool underestimated the number of deaths at 5 years by 41.7%.

Subgroup: tumour size

- for women with tumours 0 to 10 mm. (n=265), PREDICT v1.2 showed poor calibration (O:E = 2.1). The tool underestimated the number of deaths at 5 years by 52.4%;
- for women with tumours 11 to 20 mm. (n=930), PREDICT v1.2 showed poor calibration (O:E = 1.25). The tool underestimated the number of deaths at 5 years by 20%;
- for women with tumours 21 to 50 mm. (n=1229), PREDICT v1.2 showed poor calibration (O:E = 1.26). The tool underestimated the number of deaths at 5 years by 22.8%;
- for women with tumours >50 mm. (n=244), PREDICT v1.2 showed good calibration (O:E = 1.16). The tool underestimated the number of deaths at 5 years by 14%.
- for women with unknown size tumours (n=159), PREDICT v1.2 showed poor calibration (O:E = 2.44). The tool underestimated the number of deaths at 5 years by 59%.

Subgroup: nodal status

- for women with negative nodal status (n=1370), PREDICT v1.2 showed poor calibration (O:E = 1.26). The tool underestimated the number of deaths at 5 years by 20.5%;
- for women with positive nodal status (n=1431), PREDICT v1.2 showed poor calibration (O:E = 1.35). The tool underestimated the number of deaths at 5 years by 26.2%;
- for women with unknown nodal status (n=26), PREDICT v1.2 showed poor calibration (O:E = 1.75). The tool underestimated the number of deaths at 5 years by 42.9%.

Subgroup: ER status

- for women with negative ER status (n=965), PREDICT v1.2 showed good calibration (O:E = 0.82). The tool overestimated the number of deaths at 5 years by 21.2%;
- for women with unknown ER status (n=1862), PREDICT v1.2 showed poor calibration (O:E = 2.29). The tool underestimated the number of deaths at 5 years by 56.4%.

Subgroup: HER2 status

- for women with negative HER2 status (n=1773), PREDICT v1.2 showed poor calibration (O:E = 1.50). The tool underestimated the number of deaths at 5 years by 33.4%;
- for women with positive HER2 status (n=679), PREDICT v1.2 showed good calibration (O:E = 1.15). The tool underestimated the number of deaths at 5 years by 13.1%;

- for women with borderline HER2 status (n=40), PREDICT v1.2 showed poor calibration (O:E = 1.67). The tool underestimated the number of deaths at 5 years by 40%;
- for women with unknown HER2 status (n=335), PREDICT v1.2 showed good calibration (O:E = 0.88). The tool overestimated the number of deaths at 5 years by 12.9%.

Tool calibration and discrimination - 8 year all-cause mortality [PREDICT v1.0] (proxy outcome for long term all-cause mortality)

There was good moderate evidence from 1 validation study conducted in the UK with 5648 women diagnosed with invasive breast cancer that reported the following:

- for the whole cohort (N=5468), PREDICT v1.0 showed good calibration (O:E = 0.95) and good discrimination (AUC = 0.79). The tool overestimated the number of deaths at 8 years by 0.93%.

The study also reported the tool calibration according to different factors:

Subgroup: age at diagnosis

- for the women aged <35 (n=108), PREDICT v1.0 showed good calibration (O:E = 1.08) but poor discrimination (AUC = 0.70). The tool underestimated the number of deaths at 8 years by 1.85%;
- for the women aged 35 to 49 (n=1195), PREDICT v1.0 showed good calibration (O:E = 0.87) and good discrimination (AUC = 0.79). The tool overestimated the number of deaths at 8 years by 2.18%;
- for the women aged 50 to 67 (n=2393), PREDICT v1.0 showed good calibration (O:E = 0.92) and good discrimination (AUC = 0.80). The tool overestimated the number of deaths at 8 years by 0.09%;
- for the women aged 65 to 74 (n=1101), PREDICT v1.0 showed good calibration (O:E ≈ 1.00) and good discrimination (AUC = 0.76). The tool underestimated the number of deaths at 8 years by 0.45%;
- for the women aged 75+ (n=671), PREDICT v1.0 showed good calibration (O:E = 0.98) but poor discrimination (AUC = 0.72). The tool overestimated the number of deaths at 8 years by 0.6%.

Subgroup: tumour grade

- for women with grade 1 tumour (n=1017), PREDICT v1.0 showed good calibration (O:E = 1.04) and good discrimination (AUC = 0.79). The tool underestimated the number of deaths at 8 years by 0.29%;
- for women with grade 2 tumour (n=2442), PREDICT v1.0 showed good calibration (O:E = 1.04) and good discrimination (AUC = 0.77). The tool underestimated the number of deaths at 8 years by 0.61%;
- for women with grade 3 tumour (n=2009), PREDICT v1.0 showed good calibration (O:E = 0.88) and good discrimination (AUC = 0.75). The tool overestimated the number of deaths at 8 years by 3.38%.

Subgroup: tumour size

- for women with tumours <10 mm. (n=485), PREDICT v1.0 showed good calibration (O:E = 0.85) and good discrimination (AUC = 0.82). The tool overestimated the number of deaths at 8 years by 1.03%;
- for women with tumours 11 to 19 mm. (n=2136), PREDICT v1.0 showed good calibration (O:E = 0.84) and good discrimination (AUC = 0.76). The tool overestimated the number of deaths at 8 years by 1.73%;

- for women with tumours 20 to 19 mm. (n=1566), PREDICT v1.0 showed good calibration (O:E = 0.97) but poor discrimination (AUC = 0.71). The tool overestimated the number of deaths at 8 years by 0.57%;
- for women with tumours 30 to 49 mm. (n=923), PREDICT v1.0 showed good calibration (O:E = 0.98) but poor discrimination (AUC = 0.72). The tool overestimated the number of deaths at 8 years by 0.43%;
- for women with tumours 50+ mm. (n=358), PREDICT v1.0 showed poor calibration (O:E = 0.56) and poor discrimination (AUC = 0.72). The tool overestimated the number of deaths at 8 years by 3.35%.

Subgroup: nodal status

- for women with negative nodal status (n=3184), PREDICT v1.0 showed good calibration (O:E = 0.84) but poor discrimination (AUC = 0.74). The tool overestimated the number of deaths at 8 years by 1.76%;
- for women with positive nodal status (n=2284), PREDICT v1.0 showed good calibration (O:E = 1.01) and good discrimination (AUC = 0.75). The tool underestimated the number of deaths at 8 years by 0.26%.

Subgroup: ER status

- for women with negative ER status (n=1116), PREDICT v1.0 showed good calibration (O:E = 0.90) and good discrimination (AUC = 0.76). The tool overestimated the number of deaths at 8 years by 3.49%;
- for women with positive ER status (n=4352), PREDICT v1.0 showed good calibration (O:E = 0.98) and good discrimination (AUC = 0.78). The tool overestimated the number of deaths at 8 years by 0.25%.

Tool calibration and discrimination - 10 year all-cause mortality [PREDICT v1.2]

There was good quality evidence from 1 validation study conducted in the UK with 3000 women aged ≤40 years at diagnosis that reported the following:

- for the whole cohort (N=597), PREDICT v1.2 showed good calibration (O:E = 0.93). The tool overestimated the number of deaths at 10 years by 7.9%.

The study also reported the tool calibration and discrimination according to different factors:

Subgroup: age at diagnosis

- for the women aged 18 to 25 (n=8), PREDICT v1.2 showed perfect calibration (O:E = 1);
- for the women aged 26 to 30 (n=55), PREDICT v1.2 showed good calibration (O:E = 0.94). The tool overestimated the number of deaths at 10 years by 6.7%;
- for the women aged 31 to 35 (n=203), PREDICT v1.2 showed good calibration (O:E = 1.05). The tool underestimated the number of deaths at 10 years by 5%;
- for the women aged 36 to 40 (n=331), PREDICT v1.2 showed good calibration (O:E = 0.84). The tool overestimated the number of deaths at 10 years by 18.4%.

Subgroup: tumour grade

- for women with grade 1 tumour (n=31), PREDICT v1.2 showed poor calibration (O:E = 1.5). The tool underestimated the number of deaths at 10 years by 33%;
- for women with grade 2 tumour (n=200), PREDICT v1.2 showed poor calibration (O:E = 1.42). The tool underestimated the number of deaths at 10 years by 30%;
- for women with grade 3 tumour (n=351), PREDICT v1.2 showed good calibration (O:E = 0.80). The tool overestimated the number of deaths at 10 years by 25.5%;
- for women with unknown grade tumour (n=15), PREDICT v1.2 showed perfect calibration (O:E = 1).

Subgroup: tumour size

- for women with tumours 0 to 10 mm. (n=48), PREDICT v1.2 showed poor calibration (O:E = 2). The tool underestimated the number of deaths at 10 years by 50%;
- for women with tumours 11 to 20 mm. (n=221), PREDICT v1.2 showed good calibration (O:E = 0.91). The tool overestimated the number of deaths at 10 years by 9.8%;
- for women with tumours 21 to 50 mm. (n=244), PREDICT v1.2 showed good calibration (O:E = 0.99). The tool overestimated the number of deaths at 10 years by 1.3%;
- for women with tumours >50 mm. (n=54), PREDICT v1.2 showed poor calibration (O:E = 0.46). The tool overestimated the number of deaths at 10 years by 115.4%;
- for women with unknown size tumours (n=30), PREDICT v1.2 showed good calibration (O:E = 1.2). The tool underestimated the number of deaths at 5 years by 16.7%.

Subgroup: node status

- for women with negative nodal status (n=266), PREDICT v1.2 showed good calibration (O:E = 0.93). The tool overestimated the number of deaths at 10 years by 7.7%;
- for women with positive nodal status (n=327), PREDICT v1.2 showed good calibration (O:E = 0.92). The tool overestimated the number of deaths at 10 years by 8%;
- for women with unknown nodal status (n=4), PREDICT v1.2 showed perfect calibration (O:E = 1).

Subgroup: ER status

- for women with negative ER status (n=231), PREDICT v1.2 showed poor calibration (O:E = 0.68). The tool overestimated the number of deaths at 10 years by 46.9%. The tool also showed poor discrimination (AUC=0.694);
- for women with unknown ER status (n=366), PREDICT v1.2 showed poor calibration (O:E = 1.26). The tool underestimated the number of deaths at 10 years by 20.5%. The tool also showed poor discrimination (AUC=0.694).

Subgroup: HER2 status

- for women with negative HER2 status (n=327), PREDICT v1.2 showed good calibration (O:E = 0.99). The tool overestimated the number of deaths at 10 years by 1.2%. However the tool showed poor discrimination (AUC=0.724);
- for women with positive HER2 status (n=140), PREDICT v1.2 showed good calibration (O:E = 0.94). The tool overestimated the number of deaths at 10 years by 6%. However the tool showed poor discrimination (AUC=0.592);
- for women with borderline HER2 status (n=14), PREDICT v1.2 showed poor calibration (O:E = 1.25). The tool underestimated the number of deaths at 10 years by 20%;
- for women with unknown HER2 status (n=116), PREDICT v1.2 showed poor calibration (O:E = 0.62). The tool overestimated the number of deaths at 10 years by 60%.

Tool calibration and discrimination - 10 year breast cancer mortality [PREDICT v1.1 and v1.2]

There was good quality evidence from 1 validation study conducted in the UK with 1726 cases of invasive breast cancer and ER+ that reported the following:

- for the whole cohort (N=1726), both PREDICT v1.1 and PREDICT v1.2 showed good calibration (O:E = 1.13 and 1.08 respectively). The tool also showed good discrimination [AUC = 0.7611 and 0.7676 respectively – (p-value = 0.0008)].

The study also reported the tool calibration according to different factors:

Subgroup: age

- for the women aged <40 (n=67), both PREDICT v1.1 and PREDICT v1.2 showed good calibration (O:E = 1.15 and 1.07 respectively);
- for the women aged 40 to 49 (n=274), both PREDICT v1.1 and PREDICT v1.2 showed good calibration (O:E = 1.18 and 1.13 respectively);
- for the women aged 50 to 59 (n=436), both PREDICT v1.1 and PREDICT v1.2 showed good calibration (O:E = 1.18 and 1.15 respectively);
- for the women aged 60+ (n=497), both PREDICT v1.1 and PREDICT v1.2 showed good calibration (O:E = 1.06 and 1.01 respectively).

Subgroup: tumour size

- for women with tumours <10 mm (n=144), both PREDICT v1.1 and PREDICT v1.2 showed poor calibration (O:E = 0.78 and 0.78 respectively);
- for the women with tumours 10 to 19 mm (n=574), both PREDICT v1.1 and PREDICT v1.2 showed good calibration (O:E = 1.09 and 1.05 respectively);
- for the women with tumours 20 to 29 mm (n=404), both PREDICT v1.1 and PREDICT v1.2 showed poor calibration (O:E = 1.32 and 1.26 respectively);
- for the women with tumours 30 to 49 mm (n=140), both PREDICT v1.1 and PREDICT v1.2 showed good calibration (O:E = 0.95 and 0.91 respectively);
- for the women with tumours 50+ mm (n=11), both PREDICT v1.1 and PREDICT v1.2 showed poor calibration (O:E = 0.5 and 0.5 respectively).
- for the women with tumours of unknown size (n=1), both PREDICT v1.1 and PREDICT v1.2 showed perfect calibration (O:E = 1 and 1 respectively).

Subgroup: nodal status

- for women with negative nodal status (n=709), both PREDICT v1.1 and PREDICT v1.2 showed good calibration (O:E = 1.19 and 1.15 respectively);
- for the women with 1+ nodes (n=241), both PREDICT v1.1 and PREDICT v1.2 showed poor calibration (O:E = 1.23 and 1.17 respectively);
- for the women with 2 to 4+ nodes (n=184), PREDICT v1.1 showed good calibration (O:E = 1.05) and PREDICT v1.2 showed perfect calibration (O:E = 1);
- for the women with 5 to 9+ nodes (n=37), both PREDICT v1.1 and PREDICT v1.2 showed good calibration (O:E = 1.10 and 1.05 respectively);
- for the women with 10+ nodes (n=6), both PREDICT v1.1 and PREDICT v1.2 showed good calibration (O:E = 0.8 and 0.8 respectively).
- for the women with unknown nodal status (n=97), both PREDICT v1.1 and PREDICT v1.2 showed good calibration (O:E = 1.07 and 1.07 respectively).

Subgroup: grade

- for women with grade 1 breast cancer (n=235), both PREDICT v1.1 and PREDICT v1.2 showed poor calibration (O:E = 1.8 and 1.8 respectively);
- for the women with grade 2 breast cancer (n=528), both PREDICT v1.1 and PREDICT v1.2 showed good calibration (O:E = 1.16 and 1.14 respectively);
- for the women with grade 3 breast cancer (n=395), both PREDICT v1.1 and PREDICT v1.2 showed good calibration (O:E = 1.14 and 1.07 respectively);
- for the women with unknown graded (n=116), both PREDICT v1.1 and PREDICT v1.2 showed poor calibration (O:E = 0.31 and 0.31 respectively).

Subgroup: HER2 status

- for women with negative HER2 status (n=792), both PREDICT v1.1 and PREDICT v1.2 showed poor calibration (O:E = 1.35 and 1.29 respectively);

- for the women with positive HER2 status (n=77), both PREDICT v1.1 and PREDICT v1.2 showed poor calibration (O:E = 1.35 and 1.24 respectively);
- for the women with unknown HER2 status (n=405), both PREDICT v1.1 and PREDICT v1.2 showed poor calibration (O:E = 0.44 and 0.44 respectively).

Tool discrimination and collaboration - 10-year breast cancer mortality [PREDICT 2.0]

There was good evidence from 1 validation study conducted with combined data sets (N=5316) that assessed the tool calibration and discrimination of a new version of PREDICT.

The tool calibration was reported according to different factors, and segregated by ER status (total cohort data was not reported).

Subgroup: age at diagnosis (segregated by ER- and ER+)

ER-

- for the women aged 20 to 29 and negative ER status (n=92), PREDICT v2.0 showed good calibration (O:E = 0.94). The tool overestimated breast cancer mortality by 6%;
- for the women aged 30 to 39 and negative ER status (n=855), PREDICT v2.0 showed good calibration (O:E = 0.92). The tool overestimated breast cancer mortality by 9%;
- for the women aged 40 to 49 and negative ER status (n=414), PREDICT v2.0 showed good calibration (O:E = 0.98). The tool overestimated breast cancer mortality by 2%;
- for the women aged 50 to 59 and negative ER status (n=165), PREDICT v2.0 showed good calibration (O:E = 0.97). The tool overestimated breast cancer mortality by 3%;
- for the women aged 60 to 69 and negative ER status (n=117), PREDICT v2.0 showed good calibration (O:E = 0.82). The tool overestimated breast cancer mortality by 21%;
- for the women aged 70 to 79 and negative ER status (n=11), PREDICT v2.0 showed poor calibration (O:E = 0.36). The tool overestimated breast cancer mortality by 180%.

ER+

- for the women aged 20 to 29 and positive ER status (n=140), PREDICT v2.0 showed poor calibration (O:E = 0.71). The tool overestimated breast cancer mortality by 40%;
- for the women aged 30 to 39 and positive ER status (n=1633), PREDICT v2.0 showed good calibration (O:E = 0.96). The tool overestimated breast cancer mortality by 4%;
- for the women aged 40 to 49 and positive ER status (n=1063), PREDICT v2.0 showed good calibration (O:E = 0.90). The tool overestimated breast cancer mortality by 11%;
- for the women aged 50 to 59 and positive ER status (n=467), PREDICT v2.0 showed good calibration (O:E = 0.96). The tool overestimated breast cancer mortality by 4%;
- for the women aged 60 to 69 and positive ER status (n=517), PREDICT v2.0 showed good calibration (O:E = 1.08). The tool underestimated breast cancer mortality by 7%;
- for the women aged 70 to 79 and positive ER status (n=55), PREDICT v2.0 showed poor calibration (O:E = 0.38). The tool overestimated breast cancer mortality by 26%.

Subgroup: tumour size (segregated by ER- and ER+)

ER-

- for women with tumours 0 to 9 mm and negative ER status (n=96), PREDICT v2.0 showed good calibration (O:E = 0.90). The tool overestimated breast cancer mortality by 10%;
- for women with tumours 10 to 19 mm and negative ER status (n=559), PREDICT v2.0 showed good calibration (O:E = 0.92). The tool overestimated breast cancer mortality by 8%;

- for women with tumours 20 to 29 mm and negative ER status (n=524), PREDICT v2.0 showed good calibration (O:E = 0.97). The tool overestimated breast cancer mortality by 3%;
- for women with tumours 30 to 49 mm and negative ER status (n=354), PREDICT v2.0 showed good calibration (O:E = 0.99). The tool overestimated breast cancer mortality by 1%;
- for women with tumours 50+ mm and negative ER status (n=121), PREDICT v2.0 showed poor calibration (O:E = 0.75). The tool overestimated breast cancer mortality by 33%.

ER+

- for women with tumours 0 to 9 mm and positive ER status (n=352), PREDICT v2.0 showed poor calibration (O:E = 1.54). The tool underestimated breast cancer mortality by 6%;
- for women with tumours 10 to 19 mm and negative ER status (n=1428), PREDICT v2.0 showed good calibration (O:E = 1.06). The tool underestimated breast cancer mortality by 8%;
- for women with tumours 20 to 29 mm and positive ER status (n=1111), PREDICT v2.0 showed good calibration (O:E = 0.98). The tool overestimated breast cancer mortality by 0.80%;
- for women with tumours 30 to 49 mm and positive ER status (n=695), PREDICT v2.0 showed good calibration (O:E = 0.87). The tool overestimated breast cancer mortality by 15%;
- for women with tumours 50+ mm and positive ER status (n=289), PREDICT v2.0 showed poor calibration (O:E = 0.74). The tool overestimated breast cancer mortality by 35%.

Subgroup: number of positive nodes (segregated by ER- and ER+)

ER-

- for women with 0 positive nodes and negative ER status (n=937), PREDICT v2.0 showed good calibration (O:E = 1.01). The tool underestimated breast cancer mortality by 0.89%;
- for women with 1 positive node and negative ER status (n=232), PREDICT v2.0 showed good calibration (O:E = 0.86). The tool overestimated breast cancer mortality by 17%;
- for women with 2 to 4 positive nodes and negative ER status (n=300), PREDICT v2.0 showed good calibration (O:E = 0.88). The tool overestimated breast cancer mortality by 13%;
- for women with 5 to 9 positive nodes and negative ER status (n=101), PREDICT v2.0 showed good calibration (O:E = 0.96). The tool overestimated breast cancer mortality by 4%;
- for women with 10+ positive nodes and negative ER status (n=84), PREDICT v2.0 showed good calibration (O:E = 0.85). The tool overestimated breast cancer mortality by 17%.

ER+

- for women with 0 positive nodes and positive ER status (n=2085), PREDICT v2.0 showed good calibration (O:E = 0.99). The tool overestimated breast cancer mortality by 1%;
- for women with 1 positive node and positive ER status (n=675), PREDICT v2.0 showed good calibration (O:E = 0.92). The tool overestimated breast cancer mortality by 9%;
- for women with 2 to 4 positive nodes and positive ER status (n=734), PREDICT v2.0 showed good calibration (O:E = 0.96). The tool overestimated breast cancer mortality by 4%;

- for women with 5 to 9 positive nodes and positive ER status (n=245), PREDICT v2.0 showed good calibration (O:E = 0.86). The tool overestimated breast cancer mortality by 17%;
- for women with 10+ positive nodes and positive ER status (n=136), PREDICT v2.0 showed good calibration (O:E = 0.87). The tool overestimated breast cancer mortality by 15%.

Subgroup: tumour grade (segregated by ER- and ER+)

ER-

- for women with grade 1 tumour and negative ER status (n=44), PREDICT v2.0 showed good calibration (O:E = 0.96). The tool overestimated breast cancer mortality by 4%;
- for women with grade 2 tumour and negative ER status (n=183), PREDICT v2.0 showed good calibration (O:E = 0.86). The tool overestimated breast cancer mortality by 17%;
- for women with grade 3 tumour and negative ER status (n=1427), PREDICT v2.0 showed good calibration (O:E = 0.94). The tool overestimated breast cancer mortality by 7%.

ER+

- for women with grade 1 tumour and positive ER status (n=658), PREDICT v2.0 showed good calibration (O:E = 0.96). The tool overestimated breast cancer mortality by 4%;
- for women with grade 2 tumour and positive ER status (n=1730), PREDICT v2.0 showed good calibration (O:E = 0.86). The tool overestimated breast cancer mortality by 17%;
- for women with grade 3 tumour and positive ER status (n=1487), PREDICT v2.0 showed good calibration (O:E = 0.94). The tool overestimated breast cancer mortality by 7%.

Tool discrimination was also reported by ER status:

- for women with negative ER status, the tool discrimination was poor (AUC=0.696);
- however for women with positive ER status, the tool discrimination was good (AUC=0.790).

Important outcomes

Prognostic accuracy (sensitivity, specificity)

- No evidence was found for this outcome.

Predictive prognostic tool 3: Nottingham Prognostic Index (NPI)

Critical outcomes

Tool calibration - 10-year breast cancer survival

There was good quality evidence from 1 validation study conducted in the UK with 2238 women diagnosed with or treated for primary operable invasive breast cancer that reported the following:

- for women in the excellent prognosis group according to their NPI score (n=320), the tool showed good calibration (O:E = 0.98);
- for women in the good prognosis group according to their NPI score (n=475), the tool showed good calibration (O:E = 0.99);
- for women in the moderate prognosis group I according to their NPI score (n=634), the tool showed good calibration (O:E = 1.03);
- for women in the moderate prognosis group II according to their NPI score (n=489), the tool showed perfect calibration (O:E = 1.00);

- for women in the poor prognosis group according to their NPI score (n=233), the tool showed good calibration (O:E = 1.02);
- for women in the very poor prognosis group according to their NPI score (n=86), the tool showed good calibration (O:E = 0.89).

Tool discrimination

- No evidence was found for this outcome.

Important outcomes

Prognostic accuracy (sensitivity, specificity)

- No evidence was found for this outcome.

Predictive prognostic tool 4: FinProg

Critical outcomes

Tool calibration

- No evidence was found for this outcome.

Tool discrimination

- No evidence was found for this outcome.

Important outcomes

Prognostic accuracy (sensitivity, specificity)

- No evidence was found for this outcome.

Predictive prognostic tool 5: CancerMath

Critical outcomes

Tool calibration

- No evidence was found for this outcome.

Tool discrimination

- No evidence was found for this outcome.

Important outcomes

Prognostic accuracy (sensitivity, specificity)

- No evidence was found for this outcome.

Predictive prognostic tool 6: Oxford Prognostic Index (OPI)

Critical outcomes

Tool calibration and discrimination - 5-year recurrence-free survival

There was moderate quality evidence from 1 validation study with 1787 women treated for invasive ductal carcinoma that reported the following:

- for the whole cohort (N=1789), OPI showed good calibration (O:E = 1.01). The tool underestimated 5-year recurrence free survival by 0.7%. However the tool showed poor discrimination (overall C-statistic = 0.720).
- The study also reported the calibration according to different factors (tool discrimination was not reported):

Subgroup: age

- for women ≤50 years (n=1097), OPI showed good calibration (O:E = 1.03). The tool underestimated 5-year recurrence free survival by 1.92%;
- for women >50 years (n=690), OPI showed perfect calibration (O:E ≈ 1.00). The tool overestimated 5-year recurrence free survival by 0.10%.

Subgroup: tumour grade

- for women with grade 1 tumour (n=196), OPI showed good calibration (O:E = 1.06). The tool underestimated 5-year recurrence free survival by 5.15%;
- for women with grade 2 tumour (n=772), OPI showed good calibration (O:E = 1.03). The tool underestimated 5-year recurrence free survival by 2.44%;
- for women with grade 3 tumour (n=819), OPI showed good calibration (O:E = 0.98). The tool overestimated 5-year recurrence free survival by 1.04%.

Subgroup: tumour size

- for women with tumours ≤2 cm (n=954), OPI showed good calibration (O:E = 1.06). The tool overestimated 5-year recurrence free survival by 0.89%;
- for women with tumours >2 cm to ≤5 cm(n=772), OPI showed good calibration (O:E = 0.95). The tool underestimated 5-year recurrence free survival by 2.7%;
- for women with tumours >5 cm (n=61), OPI showed good calibration (O:E = 1.04). The tool overestimated 5-year recurrence free survival by 3.71%.

Subgroup: nodal status

- for women with negative nodal status (n=674), OPI showed good calibration (O:E = 1.02). The tool overestimated 5-year recurrence free survival by 1.82%;
- for women with positive nodal status (n=1113), OPI showed good calibration (O:E = 1.01). The tool underestimated 5-year recurrence free survival by 0.71%.

Subgroup: ER status

- for women with negative ER status (n=1097), OPI showed good calibration (O:E = 1.03). The tool overestimated 5-year recurrence free survival by 1.92%;
- for women with positive ER status (n=690), OPI showed perfect calibration (O:E ≈ 1.00). The tool underestimated 5-year recurrence free survival by 0.10%.

Important outcomes

Prognostic accuracy (sensitivity, specificity)

- No evidence was found for this outcome.

The committee's discussion of the evidence

Interpreting the evidence

The outcomes that matter most

The aim of this review was to determine which prognostic prediction tool is most helpful at predicting survival, and therefore identifying women who may benefit from adjuvant treatment.

The committee agreed that tool calibration and tool discrimination were the critical outcomes for decision making. This is because identifying people with a worse prognosis would guide decisions regarding the use of adjuvant treatment. In addition they also included sensitivity and specificity as important outcomes.

The quality of the evidence

This review included validation studies. The quality of the individual studies was assessed using the CASP tool for clinical prediction rule. The overall judgement of the quality was based on the consideration of the individual domains.

One study evaluated the Adjuvant! Online tool, and was assessed as high quality.

Four studies looked at PREDICT. Results were reported separately for each study, as they assessed different versions of PREDICT. Studies using older versions of the tool were rated down because they did not consider all relevant prognostic factors.

One study evaluated the NPI tool, and was assessed as high quality.

One study evaluated the OPI tool, and was rated as moderate quality. The main reason for rating down the quality of the study was because the tool is not available in clinical practice, and therefore is of limited use.

Benefits and drawbacks

The committee discussed the benefits and drawbacks of the various tools.

Four studies reported on the prognostic accuracy of PREDICT, although studies assessed different versions. The results of 1 study showed that the first version of PREDICT (v1.0) was well calibrated to estimate 5-year mortality in the whole cohort, and across different prognostic groups (including age at diagnosis, tumour grade, tumour size, nodal status and ER status). The tool also showed good discrimination for the ER positive and negative models. Likewise, the tool showed good calibration and good discrimination to estimate 8-year mortality in the whole cohort. The tool was well calibrated for all prognostic subgroups, except those with tumours over 50 mm, but it showed poor discrimination for young and old women (those <35 and 75+), those with negative nodal status, and in women with tumours over 20 mm.

Another study looked at versions v1.1 and v1.2 of PREDICT in women with invasive breast cancer and ER-positive. Results were quite similar for both versions, showing good calibration to estimate 10-year breast cancer mortality in the whole cohort and across most subgroups, the exceptions being those based on tumour size, HER2 status and grade 1 tumours. Discrimination was also good for both versions, but the authors of the study noted that discrimination significantly improved in v1.2.

Another study that looked at an updated version of PREDICT (v1.2) showed poor calibration and poor discrimination to predict all-cause mortality at 5 years in a cohort of women aged ≤40 years. The tool also showed poor calibration across most prognostic subgroups (including age at diagnosis, tumour grade, tumour size, nodal status, and negative HER2

status). The tool showed good calibration to predict all-cause mortality at 10 years in a cohort of women aged ≤40 years; however there was poor calibration for prognostic subgroups based on tumour grade, tumour size, nodal status and ER status.

A recent study evaluated the most updated model, PREDICT version 2.0 (version release in 2017). The new tool was shown to have good prognostic accuracy to estimate 10-year breast cancer mortality across most subgroups, including age, tumour size, tumour grade and number of positive nodes, independent of ER status. The committee noted the improved performance of this version of the tool among young women, however calibration was still poor in women aged 20 to 29 who were ER-positive. There was also poor calibration in women aged 70 to 79 and those with tumours greater than 50 mm, independent of ER status, and women with ER-positive tumours smaller than 10 mm.

The evidence suggested that Adjuvant! Online was a well calibrated tool to predict survival at 10 years. This was shown consistently for the total cohort of women on whom the tool was validated, and for the different subgroups (age, tumour grade, tumour size, nodal status and ER status). This supported the committee's experience in clinical practice, as they agreed it is a very useful tool. Indeed this tool was extensively used in clinical practice, however they noted this tool is no longer available. Therefore they agreed this tool could not be recommended.

The evidence suggested that the NPI is also a well calibrated tool to predict 10-year cancer survival. However, the committee agreed it has now been superseded by other tools which take into account more factors such as ER and HER2 status.

The committee also discussed the results for the OPI. The evidence included in this review suggested that this is a well calibrated tool to predict recurrence-free survival at 5 years for the total cohort of women on whom the tool was validated, and for the different subgroups (age, tumour grade, tumour size, nodal status and ER status). However they noted they were not aware of this tool, as it is not actually available in practice. Based on this they agreed it could not be recommended.

No studies were found reporting on the prognostic accuracy of CancerMath and FinProg, and the committee agreed they could not make recommendations in favour or against their use.

The committee agreed that using accurate prognostic tools helps to have more informed decision making, but noted that over-reliance on the results of a prognostic tool could result in over- or under-treatment for some people, if individual characteristics are not taken into consideration (for example significant comorbidities or age group variations).

In addition they note that as there is limited evidence by population age it is not possible to confirm the accuracy of the tool for all groups. This is because although studies report results by age groups, the sample size for young women is too small to allow sufficient statistical power.

Overall, the committee agreed that a validated prognostic tool provides important guidance in treatment, but clinical judgement should also play an important role.

Cost effectiveness and resource use

A systematic review of the economic literature was conducted but no relevant studies were identified which were applicable to this review question.

It was thought that the economic impact of recommendations made in this area would be relatively small because there is little difference in resource use between the prognostic tools (they are freely available and the time taken to complete them is similar).

The committee have recommended the use of PREDICT which is a change from the previous guideline (CG80), in which Adjuvant! Online was recommended. It might therefore

be considered a change in practice. However, in reality most professionals are already using PREDICT because adjuvant! Online is no longer available.

It is possible that the use of a different tool may have implications for the numbers of patients receiving adjuvant treatment because of differences in prognostic accuracy. Therefore there could be a cost impact associated with changes in patient management. While it is difficult to speculate fully on the direction of this effect, it was considered likely that the scale of the effect would be relatively small and that in most cases the decision on whether to use adjuvant therapy or not would be similar with PREDICT or with Adjuvant! Online.

Other factors the committee took into account

The committee agreed that this recommendation would make the same prognostic tool available to all populations nationally, and this could potentially reduce inequalities. At the time of guideline publication, PREDICT v2.0 was the version available on the PREDICT homepage (http://predict.nhs.uk/), although version 1.2 can still be accessed on the website. Although the evidence had considered previous versions of PREDICT, the committee made their recommendations based on PREDICT v2.0. If future versions of PREDICT are released, the recommendations relating to groups in whom the tool is less accurate may no longer be applicable, and this information is provided in a footnote.

However the committee recognised that the validation of the model may under-represent some ethnic groups. Similarly, this tool has not been validated in men, therefore it is not possible to know if it is applicable to them. They also noted that the very young and older ages are under-represented.

The committee pointed out that the availability of the prognostic tools affected the recommendations, therefore the long-term adoption of prognostic tools is dependent on continued support and availability.

Finally the committee highlighted this review did not include gene profiling tools as these will be covered by NICE diagnostic guidance (Gene expression profiling and expanded immunohistochemistry tests for guiding adjuvant chemotherapy decisions in early breast cancer management: MammaPrint, Oncotype DX, IHC4 and Mammostrat; DG10 update), and a link was included to this guidance.

References

Blamey 2007

Blamey, R. W., Ellis, I. O., Pinder, S. E., Lee, A. H. S., Macmillan, R. D., Morgan, D. A. L., Robertson, J. F. R., Mitchell, M. J., Ball, G. R., Haybittle, J. L., Elston, C. W. (2007) Survival of invasive breast cancer according to the Nottingham Prognostic Index in cases diagnosed in 1990-1999. European journal of cancer, 43, 1548-1555.

Campbell 2009

Campbell, H. E., Taylor, M. A., Harris, A. L., Gray, A. M. (2009) An investigation into the performance of the Adjuvant! Online prognostic programme in early breast cancer for a cohort of patients in the United Kingdom. British journal of cancer, 101, 1074-1084.

Campbell 2010

Campbell, H. E., Gray, A. M., Harris, A. L., Briggs, A. H., Taylor, M. A. (2010) Estimation and external validation of a new prognostic model for predicting recurrence-free survival for early breast cancer patients in the UK. British journal of cancer, 103, 776-786.

Candido Dos Reis 2017

Candido Dos Reis, F. J., Wishart, G. C., Dicks, E. M., Greenberg, D., Rashbass, J., Schmidt, M. K., van den Broek, A. J., Ellis, I. O., Green, A., Rakha, E., Maishman, T., Eccles, D. M., Pharoah, P. D. P. (2017) An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation. Breast Cancer Research, 19, 58.

Debray 2017

Debray, T., Damen, J., Snell, K., Ensor, J., Hooft, L., Reitsma, J. et al. (2017) A guide to systematic review and meta-analysis of prediction model performance. BMJ, 356, i6460.

Maishman 2015

Maishman, T., Copson, E., Stanton, L., Gerty, S., Dicks, E., Durcan, L., Wishart, G. C., Pharoah, P., Eccles, D. (2015) An evaluation of the prognostic model PREDICT using the POSH cohort of women aged <40 years at breast cancer diagnosis. British journal of cancer, 112, 983-991.

NICE 2009

National Institute for Health and Clinical Excellence. (2009). Early and locally advanced breast cancer: diagnosis and treatment. NICE guideline (CG80).

Riley 2015

Riley RD, Ahmed I, Debray T, Willis BH, Noordzij JP, Higgins J, Deeks JJ. (2015) Summarising and validating test accuracy results across multiple studies for use in clinical practice. Statistics in medicine. 2015 Jun 15;34(13):2081-103.

Wishart 2010

Wishart, G. C., Azzato, E. M., Greenberg, D. C., Rashbass, J., Kearins, O., Lawrence, G., Caldas, C., Pharoah, P. D. (2010) PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. [Erratum appears in Breast Cancer Research 2010;12(2):401]. Breast Cancer Research, 12, R1.

Wishart 2014

Wishart, G. C., Rakha, E., Green, A., Ellis, I., Ali, H. R., Provenzano, E., Blows, F. M., Caldas, C., Pharoah, P. D. (2014) Inclusion of KI67 significantly improves performance of the PREDICT prognostication and prediction model for early breast cancer. BMC cancer, 14, 908.

Appendices

Appendix A – Review protocols

Review protocol for 3.1 Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?

Field (based on PRISMA-P)	Content
Review question	Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?
Type of review question	Intervention review
Objective of the review	The objective of this review is to establish the role and benefit of the pathological assessment of PR in breast cancers for planning adjuvant chemotherapy. Recommendations will cover whether PR testing should occur.
Eligibility criteria – population/disease/condition/issue/domain	Adults (18 or over) with invasive breast cancer (M0)
Eligibility criteria – intervention(s)/exposure(s)/prognostic factor(s)	ER and HER2 plus PR test followed by chemotherapy as indicated based on test results
Eligibility criteria – comparator(s)/control or reference (gold) standard	ER and HER2 test followed by chemotherapy as indicated based on test results
Outcomes and prioritisation	Critical (up to 3 outcomes) Disease-free survival (MID: any statistically significant difference) Overall survival (MID: any statistically significant difference) Important but not critical Treatment-related morbidity (MID: GRADE default values) 5 and 10 year follow-up periods will be prioritised if multiple time points are reported.
Eligibility criteria – study design	Systematic reviews/meta-analyses of 'test and treat' RCTs 'Test and treat' RCTs Modified nominal group technique will be used to make recommendations regarding the appropriateness of PR testing if no published test and treat RCTs are identified
Other inclusion exclusion criteria	Foreign language studies, conference abstracts, and narrative reviews will not routinely be included.

Field (based on PRISMA-P)	Content
Proposed sensitivity/subgroup analysis, or meta-regression	N/A
Selection process – duplicate screening/selection/analysis	Sifting, data extraction, appraisal of methodological quality and GRADE assessment will be performed by the reviewing team. Quality control will be performed by the senior systematic reviewer. Dual sifting will not be performed for this question.
Data management (software)	Study sifting and data extraction will be undertaken in STAR. Pairwise meta-analyses will be performed using Cochrane Reviewer Manager (RevMan 5). GRADEpro will be used to assess the quality of evidence for each outcome.
Information sources – databases and dates	The following key databases will be searched: Cochrane Library (CDSR, DARE, CENTRAL, HTA) through Wiley, Medline & Medline in Process and Embase through OVID. Additionally we may search Web of Science and consideration will be given to subject-specific databases and used as appropriate. The search will be undertaken from 1984, when the first studies of immunohistochemical determination of progesterone receptor status were published, as opposed to updating the search from the previous guideline due to substantial change in the focus of the review question.
Identify if an update	Previous question: Does progesterone receptor status add further, useful information to that of oestrogen receptor status in patients with invasive breast cancer? Date of search: 27/02/2008 Relevant recommendation(s) from previous guideline: 1) Do not routinely assess progesterone receptor status of tumours in patients with invasive breast cancer.
Author contacts	Please see guideline in development page on the web site.
Highlight if amendment to previous protocol	For details please see Section 4.5 of Developing NICE guidelines: the manual
Search strategy – for one database	For details please see appendix B.
Data collection process – forms/duplicate	A standardised evidence table format will be used, and published as appendix D (clinical evidence tables) or appendix H (economic evidence tables).
Data items – define all variables to be collected	For details please see evidence tables in appendix D (clinical evidence tables) or appendix H (economic evidence tables).
Methods for assessing bias at outcome/study level	Standard study checklists were used to critically appraise individual studies. For details please see section 6.2 of Developing NICE guidelines: the manual

Field (based on PRISMA-P)	Content
	The risk of bias across all available evidence was evaluated for each outcome using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group http://www.gradeworkinggroup.org/
Criteria for quantitative synthesis	For details please see Section 6.4 of Developing NICE guidelines: the manual
Methods for quantitative analysis – combining studies and exploring (in)consistency	For details please see the methods chapter.
Meta-bias assessment – publication bias, selective reporting bias	For details please see Section 6.2 of Developing NICE guidelines: the manual.
Confidence in cumulative evidence	For details please see Sections 6.4 and 9.1 of Developing NICE guidelines: the manual
Rationale/context – what is known	For details please see the introduction to the evidence review in the main file.
Describe contributions of authors and guarantor	A multidisciplinary committee developed the guideline. The committee was convened by the NGA and chaired by Dr Jane Barrett in line with section 3 of Developing NICE guidelines: the manual. Staff from NGA undertook systematic literature searches, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate, and drafted the guideline in collaboration with the committee. For details please see the methods supplement.
Sources of funding/support	NGA is funded by NICE and hosted by the Royal College of Obstetricians and Gynaecologists.
Name of sponsor	NGA is funded by NICE and hosted by the Royal College of Obstetricians and Gynaecologists.
Roles of sponsor	NICE funds NGA to develop guidelines for the NHS in England.
PROSPERO registration number	N/A

GRADE, Grading of Recommendations Assessment, Development and Evaluation; M0, no distant metastases; MID, minimally important difference; N/A, not applicable; NHS, National Health Service, NICE, National Institute of Health and Care Excellence; NGA, National Guideline Alliance; RCT, randomised controlled trial

Review protocol for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?

Field (based on PRISMA-P)	What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?
Review question	What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?
Type of review question	Prediction model performance
Objective of the review	The objective of this review is to determine the accuracy of prognostic tools for predicting survival and benefit of treatment. Recommendations will aim to cover which tools should be used to aid decision adjuvant treatment planning.
Eligibility criteria – population/disease/condition/issue/domain	Adults (18 or over) with invasive breast cancer (M0) who have undergone surgery and who are candidates for adjuvant systemic therapy
Eligibility criteria – intervention(s)/exposure(s)/prognostic factor(s)	Any appropriate predictive prognostic tools, e.g., Adjuvant! Online PREDICT Nottingham Prognostic Index (NPI) FinProg CancerMath Other relevant validated tools will also be considered for inclusion
Eligibility criteria – comparator(s)/control or reference (gold) standard	N/A
Outcomes and prioritisation	Accuracy (sensitivity/specificity) (important outcome) Tool discrimination (AUC or C-statistic) (critical outcome) Tool calibration (mortality ratio or survival ratio) (critical outcome) For the following: Disease free survival Overall survival/ death At the following time points: Short-term: 5 years

Field (based on PRISMA-P)	What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?
	Long-term: 10 years Note: It was agreed with the committee that sensitivity or specificity would be considered high when sensitivity or specificity was 90% or higher, and moderate when sensitivity or specificity was between 75% and 89%. The mortality ratio is defined as the ratio of observed number of deaths in a study population and the expected number of deaths. In this review, a tool will be considered to have good calibration if the ratio ranges from 0.8 to 1.2 (as suggested by Debray 2017). Discrimination is a measure to assess how well a tool identifies people with worse survival, and it is often reported by the concordance c-statistic (also known as AUC). In this review a tool will be considered to have good discrimination if c-statistic is above 0.75 (as suggested by Debray 2017).
Eligibility criteria – study design	Systematic reviews/meta-analyses of prognostic studies Prospective cohort studies Retrospective cohort studies
Other inclusion exclusion criteria	Foreign language studies, conference abstracts, and narrative reviews will not routinely be included.
Proposed sensitivity/sub-group analysis, or meta-regression	Factors/ sub-groups to look at separately: age, tumour size, tumour grade, ER status, HER2 status and nodal involvement Accuracy of each tool will be presented separately.
Selection process – duplicate screening/selection/analysis	Sifting, data extraction, and appraisal of methodological quality will be performed by the reviewing team. Quality control will be performed by the senior systematic reviewer. Dual sifting will be performed on at least 10% of records and where possible all records as this is a prognostic review; 90% agreement is required and any discussions will be resolved through discussion and consultation with senior staff where necessary.
Data management (software)	Study sifting and data extraction will be undertaken in STAR. Meta-analysis will not be performed The CASP clinical prediction rule checklist will be used to assess the quality of included studies.

Field (based on PRISMA-P)	What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?
Information sources – databases and dates	The following key databases will be searched: Cochrane Library (CDSR, DARE, CENTRAL, HTA) through Wiley, Medline & Medline in Process and Embase through OVID. Additionally we may search Web of Science and consideration will be given to subject-specific databases and used as appropriate. The search will be undertaken from 1982 when the Nottingham Prognostic Index (the oldest of the tools identified above) was first published. Date limit: 1982 (first publication - Nottingham Prognostic Index)
Identify if an update	N/A
Author contacts	Please see guideline in development page on the web site.
Highlight if amendment to previous protocol	For details please see Section 4.5 of Developing NICE guidelines: the manual
Search strategy	For details please see appendix B.
Data collection process – forms/duplicate	A standardised evidence table format will be used, and published as appendix D (clinical evidence tables) or appendix H (economic evidence tables).
Data items – define all variables to be collected	For details please see evidence tables in appendix D (clinical evidence tables) or appendix H (economic evidence tables).
Methods for assessing bias at outcome/study level	Standard study checklists were used to critically appraise individual studies. For details please see Section 6.2 of Developing NICE guidelines: the manual The risk of bias across all available evidence was evaluated for each outcome using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group http://www.gradeworkinggroup.org/
Criteria for quantitative synthesis	For details please see Section 6.4 of Developing NICE guidelines: the manual
Methods for quantitative analysis – combining studies and exploring (in)consistency	For details please see the methods chapter.
Meta-bias assessment – publication bias, selective reporting bias	For details please see Section 6.2 of Developing NICE guidelines: the manual.
Confidence in cumulative evidence	For details please see Sections 6.4 and 9.1 of Developing NICE guidelines: the manual
Rationale/context – what is known	For details please see the introduction to the evidence review.

Field (based on PRISMA-P)	What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?
Describe contributions of authors and guarantor	A multidisciplinary committee developed the guideline. The committee was convened by the NGA and chaired by Dr Jane Barrett in line with section 3 of Developing NICE guidelines: the manual. Staff from NGA undertook systematic literature searches, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate, and drafted the guideline in collaboration with the committee. For details please see the methods chapter of the full guideline.
Sources of funding/support	NGA is funded by NICE and hosted by the Royal College of Obstetricians and Gynaecologists.
Name of sponsor	NGA is funded by NICE and hosted by the Royal College of Obstetricians and Gynaecologists.
Roles of sponsor	NICE funds NGA to develop guidelines for the NHS in England.
PROSPERO registration number	N/A

AUC, area under the curve; CASP, Critical Appraisal Skills Programme; ER, oestrogen receptor; GRADE, Grading of Recommendations Assessment, Development and Evaluation; HER2, Human epidermal growth factor receptor 2; M0, no distant metastases; MID, minimally important difference; N/A, not applicable; NHS, National Health Service, NICE, National Institute of Health and Care Excellence; NGA, National Guideline Alliance; NPI, Nottingham prognostic index; RCT, randomised controlled trial

Appendix B – Literature search strategies

Literature search strategies for 3.1 Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?

Database: Medline & Embase (Multifile)

Last searched on Embase 1974 to 2017 March 03, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present.

Date of last search: 6 March 2017.

#	Searches
1	exp breast cancer/ use oemezd
2	exp breast carcinoma/ use oemezd
3	exp medullary carcinoma/ use oemezd
4	exp intraductal carcinoma/ use oemezd
5	exp breast tumor/ use oemezd
6	exp Breast Neoplasms/ use prmz
7	exp "Neoplasms, Ductal, Lobular, and Medullary"/ use prmz
8	Carcinoma, Intraductal, Noninfiltrating/ use prmz
9	Carcinoma, Lobular/ use prmz
10	Carcinoma, Medullary/ use prmz
11	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
12	exp breast/ use oemezd
13	exp Breast/ use prmz
14	breast.tw.
15	12 or 13 or 14
16	(breast adj milk).tw.
17	(breast adj tender$).tw.
18	16 or 17
19	15 not 18
20	exp neoplasm/ use oemezd
21	exp Neoplasms/ use prmz
22	20 or 21
23	19 and 22
24	(breast$ adj5 (neoplasm$ or cancer$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or leiomyosarcoma$ or dcis or duct$ or infiltrat$ or intraduct$ or lobul$ or medullary or tubular)).tw. use oemezd
25	(mammar$ adj5 (neoplasm$ or cancer$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or leiomyosarcoma$ or dcis or duct$ or infiltrat$ or intraduct$ or lobul$ or medullary or tubular)).tw. use oemezd
26	(breast$ adj5 (neoplasm$ or cancer$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or leiomyosarcoma$ or dcis or duct$ or infiltrat$ or intraduct$ or lobul$ or medullary or tubular)).mp. use prmz

#	Searches
27	(mammar$ adj5 (neoplasm$ or cancer$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or leiomyosarcoma$ or dcis or duct$ or infiltrat$ or intraduct$ or lobul$ or medullary or tubular)).mp. use prmz
28	exp Paget nipple disease/ use oemezd
29	Paget's Disease, Mammary/ use prmz
30	(paget$ and (breast$ or mammary or nipple$)).tw.
31	23 or 24 or 25 or 26 or 27 or 28 or 29 or 30
32	11 or 31
33	Receptors, Progesterone/ use prmz
34	progesterone receptor/ use oemezd
35	((progesteron$ or progestin or PgR or PR) adj3 (status or test$ or level$ or receptor$ or expression)).ti,ab.
36	((PR adj2 positiv$) or (PR adj2 negativ$) or (PgR adj2 positiv$) or (PgR adj2 negativ$) or (progesteron$ adj2 positiv$) or (progesteron$ adj2 negativ$) or (progestin adj2 negativ$) or (progestin adj2 positiv$)).ti,ab.
37	(progesteron$ or progestin or PgR or PR).m_titl.
38	33 or 34 or 35 or 36 or 37
39	32 and 38
40	limit 39 to yr="1984 -Current"
41	limit 40 to RCTs and SRs, and general exclusions filter applied

Database: Cochrane Library via Wiley Online

Date of last search: 6 March 2017.

#	Searches
#1	MeSH descriptor: [Breast Neoplasms] explode all trees
#2	MeSH descriptor: [Neoplasms, Ductal, Lobular, and Medullary] explode all trees
#3	MeSH descriptor: [Carcinoma, Intraductal, Noninfiltrating] explode all trees
#4	MeSH descriptor: [Carcinoma, Lobular] this term only
#5	MeSH descriptor: [Carcinoma, Medullary] this term only
#6	#1 or #2 or #3 or #4 or #5
#7	MeSH descriptor: [Breast] explode all trees
#8	breast:ti,ab,kw (Word variations have been searched)
#9	#7 or #8
#10	(breast next milk):ti,ab,kw (Word variations have been searched)
#11	(breast next tender*):ti,ab,kw (Word variations have been searched)
#12	#10 or #11
#13	#9 not #12
#14	MeSH descriptor: [Neoplasms] explode all trees
#15	#13 and #14
#16	(breast* near/5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or leiomyosarcoma* or dcis or duct* or infiltrat* or intraduct* or lobul* or medullary or tubular)):ti,ab,kw (Word variations have been searched)

#	Searches
#17	(mammar* near/5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or leiomyosarcoma* or dcis or duct* or infiltrat* or intraduct* or lobul* or medullary or tubular)):ti,ab,kw (Word variations have been searched)
#18	MeSH descriptor: [Paget's Disease, Mammary] this term only
#19	(paget* and (breast* or mammary or nipple*)):ti,ab,kw (Word variations have been searched)
#20	#15 or #16 or #17 or #18 or #19
#21	#6 or #20
#22	MeSH descriptor: [Receptors, Progesterone] explode all trees
#23	((progesteron* or progestin or PgR or PR) near/3 (status or test* or level* or receptor* or expression)):ti,ab,kw (Word variations have been searched)
#24	((PR near/2 positiv*) or (PR near/2 negativ*) or (PgR near/2 positiv*) or (PgR near/2 negativ*) or (progesteron* near/2 positiv*) or (progesteron* near/2 negativ*) or (progestin near/2 negativ*) or (progestin near/2 positiv*)):ti,ab,kw (Word variations have been searched)
#25	(progesterone* or progestin or PgR or PR):ti (Word variations have been searched)
#26	#22 or #23 or #24 or #25
#27	#21 and #26

Literature search strategies for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?

Database: Medline & Embase (Multifile)

Last searched on Embase 1974 to 2017 September 20, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present.

Date of last search: 22 September 2017

#	Searches
1	exp breast cancer/ use oemezd
2	exp breast carcinoma/ use oemezd
3	exp medullary carcinoma/ use oemezd
4	exp intraductal carcinoma/ use oemezd
5	exp breast tumor/ use oemezd
6	exp Breast Neoplasms/ use prmz
7	exp "Neoplasms, Ductal, Lobular, and Medullary"/ use prmz
8	Carcinoma, Intraductal, Noninfiltrating/ use prmz
9	Carcinoma, Lobular/ use prmz
10	Carcinoma, Medullary/ use prmz
11	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
12	exp breast/ use oemezd
13	exp Breast/ use prmz
14	breast.tw.
15	12 or 13 or 14

#	Searches
16	(breast adj milk).tw.
17	(breast adj tender$).tw.
18	16 or 17
19	15 not 18
20	exp neoplasm/ use oemezd
21	exp Neoplasms/ use prmz
22	20 or 21
23	19 and 22
24	(breast$ adj5 (neoplasm$ or cancer$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or leiomyosarcoma$ or dcis or duct$ or infiltrat$ or intraduct$ or lobul$ or medullary or tubular)).tw. use oemezd
25	(mammar$ adj5 (neoplasm$ or cancer$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or leiomyosarcoma$ or dcis or duct$ or infiltrat$ or intraduct$ or lobul$ or medullary or tubular)).tw. use oemezd
26	(breast$ adj5 (neoplasm$ or cancer$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or leiomyosarcoma$ or dcis or duct$ or infiltrat$ or intraduct$ or lobul$ or medullary or tubular)).mp. use prmz
27	(mammar$ adj5 (neoplasm$ or cancer$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or leiomyosarcoma$ or dcis or duct$ or infiltrat$ or intraduct$ or lobul$ or medullary or tubular)).mp. use prmz
28	exp Paget nipple disease/ use oemezd
29	Paget's Disease, Mammary/ use prmz
30	(paget$ and (breast$ or mammary or nipple$)).tw.
31	23 or 24 or 25 or 26 or 27 or 28 or 29 or 30
32	11 or 31
33	(adjuvant$ adj (online or model$ or program$ or tool$)).tw.
34	"www.adjuvantonline.com".tw.
35	adjuvant?online$.tw.
36	(PREDICT adj2 (online or model$ or program$ or tool$ or estimat$)).tw.
37	"www.predict.nhs.uk".tw.
38	(predict adj plus).tw.
39	(prognost$ adj index).tw.
40	"Nottingham Prognostic Index".tw.
41	NPI.tw.
42	FinProg$.tw.
43	CancerMath$.tw.
44	"www.CancerMath.net".tw.
45	33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44
46	32 and 45
47	limit 46 to yr="1982 -Current"
48	decision support system/ use oemezd
49	Decision Making, Computer-Assisted/ use prmz
50	computer/ use oemezd
51	Computers/ use prmz

#	Searches
52	clinical decision support system/ use oemezd
53	Decision Support Systems, Clinical/ use prmz
54	computer program/ use oemezd
55	Software/ use prmz
56	Decision Support Techniques/ use prmz
57	*decision making/ use oemezd
58	medical decision making/ use oemezd
59	clinical decision making/ use oemezd
60	Decision Making/ use prmz
61	48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60
62	adjuvant$.tw.
63	32 and 61 and 62
64	limit 63 to yr="2007 -Current"
65	survival/ use oemezd
66	survival analysis/ use prmz
67	65 or 66
68	Internet/ use prmz
69	internet/ use oemezd
70	Databases, factual/ use prmz
71	*data base/ use oemezd
72	Online systems/ use prmz
73	online system/ use oemezd
74	Web browser/ use prmz
75	web browser/ use oemezd
76	User computer interface/ use prmz
77	computer interface/ use oemezd
78	68 or 69 or 70 or 71 or 72 or 73 or 74 or 75 or 76 or 77
79	32 and 67 and 78
80	limit 79 to yr="1982 -Current"
81	47 or 64 or 80
82	remove duplicates from 81 [Then general exclusions filter applied]

Database: Cochrane Library via Wiley Online

Date of last search: 22 September 2017.

#	Searches
#1	MeSH descriptor: [Breast Neoplasms] explode all trees
#2	MeSH descriptor: [Neoplasms, Ductal, Lobular, and Medullary] explode all trees
#3	MeSH descriptor: [Carcinoma, Intraductal, Noninfiltrating] explode all trees
#4	MeSH descriptor: [Carcinoma, Lobular] this term only
#5	MeSH descriptor: [Carcinoma, Medullary] this term only
#6	#1 or #2 or #3 or #4 or #5

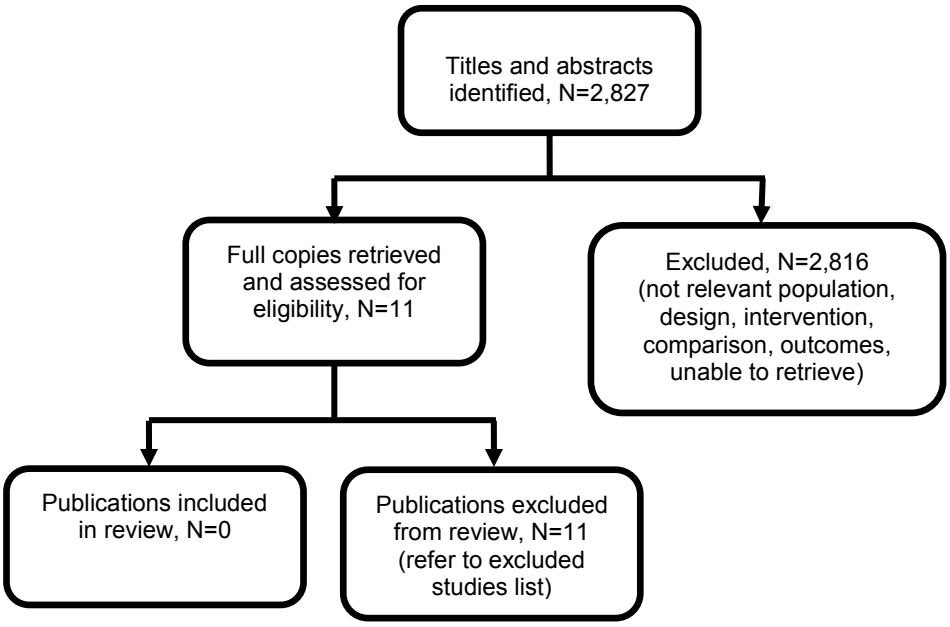
#	Searches
#7	MeSH descriptor: [Breast] explode all trees
#8	breast:ti,ab,kw (Word variations have been searched)
#9	#7 or #8
#10	(breast next milk):ti,ab,kw (Word variations have been searched)
#11	(breast next tender*):ti,ab,kw (Word variations have been searched)
#12	#10 or #11
#13	#9 not #12
#14	MeSH descriptor: [Neoplasms] explode all trees
#15	#13 and #14
#16	(breast* near/5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or leiomyosarcoma* or dcis or duct* or infiltrat* or intraduct* or lobul* or medullary or tubular)):ti,ab,kw (Word variations have been searched)
#17	(mammar* near/5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or leiomyosarcoma* or dcis or duct* or infiltrat* or intraduct* or lobul* or medullary or tubular)):ti,ab,kw (Word variations have been searched)
#18	MeSH descriptor: [Paget's Disease, Mammary] this term only
#19	(paget* and (breast* or mammary or nipple*)):ti,ab,kw (Word variations have been searched)
#20	#15 or #16 or #17 or #18 or #19
#21	#6 or #20
#22	(adjuvant* next (online or model* or program* or tool*)):ti,ab,kw (Word variations have been searched)
#23	"www.adjuvantonline.com":ti,ab,kw (Word variations have been searched)
#24	adjuvantonline*:ti,ab,kw (Word variations have been searched)
#25	(PREDICT near/2 (online or model* or program* or tool* or estimat*)):ti,ab,kw (Word variations have been searched)
#26	"www.predict.nhs.uk":ti,ab,kw (Word variations have been searched)
#27	(predict next plus):ti,ab,kw (Word variations have been searched)
#28	(prognost* next index):ti,ab,kw (Word variations have been searched)
#29	"Nottingham Prognostic Index":ti,ab,kw (Word variations have been searched)
#30	NPI:ti,ab,kw (Word variations have been searched)
#31	FinProg*:ti,ab,kw (Word variations have been searched)
#32	CancerMath*:ti,ab,kw (Word variations have been searched)
#33	"www.CancerMath.net":ti,ab,kw (Word variations have been searched)
#34	#22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33
#35	#21 and #34 Publication Year from 1982 to 2017
#36	MeSH descriptor: [Decision Making, Computer-Assisted] explode all trees
#37	MeSH descriptor: [Computers] explode all trees
#38	MeSH descriptor: [Decision Support Systems, Clinical] explode all trees
#39	MeSH descriptor: [Software] explode all trees
#40	MeSH descriptor: [Decision Support Techniques] explode all trees
#41	MeSH descriptor: [Decision Making] explode all trees
#42	#36 or #37 or #38 or #39 or #40 or #41
#43	adjuvant*:ti,ab,kw (Word variations have been searched)

#	Searches
#44	#21 and #42 and #43 Publication Year from 2007 to 2017
#45	MeSH descriptor: [Survival Analysis] explode all trees
#46	MeSH descriptor: [Internet] explode all trees
#47	MeSH descriptor: [Databases, Factual] explode all trees
#48	MeSH descriptor: [Online Systems] explode all trees
#49	MeSH descriptor: [Web Browser] explode all trees
#50	MeSH descriptor: [User-Computer Interface] explode all trees
#51	#46 or #47 or #48 or #49 or #50
#52	#21 and #45 and #51 Publication Year from 1982 to 2017
#53	#35 or #44 or #52

Appendix C – Clinical evidence study selection

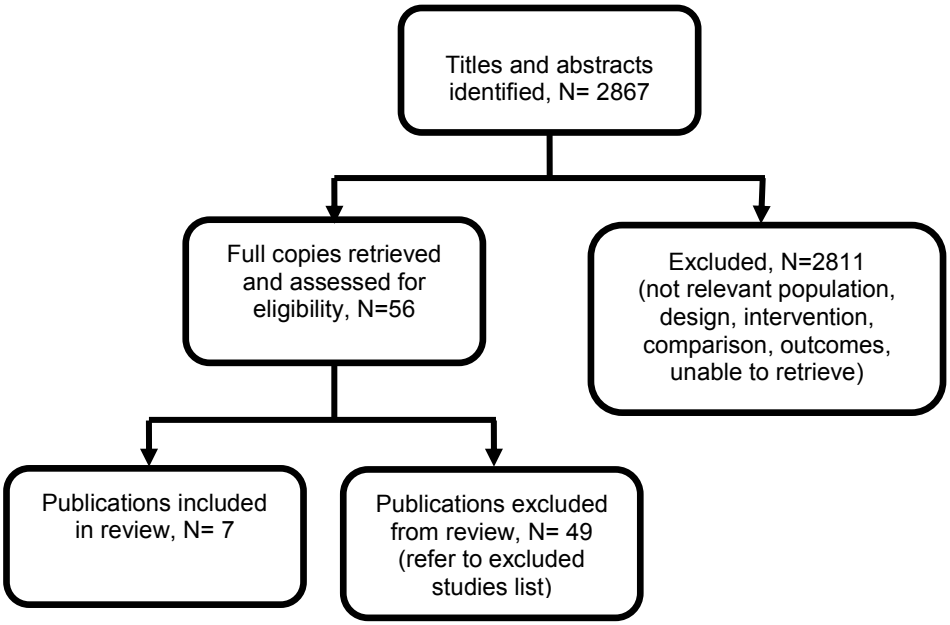
Clinical evidence study selection for 3.1 Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?

Figure 1: Flow diagram of clinical article selection for progesterone receptor testing



Clinical evidence study selection for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?

Figure 2: Flow diagram of clinical article selection for prognostic tools review



Appendix D – Clinical evidence tables

Clinical evidence tables for 3.1 Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?

There are no clinical evidence tables for this evidence review as no studies met the inclusion criteria.

Clinical evidence tables for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?

Table 10: Clinical evidence tables for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
Full citation Blamey, R. W., Ellis, I. O., Pinder, S. E., Lee, A. H. S., Macmillan, R. D., Morgan, D. A. L., Robertson, J. F. R., Mitchell, M. J., Ball, G. R., Haybittle, J. L., Elston, C. W., Survival of invasive breast cancer according to the Nottingham Prognostic Index in cases diagnosed in 1990-1999, European journal of cancer, 43, 1548-1555, 2007 Ref Id 583740 Country/ies where the study was carried out UK **Aim of the study**	**Sample size** 1980–1986 cohort: N = 892 1990–1999 cohort: N = 2238 **Characteristics** Not reported. **Inclusion criteria** Age 70 years or less Tumours of less than 5 cm diameter on clinical measurement and/or on operative histology **Exclusion criteria** Not reported.	**Prognostic tool** Nottingham Prognostic Index (NPI) Women were divided in six NPI groups: Excellent Prognostic Group (EPG) with an observed NPI range of 2.08–2.4; Good Prognostic Group (GPG) 2.42 to 63.4; Moderate I Prognostic Group (MPG I) 3.42 to 64.4; Moderate II Prognostic Group (MPG II) 4.42 to 65.4;	**Details** Nottingham Prognostic Index (NPI) Women were divided in six NPI groups: an Excellent Prognostic Group (EPG) with an observed NPI range of 2.08–2.4; Good Prognostic Group (GPG) 2.42 to 63.4; Moderate I Prognostic Group (MPG I) 3.42 to 64.4; Moderate II Prognostic Group (MPG II) 4.42 to 65.4; Poor Prognostic Group (PPG) 5.42 to 66.4; and Very Poor Prognostic Group (VPG) 6.5–6.8. **Sample selection** Consecutive women diagnosed with and treated for primary operable invasive breast cancer at Nottingham City	**Results** **Prognostic accuracy (sensitivity, specificity)** Not reported **Model calibration** Results only available for the 2000 to 2009 cohort ***10-year breast cancer survival*** Total cohort = not reported Excellent prognostic group (EPG) (n=320): Mortality ratio (%) O:E = 0.98 Good prognostic group (GPG) (n=475): Mortality ratio (%) O:E = 0.99 Moderate prognostic group 1 (MPG I) (n=634): Mortality ratio (%) O:E = 1.03	**Limitations** The quality of this study was assessed using the CASP tool for clinical prediction rule (CPR). A. Are the results valid? 1 Is the CPR clearly defined? Yes 2 The population from which the rule was derived included an appropriate spectrum of patients? Yes 3 Was the rule validated in a different group of patients? Yes 4 Were the predictor variables and the outcome evaluated in a blinded fashion? Not applicable (the outcome is mortality)

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
To report the predicted and actual survival figures for each NPI group. **Study dates** 1980 to 1986 and 1990 to 1999 **Source of funding** Not reported.		Poor Prognostic Group (PPG) 5.42 to 66.4; Very Poor Prognostic Group (VPG) 6.5–6.8.	Hospital between the years 1980–1986 inclusive (n = 892) and 1990–1999 inclusive (n = 2238) Cases in the 1980–1986 set came under the care of a single surgeon (RWB), with pathology by a single pathologist (CWE) Cases in the 1990s set were under the care of the integrated Breast Team at Nottingham City Hospital. Cases referred after an initial operation for diagnosis or following treatment carried out elsewhere were excluded. Note that authors excluded cases diagnosed in the years 1987–1989 because major changes in diagnosis and treatment were made in those years (for example the introduction of population screening, of expertise in radiology, case management by a	Moderate prognostic group 2 (MPG II) (n=489): Mortality ratio (%) O:E = 1.00 Poor prognostic group (PPG) (n=233): Mortality ratio (%) O:E = 1.02 Very poor prognostic group (VPG) (n=86): Mortality ratio (%) O:E = 0.89 **Tool discrimination** Not reported	5 Were the predictor variables and the outcome evaluates in the whole sample selected initially? Yes 6 Are the statistical methods used to construct and validate the rule clearly described? Yes B. What are the results? 7 Can the performance of the rule be calculated? No (not enough data is available to calculate sensitivity, specificity, LR+, LR+, ROC curve. Mortality ratio can be calculated) 8 How precise was the estimate of the treatment effect? The model considers the most relevant variables C. Will the results help locally? Are the results

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
			team of breast specialists in all disciplines, strict criteria for selection for breast conserving therapy, the introduction of selective local, regional and systemic adjuvant therapies). **Data collection** Women were followed up regularly and indefinitely in the hospital Primary Breast Clinic (PBC) and data on survival and recurrence recorded. At death the hospital notes are examined and deaths allocated to 'With/from breast cancer' or to 'Without known breast cancer'. Women with distant metastatic spread were allocated to the first group, even if the disease appeared to be in complete remission. **Data analysis**		applicable to the scenario? 9 Would the prediction rule be reliable and the results interpretable if used for your patient? Yes (UK population) 10 Is the rule acceptable in your case? Yes 11 Would the results of the rule modify your decision about the management of the patient or the information you can give to him/her? Yes **Indirectness** This study includes direct population (UK). **Other information**

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
			Life table survival curves for both breast cancer specific and all causes of death and for both time sets were done using SPSS version 13. Note that although in the early reports of the NPI, survival was from all causes of death, in this study the survival curves were constructed for death from breast cancer.		
Full citation Campbell, H. E., Gray, A. M., Harris, A. L., Briggs, A. H., Taylor, M. A., Estimation and external validation of a new prognostic model for predicting recurrence-free survival for early breast cancer patients in the UK, British journal of cancer, 103, 776-786, 2010 **Ref Id** 583803	**Sample size** Tool develpment N=1844 women with early invasive ductal carcinoma of the breast were used to develop the model (Churchill Hospital in London) Tool validation	**Prognostic tool** Oxford Prognostic Index (OPI)	**Details** **Sample selection** Tool development Women were consecutively diagnosed. All of them underwent surgery at the Churchill Hospital, Oxford.They were followed-up was untill	**Results** **Results reported for external validation of the tool only** **Prognostic accuracy (sensitivity, specificity)** Not reported **Tool calibration** ***5-year recurrent free survival*** Total cohort (n=1789) o RFS ratio O:E: 1.01 o Difference O-E: 0.7%	**Limitations** The quality of this study was assessed using the CASP tool for clinical prediction rule (CPR). A. Are the results valid? 1 Is the CPR clearly defined? Yes 2 The population from which the rule was derived included an appropriate spectrum of patients? Yes

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
Country/ies where the study was carried out UK **Aim of the study** To develop and validate a new prognostic tool (Oxford Prognostic Index, OPI), for predicting recurrence in women with early breast cancer. **Study dates** 1986 to 2001 **Source of funding** The Cancer Research UK and the Oxford NHS Comprehensive Biomedical Research Centre.	N=1787 women with invasive ductal carcinoma **Characteristics** Tool development Age: <35: 56 (3.04%) ≥35 to <45: 257 (13.95%) ≥45 to <55: 533 (28.94%) ≥55 to <65: 473 (25.68%) ≥65 to <75: 388 (21.06%) ≥75: 135 (7.33%) Unknown: 2 (0%) Positive nodes: 0: 1070 (60.45%) 1: 258 (14.58%) 2: 142 (8.02%)		31 January 2006, and completion rate was 89. Tool validation The ABC subset included 1789 patients from 70 hospitals from the UK. Dates 1992 to 2000. **Data collection** Not reported. **Data analysis** Tool develpment Model estimation was conducted using STATA. A parametric regression-based survival model was estimated on time from initial surgery to a first recurrent event or censoring (patients were censored when they died from causes unrelated to breast cancer without recurrence being first recorded (n¼111/1844) or were lost to follow-up	Sub-group: age • ≤50 years (n=1097) o RFS ratio O:E: 1,03 o Difference O-E: 1.92 • > 50 years (n=690) o RFS ratio O:E: 1.00 o Difference O-E: -0.10 Sub-group: tumour grade • Grade 1 (n=196) o RFS ratio O:E: 1.06 o Difference O-E: 5.15% • Grade 2 (n=772) o RFS ratio O:E: 1.03 o Difference O-E: 2.44% • Grade 3 (n=819) o RFS ratio O:E: 0.98 o Difference O-E: -1.04% Sub-group: tumour size • ≤2 cm (n=954) o RFS ratio O:E: 1.06 o Difference O-E: 4.6% • >2 cm to ≤5 cm (n=772) o RFS ratio O:E: 0.95 o Difference O-E: -3.16% • >5 cm (n=61) o RFS ratio O:E: 1.04 o Difference O-E: 2.47% Sub-group: nodal status • Negative (n=674) o RFS ratio O:E: 1.02 o Difference O-E: 1.82% • Positive (n=1113)	3 Was the rule validated in a different group of patients? Yes (the study shows results for both development and validation) 4 Were the predictor variables and the outcome evaluated in a blinded fashion? Not applicable (the outcome is mortality) 5 Were the predictor variables and the outcome evaluates in the whole sample selected initially? Yes 6 Are the statistical methods used to construct and validate the rule clearly described? Yes B. What are the results? 7 Can the performance of the rule be calculated? No (not enough data is available to calculate sensitivity,

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
	3: 92 (5.20%) 4: 52 (2.94%) 5: 33 (1.86%) 6: 35 (1.98%) 7: 21 (1.19%) 8: 16 (0.90%) 9: 10 (0.56%) 10+: 41 (2.32%) Unknown: 74 (0%) Tumour size <1 cm: 204 (11.16%) ≥1 cm and <2 cm: 644 (35.23%) ≥2 cm and <3 cm: 562 (30.74%) ≥3 cm and <4 cm: 238 (13.02%) ≥4 cm and <5 cm: 77 (4.21%) ≥5 cm: 103 (5.63%)		without any previous diagnosis of recurrence). Tool validation Details not reported	o RFS ratio O:E: 1.01 o Difference O-E: 0.71% Sub-group: ER status • Negative (n=755) o RFS ratio O:E: 1.03 o Difference O-E: 2.05% • Positive (n=1032) o RFS ratio O:E: 1.01 o Difference O-E: 0.46% **Tool discrimination** ***5-year recurrent free survival*** • Overall C = 0.720 (95%CI 0.693 to 0.746)	specificity, LR+, LR+, ROC curve. Mortality ratio can be calculated) 8 How precise was the estimate of the treatment effect? The model considers the most relevant variables C. Will the results help locally? Are the results applicable to the scenario? 9 Would the prediction rule be reliable and the results interpretable if used for your patient? Yes (UK population) 10 Is the rule acceptable in your case? No (this tool has never been made available) 11 Would the results of the rule modify your decision about the management of the patient or the information you can give to him/her? Can't tell (as indicated above this tool

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
	Unknown: 16 (0%) Tumour grade 1: 329 (18.97%) 2: 770 (44.41%) 3: 635 (36.62%) Unknown: 110 (0%) ER status Negative: 477 (33.33%) Positive: 954 (66.67%) Unknown: 413 (0%) Tool validation Not reported **Inclusion criteria** Tool validation A sub-set of women from the UK obtained from the the Adjuvant				has never been made available in clinical practice) **Indirectness** The study includes direct UK population. **Other information** Conflict of interest: not reported (however sources of funding have been reported). Results for the external validation of the tool are reported here.

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
	Breast Cancer (ABC) trial. All women had invasive ductal cancer. **Exclusion criteria** Not reported				
Full citation Campbell, H. E., Taylor, M. A., Harris, A. L., Gray, A. M., An investigation into the performance of the Adjuvant! Online prognostic programme in early breast cancer for a cohort of patients in the United Kingdom, British journal of cancer, 101, 1074-1084, 2009 **Ref Id** 583804 **Country/ies where the study was carried out** UK **Aim of the study** To evaluate the performance of the	**Sample size** N=1065 women with early breast cancer **Characteristics** Not reported **Inclusion criteria** Up to 85 years With complete data on nodal status, tumour size, and adjuvant systemic therapy People who had undergone complete local therapy Complete 10-year follow-up. **Exclusion criteria**	**Prognostic tool** Adjuvant! Online	**Details** **Sample selection** All people diagnosed with breast cancer patients consecutively between 1986 and 1996 at the Churchill Hospital in Oxford. **Data collection** Patients were followed up on an annual basis through the Cancer Intelligence Network and General Practitioners, who provided information on recurrence and survival status. Observed 10-year outcomes for each	**Results** **Prognostic accuracy (sensitivity, specificity)** Not reported **Tool calibration** ***10-year overall survival*** All population (N=1065): Mortality ratio O:E = 0.93; Difference O-E = -5.54 (p<0.01) *Sub-group: age* 20 to 35 (n=34): Mortality ratio O:E = 0.97; Difference O-E = -2.27% (n.s.) 36 to 50 (n=363): Mortality ratio O:E = 0.95; Difference O-E = -4.33% (p<0.05)	**Limitations** The quality of this study was assessed using the CASP tool for clinical prediction rule (CPR). A. Are the results valid? 1 Is the CPR clearly defined? Yes 2 The population from which the rule was derived included an appropriate spectrum of patients? Yes 3 Was the rule validated in a different group of patients? Yes 4 Were the predictor variables and the outcome evaluated in a blinded fashion? Not

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
Adjuvant! Online programme by comparing its 10-year predictions with observed outcomes in people with early breast cancer. **Study dates** 1986 to 1996 **Source of funding** NIHR Biomedical Research Centre Programme, Oxford, and by Cancer Research UK	Women with locally advanced disease (those with T3 (45 cm tumour) and T4 (tumour of any size growing into the skin or chest wall) tumours, and those with N2 (4–9 nodes involved) and N3 (10 or more lymph nodes involved) tumours) Women with metastatic disease (M1). Note: this is because Adjuvant! was developed for 'adjuvant' decision-making in those where benefit is less certain		woman were available from the Churchill Hospital data set. The programme was used to generate 10-year predictions of OS, BCSS, and EFS by entering information on each patient's age, tumour size, number of positive nodes, grade, ER status, and adjuvant systemic therapies received (types of hormone and chemotherapies). **Data analysis** Comparisons between predicted and observed outcomes (OS, BCSS, and EFS) were conducted for the whole cohort, and for clinically important subgroups. For each of these separate analyses, Kaplan–Meier survival analysis provided observed 10-year percentages. Predicted 10-year percentages	51 to 65 (n=458): Mortality ratio O:E = 0.95; Difference O-E = -4.02% (p<0.05) 66 to 75 (n=194): Mortality ratio O:E = 0.82; Difference O-E = -12.17% (p<0.01) ≥76 (n=16): Mortality ratio O:E = 0.94; Difference O-E = -3.11% (n.s.) *Sub-group: grade:* Grade 1 (n=152): Mortality ratio O:E: 0.96; Difference O-E: -3.65% (n.s.) Grade 2 (n=421): Mortality ratio O:E: 0.91; Difference O-E: -7.05% (p<0.01) Grade 3 (n=248): Mortality ratio O:E: 0.86; Difference O-E: -9.82% (p<0.01) Unknown grade (n=244): Mortality ratio O:E: 1.00; Difference O-E: 0.26% (n.s.) *Sub-group: tumour size:* 0.1 to 1 cm (n=150): Mortality ratio O:E: 0.93; Difference O-E: -6.10% (n.s.)	applicable (the outcome is mortality) 5 Were the predictor variables and the outcome evaluates in the whole sample selected initially? Yes 6 Are the statistical methods used to construct and validate the rule clearly described? Yes B. What are the results? 7 Can the performance of the rule be calculated? No (not enough data is available to calculate sensitivity, specificity, LR+, LR+, ROC curve. Mortality ratio can be calculated) 8 How precise was the estimate of the treatment effect? Yes (the model considers the most relevant variables)

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
			were given by averaging over the relevant Adjuvant! predictions. Statistical uncertainty around these differences was assessed by way of a t-test, the statistic for which was calculated by dividing the difference between predicted and observed percentages by s.e. for the observed percentages.	1.1 to 2 cm (n=471): Mortality ratio O:E: 0.92; Difference O-E: -6.57% (p<0.01) 2.1 to 5 cm (n=444): Mortality ratio O:E: 0.94; Difference O-E: -4.26% (n.s.) *Sub-group: nodal involvement* Negative (n=733): Mortality ratio O:E = 0.94; Difference O-E = -4.70% (p<0.01) Positive (n=332): Mortality ratio O:E = 0.89; Difference O-E = -7.38% (p<0.01) *Sub-group: ER status* Negative (n=261): Mortality ratio O:E = 0.97; Difference O-E = -1.93% (n.s.) Positive (n=495): Mortality ratio O:E = 0.89; Difference O-E = -9.00% (p<0.01) Unknown (n=309): Mortality ratio O:E = 0.96; Difference O-E = -3.04% (n.s.) ***Tool discrimination*** Not reported	C. Will the results help locally? Are the results applicable to the scenario? 9 Would the prediction rule be reliable and the results interpretable if used for your patient? Yes (UK population) 10 Is the rule acceptable in your case? No (this tool is not currently available) 11 Would the results of the rule modify your decision about the management of the patient or the information you can give to him/her? Yes (although as noted above this tool is no longer available) **Indirectness** This study includes direct population (UK based study). **Other information**

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
				10-year breast cancer specific survival All population (N=1058):Mortality ratio O:E: 0.95; Difference O-E: -4.53% (p<0.01) *Sub-group: age* 20 to 35 (n=34); Mortality ratio O:E = 0.99; Difference O-E = -0.67% (n.s.) 36 to 50 (n=361): Mortality ratio O:E = 0.94; Difference O-E = -4.62% (p<0.05) 51 to 65 (n=454): Mortality ratio O:E = 0.96; Difference O-E = -3.51% (n.s.) 66 to 75 (n=193): Mortality ratio O:E = 0.89; Difference O-E = -9.28% (p<0.05) ≥76 (n=16): Mortality ratio O:E = 1.08; Difference O-E = 7.04% (n.s.) *Sub-group - grade:* Grade 1 (n=152): Mortality ratio O:E: 0.99; Difference O-E: -1.29% (n.s.)	Conflict of interest: not explicitly reported

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
				Grade 2 (n=420): Mortality ratio O:E: 0.93; Difference O-E: -5.89% (p<0.01) Grade 3 (n=243): Mortality ratio O:E: 0.92; Difference O-E: -6.10 (p<0.05) Unknown grade (n=243): Mortality ratio O:E: 0.96; Difference O-E: -2.78 (n.s.) *Sub-group – tumour size:* 0.1 to 1 cm (n=148): Mortality ratio O:E: 0.92; Difference O-E: -7.95% (p<0.01) 1.1 to 2 cm (n=470): Mortality ratio O:E: 0.95; Difference O-E: -4.54% (p<0.01) 2.1 to 5 cm (n=440): Mortality ratio O:E: 0.95; Difference O-E: -3.53% (n.s.) *Sub-group: nodal involvement* Negative (n=729): Mortality ratio O:E = 0.96; Difference O-E = -3.53% (p<0.01) Positive (n=329): Mortality ratio O:E = 0.91; Difference O-E = -6.73% (p<0.01)	

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
				Sub-group: ER status Negative (n=259): Mortality ratio O:E = 0.96; Difference O-E = -2.76% (n.s.) Positive (n=491): Mortality ratio O:E = 0.92; Difference O-E = -6.62% (p<0.01) Unknown (n=308): Mortality ratio O:E = 0.96; Difference O-E = -2.74% (n.s.) **Tool discrimination** Not reported Note: mortality ratios were calculated by the NGA technical team with the data available in the study	
Full citation Maishman, T., Copson, E., Stanton, L., Gerty, S., Dicks, E., Durcan, L., Wishart, G. C., Pharoah, P., Eccles, D., An	**Sample size** N=3000 young women diagnosed with breast cancer **Characteristics**	**Prognostic tool** PREDICT version 1.2	**Details** **Sample selection** This study used data from the POSH multicentre prospective	**Results** **Prognostic accuracy (sensitivity, specificity)** Not reported	**Limitations** The quality of this study was assessed using the CASP tool for clinical prediction rule (CPR).

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
evaluation of the prognostic model PREDICT using the POSH cohort of women aged <40 years at breast cancer diagnosis, British journal of cancer, 112, 983-991, 2015 **Ref Id** 584799 **Country/ies where the study was carried out** UK **Aim of the study** To assess how well PREDICT v1.2 performs in estimating survival in a cohort of young women. **Study dates** 2000 to 2008 **Source of funding** Cancer Research UK provided funding for data collection and analysis for the POSH study. The study was sponsored by University Hospital	Young women ≤40 diagnosed with breast cancer **Inclusion criteria** Women from the POSH multicentre prospective observational cohort study **Exclusion criteria** Not reported.		observational cohort study. This study included 3000 young women diagnosed with breast cancer between 2000 and 2008 in the UK. **Data collection** The data obtained from the POSH study included: age at diagnosis, ethnicity, menopausal status, family history of breast cancer, ER, PR, and HER2 status, histology, histological grade, tumour size, number of positive lymph nodes, lymphovascular invasion status, focality, presentation, gene status, and type of adjuvant therapy. In this study, the authors report the analyses conducted on follow-up data from the POSH cohort. Outcomes included OS and BCSS at 5, 8, and 10 year (OS = time from	**Tool calibration** ***5-year all-cause mortality*** Total cohort (N=2827) Mortality ratio O:E = 1.33; Difference O-E = 25% (n=152) *Sub-group: age at diagnosis* 18 to 25 (n=40): Mortality ratio O:E = 1.4; Difference O-E = 28.6% (n=2) 26 to 30 (n=258): Mortality ratio O:E = 1.35; Difference O-E = 25.8% (n=16) 31 to 35 (n=864): Mortality ratio O:E = 1.38; Difference O-E = 27.6% (n=58) 36 to 40 (n=1665): Mortality ratio O:E = 1.30; Difference O-E = 23.2% (n=76) *Sub-group: grade* Grade 1 (n=156): Mortality ratio O:E = 1.25; Difference O-E = 20% (n=1) Grade 2 (n=929): Mortality ratio O:E = 2.40; Difference O-E = 58.4% (n=94)	A. Are the results valid? 1 Is the CPR clearly defined? Yes 2 The population from which the rule was derived included an appropriate spectrum of patients? Yes 3 Was the rule validated in a different group of patients? Yes (this tool has been validated in a number of studies, this study aims to evaluate it in a cohort of young women) 4 Were the predictor variables and the outcome evaluated in a blinded fashion? Not applicable (the outcome is mortality) 5 Were the predictor variables and the outcome evaluates in the whole sample selected initially? Yes 6 Are the statistical methods used to

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
Southampton NHS Foundation Trust.			breast cancer diagnosis to death from any cause; BCSS = time to death from breast cancer, with deaths from other causes censored at the time of last follow-up). **Data analysis** The tool calibration was calculated comparing predicted and observed mortality. The tool discrimination was calculated using the area under the receiver-operator characteristic curve (AUC) and corresponding 95% confidence intervals for 5-, 8-, and 10-year predicted all-cause mortality and breast cancer-specific mortality. Analyses were done using STATA v12.1.	Grade 3 (n=1676): Mortality ratio O:E = 1.13; Difference O-E = 11.9% (n=51) Unknown (n=66): Mortality ratio O:E = 1.71; Difference O-E = 41.7% (n=5) *Sub-group: tumour size* 0 to 10 mm (n=265): Mortality ratio O:E = 2.1; Difference O-E = 52.4% (n=22) 11 to 20 mm (n=930): Mortality ratio O:E = 1.25; Difference O-E = 20% (n=25) 21 to 50 mm (n=1229): Mortality ratio O:E = 1.26; Difference O-E = 22.8% (n=69) >50 mm (n=244): Mortality ratio O:E = 1.16: Difference O-E = 14% (n=85) Unknown (n=159): Mortality ratio O:E = 2.44; Difference O-E = 59% (n=23) *Sub-group: node status* Negative (n=1370): Mortality ratio O:E = 1.26: Difference O-E = 20.5% (n=33)	construct and validate the rule clearly described? Yes B. What are the results? 7 Can the performance of the rule be calculated? No (not enough data is available to calculate sensitivity, specificity, LR+, LR+, ROC curve. Mortality ratio can be calculated) 8 How precise was the estimate of the treatment effect? The model considers the most relevant variables C. Will the results help locally? Are the results applicable to the scenario? 9 Would the prediction rule be reliable and and the results interpretable if used for your patient? Yes (UK population, but this study shows the tool

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
				Positive (n=1431): Mortality ratio O:E = 1.35: Difference O-E = 26.2% (n=115) Unknown (n=26): Mortality ratio O:E = 1.75: Difference O-E = 42.9% (n=3) *ER status* Negative (n=965): Mortality ratio O:E = 0.82; Difference O-E = -21.2% (n=-52) Positive (n=1862): Mortality ratio O:E = 2.29; Difference O-E = 56.4% (n=204) *HER2 status* Negative (n=1773): Mortality ratio O:E = 1.50; Difference O-E = 33.4% (n=128) Positive (n=679): Mortality ratio O:E = 1.15; Difference O-E = 13.1% (n=24) Borderline (n=40): Mortality ratio O:E = 1.67; Difference O-E = 40% (n=4) Unknown (n=335): Mortality ratio O:E = 0.88; Difference O-E = -12.9% (n=-4)	is not accurate in young women) 10 Is the rule acceptable in your case? Yes (this tool cannot be used in young women) 11 Would the results of the rule modify your decision about the management of the patient or the information you can give to him/her? Yes (this tool cannot be used in young women) **Indirectness** This study includes direct population (UK based). **Other information** Conflict of interest: EC received honoraria from Roche. All other authors declare no conflict of interest. PREDICT was contacted to determine which version of PREDICT was used in

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
				10-year all-cause mortality Total cohort (N=597): Mortality ratio O:E = 0.93; Difference O-E = -7.9% (n=-12) *Sub-group: age at diagnosis* 18 to 25 (n=8): Mortality ratio O:E = 1 o Difference O-E = 0% (n=0) 26 to 30 (n=55): Mortality ratio O:E = 0.94 o Difference O-E = -6.7% (n=-1) 31 to 35 (n=203): Mortality ratio O:E = 1.05 o Difference O-E = 5% (n=3) 36 to 40 (n=331): Mortality ratio O:E = 0.84 o Difference O-E = -18.4% (n=-14) *Sub-group: grade* Grade 1 (n=31): Mortality ratio O:E = 1.5 o Difference O-E = 33% (n=1) Grade 2 (n=200): Mortality ratio O:E = 1.42 o Difference O-E = 30% (n=13)	this study (info@predict.nhs.uk)

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
				Grade 3 (n=351): Mortality ratio O:E = 0.80 o Difference O-E = -25.5% (n=-26) Unknown (n=15): Mortality ratio O:E = 1 o Difference O-E = 0% (n=0) *Sub-group: tumour size* 0 to 10 (n=48): Mortality ratio O:E = 2; Difference O-E = 50% (n=7) 11 to 20 (n=221): Mortality ratio O:E = 0.91; Difference O-E = -9.8% (n=-4) 21 to 50 (n=244): Mortality ratio O:E = 0.99; Difference O-E = -1.3% (n=-1) >50 (n=54): Mortality ratio O:E = 0.46; Difference O-E = -115.4% (n=-15) Unknown (n=30): Mortality ratio O:E = 1.2; Difference O-E = 16.7% (n=1) *Sub-group: node status* Negative (n=266): Mortality ratio O:E = 0.93; Difference O-E = -7.7% (n=-3)	

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
				Positive (n=327): Mortality ratio O:E = 0.92; Difference O-E = 8% (n=9) Unknown (n=4): Mortality ratio O:E = 1; Difference O-E = 0% (n=0) *Sub-group: ER status* Negative (n=231): Mortality ratio O:E = 0.68; Difference O-E = -46.9% (n=-30) Positive (n=366): Mortality ratio O:E = 1.26; Difference O-E = 20.5% (n=18) *Sub-group: HER2 status* Negative (n=327): Mortality ratio O:E = 0.99; Difference O-E = -1.2% (n=-1) Positive (n=140): Mortality ratio O:E = 0.94; Difference O-E = -6% (n=-3) Borderline (n=14): Mortality ratio O:E = 1.25; Difference O-E = 20% (n=1) Unknown (n=116): Mortality ratio O:E = 0.62; Difference O-E = -60% (n=-9)	

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
				Tool discrimination ***5-year all-cause mortality*** AUC ER- vs ER+ = 0.718 vs 0.730 (discrimination was better for ER+ tumours, compared to ER- tumours) ***10-year all-cause mortality*** AUC ER- vs ER+ = 0.694 vs 0.724 (discrimination was better for ER+ tumours, compared to ER- tumours) AUC HER2- vs HER2+ =0.724 vs 0.592 (discrimination was better for HER2- tumours, compared to HER2+ tumours)	
Full citation Wishart, G. C., Azzato, E. M., Greenberg, D. C., Rashbass, J., Kearins, O., Lawrence, G., Caldas, C., Pharoah, P. D., PREDICT: a new UK prognostic model that predicts	**Sample size** N=5468 patients from the West Midlands Cancer Intelligence Unit (WMCIU) **Characteristics**	**Prognostic tool** PREDICT v1.0	**Details** Model discrimination was assessed using the area under the receiver-operator-characteristic (ROC) curve (AUC) calculated for the	**Results** **Prognostic accuracy (sensitivity, specificity)** Not reported	**Limitations** The quality of this study was assessed using the CASP tool for clinical prediction rule (CPR). A. Are the results valid?

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
survival following surgery for invasive breast cancer.[Erratum appears in Breast Cancer Res. 2010;12(2):401], Breast Cancer Research, 12, R1, 2010 **Ref Id** 585712 **Country/ies where the study was carried out** UK **Aim of the study** To develop and validate a model to predict overall and breast cancer specific survival for women treated for early breast cancer. **Study dates** 1999 and 2003 **Source of funding** Educational grant from Pfizer Limited. GCW & CC receive research funding from the Cambridge NIHR Biomedical Research Centre.	Median follow-up (years): 4.85 (0.07 to 8.00) Median age at diagnosis: 58 (22 to 93) *Age, years* <35: 2% (n=108) 35 to 49: 22% (n=1,195) 50 to 64: 44% (n=2,393) 65 to 74: 20% (n=1,101) 75+: 12% (n=671) *Nodal status* 0: 58% (n=3,184) 1: 14% (n=746) 2 to 4: 14% (n=792) 5 to 9: 8% (n=451) 10+: 5% (n=295) *Tumour size, mm*		overall deaths at 8 years after diagnosis. Model calibration was calculated using a simplified goodness-of-fit method for the Cox proportional hazards mode, where observed and model-based predicted deaths at 5 and 8 years were compared. Observed and predicted deaths were compared using a standard Chi-squared test. The analyses were conducted using STATA, version 9.2. **Sample selection** The primary analysis was based on data from patients with invasive breast cancer diagnosed in East Anglia, UK between 1999 and 2003 identified by the Eastern Cancer Registration and Information Centre. The validation study was conducted using	**Tool calibration and discrimination** ***5-year all-cause mortality*** Total cohort (N=5468): Mortality ratio O:E = 0.91; Difference O-E = -1.61% *Sub-group: age* <35 (n=108); Mortality ratio O:E = 0.88; Difference O-E = -2.78% 35 to 49 (n=1195); Mortality ratio O:E = 0.83; Difference O-E = -2.68% 50 to 67 (n=2393); Mortality ratio O:E = 0.90; Difference O-E = -1.34% 65 to 74 (n=1101); Mortality ratio O:E = 0.98; Difference O-E = -0.45% 75+ (n=671); Mortality ratio O:E = 0.98; Difference O-E = -0.75% *Sub-group: grade* Grade 1 (n=1017): Mortality ratio O:E = 0.98; Difference O-E = -0.1%	1 Is the CPR clearly defined? Yes 2 The population from which the rule was derived included an appropriate spectrum of patients? Yes 3 Was the rule validated in a different group of patients? Yes (this study aims to develop and validate the tool. A different group of women was used to validate the tool) 4 Were the predictor variables and the outcome evaluated in a blinded fashion? Not applicable (the outcome is mortality) 5 Were the predictor variables and the outcome evaluates in the whole sample selected initially? Yes 6 Are the statistical methods used to construct and validate

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
	<10: 9% (n=485) 10 to 19: 39% (n=2,136) 20 to 29: 29% (n=1,566) 30 to 49: 17% (n=923) 50+: 7% (n=358) *Grade* I: 19% (n=1,017) II: 45% (n=2,442) III: 37% (n=2,009) *ER Status* ER-: 20% (n=1,116) ER+: 80% (n=4,352) **Inclusion criteria** West Midlands Cancer Intelligence Unit (WMCIU) registry data. No details reported. **Exclusion criteria** Not reported.		data from the West Midlands Cancer Intelligence Unit (WMCIU). This cohort included all women diagnosed with invasive breast cancer between 1999 and 2003. **Data collection** The following information was obtained from WMCIU database: age at diagnosis, number of lymph nodes sampled and number of lymph nodes positive (categorised as 0, 1, 2 to 4, 5 to 9, and 10+ nodes positive), tumour size (categorised as <10 mm, 10 to 19 mm, 20 to 29 mm, 30 to 49 mm, 50+ mm), histological grade (I, II, III), oestrogen receptor (ER) status (positive or negative), mode of detection (screening vs. clinical), information on local therapy (wide local excision, mastectomy, radiotherapy), and type	Grade 2 (n=2442): Mortality ratio O:E = 0.98; Difference O-E = -0.16% Grade 3 (n=2009): Mortality ratio O:E = 0.87; Difference O-E = -3.58% *Sub-group: tumour size* <10 mm (n=485): Mortality ratio O:E = 0.84; Difference O-E = -1.03% 10 to 19 mm (n=2136): Mortality ratio O:E = 0.88; Difference O-E = -2.01% 20 to 29 mm (n=1566): Mortality ratio O:E = 0.94; Difference O-E = -0.96% 30 to 49 mm (n=923): Mortality ratio O:E = 0.99; Difference O-E = -0.11% 50+ mm (n=358): Mortality ratio O:E = 0.91; Difference O-E = -3.91% *Sub-group: nodal status* Negative (n=3184): Mortality ratio O:E = 0.80; Difference O-E = -2.14%	the rule clearly described? Yes B. What are the results? 7 Can the performance of the rule be calculated? No (not enough data is available to calculate sensitivity, specificity, LR+, LR+, ROC curve. Mortality ratio can be calculated) 8 How precise was the estimate of the treatment effect? The model considers many relevant variables, but updates of this version include additional factors C. Will the results help locally? Are the results applicable to the scenario? 9 Would the prediction rule be reliable and and the results interpretable

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
			of adjuvant systemic therapy (chemotherapy, endocrine therapy, both). **Data analysis**	Positive (n=2284): Mortality ratio O:E = 0.98; Difference O-E = -0.39% *Sub-group: ER status* Negative (n=1116): Mortality ratio O:E = 0.87; Difference O-E = -4.21% AUC 0.81 (SE 0.0111) Positive (n=4352): Mortality ratio O:E = 0.95; Difference O-E = -0.69% AUC 0.75 (SE 0.0169) ***8-year all-cause mortality (proxy for long-term)*** Total cohort (N=5468): Mortality ratio O:E = 0.95; Difference O-E = -0.93% *Sub-group: age* <35 (n=108); Mortality ratio O:E = 1.08; Difference O-E = 1.85% 35 to 49 (n=1195); Mortality ratio O:E = 0.87; Difference O-E = -2.18%	if used for your patient? Yes (UK population) 10 Is the rule acceptable in your case? Yes 11 Would the results of the rule modify your decision about the management of the patient or the information you can give to him/her? Yes **Indirectness** This study includes direct population (UK based). **Other information** Conflict of interest: none For the purpose of this review, we have only considered the validation data (WMCIU cohort). Breast cancer specific deaths were not reported in detail in the published study. PREDICT was contacted to determine

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
				50 to 67 (n=2393); Mortality ratio O:E = 0.92; Difference O-E = -1% 65 to 74 (n=1101); Mortality ratio O:E = 1.00; Difference O-E = -0.09% 75+ (n=671); Mortality ratio O:E = 0.98; Difference O-E = -0.6% *Sub-group: grade* Grade 1 (n=1017): Mortality ratio O:E = 1.04; Difference O-E = 0.29% Grade 2 (n=2442): Mortality ratio O:E = 1.04; Difference O-E = 0.61% Grade 3 (n=2009): Mortality ratio O:E = 0.88; Difference O-E = -3.38% *Sub-group: tumour size* <10 mm (n=485): Mortality ratio O:E = 0.85; Difference O-E = -1.03% 10 to 19 mm (n=2136): Mortality ratio O:E = 0.84; Difference O-E = -1.73%	which version of PREDICT was used in this study (info@predict.nhs.uk)

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
				20 to 29 mm (n=1566): Mortality ratio O:E = 0.97; Difference O-E = -0.57% 30 to 49 mm (n=923): Mortality ratio O:E = 0.98; Difference O-E = -0.43% 50+ mm (n=358): Mortality ratio O:E = 0.56; Difference O-E = -3.35% *Sub-group: nodal status* Negative (n=3184): Mortality ratio O:E = 0.84; Difference O-E = -1.76% Positive (n=2284): Mortality ratio O:E = 1.01; Difference O-E = 0.26% *Sub-group: ER status* Negative (n=1116): Mortality ratio O:E = 0.90; Difference O-E = -3.49% Positive (n=4352): Mortality ratio O:E = 0.98; Difference O-E = -0.25% ***8-year all-cause mortality (proxy for long-term)***	

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
				Total cohort (N=5468): AUC (SE) = 0.79 (0.008) *Sub-group: age* <35 (n=108); AUC (SE) = 0.70 (0.057) 35 to 49 (n=1195); AUC (SE) = 0.79 (0.018) 50 to 67 (n=2393); AUC (SE) = 0.80 (0.013) 65 to 74 (n=1101); AUC (SE) =0.76 (0.018) 75+ (n=671); AUC (SE) = 0.72 (0.021) *Sub-group: grade* Grade 1 (n=1017): AUC (SE) = 0.79 (0.029) Grade 2 (n=2442): AUC (SE) = 0.77 (0.013) Grade 3 (n=2009): AUC (SE) = 0.75 (0.012) *Sub-group: tumour size* <10 mm (n=485): AUC (SE) = 0.82 (0.040)	

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
				10 to 19 mm (n=2136): AUC (SE) = 0.76 (0.018) 20 to 29 mm (n=1566): AUC (SE) = 0.71 (0.017) 30 to 49 mm (n=923): AUC (SE) = 0.72 (0.018) 50+ mm (n=358): AUC (SE) = 0.72 (0.027) *Sub-group: nodal status* Negative (n=3184): AUC (SE) = 0.74 (0.015) Positive (n=2284): AUC (SE) = 0.75 (0.011) *Sub-group: ER status* Negative (n=1116): AUC (SE) = 0.76 (0.016) Positive (n=4352): AUC (SE) = 0.78 (0.010)	
Full citation Wishart, G. C., Rakha, E., Green, A., Ellis, I., Ali, H. R., Provenzano, E., Blows, F. M., Caldas, C.,	**Sample size** Data for 2232 cases of invasive breast cancer treated in Nottingham - 506 node-negative	**Prognostic tool** PREDICT v1.1 and v1.2	**Details** **Sample selection** **Data collection**	**Results** **Results for PREDICT v1.1** **Prognostic accuracy (sensitivity, specificity)**	**Limitations** The quality of this study was assessed using the CASP tool for clinical prediction rule (CPR).

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
Pharoah, P. D., Inclusion of KI67 significantly improves performance of the PREDICT prognostication and prediction model for early breast cancer, BMC cancer, 14, 908, 2014 **Ref Id** 585718 **Country/ies where the study was carried out** UK **Aim of the study** To incorporate the prognostic effect of KI67 status in a new version of PREDICT(v1.2), and to compare performance with the previous version of PREDICT that includes HER2 status (v1.1). **Study dates** 1989 to 1998 **Source of funding** Cancer Research UK grant (C490/A10124).	cases were excluded, so data from n=1726 people was included in the study **Characteristics** Not reported. **Inclusion criteria** No details reported. **Exclusion criteria** Not reported.		The data was obtained from the Nottingham dataset. This included: age at diagnosis, histological grade, tumour size, number of positive lymph nodes, ER status, HER2 status, KI67 and type of adjuvant systemic therapy (none, chemotherapy, endocrine therapy, both). The missing data was replaced with the mean for that variable. **Data analysis** The tool calibration was evaluated within quintile of predicted mortality. A goodness-of-fit test was carried out by using a χ2-test based on the observed and predicted number of events within each quintile. The tool discrimination was calculated using	Not reported **Tool calibration** ***10-year breast cancer mortality*** Total cohort (N=1726): BC mortalityratio O:E = 1.13 *Sub-group: age* <40 (n=67): BC mortality ratio O:E = 1.15 40 to 49 (n=274): BC mortality ratio O:E = 1.18 50 to 59 (n=436): BC mortality ratio O:E = 1.18 60+ (n=497): BC mortality ratio O:E = 1.06 *Sub-group: tumour size* <10 (n=144): BC mortality ratio O:E = 0.78 10 to 19 (n=574): BC mortality ratio O:E = 1.09 20 to 29 (n=404): BC mortality ratio O:E = 1.32	A. Are the results valid? 1 Is the CPR clearly defined? Yes 2 The population from which the rule was derived included an appropriate spectrum of patients? Yes 3 Was the rule validated in a different group of patients? Yes (this study aims to validate a new version of the tool) 4 Were the predictor variables and the outcome evaluated in a blinded fashion? Not applicable (the outcome is mortality) 5 Were the predictor variables and the outcome evaluates in the whole sample selected initially? Yes 6 Are the statistical methods used to construct and validate the rule clearly described? Yes

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
			the area under the receiver-operator-characteristic curve (AUC), for 10-year mortality.	30 to 49 (n=140): BC mortality ratio O:E = 0.95 50+ (n=11): BC mortality ratio O:E = 0.5 Missing (n=1): BC mortality ratio O:E = 1 *Sub-group: node status* Negative (n=709): BC mortality ratio O:E = 1.19 1+ (n=241): BC mortality ratio O:E = 1.23 2 to 4+ (n=184): BC mortality ratio O:E = 1.05 5 to 9+ (n=37): BC mortality ratio O:E = 1.10 10+ (n=6): BC mortality ratio O:E = 0.8 Missing (n=97): BC mortality ratio O:E = 1.07 *Sub-group: grade* Grade 1 (n=235): BC mortality ratio O:E = 1.8 Grade 2 (n=528): BC mortality ratio O:E = 1.16	B. What are the results? 7 Can the performance of the rule be calculated? No (not enough data is available to calculate sensitivity, specificity, LR+, LR+, ROC curve. Mortality ratio can be calculated) 8 How precise was the estimate of the treatment effect? The updated model considers additional factors C. Will the results help locally? Are the results applicable to the scenario? 9 Would the prediction rule be reliable and and the results interpretable if used for your patient? Yes (UK population) 10 Is the rule acceptable in your case? Yes

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
				Grade 3 (n=395): BC mortality ratio O:E = 1.14 Missing grade (n=116): BC mortality ratio O:E = 0.31 *Sub-group: HER2 status* Negative (n=792): BC mortality ratio O:E = 1.35 Positive (n=77): BC mortality ratio O:E = 1.35 Missing (n=405): BC mortality ratio O:E = 0.44 **Tool discrimination** AUC = 0.7611 (CI not reported) **Results for PREDICT v1.2** **Prognostic accuracy (sensitivity, specificity)** Not reported **Tool calibration** ***10-year breast cancer mortality***	11 Would the results of the rule modify your decision about the management of the patient or the information you can give to him/her? Yes **Indirectness** This study includes direct population (UK based). **Other information** Conflict of interest: none PREDICT was contacted to determine which versions of PREDICT were used in this study (info@predict.nhs.uk)

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
				Total cohort (N=1726): BC mortality ratio O:E = 1.08 *Sub-group: age* <40 (n=67): BC mortality ratio O:E = 1.07 40 to 49 (n=274): BC mortality ratio O:E = 1.13 50 to 59 (n=436): BC mortality ratio O:E = 1.15 60+ (n=497): BC mortality ratio O:E = 1.01 *Sub-group: tumour size* <10 (n=144): BC mortality ratio O:E = 0.78 10 to 19 (n=574): BC mortality ratio O:E = 1.05 20 to 29 (n=404): BC mortality ratio O:E = 1.26 30 to 49 (n=140): BC mortality ratio O:E = 0.91 50+ (n=11): BC mortality ratio O:E = 0.5 Missing (n=1): BC mortality ratio O:E = 1	

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
				Sub-group: node status Negative (n=709): BC mortality ratio O:E = 1.15 1+ (n=241): BC mortality ratio O:E = 1.17 2 to 4+ (n=184): BC mortality ratio O:E = 1 5 to 9+ (n=37): BC mortality ratio O:E = 1.05 10+ (n=6): BC mortality ratio O:E = 0.8 Missing (n=97): BC mortality ratio O:E = 1.07 *Sub-group: grade* Grade 1 (n=235): BC mortality ratio O:E = 1.8 Grade 2 (n=528): BC mortality ratio O:E = 1.14 Grade 3 (n=395): BC mortality ratio O:E = 1.07 Missing grade (n=116): BC mortality ratio O:E = 0.31 *Sub-group: HER2 status*	

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
				Negative (n=792): BC mortality ratio O:E = 1.29 Positive (n=77): BC mortality ratio O:E = 1.24 Missing (n=405): BC mortality ratio O:E = 0.44 **Tool discrimination** AUC = 0.7676 (CI not reported)	
Full citation Candido Dos Reis, F. J., Wishart, G. C., Dicks, E. M., Greenberg, D., Rashbass, J., Schmidt, M. K., van den Broek, A. J., Ellis, I. O., Green, A., Rakha, E., Maishman, T., Eccles, D. M., Pharoah, P. D. P., An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation, Breast Cancer ResearchBreast Cancer Res, 19, 58, 2017	**Sample size** N tool development = 5738 (ECRIC database) N validations study BCOS: n=981 NTBCS: n=1726 POSH: n=2609 **Characteristics** Not reported.	**Prognostic tool** PREDICT v2.0 Note: This validation study reports data for PREDICT v2.0 and v1. Data for v1 was not used in the analysis as for many of the cases in the validation data the authors did not have either HER2 status or KI67 status	**Details** **Sample selection** For the tool development, data was obtained from the ECRIC database (n=5738), which includes patients with invasive breast cancer diagnosed in East Anglia, UK, between 1999 and 2003. For the validation study, data consisted in	**Results** **Prognostic accuracy (sensitivity, specificity)** Not reported **Tool calibration** ***10-year breast cancer specific mortality*** Total cohort (n=11272): BC mortality ratio = 0.95; difference O:E = -5% (p-value = 0.027) *Sub-group: age at diagnosis*	**Limitations** The quality of this study was assessed using the CASP tool for clinical prediction rule (CPR). A. Are the results valid? 1 Is the CPR clearly defined? Yes 2 The population from which the rule was derived included an appropriate spectrum of patients? Yes 3 Was the rule validated in a different group of patients? Yes (this study

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
Ref Id 657670 **Country/ies where the study was carried out** UK **Aim of the study** To develop and validate a new version of PREDICT (v2.0). **Study dates** Tool development: single dataset: East Anglia Cancer Registration and Information Centre (ECRIC), between 1999 to 2003 Validation study: combined datasets: 1) the Nottingham/ Tenovus Breast Cancer Study (NTBCS) between 1989 and 1998; 2) the Breast Cancer Outcome Study of Mutation Carriers (BCOS) between 1990 and 2000; and 3) the Prospective study of Outcomes in Sporadic and Hereditary	**Inclusion criteria** Not reported. **Exclusion criteria** Not reported.		combined data sets from the : 1) the NTBCS study (n=1726); included patients treated in Nottingham from 1989 to 1998. 506 node-negative cases were initially excluded because of inadequate axillary node staging (initial N=2232, but 506 node negative cases were excluded because of inadequate node staging) (ER-negative, n = 452; ER-positive, n = 1274) 2) the BCOS study (n=981) used data from a cohort of consecutive females diagnosed at <50 years of age with invasive breast cancer between 1990 and 2000, identified through medical registries of participating hospitals or the Netherlands Cancer Registry 3) the POSH study (n=2609) included young women	*ER-* 20 to 29 (n=92): BC mortality ratio = 0.94; difference O:E = -6% (p-value = 0.76) 30 to 39 (n=855): BC mortality ratio = 0.92; difference O:E = -9% (p-value = 0.18) 40 to 49 (n=414): BC mortality ratio = 0.98; difference O:E = -2% (p-value = 0.83) 50 to 59 (n=165): BC mortality ratio = 0.97; difference O:E = -3% (p-value = 0.85) 60 to 69 (n=117): BC mortality ratio = 0.82; difference O:E = -21% (p-value = 0.32) 70 to 79 (n=11): BC mortality ratio = 0.36; difference O:E = -180% (p-value = 0.28) *ER+* 20 to 29 (n=140): BC mortality ratio = 0.71; difference O:E = -40% (p-value = 0.047) 30 to 39 (n=1633): BC mortality ratio = 0.96; difference O:E = -4% (p-value = 0.48)	aims to validate a new version of the tool) 4 Were the predictor variables and the outcome evaluated in a blinded fashion? Not applicable (the outcome is mortality) 5 Were the predictor variables and the outcome evaluates in the whole sample selected initially? Yes 6 Are the statistical methods used to construct and validate the rule clearly described? Yes B. What are the results? 7 Can the performance of the rule be calculated? No (not enough data is available to calculate sensitivity, specificity, LR+, LR+, ROC curve. Mortality ratio can be calculated)

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
breast cancer (POSH) between 2000 and 2008 **Source of funding** The BCOS study was funded by the Netherlands Cancer Institute (NKI2007-3839). The POSH study was funded by Cancer Research UK (C1275/A9896, C1275/A11699, and C1275/A15956) and Breast Cancer Now (2005Nov63). PDPP was funded by the National Institute for Health Research Biomedical Research Centre at the University of Cambridge.			diagnosed with breast cancer in the United Kingdom between 2000 and 2008 **Data collection** The primary analysis was conducted using data from the ECRIC database. This included: age at diagnosis, number of lymph nodes sampled and number of lymph nodes positive, tumour size, histological grade, ER status, mode of detection (screening vs. clinical), information on local therapy (wide local excision, mastectomy, radiotherapy), and type of adjuvant systemic therapy (chemotherapy, endocrine therapy, both). n=1977 (34%) had less than 10 years of potential follow-up. The validation analysis included: 1) data from the NTBCS study, where	40 to 49 (n=1063): BC mortality ratio = 0.90; difference O:E = -11% (p-value = 0.16) 50 to 59 (n=467): BC mortality ratio = 0.96; difference O:E = -4% (p-value = 0.77) 60 to 69 (n=517): BC mortality ratio = 1.08; difference O:E = 7% (p-value = 0.53) 70 to 79 (n=55): BC mortality ratio = 0.38; difference O:E = -26% (p-value = 0.54) *Sub-group: tumour size* *ER-* 0 to 9 mm (n=96): BC mortality ratio = 0.90; difference O:E = -10% (p-value = 0.73) 10 to 19 mm (n=559): BC mortality ratio = 0.92; difference O:E = -8% (p-value = 0.41) 20 to 29 mm (n=524): BC mortality ratio = 0.97; difference O:E = -3% (p-value = 0.72)	8 How precise was the estimate of the treatment effect? The updated model considers additional factors C. Will the results help locally? Are the results applicable to the scenario? 9 Would the prediction rule be reliable and and the results interpretable if used for your patient? Yes (UK population) 10 Is the rule acceptable in your case? Yes 11 Would the results of the rule modify your decision about the management of the patient or the information you can give to him/her? Yes **Indirectness** This study includes a mixed population (38%

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
			participants were followed at 3-month intervals initially, then at 6-month intervals, and then annually for a median period of 111 months (range 4–211 months). Data was obtained prospectively. For those who were lost to follow-up, hospital notes were checked; 2) the BCOS study Data included: tumour size, nodal status, receipt of adjuvant systemic therapy, and follow-up. Follow-up data were obtained from the medical registries from the participating hospitals. 3) the POSH study, that included information obtained in the POSH cohort included age at diagnosis, histological grade, tumour size, number of positive lymph nodes, ER status, adjuvant chemotherapy, chemotherapy regimen	30 to 49 mm (n=354): BC mortality ratio = 0.99; difference O:E = -1% (p-value = 0.91) 50+ mm (n=121): BC mortality ratio = 0.75; difference O:E = -33% (p-value = 0.04) *ER+* 0 to 9 mm (n=352): BC mortality ratio = 1.54; difference O:E = 35% (p-value = 0.024) 10 to 19 mm (n=1428): BC mortality ratio = 1.06; difference O:E = 6% (p-value = 0.46) 20 to 29 mm (n=1111): BC mortality ratio = 0.98; difference O:E = -2% (p-value = 0.80) 30 to 49 mm (n=695): BC mortality ratio = 0.87; difference O:E = -15% (p-value = 0.07) 50+ mm (n=289): BC mortality ratio = 0.74; difference O:E = -35% (p-value = 0.00) *Sub-group: nodes positive* *ER-*	The Netherlands, 62% UK). **Other information** Conflict of interest: none PREDICT was contacted to determine which versions of PREDICT were used in this study (info@predict.nhs.uk)

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
			and adjuvant hormone therapy. 4) data from the ECRIC database (primary analysis) **Data analysis** Model calibration. was calculated for 10-year predicted breast cancer specific mortality and other mortality using the current online version of PREDICT (v1.3). It was obtained by comparing the predicted mortality estimates from each model with the observed mortality. This was done for the complete data set, and within strata of other prognostic variables. Model discrimination was calculated using the AUC calculated for 10-year mortality. The comparison between version 2.0 and version 1 was made using the method of DeLong.	0 (n=937): BC mortality ratio = 1.01; difference O:E = 1% (p-value = 0.89) 1 (n=232): BC mortality ratio = 0.86; difference O:E = -17% (p-value = 0.23) 2 to 4 (n=300): BC mortality ratio = 0.88; difference O:E = -13% (p-value = 0.19) 5 to 9 (n=101): BC mortality ratio = 0.96; difference O:E = -4% (p-value = 0.77) 10+ (n=84): BC mortality ratio = 0.85; difference O:E = -17% (p-value = 0.28) *ER+* 0 (n=2085): BC mortality ratio = 0.99; difference O:E = -1% (p-value = 0.85) 1 (n=675): BC mortality ratio = 0.92; difference O:E = -9% (p-value = 0.39) 2 to 4 (n=734): BC mortality ratio = 0.96; difference O:E = -4% (p-value = 0.63)	

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
			All analyses were carried out using Stata version 14 software (StataCorp, College Station, TX, USA).	5 to 9 (n=245): BC mortality ratio = 0.86; difference O:E = -17% (p-value = 0.14) 10+ (n=136): BC mortality ratio = 0.87; difference O:E = -15% (p-value = 0.25) *Sub-group: tumour grade* *ER-* 1 (n=44): BC mortality ratio = 0.96; difference O:E = -4% (p-value = 0.91) 2 (n=183): BC mortality ratio = 0.86; difference O:E = -17% (p-value = 0.33) 3 (n=1427): BC mortality ratio = 0.94; difference O:E = -7% (p-value = 0.19) *ER+* 1 (n=658): BC mortality ratio = 0.86; difference O:E = -16% (p-value = 0.43) 2 (n=1730): BC mortality ratio = 0.95; difference O:E = -5% (p-value = 0.44)	

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
				3 (n=1487): BC mortality ratio = 0.93; difference O:E = -7% (p-value = 0.17) ***10-year all-cause mortality*** Total cohort (n=11272): Mortality ratio = 0.99; difference O:E = -4% (p-value = 0.023) **Tool discrimination** Combined dataset: ER-: AUC = 0.696 ER+: AUC = 0.760 All population: AUC = 0.752 ECRIC dataset: ER-: AUC = 0.726 ER+: AUC = 0.796 All population: AUC = 0.805 NTBCS dataset:	

Study details	Number of participants and participants characteristics	Prognostic tool	Methods	Outcomes and results	Comments
				ER-: AUC = 0.680 ER+: AUC = 0.790 All population: AUC = 0.772 POSH dataset: ER-: AUC = 0.696 ER+: AUC = 0.760 All population: AUC = 0.752	

ABC, adjuvant breast cancer; AUC, area under the curve; BC, breast cancer; BCOS, Breast Cancer Outcomes Simulator; BCSS, breast cancer specific survival; CASP, Critical Appraisal Skills Programme; CPR, clinical prediction rule; ECRIC, East Anglia cancer registration and information centre; EFS, event free survival; ER, oestrogen receptor; HER2, Human epidermal growth factor receptor 2; NIHR, National Institute for Health Research; NPI, Nottingham Prognostic Index; NTBC, Nottingham Tenovus Breast Cancer; OPI, Oxford Prognostic Index; OS, overall survival; POSH, Prospective study of Outcomes in Sporadic versus Hereditary breast cancer; PR, progesterone receptor; SE, standard error; UK, United Kingdom; WMCIU, West Midlands Cancer Intelligence Unit

Appendix E – Forest plots

Forest plots for 3.1 Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?

There are no forest plots for this evidence review as no studies met the inclusion criteria.

Forest plots for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?

Forest plots are not applicable to this review as no meta-analysis was undertaken.

Appendix F – GRADE tables

GRADE tables for 3.1 Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?

There are no GRADE tables for this evidence review as no studies met the inclusion criteria.

GRADE tables for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?

There are no GRADE tables for this evidence review as GRADE is not appropriate to assess the quality of evidence for prediction model performance reviews.

Appendix G – Economic evidence study selection

Economic evidence study selection for 3.1 Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?

See Supplement 1: Health economics literature review for details of economic study selection.

Economic evidence study selection for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?

See Supplement 1: Health economics literature review for details of economic study selection.

Appendix H – Economic evidence tables

Economic evidence tables for 3.1. Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?

No economic evidence was identified for this review question.

Economic evidence tables for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?

No economic evidence was identified for this review question.

Appendix I – Health economic evidence profiles

Health economic evidence profiles for 3.1 Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?

No economic evidence was identified for this review question.

Health economic evidence profiles for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?

No economic evidence was identified for this review question.

Appendix J – Health economic analysis

Health economic analysis for 3.1 Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?

No health economic analysis was carried out for this review question.

Health economic analysis for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?

No health economic analysis was carried out for this review question.

Appendix K – Excluded studies

Excluded studies for 3.1 Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?

Clinical studies

Excluded studies - RQ3.1 Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?	
Study	**Reason for exclusion**
Bauer, K., Parise, C., Caggiano, V., Use of ER/PR/HER2 subtypes in conjunction with the 2007 St Gallen Consensus Statement for early breast cancer, BMC Cancer, 10 (no pagination), 2010	Retrospective cohort
Chen, J., Jiang, P., Wang, H. J., Zhang, J. Y., Xu, Y., Guo, M. H., Zhang, B., Tang, C. Y., Cao, H. Y., Wang, S., The efficacy of molecular subtyping in predicting postoperative recurrence in breast-conserving therapy: a 15-study meta-analysis, World journal of surgical oncology, 12, 212, 2014	Does not include 'test and treat' studies
Clark, Gm, McGuire, Wl, Hubay, Ca, Pearson, Oh, Marshall, Js, Progesterone receptors as a prognostic factor in Stage II breast cancer, The New England journal of medicine, 309, 1343-7, 1983	Not 'test and treat' design
Collett, K., Skjaerven, R., Maehle, B. O., The prognostic contribution of estrogen and progesterone receptor status to a modified version of the Nottingham Prognostic Index, Breast cancer research and treatment, 48, 1-9, 1998	Non-RCT
Duffy, M. J., Predictive markers in breast and other cancers: A review, Clinical Chemistry, 51, 494-503, 2005	Non-systematic review
Duffy, M. J., Crown, J., A personalized approach to cancer treatment: How biomarkers can help, Clinical Chemistry, 54, 1770-1779, 2008	Non-systematic review
Hammond, M. E. H., Hayes, D. F., Dowsett, M., Allred, D. C., Hagerty, K. L., Badve, S., Fitzgibbons, P. L., Francis, G., Goldstein, N. S., Hayes, M., Hicks, D. G., Lester, S., Love, R., Mangu, P. B., McShane, L., Miller, K., Osborne, C. K., Paik, S., Perlmutter, J., Rhodes, A., Sasano, H., Schwartz, J. N., Sweep, F. C. G., Taube, S., Torlakovic, E. E., Valenstein, P., Viale, G., Visscher, D., Wheeler, T., Williams, R. B., Wittliff, J. L., Wolff, A. C., American society of clinical oncology/college of American pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version), Archives of Pathology and Laboratory Medicine, 134, e48-e72, 2010	Guideline
Harris, L., Fritsche, H., Mennel, R., Norton, L., Ravdin, P., Taube, S., Somerfield, M. R., Hayes, D. F., Bast Jr, R. C., American society of clinical oncology 2007 update of recommendations for the use of tumor markers in breast cancer, Journal of Clinical Oncology, 25, 5287-5312, 2007	Guideline
Henry, N. L., Somerfield, M. R., Abramson, V. G., Allison, K. H., Anders, C. K., Chingos, D. T., Hurria, A., Openshaw, T. H., Krop, I. E., Role of patient and disease factors in adjuvant systemic therapy	Review of guideline

Excluded studies - RQ3.1 Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?	
Study	**Reason for exclusion**
decision making for early-stage, operable breast cancer: American society of clinical oncology endorsement of cancer care Ontario guideline recommendations, Journal of clinical oncology, 34, 2303-2311, 2016	
Punglia,R.S., Kuntz,K.M., Winer,E.P., Weeks,J.C., Burstein,H.J., The impact of tumor progesterone receptor status on optimal adjuvant endocrine therapy for postmenopausal patients with early-stage breast cancer: A decision analysis, Cancer, 106, 2576-2582, 2006	Non-RCT
Williams, C., Brunskill, S., Altman, D., Briggs, A., Campbell, H., Clarke, M., Glanville, J., Gray, A., Harris, A., Johnston, K., Lodge, M., Cost-effectiveness of using prognostic information to select women with breast cancer for adjuvant systemic therapy, Health Technology Assessment, 10, 1-153, 2006	Does not include any 'test and treat' studies

RCT, randomised controlled trial

Economic studies for 3.1 Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?

See Supplement 1: Health economics literature review for list of excluded economic studies.

Excluded studies for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?

Clinical studies

Excluded studies for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?	
Study	**Reason for exclusion**
Aaltomaa, S., Lipponen, P., Prognostic factors in breast-cancer (review), International Journal of Oncology, 1, 153-9, 1992	Aim not relevant (prediction of lymph node involvement).
Albergaria, A., Ricardo, S., Milanezi, F., Carneiro, V., Amendoeira, I., Vieira, D., Cameselle-Teijeiro, J., Schmitt, F., Nottingham Prognostic Index in triple-negative breast cancer: a reliable prognostic tool?, BMC cancer, 11, 299, 2011	Not relevant population (Brazil), and no relevant outcomes reported.
Anwar, K., Edmiston, K., Khan, A., Walsh, W., To compare the results of Adjuvant Online and Oncotype DX in estimating risk for relapse in hormone receptor positive stage I breast cancer patients, Journal of clinical oncology, 26, 22069, 2008	No relevant outcomes reported.
Asano, J., Hirakawa, A., Hamada, C., Assessing the prediction accuracy of cure in the Cox proportional hazards cure model: An application to breast cancer data, Pharmaceutical Statistics, 13, 357-363, 2014	Aim not relevant.
Balslev, I., Axelsson, C. K., Zedeler, K., Rasmussen, B. B., Carstensen, B., Mouridsen, H. T., The Nottingham Prognostic Index applied to 9,149 patients from the studies of the Danish Breast Cancer Cooperative Group (DBCG), Breast cancer research and treatment, 32, 281-290, 1994	Not relevant population (Denmark).
Blamey, R. W., Pinder, S. E., Ball, G. R., Ellis, I. O., Elston, C. W., Mitchell, M. J., Haybittle, J. L., Reading the prognosis of the individual with breast cancer, European journal of cancer, 43, 1545-1547, 2007	No relevant outcomes reported.
Boland, G. P., Chan, K. C., Knox, W. F., Roberts, S. A., Bundred, N. J., Value of the Van Nuys Prognostic Index in prediction of recurrence of ductal carcinoma in situ after breast-conserving surgery, British Journal of Surgery, 90, 426-432, 2003	No relevant outcomes reported.
Chollet, P., Amat, S., Belembaogo, E., Cure, H., De Latour, M., Dauplat, J., Le Bouedec, G., Mouret-Reynier, M. A., Ferriere, J. P., Penault-Llorca, F., Is Nottingham prognostic index useful after induction chemotherapy in operable breast cancer?, British journal of cancer, 89, 1185-1191, 2003	Not relevant population (France), and no relevant outcomes reported.
Cufer, T., Which tools can I use in daily clinical practice to improve tailoring of treatment for breast cancer? The 2007 St Gallen guidelines and/or Adjuvant! Online, Annals of Oncology, 19, vii41-vii45, 2008	Narrative paper.

Excluded studies for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?	
Study	**Reason for exclusion**
De Glas, N. A., Bastiaannet, E., Engels, C. C., De Craen, A. J. M., Putter, H., Van De Velde, C. J. H., Hurria, A., Liefers, G. J., Portielje, J. E. A., Validity of the online PREDICT tool in older patients with breast cancer: A population-based study, British journal of cancer, 114, 395-400, 2016	Not relevant population (The Netherlands).
De Glas, N. A., Van de Water, W., Engelhardt, E. G., Bastiaannet, E., De Craen, A. J. M., Kroep, J. R., Putter, H., Stiggelbout, A. M., Weijl, N. I., Van de Velde, C. J. H., Portielje, J. E. A., Liefers, G. J., Validity of adjuvant! Online program in older patients with breast cancer: A population-based study, The Lancet Oncology, 15, 722-729, 2014	Not relevant population (The Netherlands).
De Mascarel, I., Bonichon, F., Macgrogan, G., De Lara, C. T., Avril, A., Picot, V., Durand, M., Mauriac, L., Trojani, M., Coindre, J. M., Application of the Van Nuys prognostic index in a retrospective series of 367 ductal carcinomas in situ of the breast examinated by serial macroscopic sectioning: Practical considerations, Breast cancer research and treatment, 61, 151-159, 2000	No relevant outcomes reported.
D'Eredita, G., Giardina, C., Martellotta, M., Natale, T., Ferrarese, F., Prognostic factors in breast cancer: The predictive value of the Nottingham Prognostic Index in patients with a long-term follow-up that were treated in a single institution, European journal of cancer, 37, 591-596, 2001	Not relevant population (Italy).
Down, S. K., Lucas, O., Benson, J. R., Wishart, G. C., Effect of PREDICT on chemotherapy/trastuzumab recommendations in HER2-positive patients with early-stage breast cancer, Oncology Letters, 8, 2757-2761, 2014	Aim not relevant.
Engelhardt, E. G., van den Broek, A. J., Linn, S. C., Wishart, G. C., Rutgers, E. J. T., van de Velde, A. O., Smit, Vthbm, Voogd, A. C., Siesling, S., Brinkhuis, M., Seynaeve, C., Westenend, P. J., Stiggelbout, A. M., Tollenaar, Raem, van Leeuwen, F. E., van 't Veer, L. J., Ravdin, P. M., Pharaoh, P. D. P., Schmidt, M. K., Accuracy of the online prognostication tools PREDICT and Adjuvant! for early-stage breast cancer patients younger than 50 years, European Journal of CancerEur J Cancer, 78, 37-44, 2017	Mixed population (The Netherlands).
Fong, Y., Evans, J., Brook, D., Kenkre, J., Jarvis, P., Gower-Thomas, K., The Nottingham Prognostic Index: five- and ten-year data for all-cause survival within a screened population, Annals of the Royal College of Surgeons of England, 97, 137-139, 2015	No relevant outcomes reported.
Green, A. R., Soria, D., Powe, D. G., Nolan, C. C., Aleskandarany, M., Szasz, M. A., Tokes, A. M., Ball, G. R., Garibaldi, J. M., Rakha, E. A., Kulka, J., Ellis, I. O., Nottingham prognostic index plus (NPI+) predicts risk of distant metastases in primary breast cancer [Erratum: Breast Cancer Res Treat 2016; 159(1); 199], Breast cancer research and treatment, 157, 65-75, 2016	No relevant outcomes reported.
Hajage, D., de Rycke, Y., Bollet, M., Savignoni, A., Caly, M., Pierga, J. Y., Horlings, H. M., Van de Vijver, M. J., Vincent-Salomon, A., Sigal-Zafrani, B., Senechal, C., Asselain, B., Sastre, X., Reyal, F., External validation of Adjuvant! Online breast cancer prognosis tool. Prioritising recommendations for improvement, PLoS ONE [Electronic Resource], 6, e27446, 2011	Not relevant population (French and Dutch data set).

Excluded studies for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?	
Study	**Reason for exclusion**
Hajage, D., De Rycke, Y., Bollet, M., Savignoni, A., Caly, M., Pierga, J. Y., Horlings, H. M., Van De Vjver, M. J., Vincent-Salomon, A., Sigal, B., Senechal, C., Asselain, B., Sastre, X., Reyal, F., External validation of adjuvant! Online breast cancer prognosis tool. Improvement is still needed, Cancer Research. Conference: 34th Annual CTRC AACR San Antonio Breast Cancer Symposium. San Antonio, TX United States. Conference Start, 71, 2011	Not relevant population (France and The Netherlands).
Hasby, E. A., Khalifa, R. A., Expression of CD74 in invasive breast carcinoma: Its relation to Nottingham Prognostic Index, hormone receptors, and HER2 immunoprofile, Tumori, 103, 193-203, 2017	No relevant outcomes.
Kelley, L., Silverstein, M., Guerra, L., Analyzing the risk of recurrence after mastectomy for DCIS: A new use for the USC/Van nuys prognostic index, Annals of surgical oncology, 18, 459-462, 2011	No relevant outcomes reported.
Kraft Rovere, R., Dos Santos Borges, G., Staak Jr, M. C., Colchon, P. H., Rebello, J. R., Adjuvant! Online: Mind the gap!, Klinicka Onkologie, 26, 110-113, 2013	Not relevant population (Brazil).
Kwon, J., Eom, K. Y., Koo, T. R., Kim, B. H., Kang, E., Kim, S. W., Kim, Y. J., Park, S. Y., Kim, I. A., A prognostic model for patients with triple-negative breast cancer: Importance of the modified nottingham prognostic index and age, Journal of Breast Cancer, 20, 65-73, 2017	No outcomes of interest reported (correlations only).
Laas, E., Mallon, P., Delomenie, M., Gardeux, V., Pierga, J. Y., Cottu, P., Lerebours, F., Stevens, D., Rouzier, R., Reyal, F., Are we able to predict survival in ER-positive HER2-negative breast cancer ? A comparison of web-based models, British journal of cancer, 112, 912-917, 2015	Not relevant population (data set from USA and Canada).
Lambertini, M., Pinto, A. C., Ameye, L., Jongen, L., Del Mastro, L., Puglisi, F., Poggio, F., Bonotto, M., Floris, G., Van Asten, K., Wildiers, H., Neven, P., De Azambuja, E., Paesmans, M., Azim, H. A., The prognostic performance of Adjuvant! Online and Nottingham Prognostic Index in young breast cancer patients, British journal of cancer, 115, 1471-1478, 2016	Not relevant population (Belgian and Italian referral institutions).
Liu, M. T., Huang, W. T., Wang, A. Y., Huang, C. C., Huang, C. Y., Chang, T. H., Pi, C. P., Yang, H. H., Prediction of outcome of patients with metastatic breast cancer: evaluation with prognostic factors and Nottingham prognostic index, Supportive care in cancer, 18, 1553-64, 2010	No relevant outcomes reported.
Lundin, J., The Nottingham Prognostic Index - from relative to absolute risk prediction, European journal of cancer, 43, 1498-1500, 2007	Editorial comment
Maishman, T., Copson, E., Stanton, L., Gerty, S., Dicks, E., Durcan, L., Wishart, G., Pharoah, P., Eccles, D., A review of the online prognositc model predict using the POSH cohort (women aged <=40 years at breast cancer diagnosis), Trials. Conference: 3rd International Clinical Trials Methodology Conference. United Kingdom, 16, 2015	Conference abstract. Full published paper included in the review.

Excluded studies for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?	
Study	**Reason for exclusion**
Miao, H., Hartman, M., Verkooijen, H. M., Taib, N. A., Wong, H. S., Subramaniam, S., Yip, C. H., Tan, E. Y., Chan, P., Lee, S. C., Bhoo-Pathy, N., Validation of the CancerMath prognostic tool for breast cancer in Southeast Asia, BMC cancer, 16, 820, 2016	Not relevant population (validation study in Asiatic population).
Michaelson, J. S., Chen, L. L., Bush, D., Fong, A., Smith, B., Younger, J., Improved web-based calculators for predicting breast carcinoma outcomes, Breast cancer research and treatment, 128, 827-835, 2011	Not relevant population (USA).
Mook, S., Schmidt, M. K., Rutgers, E. J., van de Velde, A. O., Visser, O., Rutgers, S. M., Armstrong, N., van't Veer, L. J., Ravdin, P. M., Calibration and discriminatory accuracy of prognosis calculation for breast cancer with the online Adjuvant! program: a hospital-based retrospective cohort study, The Lancet Oncology, 10, 1070-1076, 2009	Not relevant population (Dutch).
Okugawa, H., Yamamoto, D., Uemura, Y., Sakaida, N., Yamada, M., Tanaka, K., Kamiyama, Y., Prognostic factors in breast cancer: The value of the Nottingham Prognostic Index for patients treated in a single institution [Erratum: Surgery Today 2009; 39(8): 738], Surgery Today, 35, 907-911, 2005	Not relevant population (Japan), and no relevant outcomes reported. . Note that this paper has been retracted, as Substantial portions of this article were found to have been published previously by Dâ□™Eredità G. et al.
Olivotto, I. A., Bajdik, C. D., Ravdin, P. M., Speers, C. H., Coldman, A. J., Norris, B. D., Davis, G. J., Chia, S. K., Gelmon, K. A., Population-based validation of the prognostic model ADJUVANT! for early breast cancer, Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 23, 2716-2725, 2005	Not relevant population (Canada).
Ozanne, E. M., Braithwaite, D., Sepucha, K., Moore, D., Esserman, L., Belkora, J., Sensitivity to input variability of the adjuvant! Online breast cancer prognostic model, Journal of clinical oncology, 27, 214-219, 2009	No relevant outcomes reported.
Paridaens, R. J., Gelber, S., Cole, B. F., Gelber, R. D., Thurlimann, B., Price, K. N., Holmberg, S. B., Crivellari, D., Coates, A. S., Goldhirsch, A., Adjuvant! Online estimation of chemotherapy effectiveness when added to ovarian function suppression plus tamoxifen for premenopausal women with estrogen-receptor-positive breast cancer, Breast Cancer Research & TreatmentBreast Cancer Res Treat, 123, 303-10, 2010	Not relevant population (IBCSG trial', Switzerland, Sweden, Italy, Australia).
Plakhins, G., Irmejs, A., Gardovskis, A., Subatniece, S., Liepniece-Karele, I., Purkalne, G., Teibe, U., Trofimovics, G., Miklasevics, E., Gardovskis, J., Underestimated survival predictions of the prognostic tools Adjuvant! Online and PREDICT in BRCA1-associated breast cancer patients, Familial Cancer, 12, 683-689, 2013	Not relevant population (Latvia).
Puente, J., Lopez-Tarruella, S., Ruiz, A., Lluch, A., Pastor, M., Alba, E., De La Haba, J., Ramos, M., Cirera, L., Anton, A., Llombart, A., Plazaola, A., Fernandez-Aramburo, A., Sastre, J., Diaz-Rubio, E., Martin, M., Practical prognostic index for patients with metastatic recurrent breast cancer:	Not relevant population (Spain).

Excluded studies for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?	
Study	**Reason for exclusion**
Retrospective analysis of 2,322 patients from the GEICAM Spanish El Alamo Register, Breast cancer research and treatment, 122, 591-600, 2010	
Quintyne, K. I., Woulfe, B., Coffey, J. C., Gupta, R. K., Correlation between Nottingham Prognostic Index and Adjuvant! Online Prognostic tools in patients with early-stage breast cancer in Mid-Western ireland, Clinical breast cancer, 13, 233-238, 2013	No relevant outcomes reported.
Rejali, M., Tazhibi, M., Mokarian, F., Gharanjik, N., Mokarian, R., The Performance of the Nottingham Prognosis Index and the Adjuvant Online Decision Making Tool for Prognosis in Early-stage Breast Cancer Patients, International Journal of Preventive Medicine, 6, 93, 2015	Not relevant population (Iran).
Sauerbrei, W., Hubner, K., Schmoor, C., Schumacher, M., Validation of existing and development of new prognostic classification schemes in node negative breast cancer, Breast cancer research and treatment, 42, 149-163, 1997	No relevant outcomes reported.
Schmidt, M., Victor, A., Bratzel, D., Boehm, D., Cotarelo, C., Lebrecht, A., Siggelkow, W., Hengstler, J. G., Elsasser, A., Gehrmann, M., Lehr, H. A., Koelbl, H., Von Minckwitz, G., Harbeck, N., Thomssen, C., Long-term outcome prediction by clinicopathological risk classification algorithms in node-negative breast cancer - Comparison between Adjuvant!, St Gallen, and a novel risk algorithm used in the prospective randomized Node-Negative-Breast Cancer-3 (NNBC-3) trial, Annals of Oncology, 20, 258-264, 2009	Not relevant population (Germany).
Serrero, G., Hawkins, D. M., Bejarano, P. A., Ioffe, O., Tkaczuk, K. R., Elliott, R. E., Head, J. F., Phillips, J., Godwin, A. K., Weaver, J., Hicks, D., Yue, B., Improvement in risk predictive value of Nottingham prognostic index by determining GP88 tumor tissue expression for estrogen receptor positive breast cancer patients, Cancer Research. Conference: 39th Annual CTRC AACR San Antonio Breast Cancer Symposium. United States, 77, 2017	Conference abstract.
Suen, D., Chow, L. W. C., Prognostic contribution of the HER-2 oncogene overexpression to the Nottingham Prognostic Index in breast cancer, Biomedicine and Pharmacotherapy, 60, 293-297, 2006	No relevant outcomes reported.
Van Belle, V., Decock, J., Hendrickx, W., Brouckaert, O., Pintens, S., Moerman, P., Wildiers, H., Paridaens, R., Christiaens, M. R., Van Huffel, S., Neven, P., Short-Term Prognostic Index for Breast Cancer: NPI or Lpi, Pathology Research International, 2011, 918408, 2010	Not relevant population (Belgium).
Ward, S., Scope, A., Rafia, R., Pandor, A., Harnan, S., Evans, P., Wyld, L., Gene expression profiling and expanded immunohistochemistry tests to guide the use of adjuvant chemotherapy in breast cancer management: A systematic review and cost-effectiveness analysis, Health Technology Assessment, 17, V-302, 2013	No relevant tools. HTA.

Excluded studies for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?	
Study	**Reason for exclusion**
Wishart, G. C., Bajdik, C. D., Azzato, E. M., Dicks, E., Greenberg, D. C., Rashbass, J., Caldas, C., Pharoah, P. D. P., A population-based validation of the prognostic model PREDICT for early breast cancer, European Journal of Surgical Oncology, 37, 411-417, 2011	Not relevant population (Canada).
Wishart, G. C., Bajdik, C. D., Dicks, E., Provenzano, E., Schmidt, M. K., Sherman, M., Greenberg, D. C., Green, A. R., Gelmon, K. A., Kosma, V. M., Olson, J. E., Beckmann, M. W., Winqvist, R., Cross, S. S., Severi, G., Huntsman, D., Pylkas, K., Ellis, I., Nielsen, T. O., Giles, G., Blomqvist, C., Fasching, P. A., Couch, F. J., Rakha, E., Foulkes, W. D., Blows, F. M., Begin, L. R., Van, T. Veer L. J., Southey, M., Nevanlinna, H., Mannermaa, A., Cox, A., Cheang, M., Baglietto, L., Caldas, C., Garcia-Closas, M., Pharoah, P. D. P., PREDICT Plus: Development and validation of a prognostic model for early breast cancer that includes HER2, British journal of cancer, 107, 800-807, 2012	Not relevant population (data from multiple countries).
Wong, H. S., Subramaniam, S., Alias, Z., Taib, N. A., Ho, G. F., Ng, C. H., Yip, C. H., Verkooijen, H. M., Hartman, M., Bhoo-Pathy, N., The predictive accuracy of PREDICT: A personalized decision-making tool for southeast Asian women with breast cancer, Medicine (United States), 94, e593, 2015	Not relevant population (Southeast Asia).
Wu, X., Ye, Y., Barcenas, C. H., Chow, W. H., Meng, Q. H., Chavez-MacGregor, M., Hildebrandt, M. A., Zhao, H., Gu, X., Deng, Y., Wagar, E., Esteva, F. J., Tripathy, D., Hortobagyi, G. N., Personalized Prognostic Prediction Models for Breast Cancer Recurrence and Survival Incorporating Multidimensional Data, J Natl Cancer InstJournal of the National Cancer Institute, 109, 2017	Not relevant population (USA).

HTA, Health Technology Assessment; IBCSG, International Breast Cancer Study Group

Economic studies for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?

See Supplement 1: Health economics literature review for list of excluded economic studies.

Appendix L – Research recommendations

Research recommendations for 3.1 Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?

No research recommendations were made for this review question.

Research recommendations for 3.2 What predictive prognostic tools, excluding gene profiling tests, should be used for determining adjuvant systemic therapy?

No research recommendations were made for this review question.

Appendix M – 3.1 Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?

Nominal group technique questionnaire for progesterone receptor testing

Name:										
Prognosis										
	Strongly disagree								Strongly agree	Insufficient knowledge
Positive progesterone receptor status is associated with favourable prognosis	1	2	3	4	5	6	7	8	9	☐
Comments:										
Tumours that are negative for progesterone receptors have a worse prognosis	1	2	3	4	5	6	7	8	9	☐
Comments:										
Endocrine therapy										

	Strongly disagree								Strongly agree	Insufficient knowledge
Positive progesterone receptor status predicts response to endocrine therapy in people with oestrogen receptor negative breast cancer	1	2	3	4	5	6	7	8	9	☐
Comments:										
Endocrine therapy should be offered to individuals whose tumour is oestrogen receptor negative, but progesterone receptor positive	1	2	3	4	5	6	7	8	9	☐
Comments:										
Positive progesterone receptor status is indicative of benefit from endocrine therapy	1	2	3	4	5	6	7	8	9	☐
Comments:										
Tumours that are negative for progesterone receptors are less responsive to endocrine therapy	1	2	3	4	5	6	7	8	9	☐
Comments:										
Combined measurement of oestrogen receptor status and progesterone receptor status more accurately predicts benefit from adjuvant hormone treatment for invasive breast cancer than oestrogen receptor status alone	1	2	3	4	5	6	7	8	9	☐
Comments:										

Progesterone receptor status does not predict benefit of endocrine therapy in oestrogen receptor negative tumours	1	2	3	4	5	6	7	8	9	☐
Comments:										
Chemotherapy										
	Strongly disagree								Strongly agree	Insufficient knowledge
For breast cancer where the benefit of chemotherapy is borderline, it should be offered if individuals have progesterone receptor negative breast cancer	1	2	3	4	5	6	7	8	9	☐
Comments:										
Progesterone receptor negative tumours should be considered high risk and, therefore, candidates for chemotherapy	1	2	3	4	5	6	7	8	9	☐
Comments:										
Assessment of PR status										
	Strongly disagree								**Strongly agree**	**Insufficient knowledge**

Progesterone receptor status does not provide useful information in oestrogen receptor positive breast cancer	1	2	3	4	5	6	7	8	9	☐
Comments:										
Progesterone receptor status is relevant when making decisions regarding adjuvant therapy	1	2	3	4	5	6	7	8	9	☐
Comments:										
Progesterone receptor status should be assessed in all newly diagnosed invasive breast cancers	1	2	3	4	5	6	7	8	9	☐
Comments:										
Progesterone receptor status should be assessed in invasive breast cancer only if the results would influence treatment planning	1	2	3	4	5	6	7	8	9	☐
Comments:										
Progesterone receptor status should be assessed only in selected cases (e.g., oestrogen receptor negative cancer, cases with borderline benefit of chemotherapy, assessing eligibility for clinical trials)	1	2	3	4	5	6	7	8	9	☐
Comments:										
Progesterone receptor status of tumours in patients with invasive breast cancer should not be routinely assessed	1	2	3	4	5	6	7	8	9	☐
Comments:										

Re-rated statements (Round 2)										
	Strongly disagree								Strongly agree	Insufficient knowledge
Progesterone receptor status provides additional information to oestrogen receptor status that may be benefical when considering the likely benefit of adjuvant hormone treatment for invasive breast cancer	1	2	3	4	5	6	7	8	9	☐
Comments:										
Negative progesterone receptor status is one factor that may increase the benefit from chemotherapy and may increase the likelihood chemotherapy is offered in borderline cases	1	2	3	4	5	6	7	8	9	☐
Comments:										

Appendix N - 3.1 Is there a benefit of progesterone receptor (PR) testing for adjuvant chemotherapy planning?

Nominal group technique results

Table 11: Nominal group technique consensus ratings for progesterone receptor testing

Area	Statement no.	Statement	Agreement (%)	Action
Prognosis	1	Positive progesterone receptor status is associated with favourable prognosis	80	Used to inform recommendation
	2	Tumours that are negative for progesterone receptors have a worse prognosis	100	Used to inform recommendation
Endocrine therapy	3	Positive progesterone receptor status predicts response to endocrine therapy in people with oestrogen receptor negative breast cancer	11	Discarded as less than 60% agreement
	4	Endocrine therapy should be offered to individuals whose tumour is oestrogen receptor negative, but progesterone receptor positive	22	Discarded as less than 60% agreement
	5	Positive progesterone receptor status is indicative of benefit from endocrine therapy	33	Discarded as less than 60% agreement
	6	Tumours that are negative for progesterone receptors are less responsive to endocrine therapy	45	Discarded as less than 60% agreement
	7	Combined measurement of oestrogen receptor status and progesterone receptor status more accurately predicts benefit from adjuvant hormone treatment for invasive breast cancer than oestrogen receptor status alone	64	Re-drafted and re-rated.
	8	Progesterone receptor status does not predict benefit of endocrine therapy in oestrogen receptor negative tumours	18	Discarded as less than 60% agreement
Chemotherapy	9	For breast cancer where the benefit of chemotherapy is borderline, it should be offered if individuals have progesterone receptor negative breast cancer	40	Re-drafted and re-rated despite low agreement (<60%) due

Area	Statement no.	Statement	Agreement (%)	Action
				to committee comments
	10	Progesterone receptor negative tumours should be considered high risk and, therefore, candidates for chemotherapy	27	Discarded as less than 60% agreement
Assessment of PR status	11	Progesterone receptor status does not provide useful information in oestrogen receptor positive breast cancer	27	Discarded as less than 60% agreement
	12	Progesterone receptor status is relevant when making decisions regarding adjuvant therapy	82	Used to inform recommendation
	13	Progesterone receptor status should be assessed in all newly diagnosed invasive breast cancers	82	Used to inform recommendation
	14	Progesterone receptor status should be assessed in invasive breast cancer only if the results would influence treatment planning	27	Comments used to inform recommendation
	15	Progesterone receptor status should be assessed only in selected cases (e.g., oestrogen receptor negative cancer, cases with borderline benefit of chemotherapy, assessing eligibility for clinical trials)	27	Comments used to inform recommendation
	16	Progesterone receptor status of tumours in patients with invasive breast cancer should not be routinely assessed	18	Discarded as less than 60% agreement
Re-rated statements	7 (round 2)	Progesterone receptor status provides additional information to oestrogen receptor status that may be beneficial when considering the likely benefit of adjuvant hormone treatment for invasive breast cancer	90	Used to inform recommendation
	9 (round 2)	Negative progesterone receptor status is one factor that may increase the benefit from chemotherapy and may increase the likelihood chemotherapy is offered in borderline cases	100	Used to inform recommendation