Evidence Assessment and Analysis Report commissioned by the NIHR HTA Programme on behalf of the National Institute for Health and Care Excellence – Protocol Guide for Diagnostics Assessment Programme

Tumour profiling tests to guide treatment decisions in people with breast cancer (update of DG10) HTA Reference No. 16/30/03

Final amended version, 28/02/17

1. Name of External Assessment Group (EAG) and project lead

ScHARR, University of Sheffield
Health Economic and Decision Science
The University of Sheffield
Regent Court
30 Regent Street
Sheffield
S1 4DA

Project Lead
Name: *Sue Ward*
Post held *Senior Research Fellow*
Health Economic and Decision Science
The University of Sheffield
ScHARR
Regent Court
30 Regent Street
Sheffield
S1 4DA
Telephone number and extension: *(+44) (0) 114 2220816*
Fax number, E-mail address: *(+44) (0) 114 2724095, s.e.ward@sheffield.ac.uk*

2. Plain English Summary

Breast cancer is the most commonly diagnosed cancer and the third most common cause of cancer related deaths in the UK. In 2014, 46,085 women and 332 men were diagnosed with breast cancer in England, while 11,360 women and 73 men died from breast cancer in the UK. Treatment usually involves surgery to remove the tumour and any involved lymph nodes. This may be followed by one or more of the following: radiotherapy, endocrine (hormone) therapy, trastuzumab and/or chemotherapy.

There are various prognostic tools that help patients and clinicians make treatment decisions by predicting the risk of the disease coming back (recurring) after surgery. These include the Nottingham Prognostic Index (NPI), PREDICT and Adjuvant! Online. They predict the risk of recurrence, based on pathological information (e.g. tumour size, grade and lymph node status for NPI), plus other factors including oestrogen receptor (ER) status, age and co-morbidity for Adjuvant! Online and PREDICT. However, it has been suggested that these clinical tools do not predict recurrence and response to treatment particularly well for some patients (Paik, 2007[1]). This presents a challenge to clinicians in making decisions relating to whether or not to recommend the use adjuvant chemotherapy (chemotherapy after surgery) in people with early stage breast cancer (Stages I, II (A or B) and IIIA[2]).

Tumour profiling, using either gene expression profiling or protein expression profiling (with immunohistochemistry), seeks to identify genes or proteins that may be helpful in assessing disease prognosis and guiding therapy. Improved information on a patient's risk of recurrence (i.e. prognostic risk) and/or likely response to chemotherapy (i.e. predictive benefit) may help target chemotherapy at those patients who will benefit the most. Avoiding chemotherapy in patients at low risk of recurrence, and who would therefore obtain limited benefit, avoids the unpleasant side effects of chemotherapy and reduces expenditure on both the chemotherapy itself and the treatment of these side effects. It is therefore important to understand the benefits offered by these tumour profiling tests compared with existing prognostic tools and whether or not they represent a good use of National Health Service (NHS) resources.

A previous systematic review and economic evaluation, published in 2013, considered the clinical effectiveness and cost-effectiveness of tumour profiling tests compared with current prognostic tools in guiding the use of adjuvant chemotherapy in women with early breast cancer in England and Wales. This report informed the NICE decision to approve the use of Oncotype DX as an option for guiding adjuvant chemotherapy decisions for people with ER positive, lymph node negative and human epidermal growth factor receptor 2 (HER2) negative early breast cancer assessed to be at intermediate risk of recurrence of breast cancer after surgery.

This review aims to update the previous report by systematically evaluating the most recent evidence on the use of five tumour profiling tests to guide adjuvant chemotherapy treatment decisions in early breast cancer management, and by conducting an economic evaluation to determine if these tests represent good value for money to the NHS.

3. Decision problem

3.1 Purpose of the decision to be made

Do tumour profiling tests used for guiding adjuvant chemotherapy decision in patients with early stage breast cancer represent a clinical- and cost-effective use of NHS resources?

This project will update the systematic review and cost-effectiveness analysis[3] that informed Nice Diagnostics Guidance 10 (DG10[4]).

3.2 Clear definition of the intervention

Five tests have been identified by NICE and will be included in this assessment. The use of these interventions should be considered in combination with current decision making. The five tests are summarised in Table 1.

EndoPredict (Myriad Genetics) is a Conformité Européene (CE) marked assay that is designed to predict the likelihood of metastases developing within 10 years of initial diagnosis. The test is intended for use in pre- and post-menopausal women with early stage breast cancer with all of the following clinical features:

- oestrogen receptor (ER)-positive
- human epidermal growth factor receptor 2 (HER2)-negative
- lymph node (LN)-negative or LN-positive (up to 3 positive nodes).

EndoPredict measures the expression of 12 genes: 3 proliferation associated genes, 5 hormone receptor associated genes, 3 reference (normalisation) genes and 1 control gene.

EndoPredict requires RNA samples extracted from formalin-fixed, paraffin-embedded (FFPE) breast cancer tissue. The test can be performed in a local laboratory using a VERSANT kPCR AD module (Siemens Healthcare Diagnostics). Alternatively, FFPE samples can be submitted to a Myriad Genetics pathology laboratory in Munich that is accredited by the Deutsche Akkreditierungsstelle, a national accreditation body for Germany.

The test process involves using a reverse transcription-quantitative polymerase chain reaction (RT-qPCR), in which target messenger RNAs are reverse transcribed, amplified and simultaneously detected. The raw data are then exported to online evaluation software (EndoPredict Report Generator) which performs a quality check and calculates the EP score and the EPclin score.

The EP score is a number on a scale between 0 and 15. An EP score of less than 5 indicates low risk of distant disease recurrence reoccurring in the next 10 years. An EP score of 5 or more indicates a high risk of distant disease recurrence in the next 10 years.

The EPclin score is calculated by adding clinical data about tumour size and nodal status to the EP score. From the EPclin score, the probability of metastasis formation within 10 years is estimated, assuming 5 years of hormonal treatment. If the EPclin 10 year risk is less than 10% the patient is classed as low risk for metastases recurring in the next 10 years. If the EPclin 10 year risk is 10% or greater the patient is classed as high risk for metastases recurring in the next 10 years.

It takes approximately 2 days to obtain the test results if the test is done in-house. If samples are sent away for testing the turnaround time for the central service is 4 to 5 working days.

MammaPrint (Agendia) is a CE marked microarray that is designed to assess the risk of distant recurrence within 5 years and whether a woman would benefit from chemotherapy. The test is intended for use in pre- and post-menopausal women with stage I or II breast cancer with the following clinical features:

- tumour size less than or equal to 5cm,
- LN-negative or LN-positive (up to 3 positive nodes)

The test can be used irrespective of ER and HER2 status, that is, it can be used for tumours that are ER-negative or ER-positive, and HER2-negative or HER2-positive.

MammaPrint measures the expression of 70 genes, including genes associated with 7 different parts of the metastatic pathway: growth and proliferation, angiogenesis, local invasion, entering the circulation, survival in the circulation, entering organs from the circulation, and adaption to the microenvironment at a secondary site.

The MammaPrint test is offered as an off-site service. In Europe, samples are sent for analysis at the Agendia laboratory in Amsterdam, the Netherlands. The test requires a FFPE breast cancer tissue sample from a surgical specimen or core needle biopsy.

The test process involves isolation of RNA from FFPE sample followed by reverse transcription of the RNA to get complementary deoxyribonucleic acid (cDNA). The cDNA is amplified and labelled before being hybridized (bound) to the diagnostic microarray. The microarray is

washed and then scanned using an Agilent DNA microarray scanner. The scan file is analysed using Agilent Feature Extraction Software and an algorithm is used to calculate the correlation of the sample profile to a "Low Risk" template profile on a scale of -1.000 to +1.000 with a cut off at 0. The threshold was set such that women with a low risk result have a 10% risk of developing distant metastases over the next 10 years without any adjuvant hormone or chemotherapy.

Test results are available to healthcare professionals within 10 days of submitting the sample.

Oncotype DX Breast Recurrence Score (Genomic Health) is designed to quantify the 10-year risk of distant recurrence and predict the likelihood of chemotherapy benefit. The test also reports the underlying tumour biology: ER, progesterone receptor (PR) and HER2 status. The test is intended for use in pre- and post-menopausal women with stage I or II breast cancer that has the following clinical features:

- LN-negative or LN-positive (up to 3 positive nodes),
- ER-positive
- HER2-negative.

Oncotype DX quantifies the expression of 21 genes. Of these, 16 are cancer-related genes correlated with distance recurrence-free survival, and 5 are reference genes for normalising the expression of the cancer-related genes. This information is used to calculate the breast recurrence score.

Oncotype DX is offered as a test service to the NHS. Samples are processed centrally at the Genomic Health Inc. laboratory in the US, which is accredited by the American Association for Laboratory Accreditation and the College of American Pathologists. The test requires a FFPE breast cancer tissue sample from a biopsy or surgical resection, which can be sent as a paraffin embedded block or as 15 unstained charged slides. The test process is based on RT-qPCR.

The test gives a recurrence score of between 0 and 100, which is used to quantify the 10 year risk of distant recurrence, assuming 5 years of hormonal (endocrine) therapy. A score below 18 indicates low risk of distant recurrence; a score between 18 and 30 indicates intermediate risk; and a score of 31 or more indicates high risk. The recurrence score may also predict the benefit of chemotherapy.

The breast recurrence score can be combined with clinical and pathological factors using the recurrence score-pathology-clinical (RSPC) calculator; however, this calculator has not been validated.

The Oncotype DX results are typically reported within 7 to 10 calendar days after the sample is received at the laboratory.

Prosigna (NanoString Technologies) is a CE marked assay designed to predict distant recurrence-free survival at 10 years. The test is intended for use in postmenopausal women with early stage breast cancer that is:

- LN-negative or LN-positive (up to 3 positive nodes)
- ER-positive
- HER2-negative.

Prosigna is based on the PAM50 gene signature. It measures the expression levels of 50 genes used for the intrinsic subtype classification algorithm. It also measures the expression of 8 housekeeping genes used for signal normalisation, 6 positive controls, and 8 negative controls.

The test requires RNA extracted from a FFPE breast tumour tissue sample and is done using the NanoString nCounter analysis system. The test process involves fluorescent probe pairs that hybridise to the mRNA, the fluorescence is then detected by the nCounter Digital Analyser.

Prosigna classifies samples into the following breast cancer subtypes based on their PAM50 gene expression signatures: luminal A, luminal B, HER2-enriched or basal-like. Risk of distant recurrence within 10 years, assuming 5 years of hormonal treatment, is then derived from an algorithm which is based on the results of the PAM50 gene signature, breast cancer subtype, tumour size, nodal status and proliferation score. The proliferation score is determined by evaluating multiple genes associated with the proliferation pathway. The risk of recurrence score is provided as a numerical score on a 0 to 100 scale that estimates the probability of distant recurrence over 10 years. Based on this score and the nodal status, samples are classified into risk categories:

- Node negative: low risk (0 to 40), intermediate risk (41 to 60), or high risk (61 to 100)

- Node positive (up to 3 positive nodes): low risk (0 to 15), intermediate risk (16 to 40), or high risk (41 to 100)

Immunohistochemistry 4 (IHC4) test is a laboratory developed test which combines the results of 4 immunohistochemistry measured parameters with clinical and pathologic features. It is sometimes called the IHC4+C test. It is designed to quantify the risk of distant disease recurrence of breast cancer patients, assuming 5 years of endocrine therapy. The test is intended for use in post-menopausal women with early stage breast cancer with the following clinical features:

- ER-positive
- LN-negative or LN-positive (up to 3 positive nodes)

The components of the test are 4 immunohistochemical assays: oestrogen receptor (ER), progesterone receptor (PR), HER2 and the proliferation marker Ki67. The IHC4 test is currently used within the Royal Marsden Breast Cancer Unit service, but the test could be run in local NHS laboratories if quality assurance programmes for the individual assays are in place. It uses FFPE breast tumour tissue samples and immunohistochemistry techniques that are universally available in NHS pathology departments. ER and HER2 markers are commonly measured in NHS laboratories. Whilst PR and Ki67 markers are not routinely measured in breast tumour tissue samples, the assays are commonly available for use if needed. The quantitative assessment of Ki67 required for the IHC4 test is not currently performed in most NHS laboratories and therefore further training for pathologists and biomedical scientists is likely to be needed.

The IHC4 test has an algorithm that calculates a risk score for distant recurrence based on the results of the 4 assays and clinical factors such as: age, nodal status, tumour size, and grade. The algorithm has been published and validated[5] and is freely available, and a calculator is available for use on request. A distant recurrence score of less than 10% is categorised as low risk for distant recurrence at 10 years; a score of 10% or more but less than 20% is categorised as intermediate risk, and a score of 20% or more is categorised as high risk for distant recurrence at 10 years.

At the Royal Marsden NHS Foundation Trust the test is processed with an average estimated turnaround time of 1 week, however results may be made available in 2 working days if they are required urgently.

Table 1: Summary of technologies

Test	EndoPredict	MammaPrint	Oncotype DX	Prosigna	IHC4
Manufacturer	Myriad	Agendia	Genomic Health	NanoString	-
Purpose	Recurrence risk	Recurrence risk and chemotherapy benefit	Recurrence risk and chemotherapy benefit	Intrinsic subtype and recurrence risk	Recurrence risk
Description	12 gene assay (8 cancer genes; RT-qPCR) + clinical factors	70 gene array (microarray)	21 gene assay (16 cancer genes; RT-qPCR)	50 gene assay (50 cancer genes; direct mRNA counting) + clinical factors	4 IHC tests (ER, PR, HER2, Ki67) + clinical factors
Testing location	Local laboratory or test service (Germany)	Test service (the Netherlands)	Test service (US)	Local laboratory or test service (UK)	Local laboratory
Stage	Early stage	Early stage (stage I or II)	Early stage (stage I or II)	Early stage (stage I to IIIA)	Early stage
Lymph node status	LN- and LN+ (up to 3 positive)	LN- or LN+ (up to 3 positive)	LN- or LN+ (up to 3 positive)	LN- and LN+	LN- and LN+ (1 to 3 positive nodes)
Hormone receptor status	ER+	ER+ or ER-	ER+	ER+	ER+
HER2 status	HER2-	HER2- or HER2+	HER2-	HER2-	HER2- or HER 2+
Menopausal status	Pre- and post-menopausal	Pre- and post-menopausal	Pre- and post-menopausal	Post-menopausal	Post-menopausal
Test result	Low risk, high risk	Low risk, high risk	Low risk, intermediate risk, high risk	Low risk, intermediate risk, high risk Intrinsic subtype	Low risk, intermediate risk, high risk
Assumptions	Scores assume 5 years of hormonal treatment	Assumes no therapy	Score assumes 5 years of hormonal treatment	Score assumes 5 years of hormonal treatment	Score assumes 5 years of hormonal treatment

Abbreviations: ER+/- oestrogen receptor positive or negative; LN+/- lymph node positive or negative; PR Progesterone receptor; HER2 human epidermal growth factor receptor; IHC immunohistochemistry

3.3 Populations and relevant subgroups

People with ER-positive (and/or PR-positive), HER2-negative, early stage breast cancer (stages I or II) with 0 to 3 positive lymph nodes.

Whilst early stage breast cancer can be defined as Stages I, IIA, IIB and IIIA[2] (See Appendix 9.1 for definitions), the focus of this assessment is patients with Stages I and II disease. Clinical advice received during the scoping workshops indicated that patients with Stage IIIA disease would routinely receive chemotherapy, as the risk of recurrence is high, and as such tumour profiling would not be useful.

Subgroups

Where data permits, the following subgroups may be considered:

- People with lymph node negative cancer; people with micro-metastases in the lymph nodes; and people with 1 to 3 positive lymph nodes
- Premenopausal women and postmenopausal women
- People predicted to be at low, intermediate or high risk using a risk assessment tool, or using clinical and pathological features
- Males and females
- People of different ethnicities

3.4 Place of the intervention in the treatment pathway(s)

Tests will be used in the secondary or tertiary care setting to make decisions about adjuvant chemotherapy treatment. Tests predicting the risk of recurrence in a specific population are likely to be used after surgery, in conjunction with other information available about tumour size, grade etc, to guide the use of adjuvant chemotherapy. Use of these tests in the neoadjuvant therapy setting (where chemotherapy would be given as a first step to shrink the tumour before surgery) will not be evaluated.

3.5 Relevant comparators

Current decision making, which may include any tool, or clinical and pathological features, used to assess risk. Clinicopathological tools used in current practice include PREDICT, the Nottingham Prognostic Index and Adjuvant! Online.

3.6 Key factors to be addressed (e.g. clinical and cost outcomes, further considerations, problematic factors)

Outcomes

Relevant outcomes include the following:

Intermediate measures:

- Time to test results
- Analytical validity (where applicable, see Section 4.4)
- Prognostic ability (e.g. calibration, discrimination, re/classification etc)
- Ability to predict benefit from chemotherapy
- Impact of test results on decision making

Clinical outcomes:

- Disease free survival
- Overall survival
- Distant recurrence
- Disease-related morbidity and mortality
- Chemotherapy-related morbidity and mortality

Patient-reported outcomes:

- Health related quality of life
- Anxiety

Costs will be considered from an NHS and Personal Social Services perspective. The cost-effectiveness of interventions will be expressed in terms of incremental cost per quality-adjusted life year. Costs for consideration may include:

- Costs of treating breast cancer, including: drug cost, administration cost, outpatient appointments, and treatment of adverse events
- Costs of the tests, including equipment costs and reagents when relevant
- Costs of staff and associated training

Issues for consideration

The following issues will require careful consideration:

There may be few studies directly comparing the tests head to head, or to some comparators, e.g. PREDICT.

The use of clinical & pathological factors alongside the tests will need to be considered in terms of how they are used to target patients to receive tests. The most challenging decisions are for the patients who are categorised at intermediate risk by existing prognostic tools, where the decision to undergo chemotherapy or not is most uncertain and additional information would be most beneficial. Existing prognostic tools e.g. NPI or PREDICT may be used to identify subgroups of patients. For instance NPI identifies a group of patients at intermediate risk, with a NPI score >3.4 and <= 5.4. PREDICT calculates the absolute 10-year survival benefit from chemotherapy. The Cambridge Breast Unit (UK) uses this to guide decision making for adjuvant chemotherapy: <3% no chemotherapy; 3-5% chemotherapy discussed as a possible option; >5% chemotherapy recommended. Adjuvant Online is temporarily disabled whilst the tool is being updated to reflect the most recent information. It is unclear how long this will be the case and to what extent the tool will have changed following this work. Clinical advice will be sought to identify the most commonly used tool(s) and clarify how these are used to identify subgroups of patients.

Clinical and pathological factors are also used alongside the results provide by the tumour profiling tests intervention to guide therapy (either incorporated formally within the test or informally in addition to the test results). This will also be considered, where evidence allows.

The impact of recent changes in practice for treatment of early breast cancer e.g. the use of bisphosphonates, the use of extended endocrine therapy (up to 10 years), and the use of aromatase inhibitors in place of tamoxifen, will impact on the baseline risk of recurrence for these patients, but is unlikely to be reflected in the historic evidence base.

The proportion of patients with early breast cancer receiving chemotherapy varies widely between countries; this is likely to impact on the outcome relating to changes in chemotherapy use when using the tests. UK –specific data will therefore be the most relevant in this instance.

3.7 Areas of agreement at the scoping workshop that are outside the scope of the evaluation and therefore do not require any detailed assessment (*e.g. key factors for which evidence is already accepted*).

Areas which will be excluded from this appraisal:

- HER2-positive population
- Stage III
- Chemotherapy in the neoadjuvant setting
- Impact on use or benefit of endocrine therapy

4. Report methods for assessing the outcomes arising from the use of the interventions

A systematic review of the evidence relating to this assessment will be undertaken. The review will be conducted following the general principles recommended in CRD's guidance,[6] the PRISMA statement,[7] the NICE Diagnostic Assessment Programme manual.[8] and the Cochrane Prognosis Methods Group.[9]

This systematic review will update a previous systematic review (Ward et al 2013[3]) conducted for DG10.[4] This review covered Oncotype DX, MammaPrint, IHC4 and PAM50. Prosigna was developed based on the PAM50 gene signature after the completion of Ward et al 2013, and shares a common evidence base with PAM50 up to this point. Endopredict was not included in Ward et al 2013,[3] and we will conduct a *de novo* systematic review with no date limits for this technology. Evidence relating to clinical validity and clinical utility will be considered, and analytical validity reviewed where necessary using rapid review methods (see Section 4.4). Whilst some tests were developed only to prognosticate disease recurrence, the intended use in the decision problem is to guide chemotherapy treatment decisions. As such, whilst both evidence relating to prognosis of disease recurrence and prediction of chemotherapy benefit will be eligible for inclusion in the review, the evidence base will be critically appraised and interpreted with respect to the decision problem.

An anticipated complication of the evidence base is that there may be few studies that are conducted in England, or that mirror English clinical practice. This is important, as chemotherapy prescription practices are disparate across countries, meaning a comparator of "normal practice" in a given study may have low relevance to our decision problem. This is most likely to affect studies that assess changes to treatment decisions. As such, and given the potentially very large evidence base for this appraisal, we propose to initially map the evidence base in terms of key study characteristics (location; population; intervention; comparator; outcome types) and if the evidence base is very large, we will further select studies with respect to a) highest relevance to practice in England (where this is likely to affect the relevance of study results, and to be determined in consultation with clinical advisors to the project); b) levels of evidence (with reference to published hierarchies[10, 11, 16]); and c) sample size, ensuring the best available evidence in terms of internal and external validity are included. The process of further study selection will be conducted in consultation with NICE, and reasons for selection documented.

The inclusion criteria for the review are detailed in Sections 4.1 to 4.5. The title and abstract of each record retrieved by the search strategy (Section 4.6) will be assessed against the inclusion criteria of the review, and irrelevant records excluded. The full text of remaining records will

be obtained and assessed against the inclusion criteria. Study selection will be conducted by one reviewer. Any studies which give rise to uncertainty will be reviewed by a second reviewer with involvement of a third reviewer when necessary. A 10% sample of the records retrieved by electronic searches will be checked by a second reviewer, and a check of all retrieved titles undertaken where the Kappa statistic for agreement between reviewers is <0.7.

Inclusion criteria

4.1 Population

Studies that selected people with ER-positive (and/or PR-positive), HER2-negative, early stage breast cancer (stages I or II) with 0 to 3 positive lymph nodes.

Studies that recruit a wider population will be included where data is reported for the relevant subgroup separately. Studies in early stage breast cancer (but not restricted to stages I or II) will be included in the review and their contribution to outcome heterogeneity considered. Where studies include patients who are non-early stage or who are otherwise out of scope, and no subgroup data are available, the following rule will be applied: if the percentage of patients out of scope is ≤20% then the study will be included (but excluded in sensitivity analyses to assess impact on conclusions), while if >20% are out of scope the study will be excluded.

Where data permits, the following subgroups may be considered:

- People with lymph node negative cancer; people with micro-metastases in the lymph nodes; and people with 1 to 3 positive lymph nodes
- Premenopausal women and postmenopausal women
- People predicted to be at low, intermediate or high risk using a risk assessment tool, or using clinical and pathological features
- Males and females
- People of different ethnicities.

4.2 Interventions

The following interventions identified in the NICE scope will be included:

- EndoPredict
- MammaPrint
- Oncotype DX Breast Recurrence Score
- Prosigna
- IHC4

Three of the tests included in the scope incorporate clinical and pathological features into the test results (EndoPredict, Prosigna and IHC4). However, evidence may be available on test results or versions of the test which do not formally incorporate clinical and pathological features. The other 2 tests do not formally include clinical and pathological features (MammaPrint and Oncotype DX Breast Recurrence Score). However, evidence may exist in which additional algorithms have been used to formally incorporate clinical and pathological features. Where such studies are identified, these will be included in the review, but will be grouped, synthesised and interpreted separately.

Studies using the interventions alone or in conjunction with clinical practice will be included. Current clinicopathological tools used in England include PREDICT, the Nottingham Prognostic Index and Adjuvant! Online.

4.3 Comparators

Comparative studies will be included that have current decision making as the comparator. This may include any tool, or clinical and pathological features, used to assess risk. Clinicopathological tools used in current practice in England include PREDICT, the Nottingham Prognostic Index and Adjuvant! Online (although not currently available).

Studies that present a head to head comparison between interventions will be also be included. Other study designs eligible for inclusion may not include a comparator, e.g. retrospective analyses.

4.4 Outcomes

Outcomes include:

Intermediate measures:

- Time to test results
- Analytical validity (where applicable, see below)
- Prognostic ability (e.g. calibration, discrimination, re/classification etc)
- Ability to predict benefit from chemotherapy
- Impact of test results on decision making

Clinical outcomes:

- Disease free survival

- Overall survival
- Distant recurrence
- Disease-related morbidity and mortality
- Chemotherapy-related morbidity and mortality

Patient-reported outcomes:

- Health related quality of life
- Anxiety

Where possible, evidence on analytical validity will be summarised with reference to existing high quality systematic reviews. For tests where analytical validity has not been previously reviewed, a rapid review with narrative synthesis will be conducted, and formal quality assessment will not be undertaken.

4.5 Study design

The highest level of evidence for the guiding of treatment decisions would be a randomised controlled trial (or systematic review thereof) that randomises patients to either treatment guided with the intervention, or treatment guided according to usual practice. However, based on our scoping searches, and given that many of the interventions were designed as prognostic not predictive tools, and given the difficulties with the length of follow-up required and powering[10], we do not anticipate identifying many or any such studies. Consequently, broad inclusion criteria for study design will be applied in this review, to capture the diversity of evidence available. Any study purporting to analyse the clinical validity (the ability of the test to reliably and accurately predict the clinically defined disorder or phenotype of interest[11]) or clinical utility (the ability of the test to improve measurable clinical outcomes, and its usefulness and added value to patient management[11]) of the interventions will be eligible for inclusion. Analytical validity (ability to accurately and reliably measure the genotype (or analyte) of interest in the clinical laboratory, and in specimens representative of the population of interest[11]) studies will be included where applicable (see Section 4.4).

Systematic reviews identified during study selection will be used to check for additional studies, and used in data extraction (see Section 4.7).

Studies not published in English language will be included if sufficient PICOS data can be extracted from non-English language full-texts, or from an existing English language abstract. Studies excluded on the basis of language will be listed separately. Non-peer-reviewed reports

or abstracts will only be included if the data are presented in a succinct and accessible manner (e.g. a manuscript prepared for submission to a journal), if sufficient methodological details are reported to allow critical appraisal of the study quality, and if results are reported in sufficient detail.

4.6 Search strategy

The search strategy for the systematic review will comprise the following main elements:

- Searching of electronic databases
- Contact with experts in the field
- Scrutiny of bibliographies of retrieved papers
- Citation searching.

The databases, trials registers and websites that will be searched include the following:

- MEDLINE and MEDLINE in Process (for latest publications)
- EMBASE
- The Cochrane Library (CDSR, DARE, CENTRAL, and HTA)
- Web of Science Citation Index Expanded
- Web of Science Conference Proceedings Citation Index
- WHO International Clinical Trials Registry Platform
- American Society of Clinical Oncology (ASCO)
- European Society for Medical Oncology (ESMO).

Search terms will include both product names and any alternative names for each of the intervention tests, combined with search terms for breast cancer. Manufacturer website publication lists will also be searched for potentially relevant studies. A draft MEDLINE search strategy is included in Appendix 9.2.

The clinical and cost effectiveness searches will be limited by date from January 2011 to present for the interventions MammaPrint, IHC4, Oncotype DX and Prosigna. This is the date when searches in the published diagnostic guidance (DG10[4]) were last conducted. Searches for EndoPredict will not be limited by date.

Reference lists of included papers will be assessed for additional relevant studies. Where necessary and where time allows, authors of eligible studies will be contacted for further information (e.g. full text of citations listed ahead of print). All searches will be limited to human studies. No limits relating to study design will be applied.

4.7 Data extraction strategy

A data extraction form will be constructed in Excel, piloted using two examples of each study design, and amended as required. It may be necessary to design different forms for different study designs. The CHARMS checklist (see appendix 9.3) and the CRD handbook[6] will be consulted to select relevant fields for piloting.

Data will be extracted by one reviewer, and checked by a second. Any disagreements will be resolved through discussion and consultation with a third reviewer where necessary. If time constraints allow, attempts will be made to contact authors for any missing data that is essential to the review. Data from multiple publications of the same study will be extracted as a single study. Where data are reported in an existing high quality (according to the checklist "Assessing the Methodological Quality of Systematic Reviews" (AMSTAR[12])) systematic review (e.g. Ward et al 2013; or any subsequent reviews identified via searches), all relevant data will be extracted from the review. If necessary, additional data will be extracted from the original papers.

4.8 Quality assessment strategy

Studies will be assessed using quality assessment tools relevant to the study design. Tools may be adapted or abbreviated to the specifics of this review, due to time and resource constraints.

For studies that evaluate the clinical utility of the interventions to guide treatment decisions, and use an RCT design, quality will be assessed using the Cochrane Risk of Bias tool.[13]

For studies that develop and/or validate the tests as prediction models, quality will be assessed using the current draft of the Prediction model study Risk Of Bias Assessment Tool (PROBAST) (personal communication, January 2017, Dr Robert Wolff). The PROBAST tool has been developed specifically for use in systematic reviews of prediction models by the Cochrane Prognosis Methods Group.[9] Whilst this tool is not yet validated or published, it has been designed using robust methods including 42 topic experts and a Delphi process,[14] and is freely available from the lead author (Dr Robert Wolff).

Any studies that do not fit into the above categories will be assessed using an alternative published tool relevant to the study design, such as the ROBINS tool (Risk Of Bias In Non-randomised Studies - of Interventions),[15] Systematic reviews will be assessed using AMSTAR.[12]

Studies will be quality assessed by one reviewer, with scores checked by a second. Any disagreements will be resolved through discussion and consultation with a third reviewer where necessary.

The impact of the quality of studies on the evidence base will be evaluated through sensitivity analyses in meta-analysis, or through narrative synthesis of the results.

4.9 Methods of analysis/synthesis

Interpretation of the evidence base will be conducted with reference to published hierarchies for predictive studies[10, 11, 16], and with regard to the ability of the study design to adequately address the decision problem.

For each intervention, studies will be ordered according to population, comparator, outcomes and study design. A narrative synthesis will be conducted, drawing on existing high quality systematic reviews where possible.

The ability to make meaningful inferences will depend on the data that is extracted.

We cannot describe the specific analyses that may be performed until we are familiar with the data that is extracted and how will be used. We will use appropriate statistical models to synthesise the data, where feasible, and allow meaningful inferences to be made about the impact of the prediction models and the benefit of treatment.

Any comparisons will acknowledge clinical, methodological and statistical heterogeneity.

4.10 Methods for estimating quality of life – if possible and relevant for the systematic review in question

Quality of life estimates reported within the clinical literature included in this review will be collated as part of the systematic review, whilst data in the cost-effectiveness literature will be identified as part of Section 5.

5 Report methods for synthesising evidence of cost effectiveness

5.1 Identifying and systematically reviewing published cost-effectiveness studies

A systematic review of the existing literature studying the cost effectiveness of the five identified tests to guide selection of chemotherapy regimes in breast cancer management will be undertaken.

The search strategy will comprise the following main elements:

- Searching of electronic databases
- Contact with experts in the field
- Scrutiny of bibliographies of retrieved papers

The databases that will be searched include the following:

- MEDLINE and MEDLINE in Process (for latest publications);
- EMBASE;
- Centre for Reviews and Dissemination (HTA and NHSEED)
- Web of Science Citation Index Expanded;
- Web of Science Conference Proceedings Citation Index

Where applicable, cost-effectiveness studies will be identified using an economic search filter. In addition, relevant cost papers identified from the clinical effectiveness searches will be included in the economic review.

The cost effectiveness searches will be limited by date from January 2011 to present for the interventions MammaPrint, IHC4, Oncotype DX and Prosigna (previously PAM50). This is the date when searches in the published diagnostic guidance (DG10) were last conducted. Searches for EndoPredict will not be limited by date.

Citation searches of key included studies will also be undertaken.

Additional searches, for example to inform the decision-analytic model, where required in the course of the project, will be undertaken through consultation between the ScHARR team.

5.2 Evaluation of costs and cost effectiveness

Only full economic evaluations published in English addressing the cost-effectiveness of the five tests compared with NPI, Adjuvant! Online, PREDICT (or any adaptations of these tools in clinical practice) or comparing one test against each other will be critically appraised using published checklists. Cost-effectiveness studies that compared tests with other guidelines such as St Gallen, the National Comprehensive Cancer Network (NCCN) and NIH guidelines will be excluded from the review because of time and resource constraints as these comparators are not directly relevant to the UK context, but such studies will be scanned by the reviewers to inform the model development.. Existing cost effectiveness analyses may also be used to

identify sources of evidence to inform structural assumptions and parameter values for the External Assessment Group model.

The quality of identified cost-effectiveness studies will be assessed against a critical appraisal checklist adapted from the Drummond (Drummond 1996) and Eddy (Eddy 1985) checklists (Appendix 9.4).

5.3 Development of a health economic model

Tumour profiling tests aim to improve the use of chemotherapy in breast cancer by stratifying patients and identifying those patients who are at high risk of recurrence and/or will gain most benefit from chemotherapy. These tests may report information on breast cancer sub-types and/or risk of recurrence/chemotherapy benefit. The focus will be on the risk of recurrence and chemotherapy benefit. Tests predicting the risk of recurrence in a specific population are likely to be used after surgery, in conjunction with other information available such as tumour size and grade to guide the use of adjuvant therapy.

The objective of the economic evaluation will be to explore the cost effectiveness of tests in the adjuvant chemotherapy setting. The cost effectiveness of these tests in the neo-adjuvant setting will not be evaluated. For three of the tests (Oncotype DX, IHC4 and Mammaprint) prior economic evaluations exist as part of NICE Diagnostic Guidance 10 and these will be reviewed and updated as appropriate. The structure of the model may need to be adapted to take account of new evidence or new comparators. For Prosigna and EndoPredict a new evaluation will be undertaken, if sufficient evidence exists. The feasibility of modelling any individual test will be dependent on the level of evidence available, the robustness of data and time constraints within the project.

Tests that do not have fully reported external validation studies (i.e. validation on an independent dataset) will not be included in the economic evaluation. Evidence identified by the systematic review of clinical evidence on the impact of the new tests on adjuvant chemotherapy treatment decisions, compared with current clinical practice in England (e.g. PREDICT or NPI) will be used in the economic model. Tests validated for use in predicting chemotherapy benefit will be distinguished from those using information on prognosis.

Both predictive (of treatment response) and prognostic (of risk recurrence) information may be used to inform chemotherapy decisions. Therefore, the EAG will seek to undertake economic evaluation of tests that provide either or both types of information if suitable evidence allows.

We anticipate that the most commonly used comparators for predicting the risk of recurrence after surgery to guide the use of chemotherapy in England are PREDICT and NPI. Adjuvant! Online is not currently available. We will seek clinical advice in terms of the most appropriate comparator(s) to be used in the economic model, where data permits

The primary outcome from the model will be an estimate of the incremental cost per additional quality-adjusted life year (QALY) gained associated with the use of tests to improve the use of chemotherapy in breast cancer. Secondary outcomes (health benefits) will also be presented. Costs and benefits will be captured using a lifetime horizon and modelled in line with the NICE Diagnostic Assessment Programme Manual (NICE, 2011[8]). The model will adopt the perspective of the UK NHS and personal social services with costs and benefits discounted at an annual rate of 3.5%. Modelling assumptions will be taken from the literature, supplemented by clinical expert opinion. Quality of life data identified from the systematic review of clinical evidence or from the identified cost effectiveness papers and/or any recent systematic reviews of quality of life in breast cancer will be used to generate the quality adjustment weights required to estimate QALYs. Costs will be derived from national sources (e.g. NHS reference costs, national unit costs, British National Formulary) and data provided by the manufacturers.

It is anticipated that there will be differences in the level and quality of evidence supporting each of the tests. Combining evidence from different studies, based on different methodologies and with different patient characteristics will limit the conclusions that could be drawn from any comparisons that could be made between the analyses. It may therefore be more appropriate to perform separate analyses for each test using the best direct sources of data available for each test; in this case it would not be appropriate to directly compare these analyses. An incremental analysis will be included, if appropriate and if evidence allows.

In the base case analysis, tests will be assessed in line with their intended use (see table 1). EndoPredict, Prosigna and IHC4 incorporate clinical and pathological features into the test results. However, evidence may be available on test results or versions of the test which do not formally incorporate clinical and pathological features. MammaPrint and Oncotype DX do not formally include clinical and pathological features. However, evidence may exist in which additional algorithms have been used to formally incorporate clinical and pathological features. Where such studies are identified, the impact of this will be explored in sensitivity analyses, where appropriate and feasible.

It is anticipated that there may be limited evidence for some of the parameters that will be included in the economic model. Therefore, the uncertainty around the parameter estimates will

be modelled to take this into account. Deterministic sensitivity analysis will be undertaken to explore the sensitivity of the results to variations in specific input parameters. A range of scenarios will be presented, if needed, varying key model assumptions to identify assumptions that have most impact on the ICER. Results will be presented for important subgroups for which sufficient evidence exists. Probabilistic sensitivity analysis will be carried out using Monte Carlo simulation. The uncertainty in each parameter will be represented using a probability distribution, with correlation between parameters maintained if identified. The decision uncertainty will be presented as the probability that each intervention is the most cost-effective for a given cost-effectiveness threshold. Cost-effectiveness acceptability curves will also be presented to illustrate graphically the decision uncertainty.

6 Handling information from the companies

All data submitted by the manufacturers/sponsors will be considered if received by the EAG no later than Monday 10 April. Data arriving after this date may not be considered.

If the data meet the inclusion criteria for the review they will be extracted and quality assessed in accordance with the procedures outlined in this protocol.

Any 'commercial in confidence' data provided by a manufacturer and specified as such will be highlighted in blue and underlined in the assessment report (followed by an indication of the relevant company name e.g. in brackets). Any 'academic in confidence' data provided by the manufacturer, and specified as such, will be highlighted in yellow and underlined in the assessment report. Any confidential data used in the cost-effectiveness model will also be highlighted.

A version of the economic model with confidential information redacted or replaced with dummy data will be provided.

7 Competing interests of authors

None

8 Timetable/milestones

Milestone	**Date to be completed**
Final date for Manufacturer/sponsor data submissions	10th April
Progress Report	30th May
Draft Assessment Report	25th July
Final Report to NICE	22th August

9 Appendices

9.1 Table of breast cancer stages, compiled from Cancer Research UK[2] and National Breast Cancer Foundation[17]

		Tumour size	Lymph nodes	Spread	
	Stage 0	NR	0	Not spread beyond the tissue of origin	
Early stage invasive breast cancer	Stage 1a	≤2cm	0	Not spread beyond breast	Microscopic invasion of tissue outside the lining of the duct or lobule, but not >1mm
	Stage 1b	0 (ie no tumor) to ≤2cm	0.2-2mm groups of cells in lymph nodes		
	Stage 2a	0 to ≤2cm	>2mm in 1-3 axillary or breast bone lymph nodes		
		>2<5cm	0		
	Stage 2b	>2<5cm	0.2-2mm groups of cells in lymph nodes		
			1-3 axillary or breast bone lymph nodes		
		>5cm	0		
	Stage 3a	Any size, or none	4-9 axillary or breast bone lymph nodes		
		>5cm	0.2-2mm in lymph nodes		
		>5cm	1-3 axillary or breast bone lymph nodes		
	Stage 3b	Any size, or none	≤9 axillary or breast bone lymph nodes,	Chest wall and/or skin, causing swelling or ulcer	
	(inflammatory breast cancer)			Reddening of large portion of skin; warm and may be swollen; cancer cells spread to	

				lymph nodes/skin	
	Stage 3c	Any size, or none.	≥10 axillary LNs10 axillary LNs OR spread to collarbone LNs OR spread to axillary AND breastbone LNs	Chest wall and/or skin, causing swelling or ulceration	
	Stage 4	Any size	Any	Metastasised to other parts of the body	

9.2 Draft search strategy

Search for interventions

EndoPredict date limits (statements 10-13): none
MammaPrint, IHC4, Oncotype DX and Prosigna intervention date limits (statements 14-40): 2011-present

Epub Ahead of Print, In-Process & Other Non-Indexed Citations, MEDLINE(R) Daily and MEDLINE (Ovid)

1 exp Breast Neoplasms/
2 exp mammary neoplasms/
3 exp "Neoplasms, Ductal, Lobular, and Medullary"/
4 exp breast/
5 exp neoplasms/
6 4 and 5
7 (breast* adj5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullary)).mp.
8 (mammar* adj5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)).mp.
9 1 or 2 or 3 or 6 or 7 or 8
10 endopredict.mp.
11 myriad genetics.mp.
12 ep score.mp.
13 epclin score.mp.
14 mammaPrint.mp.
15 70-gene.mp.
16 gene70.mp.
17 gene?seventy.mp.
18 seventy?gene.mp.
19 amsterdam profile.mp.
20 oncotype.mp.
21 oncotype dx.mp.
22 21-gene.mp.
23 gene21.mp.
24 gene?twentyone.mp.
25 twentyone?gene.mp.
26 ghi recurrence score.mp.
27 ghi-rs.mp.
28 92-gene.mp.
29 gene92.mp.
30 gene?ninetytwo.mp.
31 ninetytwo?gene.mp.
32 (rct-pcr adj5 '21').mp.
33 prosigna.mp.
34 pam50.mp.
35 50-gene.mp.
36 gene50.mp.
37 gene?fifty.mp.
38 fifty?gene.mp.
39 breast bioclassifier.mp.
40 ihc4.mp.
41 or/10-13
42 or/14-40

43 41 or (9 and 42)
44 limit 43 to yr="2011 -Current"

9.3 CHARMS checklist for data extraction[16]

CHARMS 2014 Relevant items to extract from individual studies in a systematic review of prediction models

Domain	Key items	Reported on page #
SOURCE OF DATA	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	
PARTICIPANTS	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria)	
	Participant description	
	Details of treatments received, if relevant	
	Study dates	
OUTCOME(S) TO BE PREDICTED	Definition and method for measurement of outcome	
	Was the same outcome definition (and method for measurement) used in all patients?	
	Type of outcome (e.g., single or combined endpoints)	
	Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)?	
	Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)?	
	Time of outcome occurrence or summary of duration of follow-up	
CANDIDATE PREDICTORS (OR INDEX TESTS)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics)	
	Definition and method for measurement of candidate predictors	
	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment initiation)	
	Were predictors assessed blinded for outcome, and for each other (if relevant)?	
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)	
SAMPLE SIZE	Number of participants and number of outcomes/events	
	Number of outcomes/events in relation to the number of candidate predictors (Events Per Variable)	
MISSING DATA	Number of participants with any missing value (include predictors and outcomes)	
	Number of participants with missing data for each predictor	
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)	
MODEL DEVELOPMENT	Modelling method (e.g., logistic, survival, neural network, or machine learning techniques)	
	Modelling assumptions satisfied	
	Method for selection of predictors **for inclusion** in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)	
	Method for selection of predictors **during multivariable modelling** (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike Information Criterion)	
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized estimation)	
MODEL PERFORMANCE	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination (C-statistic, D-statistic, log-rank) measures with confidence intervals	
	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used	
MODEL EVALUATION	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)	
	In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)	
RESULTS	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)	
	Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance	
	Comparison of the distribution of predictors (including missing data) for development and validation datasets	
INTERPRETATION AND DISCUSSION	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)	
	Comparison with other studies, discussion of generalizability, strengths and limitations.	

9.4 Critical appraisal checklist for economic evaluations using key components of the British Medical Journal checklist for economic evaluations (Drummond & Jefferson 1996) together with the Eddy checklist on mathematical models employed in technology assessments (Eddy 1985)

Reference ID		
Title		
Authors		
Year		
Modelling assessments should include:		**Yes/No**
1	A statement of the problem;	
2	A discussion of the need for modelling vs. alternative methodologies	
3	A description of the relevant factors and outcomes;	
4	A description of the model including reasons for this type of model and a specification of the scope including; time frame, perspective, comparators and setting. *Note: n=number of health states within sub-model*	
5	A description of data sources (including subjective estimates), with a description of the strengths and weaknesses of each source, with reference to a specific classification or hierarchy of evidence;	
6	A list of assumptions pertaining to: the structure of the model (e.g. factors included, relationships, and distributions) and the data;	
7	A list of parameter values that will be used for a base case analysis, and a list of the ranges in those values that represent appropriate confidence limits and that will be used in a sensitivity analysis;	
8	The results derived from applying the model for the base case;	
9	The results of the sensitivity analyses; unidimensional; best/worst case; multidimensional (Monte Carlo/parametric); threshold.	
10	A discussion of how the modelling assumptions might affect the results, indicating both the direction of the bias and the approximate magnitude of the effect;	
11	A description of the validation undertaken including; concurrence of experts; internal consistency; external consistency; predictive validity.	
12	A description of the settings to which the results of the analysis can be applied and a list of factors that could limit the applicability of the results;	
13	A description of research in progress that could yield new data that could alter the results of the analysis	

Additional information that is needed by NETSCC, HTA and NICE.

Please send this as a WORD document when you submit your protocol to

Htatar@nihr.ac.uk

Details of EAG

Ward, Sue, Ms

Senior Research Fellow, Health Economics and Decision Science

ScHARR, University of Sheffield

Regent Court

30 Regent Street

Sheffield

S1 4DA

Telephone number: (+44) (0)114 2220816

E-mail address: s.e.ward@sheffield.ac.uk

Harnan, Sue, Ms

Senior Research Fellow, Health Economics and Decision Science

ScHARR, University of Sheffield

+44 (0)114 2220869

s.harnan@sheffield.ac.uk

Cooper, Katy, Dr

Senior Research Fellow, Health Economics and Decision Science

ScHARR, University of Sheffield

+44 (0)114 2220773

k.l.cooper@sheffield.ac.uk

Bessey, Alice, Ms

Research Associate

ScHARR, University of Sheffield

+44 (0)114 2220814

a.r.bessey@sheffield.ac.uk

Stevens, John, Dr
Reader in Decision Science, Health Economics and Decision Science
ScHARR, University of Sheffield
+44 (0) 114 222 6396
j.w.stevens@sheffield.ac.uk

Wong, Ruth, Dr
Information Specialist, Health Economics and Decision Science
ScHARR, University of Sheffield
+44 (0)114 222 0797
ruth.wong@sheffield.ac.uk

Shippam, Andrea, Ms
Programme Manager
ScHARR, University of Sheffield
+44 (0)114 2220693
a.shippam@sheffield.ac.uk

Please indicate to whom you wish all correspondence to be addressed

Ms Sue Ward, copying in Andrea Shippam

Timetable/milestones

Please confirm the date that you will submit:

- Progress report (to NETSCC, HTA who forward it to NICE within 24 hr): 30th May 2017
- Assessment report (simultaneously to NICE and NETSCC, HTA): 22nd August 2017

References

1. Paik S. Molecular assays to predict prognosis of breast cancer. *Clinical Advances in Hematology & Oncology* 2007;5:681-2.
2. Cancer Research UK. Number stages of breast cancer. *Journal* 2017. http://about-cancer.cancerresearchuk.org/about-cancer/breast-cancer/stages-types-grades/number-stages
3. Ward S, Scope A, Rafia R, Pandor A, Harnan S, Evans P*, et al.* 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 (Winchester, England)* 2013;17:1-302.
4. National Institute for Health and Care Excellence. Gene expression profiling and expanded immunohistochemistry tests for guiding adjuvant chemotherapy decisions in early breast cancer management: MammaPrint, Oncotype DX, IHC4 and Mammostrat. 2013, https://www.nice.org.uk/guidance/dg10.
5. Bartlett JM, Christiansen J, Gustavson M, Rimm DL, Piper T, van de Velde CJ*, et al.* Validation of the IHC4 Breast Cancer Prognostic Algorithm Using Multiple Approaches on the Multinational TEAM Clinical Trial. *Archives of pathology & laboratory medicine* 2016;140:66-74.
6. Centre for Reviews Dissemination. Systematic review: CRD's guidance for undertaking reviews in health care. CRD, University of York; 2009. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]; 2011.
7. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA*, et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339.
8. National Institute for Health and Care Excellence. Diagnostics Assessment Programme Manual. 2011, https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-diagnostics-guidance/Diagnostics-assessment-programme-manual.pdf.
9. Cochrane Prognosis Methods Group. http://methods.cochrane.org/prognosis/.
10. Michiels S, Ternès N, Rotolo F. Statistical controversies in clinical research: prognostic gene signatures are not (yet) useful in clinical practice. *Annals of Oncology* 2016;27:2160-7.
11. Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N*, et al.* The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: methods of the EGAPP Working Group. *Genetics in Medicine* 2009;11:3-14.
12. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C*, et al.* Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology* 2007;7:10.
13. Cochrane Handbook for Systematic Reviews of Interventions. Chapter 8: Assessing risk of bias in included studies; 2011.
14. Wolff R, Whiting P, Mallett S, Riley R, Westwood M, Kleijnen J*, et al.* PROBAST. 2015, https://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0ahUKEwjv-ZWs6uXRAhWJDcAKHUkbDNYQFggjMAE&url=http%3A%2F%2Fs371539711.initial-website.co.uk%2Fapp%2Fdownload%2F5811143946%2FGIN%2B2015%2B-%2BPROBAST.pdf&usg=AFQjCNGklEDSPD4i_cAwHlVGn6axxG4whA&sig2=1_SEN19oEpja4g5HkciF0Q&cad=rja.
15. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M*, et al.* ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355.

16. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, *et al.* Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLOS Medicine* 2014;11:e1001744.
17. National Breast Cancer Foundation Inc. Stages 0 and 1. *Journal* 2017. http://www.nationalbreastcancer.org/breast-cancer-stage-0-and-stage-1