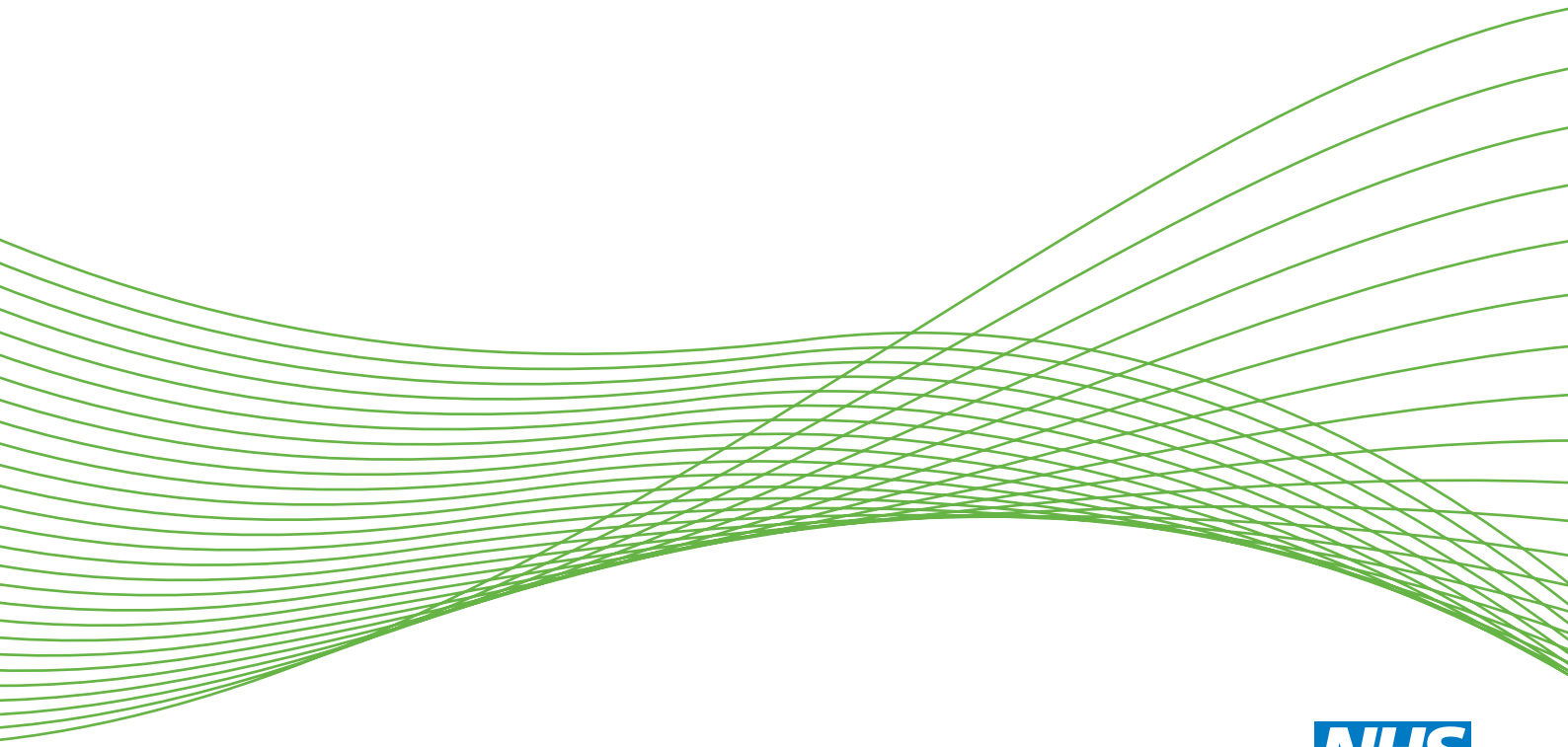
HEALTH TECHNOLOGY ASSESSMENT

VOLUME 17 ISSUE 44 OCTOBER 2013
ISSN 1366-5278

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

S Ward, A Scope, R Rafia, A Pandor, S Harnan, P Evans and L Wyld



NHS
National Institute for Health Research

DOI 10.3310/hta17440

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

S Ward,[1]* A Scope,[1] R Rafia,[1] A Pandor,[1] S Harnan,[1] P Evans[1] and L Wyld[2]

[1]School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK
[2]Royal Hallamshire Hospital, Sheffield, UK

*Corresponding author

Declared competing interests of authors: none

Published October 2013
DOI: 10.3310/hta17440

This report should be referenced as follows:

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 Technol Assess*, 2013;**17**(44).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (*SciSearch*®) and *Current Contents*®/ Clinical Medicine.

Health Technology Assessment

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Five-year impact factor: 5.804

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index and is assessed for inclusion in the Database of Abstracts of Reviews of Effects.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: www.hta.ac.uk/

This report

The research reported in this issue of the journal was commissioned and funded by the HTA programme on behalf of NICE as project number 10/125/01. The protocol was agreed in May 2011. The assessment report began editorial review in October 2011 and was accepted for publication in March 2012. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

© Queen's Printer and Controller of HMSO 2013. This work was produced by Ward *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Editor-in-Chief of *Health Technology Assessment* and NIHR Journals Library

NIHR Journals Library Editors

Abstract

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

S Ward,[1]* A Scope,[1] R Rafia,[1] A Pandor,[1] S Harnan,[1] P Evans[1] and L Wyld[2]

[1]School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK
[2]Royal Hallamshire Hospital, Sheffield, UK

*Corresponding author

Background: Gene expression profiling (GEP) and expanded immunohistochemistry (IHC) tests aim to improve decision-making relating to adjuvant chemotherapy for women with early breast cancer.
Objective: The aim of this report is to assess the clinical effectiveness and cost-effectiveness of nine GEP and expanded IHC tests compared with current prognostic tools in guiding the use of adjuvant chemotherapy in patients with early breast cancer in England and Wales. The nine tests are BluePrint, Breast Cancer Index (BCI), IHC4, MammaPrint, Mammostrat, NPI plus (NPI+), OncotypeDX, PAM50 and Randox Breast Cancer Array.
Data sources: Databases searched included MEDLINE, MEDLINE In-Process & Other Non-Indexed Citations, EMBASE and The Cochrane Library. Databases were searched from January 2009 to May 2011 for the OncotypeDX and MammaPrint tests and from January 2002 to May 2011 for the other tests.
Review methods: A systematic review of the evidence on clinical effectiveness (analytical validity, clinical validity and clinical utility) and cost-effectiveness was conducted. An economic model was developed to evaluate the cost-effectiveness of adjuvant chemotherapy treatment guided by four of the nine test (OncotypeDX, IHC4, MammaPrint and Mammostrat) compared with current clinical practice in England and Wales, using clinicopathological parameters, in women with oestrogen receptor-positive (ER+), lymph node-negative (LN–), human epidermal growth factor receptor type 2-negative (HER2–) early breast cancer.
Results: The literature searches for clinical effectiveness identified 5993 citations, of which 32 full-text papers or abstracts (30 studies) satisfied the criteria for the effectiveness review. A narrative synthesis was performed. Evidence for OncotypeDX supported the prognostic capability of the test. There was some evidence on the impact of the test on decision-making and to support the case that OncotypeDX predicts chemotherapy benefit; however, few studies were UK based and limitations in relation to study design were identified. Evidence for MammaPrint demonstrated that the test score was a strong independent prognostic factor, but the evidence is non-UK based and is based on small sample sizes. Evidence on the Mammostrat test showed that the test was an independent prognostic tool for women with ER+, tamoxifen-treated breast cancer. The three studies appeared to be of reasonable quality and provided data from a UK setting (one study). One large study reported on clinical validity of the IHC4 test, with IHC4 score a highly significant predictor of distant recurrence. This study included data from a UK setting and appeared to

be of reasonable quality. Evidence for the remaining five tests (PAM50, NPI+, BCI, BluePrint and Randox) was limited. The economic analysis suggests that treatment guided using IHC4 has the greatest potential to be cost-effective at a £20,000 threshold, given the low cost of the test; however, further research is needed on the analytical validity and clinical utility of IHC4, and the exact cost of the test needs to be confirmed. Current limitations in the evidence base produce significant uncertainty in the results. OncotypeDX has a more robust evidence base, but further evidence on its impact on decision-making in the UK and the predictive ability of the test in an ER+, LN–, HER– population receiving current drug regimens is needed. For MammaPrint and Mammostrat there were significant gaps in the available evidence and the estimates of cost-effectiveness produced were not considered to be robust by the External Assessment Group.

Limitations: Methodological weaknesses in the clinical evidence base relate to heterogeneity of patient cohorts and issues arising from the retrospective nature of the evidence. Further evidence is required on the clinical utility of all of the tests and on UK-based populations. A key area of uncertainty relates to whether the tests provide prognostic or predictive ability.

Conclusions: The clinical evidence base for OncotypeDX is considered to be the most robust. The economic analysis suggested that treatment guided using IHC4 has the most potential to be cost-effective at a threshold of £20,000; however, the evidence base to support IHC4 needs significant further research.

Study registration: PROSPERO 2011:CRD42011001361, available from www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42011001361.

Funding: The National Institute for Health Research Health Technology Assessment programme.

Contents

Glossary

Technical terms and abbreviations are used throughout this report. The meaning is usually clear from the context, but a glossary is provided for the non-specialist reader.

Adjuvant! Online A computer program designed to provide estimates of the benefits of adjuvant endocrine therapy and chemotherapy.

Adjuvant therapy Adjuvant therapy is treatment that is given in addition to the primary (initial) treatment. It is designed to help reach the primary treatment goal (e.g. disease eradication). Adjuvant therapy for cancer usually refers to chemotherapy, hormonal therapy or radiotherapy when administered after primary surgery to help decrease the risk of the cancer recurring (coming back).

Amplification In genetics, an increase in the frequency of replication of a deoxyribonucleic acid segment.

Analytical validity The ability of the test to accurately and reliably measure the expression of messenger ribonucleic acid or proteins by breast cancer tumour cells.

Axillary lymph nodes Located in the armpit area, they receive lymph fluid from the arm, breast and ipsilateral (same side) upper torso.

Chemotherapy The use of medication(s) (drugs) that are toxic to cancer cells, given with the aim of killing the cells or preventing or slowing their growth.

Clinical utility The utility of the test in relation to harm, impact on clinical decision-making, evidence of improvement in outcomes and health-care costs.

Clinical validity The degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes.

Cohort study A study that follows groups of people with and without the condition of interest over time to study outcomes.

Endocrine therapy Treatment of cancer by removing and/or blocking the effects of hormones that stimulate the growth of cancer cells.

External Assessment Group An independent group of researchers commissioned by the National Institute for Health and Care Excellence to review the evidence on a group of technologies. The External Assessment Group includes researchers who assess the quality of studies on the treatments, and health economists who look at whether or not the treatments are good value for money. The Diagnostics Assessment Committee bases its discussions on the diagnostics assessment report produced by the External Assessment Group.

Gene expression Gene expression refers to the translation of the information encoded in a gene into an ribonucleic acid (RNA) transcript. Expressed transcripts include messenger RNAs, which are translated into proteins, as well as other types of RNA, such as transfer RNA, ribosomal RNA, micro RNA and non-coding RNA, that are not translated into protein. Gene expression is a highly specific process by which cells switch genes on and off in a timely manner,

according to their state. The study of mRNA expression in a cell is an indirect way to study the protein counterpart.

Gene expression profiling This term refers to any genomic techniques that measure the fraction of the genes that are expressed in a specific sample. It refers to techniques that allow the assessment of the expression of more than one gene at a time, such as microarray analysis after the use of real-time reverse transcription-polymerase chain reaction to amplify levels of genetic material to measureable levels.

Grading Assessing the degree of aggressiveness of a malignant tumour based usually on the appearance of its cells under the microscope.

Histology An examination of the cellular characteristics of a tissue using a microscope.

Hazard ratio The hazard ratio (HR) is an estimate of the ratio of the hazard rate in two groups. It is broadly equivalent to relative risk and is useful when the risk is not constant over a given period as it uses information collected at different times. The term is typically used in the context of survival over time. If the HR = 0.5 then the relative risk of dying (or some other health event) in one group is half the risk of dying in the other group.

Hormone receptor Protein molecules with a specific conformation that bind to hormones in the cell's environment and trigger hormone-dependent changes in the cell's behaviour.

Human epidermal growth factor receptor A molecule on the surface of a cell that interacts with a specific growth factor and helps to control how rapidly the cells grow.

IHC4 The IHC4 test uses immunohistochemistry technology to assess the levels of four key proteins in a breast cancer sample – oestrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor type 2 (HER2) and Ki-67. The final algorithm for IHC4 calculates a risk score for distant recurrence based on these four proteins in addition to classical clinical and pathological variables (composite risk score IHC4 + clinical score termed IHC4 in our report).

Immunohistochemistry A technique that uses antibodies to identify specific molecules in tissues, which are examined and scored by a pathologist using a microscope.

Ki-67 Antigen KI-67 is a molecule that can be easily detected in growing cells in order to gain an understanding of the rate at which the cells within a tumour are growing.

Lymph nodes Small bean-shaped glands that are part of the lymphatic system. White blood cells in the lymph nodes attack bacteria and viruses as they pass through the node.

Malignant Cancerous cells that can invade into nearby tissue and spread to other parts of the body.

Mammography The process of taking a mammogram – a soft-tissue radiograph of the breast – which may be used to evaluate a lump or which may be used as a screening test in women with no signs or symptoms of breast cancer.

Mastectomy Surgical removal of the breast.

Metastases Deposits of cancer in the body at a site distant from the primary site.

Nottingham Prognostic Index The Nottingham Prognostic Index (NPI) is a composite prognostic parameter involving both time-dependent factors and aspects of biological aggressiveness. The NPI score is based on a mix of grade, lymph node involvement and tumour size. To calculate the score, add numerical grade (1, 2 or 3), lymph node score (negative = 1, one to three nodes = 2, more than three nodes = 3) and 0.2 × tumour size in cm. Patients can be divided into three prognostic groups on the basis of the NPI score: a good prognostic group (NPI < 3.4), a moderate prognostic group (3.4 < NPI < 5.4) and a poor prognostic group (NPI > 5.4).

Polymerase chain reaction The polymerase chain reaction (PCR) is a molecular biology technique for isolating and exponentially amplifying a deoxyribonucleic acid sequence of interest in vitro by enzymatic replication. This technique has been extensively modified to perform a wide array of tasks. It is a common tool in medical and biological research. PCR is now used to obtain the sequence of genes, diagnose hereditary diseases, identify genetic fingerprints (forensic medicine), detect infectious diseases and create transgenic organisms. Coupled to reverse transcription it is used to amplify ribonucleic acid molecules.

Predictive molecular markers A molecule that is assessed to predict the likely response to a specific treatment, for example oestrogen receptor to predict the likely response to endocrine therapy.

Prognosis A prediction of the likely outcome or course of a disease; the chance of recovery, recurrence or death.

Prognostic factors Disease characteristics that are correlated with the course of the disease and which are used to predict the likely outcomes.

Reverse transcription-polymerase chain reaction The reverse transcription-polymerase chain reaction (RT-PCR) is a variant of PCR, a laboratory technique commonly used in molecular biology to generate many copies of a deoxyribonucleic acid (DNA) sequence using a process termed 'amplification'. In RT-PCR the ribonucleic acid strand of interest is first reverse transcribed into its DNA complement (complementary DNA, or cDNA) using the enzyme reverse transcriptase, and the resulting cDNA is amplified using traditional or real-time PCR.

Staging Clinical description of the size and spread of a patient's tumour, allocated by internationally agreed categories.

Systemic therapy/treatment Medicine, usually given by mouth or injection, to treat the whole body rather than targeting one specific area.

Transcription In genetics, the process by which genetic information on a strand of deoxyribonucleic acid is used to synthesise a strand of complementary ribonucleic acid.

Translation In genetics, the process by which a messenger ribonucleic acid molecule specifies the linear sequence of amino acids on a ribosome for protein synthesis.

List of abbreviations

AC	doxorubicin and cyclophosphamide
AIC	academic-in-confidence
AML	acute myeloid leukaemia
AST	adjuvant systemic treatment
ATAC	Arimidex, Tamoxifen, Alone or in Combination trial
BCA	Breast Cancer Array
BCI	Breast Cancer Index
BCSD	breast cancer-specific death
BCSS	breast cancer-specific survival
BNF	*British National Formulary*
BSA	body surface area
CAF	cyclophosphamide, doxorubicin and fluorouracil
CBO	Dutch Institute for Healthcare Improvement
cDNA	complementary deoxyribonucleic acid
CEAC	cost-effectiveness acceptability curve
CHF	congestive heart failure
CG	Clinical Guideline
CI	confidence interval
CIC	commercial-in-confidence
CISH	chromogenic in situ hybridisation
CMF	cyclophosphamide, methotrexate and 5-fluorouracil
DARE	Database of Abstracts of Reviews of Effects
DCIS	ductal carcinoma in situ
DCS	Decisional Conflict Scale
DDFS	distant disease-free survival
DFS	disease-free survival
DMFS	distant metastasis-free survival
DNA	deoxyribonucleic acid
DRFI	distant recurrence-free interval
DRFS	distant recurrence-free survival
EAG	External Assessment Group
EBCTCG	Early Breast Cancer Trialists' Collaborative Group
ECRIC	Eastern Cancer Registration and Information Centre
EGFR	epidermal growth factor receptor
EORTC	European Organisation for Research and Treatment in Cancer
EQ-5D	European Quality of Life-5 Dimensions
ER	oestrogen receptor (ER+ is ER positive and ER– is ER negative)
FEC	5-fluorouracil, epirubicin and cyclophosphamide
FEC-D	5-flurouracil, epirubicin and cyclophosphamide-docetaxel
FEC-P	5-fluorouracil, epirubicin, cyclophosphamide and paclitaxel
FFPE	formalin-fixed paraffin-embedded
G-CSF	granulocyte colony-stimulating factor
GEP	gene expression profiling
HER2	human epidermal growth factor receptor type 2
HR	hazard ratio
HTA	Health Technology Assessment
ICER	incremental cost-effectiveness ratio
IHC	immunohistochemistry

LCIS	lobular carcinoma in situ
LN	lymph node
LR-$_{X}^{2}$	likelihood ratio chi-square
MF	methotrexate and 5-fluorouracil
MGI	Molecular Grade Index
MINDACT	Microarray In Node-negative Disease may Avoid ChemoTherapy
mRNA	messenger ribonucleic acid
NCCN	National Comprehensive Cancer Network
NICE	National Institute for Health and Care Excellence
NIH	National Institutes of Health
NPI	Nottingham Prognostic Index
NPI+	Nottingham Prognostic Index Plus
NPV	negative predictive value
NSABP	National Surgical Adjuvant Breast and Bowel Project
OHTA	Ontario Health Technology Assessment
OPTIMA	Optimal Personalised Treatment of breast cancer using Multi-parameter Analysis
OS	overall survival
PR	progesterone receptor
PPV	positive predictive value
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analysis
PSA	probabilistic sensitivity analysis
PSSRU	Personal Social Services Research Unit
QALY	quality-adjusted life-year
RCT	randomised controlled trial
RFI	recurrence-free interval
RFS	recurrence-free survival
RNA	ribonucleic acid
RS	recurrence score
RSPC	integration of RS and clinicopathological factors
RT-PCR	reverse transcription-polymerase chain reaction
SE	standard error
SEER	Surveillance, Epidemiology and End Results
STAI	State-Trait Anxiety Inventory
TAILORx	Trial Assigning Individualized Options for Treatment
TC	docetaxel and cyclophosphamide
TNM	tumour, nodes, metastases classification system for cancer stage of the UICC
TTDR	time to distant recurrence
WMCIU	West Midland Cancer Intelligence Unit

Note

This monograph is based on the Diagnostic Assessment Report produced for NICE. The full report contained a considerable number of data that were deemed commercial-in-confidence and/or academic-in-confidence. The full report was used by the Appraisal Committee at NICE in their deliberations. The full report with each piece of commercial-in-confidence and/or academic-in-confidence data removed and replaced by the statement 'commercial-in-confidence and/or academic-in-confidence information (or data) removed' is available on the NICE website: www.nice.org.uk.

The present monograph presents as full a version of the report as is possible while retaining readability, but some sections, sentences, tables and figures have been removed. Readers should bear in mind that the discussion, conclusions and implications for practice and research are based on all the data considered in the original full NICE report.

Scientific summary

Background

Prognostic tools such as the Nottingham Prognostic Index (NPI) and Adjuvant! Online are currently used in the UK to assist decision-making relating to adjuvant chemotherapy for women with early breast cancer at intermediate or high risk of recurrence following primary surgery. These tools use pathological parameters, for example tumour size, grade and lymph node status in the case of NPI, with the addition of oestrogen receptor (ER) status, age and comorbidity for Adjuvant! Online. Such tools are imperfect and some women with early breast cancer may be over- or undertreated, resulting in unnecessary use of chemotherapy for some women or avoidable deaths in women for whom chemotherapy was withheld.

Gene expression profiling (GEP) and expanded immunohistochemistry (IHC) (or protein expression) tests aim to improve the targeting of chemotherapy by more accurately identifying patients who will gain most benefit from it. These tests either aim to more accurately measure the risk of cancer recurrence by incorporating a wider range of biomarkers than standard clinicopathological algorithms or seek to identify breast cancer subtypes, which provide information on recurrence risk.

Nine tests were included in this assessment, as per the National Institute for Health and Care Excellence (NICE) scope. Six use GEP technology: the Randox Breast Cancer Array (Randox Laboratories, Crumlin, UK), MammaPrint® (Agendia, Amsterdam, the Netherlands), BluePrint™ (Agendia, Amsterdam, the Netherlands), the PAM50 gene expression assay (ARUP Laboratories, Salt Lake City, UT, USA), OncotypeDX™ (Genomic Health Inc., Redwood City, CA, USA) and the Breast Cancer Index℠ (bioTheranostics Inc., San Diego, CA, USA); and three use IHC technology: IHC4 [The National Institute for Health Research (NIHR) Specialist Biomedical Research Centre (BRC) for Cancer is a partnership between The Royal Marsden NHS Foundation Trust and The Institute of Cancer Research (ICR); see http://www.cancerbrc.org/Highlights/Breast_Cancer_highlights/index.shtml], Mammostrat® (Clarient Inc., Aliso Viejo, CA) and NPI plus (NPI+) (University of Nottingham, Nottingham, UK).

Objective

The objective of this study was to evaluate the clinical effectiveness and cost-effectiveness of GEP and expanded IHC tests compared with existing prognostic tools in guiding the use of adjuvant chemotherapy in women with early breast cancer in England and Wales.

Methods

A systematic review of the evidence on the clinical effectiveness of nine GEP and expanded IHC tests to guide the use of chemotherapy in breast cancer management was conducted. For two of the tests (OncotypeDX and MammaPrint) the review updated two existing systematic reviews. Several electronic databases (including MEDLINE, MEDLINE In-Process & Other Non-Indexed Citations, EMBASE and The Cochrane Library) were searched from January 2002 to May 2011 (for the OncotypeDX and MammaPrint tests searches were conducted from January 2009).

Outcome measures included analytical validity, clinical validity and clinical utility. The study by Altman (2001) was used to assess the methodological quality of included studies (Altman D. Systematic reviews in health care: systematic reviews of evaluations of prognostic variables. *BMJ* 2001;**323**:224–8).

A systematic review of economic evaluations was also undertaken. In addition, two economic evaluations were submitted by Genomic Health and Clarient for the use of OncotypeDX and Mammostrat in the UK respectively.

A probabilistic model was developed by the External Assessment Group (EAG) using a lifetime horizon. Following a review of the evidence available, only four of the nine tests were included in the economic evaluation. Analysis was undertaken for women with ER-positive (ER+), lymph node-negative (LN–) and human epidermal growth factor receptor type 2-negative (HER2–) early breast cancer from a NHS perspective. These tests were assessed as an addition to existing prognostic tools. A subgroup analysis was conducted in women with a NPI score ≤ 3.4 and women with a NPI score > 3.4. The model used UK-specific data where possible.

In the comparator arm of the economic model, the proportion of patients receiving chemotherapy under current practice was informed by cancer registry data, reflecting the use of current prognostic tools such as NPI and Adjuvant! Online to guide the use of chemotherapy. In the intervention arm the targeting of patients to receive chemotherapy was dependent on the classification of risk by the new test. The natural history of breast cancer was then simulated using a cohort state transition model, taking into account the reduction in the risk of recurrence associated with chemotherapy. Evidence for the benefit of chemotherapy (reduction in the risk of recurrence) by risk group for the new tests was taken directly from the studies identified through the systematic review of the literature, despite the identified limitations of the studies. Patients were able to move between five possible health states – recurrence free, distant recurrence, local recurrence, long-term adverse events and death (from breast cancer, adverse events or other causes). Results were reported in terms of cost per quality-adjusted life-year (QALY).

Results

Nature, description and quality of the available evidence

The literature searches identified 5993 citations, of which 32 full-text papers or abstracts (representing 30 studies) were included in the review. Supplementary information submitted by the manufacturers was also presented. This evidence was summarised but was not subjected to the systematic review process. Additional studies that did not meet the inclusion criteria for the systematic review were used to populate the economic model.

The study populations were generally heterogeneous in the nature of their inclusion criteria although the majority of evidence examined ER+, LN– populations. Most studies included a small number of participants, although a few studies included over 1000 patients. Follow-up was short or not reported for a large number of studies. Only six studies were specific to a UK population (three for OncotypeDX, one for NPI+, one for IHC4 and one for Mammostrat).

Summary of the benefits and risks of gene expression profiling and expanded immunohistochemistry tests

OncotypeDX

Clinical

Previous systematic reviews OncotypeDX was reported to be furthest along the validation pathway. In terms of clinical validity these reviews reported evidence that the OncotypeDX recurrence score was significantly correlated with disease-free-survival and overall survival. One study was reported that reported a significant benefit from the use of chemotherapy in the OncotypeDX high-risk group, although it was highlighted within the review that the study may have been subject to bias.

Current review The current review identified 12 additional studies on the OncotypeDX test. Further larger studies have now reported, which support the prognostic capability of the OncotypeDX test. In particular, one large-scale UK-based study, in postmenopausal women with ER+, LN– early breast cancer, reported that an increase in risk score was significantly associated with an increased risk of distant recurrence. Furthermore, the evidence base has been extended to include the LN+ population, and there are the beginnings of an evidence base for the validity of OncotypeDX in different populations such as Japanese patients. Four studies presented evidence on the impact of OncotypeDX on clinical decision-making, indicating that the use of OncotypeDX leads to changes in decision-making for between 31.5% and 38% of patients. However, only one of these studies was UK based and limitations in relation to study design were identified.

Economic

Two economic studies were identified. Both studies compared the use of OncotypeDX with Adjuvant! Online. These studies were non-UK studies and were not considered generalisable to the UK setting. The economic evaluation submitted by Genomic Health estimated the incremental cost for treatment guided using OncotypeDX to be £6232 per QALY gained compared with current clinical practice in the UK, although a number of limitations with regard to the analysis were highlighted.

A de novo economic model was built by the EAG and estimated the cost per QALY gained to be £29,502 compared with current clinical practice, assuming that the test was offered to all woman with ER+, LN–, HER2– early breast cancer, under our base-case assumptions. This analysis assumed OncotypeDX to be predictive of the benefit of chemotherapy, based on evidence from the Paik *et al.* study, although weaknesses relating to this study are highlighted. (Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, *et al.* Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 2006;**24**:3726–34.) A subgroup analysis was performed and showed that the incremental cost-effectiveness ratio (ICER) for OncotypeDX compared with current clinical practice was reduced to £9774 per QALY gained if OncotypeDX was to be offered to women with a (NPI > 3.4) only. Compared with current clinical practice, OncotypeDX had a 12.4% (all women) and 91.6% (NPI > 3.4) probability of being considered cost-effective when using a threshold of £20,000 per QALY gained respectively, although the quality of the data in the model was considered relatively weak. Key areas of uncertainly relate to assumptions about the benefits of chemotherapy in terms of relative risk reduction by risk group, the risk of recurrence over time and the impact of the new test on decision-making. The ICER increased substantially and was greater than £20,000 per QALY gained for both analyses when assuming the same relative reduction in the risk of recurrence from chemotherapy for all patients irrespective of the OncotypeDX recurrence score classification, that is, assuming that the test is prognostic only.

MammaPrint

Clinical

Previous systematic reviews There is a range of studies evaluating the prognostic ability of MammaPrint in heterogeneous populations; however, the previous reviews indicated that evidence relating to the clinical validity of MammaPrint was not always conclusive or supportive of the prognostic value of the test. In terms of clinical utility, the previous reviews identified one non-UK study which suggested that MammaPrint had an impact on clinical decision-making.

Current review Our review identified seven additional studies on the MammaPrint test. Four studies reported that the MammaPrint score is a strong independent prognostic factor and may provide additional value to standard clinicopathological measures, although the populations in all of these studies were relatively small. Six non-UK studies evaluated the clinical utility of MammaPrint. Five of the studies reported on test reclassification against currently used guidelines and one reported that treatment advice for 40% of patients would change, assuming that all patients classified as high risk and no patients classified as low risk would receive chemotherapy. However, none of the studies provided evidence of actual changes in treatment decisions following introduction of the test. A study on the benefit of chemotherapy by MammaPrint risk group was identified but omitted from the systematic review because it was based on a pooled analysis of six primary studies.

Economic

An analysis was carried out by the EAG to evaluate the use of MammaPrint in England and Wales but because of the limitations in the evidence available this was considered exploratory only and no base-case ICER was presented.

PAM50

Clinical

The evidence base for PAM50 is still relatively immature. The current review identified two analytical validity studies (reported in abstract form only) comparing the PAM50 test with standard IHC measurements. Four studies evaluated the clinical validity of PAM50; two of these are as yet unpublished. No evidence on clinical utility was identified.

Economic

The EAG did not model treatment guided using PAM50 because of gaps in the evidence base.

Mammostrat

Clinical

The current review identified three studies that provided data to support the use of the Mammostrat test as an independent prognostic tool for women with ER+, tamoxifen-treated breast cancer. Although the evidence base for the Mammostrat test is relatively immature, these studies included a large sample size, appeared to be of reasonable quality and provided data from a UK setting (one study). One study was identified for clinical utility but limitations were identified relating to this study.

Economic

The EAG conducted an exploratory analysis using the same model structure as for the OncotypeDX evaluation and unpublished data from a small sample from a non-UK population; however, because of the limitations in the evidence base, any conclusions drawn from this analysis are subject to significant uncertainty.

IHC4

Clinical

No studies on analytical validity of the test were identified. The current review identified one study on the clinical validity of IHC4, which reports that the IHC4 score is a highly significant predictor of distant recurrence. This study was based on a large sample size and detailed the development of the test in one cohort and the external validation of the test in an independent cohort. The study also reported evidence comparing IHC4 with OncotypeDX. The review did not identify any published evidence on the clinical utility of IHC4 in terms of its impact on treatment decisions or its ability to predict chemotherapy benefit by risk group.

Economic

The EAG evaluated the cost-effectiveness of IHC4 in parallel with that of OncotypeDX as there was direct evidence between the two tests in a UK population from the same data source used to evaluate the cost-effectiveness of OncotypeDX. The IHC4 test was predicted to be dominant compared with current clinical practice in patients with ER+, LN–, HER2– early breast cancer, providing more QALYs at a lower cost. An incremental analysis was conducted comparing OncotypeDX, IHC4 and current clinical practice. When the treatment decision using OncotypeDX was compared with that using IHC4, the ICER for OncotypeDX increased to £64,111 per QALY gained if the tests were to be offered to all women and £31,125 if the tests were to be offered to women with a NPI >3.4 only. IHC4 was predicted to remain dominant assuming the test to be prognostic only, that is, all women receiving chemotherapy derive the same relative benefit in terms of reduction in distant recurrences. However, because the evidence base for IHC4 is less developed than that for OncotypeDX, additional assumptions were required and the results are subject to greater uncertainty.

Nottingham Prognostic Index plus, Breast Cancer Index, BluePrint and Randox Breast Cancer Array

Clinical

Based on the limited available data identified for these tests, no firm conclusions can be drawn about their analytical validity, clinical validity (prognostic ability) and clinical utility. Further evidence on the prognostic and predictive ability of all of these tests is required.

Economic

No studies were identified in the systematic review of the economic literature. The EAG did not model treatment guided using these tests because of significant gaps in the evidence base.

Discussion

Strengths and limitations of the analyses and uncertainties

Clinical

Two of the tests (OncotypeDX and MammaPrint) have a reasonably large evidence base, although there are some methodological weaknesses relating to this evidence in terms of heterogeneity of patient cohorts and issues arising from the retrospective nature of the evidence, such as the relevance of the evidence to current methods of diagnosis, treatment and standards of care. The evidence base for OncotypeDX is considered to be the most robust. The MammaPrint evidence is typically based on observational data (small cohort studies) rather than randomised data, increasing the risk of selection bias. Both IHC4 and Mammostrat present early evidence of the prognostic ability of the tests based on large UK-based validation cohorts. Further evidence is required on the clinical utility of all of these tests, and on UK-based populations where this is

not currently available. The evidence base for the remaining five tests has significant gaps and is considered less robust.

Economic

Four of the nine tests were included in the economic evaluation by the EAG. The model used UK-specific evidence where possible, including the baseline use of chemotherapy, the risk of distant recurrence/recurrence and reclassification with the new test, so that its conclusions would be relevant to the UK setting. Our analysis focused on patients with ER+, LN–, HER2– early breast cancer as use of the tests in this population is supported by the most robust clinical evidence. Women with a NPI ≤ 3.4 and women with a NPI > 3.4 were modelled separately to account for the prognostic value of the current treatment decision based on clinicopathological parameters and to allow a scenario assuming that the test was offered to a subgroup of the population at intermediate risk to be conducted.

However, there are significant limitations with regard to the economic analyses. Results of all of the analyses have to be interpreted with caution and the results cannot be compared directly between tests. Given that no studies following patients from initial diagnosis through to final health outcomes were identified for any of the tests, the economic model needed to combine clinical data from several different sources in order to model how the results from the new tests translate into final outcomes in the form of QALYs. This resulted in significant uncertainties that were not adequately captured with the probabilistic sensitivity analysis – data used in the model were not always based on UK populations and were not always specifically based on the ER+, LN–, HER2– population of interest. Differences in the age of the study populations and the endocrine and chemotherapy regimens used in the studies compared with those in the model introduced further uncertainty. One key area of uncertainty is whether the tests are prognostic or predictive of the benefit of chemotherapy (i.e. do they allow identification of high-risk patients who would derive a greater relative benefit from chemotherapy). The ICER was very sensitive to this assumption. There were particular concerns relating to the studies used to estimate the benefit associated with chemotherapy for patients categorised by risk group by the new tests, in relation to both the study design and the populations included in these studies. The evidence base on the impact of the new tests on the selection of patients to receive chemotherapy was also lacking or not considered generalisable to the UK population. Univariate sensitivity analyses indicated that the ICER was sensitive to these assumptions.

A greater number of assumptions were required to model IHC4 compared with OncotypeDX because of data limitations for IHC4. There were more significant gaps in the evidence for MammaPrint and Mammostrat, and any conclusions that can be drawn from these exploratory analyses are subject to considerable uncertainty.

Conclusions

The OncotypeDX and MammaPrint tests have a reasonably large evidence base, although there are some methodological weaknesses relating to this evidence in terms of heterogeneity of patient cohorts and the use of retrospective data. The evidence base for OncotypeDX is considered to be the most robust. Two of the tests (IHC4 and Mammostrat) have presented early evidence of the prognostic ability of the tests, based on large UK-based validation cohorts, but further research is required. The clinical utility evidence for GEP and expanded IHC tests is limited by the lack of large prospective studies in UK populations. PAM50, BluePrint, Breast Cancer Index, NPI+ and Randox Breast Cancer Array have only limited clinical evidence to date.

The economic analysis suggests that the use of the new tests may result in small increases in QALYs compared with currently used prognostic tools, but current limitations in the evidence base introduce significant uncertainty in the results. A key area of uncertainty is whether tests are prognostic only or identify high-risk patients who will benefit more relatively from chemotherapy (from reductions in the risk of recurrence) than low-risk patients. The economic analyses suggested that, of the four tests considered, treatment guided using IHC4 has the greatest potential to be cost-effective at a £20,000 threshold, given the low cost of the test; however, further evidence on IHC4 is needed and the exact cost of using the test in the NHS needs to be investigated further. Although the OncotypeDX test has been shown to have the potential to be cost-effective at the £20,000 threshold for patients with a NPI > 3.4, further evidence is needed on the impact on decision-making in the UK and to clarify the predictive ability of the test specifically in an ER+, LN–, HER– population receiving current endocrine and chemotherapy regimens.

Implications for service provision

The impact of sending large numbers of samples to central testing facilities for pathology services, in terms of tissue tracking, pathologist and technical staff time, data input on receipt, etc., would need to be explored. Tests requiring the use of fresh tissue require a major change in practice with regard to the handling of tissue, with significant implications for service configuration and costs. The addition of expanded IHC-based tests is likely to fit more easily with current practice in the NHS. Quality assurance issues would need to be addressed, for example for the Ki-67 component of the IHC4 test, before these tests could be considered for use in clinical practice in the NHS.

The main research priorities relate to the reliability and reproducibility of the IHC4 test, along with further evidence of the prognostic ability of IHC4 compared with NPI and Adjuvant! Online. Further evidence on the predictive ability of all of the tests is also required. In addition, evidence to improve the understanding of the impact of these tests (for tests that provide a risk score/category and tests that provide subtype information only) on the management of patients in a UK population is urgently needed.

Study registration

This study is registered as PROSPERO 2011:CRD42011001361, available from www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42011001361.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Chapter 1

Background and definition of the decision problem

Condition and aetiology

Breast cancer is the most commonly diagnosed cancer in women in England and Wales. In 2009 there were 42,305 new cases diagnosed. Treatment usually involves surgery to remove the primary tumour and any involved lymph nodes; this may be followed by radiation therapy, endocrine therapy and/or chemotherapy with or without trastuzumab depending on tumour and patient variables.

Aetiology, pathology and prognosis

Aetiology

The causes of breast cancer are not completely understood. A range of risk factors have been identified including genetic, hormonal and lifestyle factors.[1]

It has been estimated that 12% of women with breast cancer have one affected family member and 1% have two or more affected family members.[2] Genetic predisposition is mediated by high-penetrance genes such as breast cancer 1 gene (*BRCA1*) and breast cancer 2 gene (*BRCA2*), responsible for around 80–90% of hereditary cancers, and low-penetrance genes, which confer increased and decreased risk.[1]

Environmental and lifestyle factors as well as genetic factors influence breast cancer risk. Asian migrants to the West have increased levels of risk compared with the indigenous population, whereas Asian Americans born in the West have incidence rates approximating the US average.[3]

Lifestyle and environmental factors thought to increase risk include hormonal factors such as taking the oral contraceptive pill or hormone replacement therapy, higher age of menopause, early age of menarche, late age of first birth and not giving birth. Factors that decrease risk include higher folate intake, higher number of pregnancies, breastfeeding and younger age at first birth.[1]

Obesity increases the risk of breast cancer in postmenopausal women.[4] The picture is less clear for premenopausal women, in whom risk may be lower but prognosis poorer. Physical activity in adolescence and young adulthood confers a decreased risk of breast cancer,[5] which may be mediated hormonally.

Pathology

Breast cancer starts with genetic changes in a single cell or a small group of cells in the epithelia of the ducts or the lobules of the breast. The genetic change allows cells to reproduce uncontrollably, creating a tumour. Tumours that have not yet spread to surrounding tissue are known as carcinoma in situ and may be ductal (DCIS) or lobular (LCIS). Once spread to surrounding tissue begins, a tumour is known as invasive. More rapid growth and spread occurs once a blood supply is secured. Cancer spreads via the lymphatic system or the bloodstream.

Lymphatic spread is usually first to the axillary lymph nodes. Spread via the bloodstream can lead to distant metastases in the bone or viscera that are incurable.

The presence or absence of axillary metastases is a key indicator of stage of disease and prognosis, and adjuvant therapy is planned, in part, based on their presence and extent.[6] They are caused when a single or small number of cells detach from the main tumour, travel via the lymphatic system and establish themselves in the tissue of the lymph nodes. Axillary metastases occur in approximately 41% of cases[7] and prognosis is better when there is no axillary spread. When metastases are present, axillary clearance is indicated to prevent further spread and ensure local disease control.

Prognosis

Overall, 5-year, age-standardised breast cancer survival rates are around 80%.[8] Survival varies with age (*Table 1*) and stage of disease (*Table 2*).[9]

Other factors can affect prognosis. Clinicians may use tools such as the Nottingham Prognostic Index (NPI),[10] which takes into account grade as well as size and spread, or Adjuvant! Online,[11] which uses patient data such as age, tumour size, nodal involvement, hormonal receptor status and histological grade to predict disease course and treatment options. Good prognosis is associated with small tumour size, lymph node-negative (LN–) status, younger age, oestrogen receptor-positive (ER+) status and progesterone receptor-positive (PR+) status. Human epidermal growth factor receptor type 2 (HER2) overexpression is associated with poor prognosis.

Epidemiology and incidence

Incidence varies most with gender. Women are far more likely to get breast cancer than men. For both genders, incidence varies with age (*Table 3*). Just over 80% of cases occur in women aged ≥ 50 years. In England and Wales, 2006 data demonstrate highest rates for women in the 60- to 70-year age range.[12]

Incidence also varies with ethnicity. Asian, Chinese and black ethnic groups and those with mixed heritage have a lower incidence than the white ethnic group in England. Compared with the white group the rate ratios are 0.65, 0.75, 0.49 and 0.58 respectively.[13]

In both England[14] and Wales[15] those who are classed as most deprived have a lower incidence of breast cancer. However, there is some evidence to suggest that the trend for mortality is reversed, with better survival for those from more affluent areas. It is unclear why this is but

TABLE 1 Five-year survival rates for women in England diagnosed during 2001–6

	Age (years)					
	15–39	40–49	50–59	60–69	70–79	80–99
5-year survival rate (%)	81	86	89	87	78	64

TABLE 2 Five-year survival rates for women diagnosed in the West Midlands from 1985 to 1989 followed up to the end of 1999, as at January 2002

	Stage of disease			
	I	II	III	IV
5-year survival rate (%)	88	69	43	12

TABLE 3 Incidence per 100,000 for England and Wales by age group and gender, 2006

	Age (years)													
	0–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85+
Women														
Wales	0	2	21	64	123	186	256	286	324	328	254	201	199	213
England	0	8	20	53	141	185	270	274	321	327	252	190	183	202
Men														
Wales	0	0	0	0	0	1	1	1	1	2	0	2	1	0
England	0	0	0	0	0	1	2	2	2	0	5	1	3	2

it may be due to lower levels of screening compliance, worse overall general health status and lower levels of treatment because of limited access to health care[16] and poorer compliance with treatment regimens.

Significance in terms of ill-health (burden of disease)

Breast cancer is the second largest cause of cancer death in women after lung cancer, with an age-standardised mortality rate of 26 per 100,000 women. In 2008 this constituted 10,716 deaths for women in England and Wales.[17]

Measurement of disease

Breast cancer has few obvious symptoms and can easily go undetected for a few years. Among the more noticeable symptoms are a palpable lump in the breast, a change in breast shape and skin appearance or changes to the nipple such as inversion, a rash or discharge.

A suspicious breast mass may be identified through screening or through presentation to a GP. Women between the ages of 50 and 70 years are routinely invited to attend regular screening (age range in the UK is changing to 47–73 years between 2010 and 2013). Screening is thought to have reduced breast cancer deaths in the 50–69 years age category by an estimated 6.4% in addition to the effects of tamoxifen, chemotherapy and earlier presentation outside of screening.[18] Screening increases the proportion of tumours detected in the early, more curable stages.

The breast mass and axillary areas are investigated clinically by palpation and mammography or ultrasound for younger women, and the status of the tumour confirmed by histology of biopsied tissue. Staging of the disease depends on tumour size, the number of involved lymph nodes and the presence or absence of distant metastases. Tumour size and axillary metastases can be estimated by clinical examination and imaging techniques, but definitive status is achieved through surgery. Those with small tumours and no axillary metastases have the best prognosis, whereas those with distant metastases are considered incurable.

Current methods for staging of breast cancer

Three main factors are used to stage breast cancer – tumour size, metastases to the regional lymph nodes and distant metastases. The tumour/node/metastases (TNM) staging system was developed and is maintained by the Union Internationale Contre le Cancer (UICC)[19] and the American Joint Committee on Cancer (AJCC).[20] T stage is classified according to the size of the tumour and degree of local infiltration; N stage is classified according to the number and location of metastases to the lymph nodes in the axilla, between the ribs (internal mammary nodes) and above or below the collarbone (supraclavicular and infraclavicular nodes); and M stage is classified by the presence of metastases beyond the breast and regional lymph nodes (*Table 4*).

The overall TNM stage of the cancer is defined as in *Table 5*. Early breast cancer is generally defined as cancer that has not spread beyond the breast or the ipsilateral axillary lymph nodes and which is confined to stages I, II or IIIA.[21–23]

TABLE 4 Descriptions of T, N and M stages

	Description
T: tumour stage	
Tx	Primary tumour cannot be assessed
T0	No evidence of primary tumour
Tis	Carcinoma in situ
T1	Tumour ≤ 2 cm across
T2	Tumour 2–5 cm across
T3	Tumour > 5 cm across
T4	Tumour of any size with direct extension to skin or chest wall, or inflammatory breast cancer
N: lymph node stage	
Nx	Nodal stage cannot be assessed
N0	No metastases to any ipsilateral lymph nodes
N1	Metastases to one to three axillary nodes or axillary nodes that are mobile
N2	Metastases to four to nine axillary nodes, or axillary nodes that are fixed to one another or other structures, or clinically apparent metastases to internal mammary nodes
N3	Metastasis to nodes above or below the collarbone (supraclavicular/infraclavicular), or to both axillary and internal mammary nodes, or to 10+ axillary nodes
M: metastasis stage	
Mx	Presence of metastases cannot be assessed
M0	No distant metastases
M1	Distant metastases

Sources: Cancer Research UK[21] and American Cancer Society.[24]

TABLE 5 Summary of TNM stages

Stage	T	N	M
0 (DCIS/LCIS)	Tis	N0	M0
I	T1	N0	M0
IIA	T0–1	N1	M0
	T2	N0	M0
IIB	T2	N1	M0
	T3	N0	M0
IIIA	T0–2	N2	M0
	T3	N1–2	M0
IIIB	T4	N0–2	M0
IIIC	T(any)	N3	M0
IV	T(any)	N (any)	M1

Sources: Cancer Research UK[21] and American Cancer Society.[24]

Current service provision

Management of early breast cancer

Patients diagnosed with early breast cancer currently follow the diagnosis/treatment pathway described in *Figure 1*.

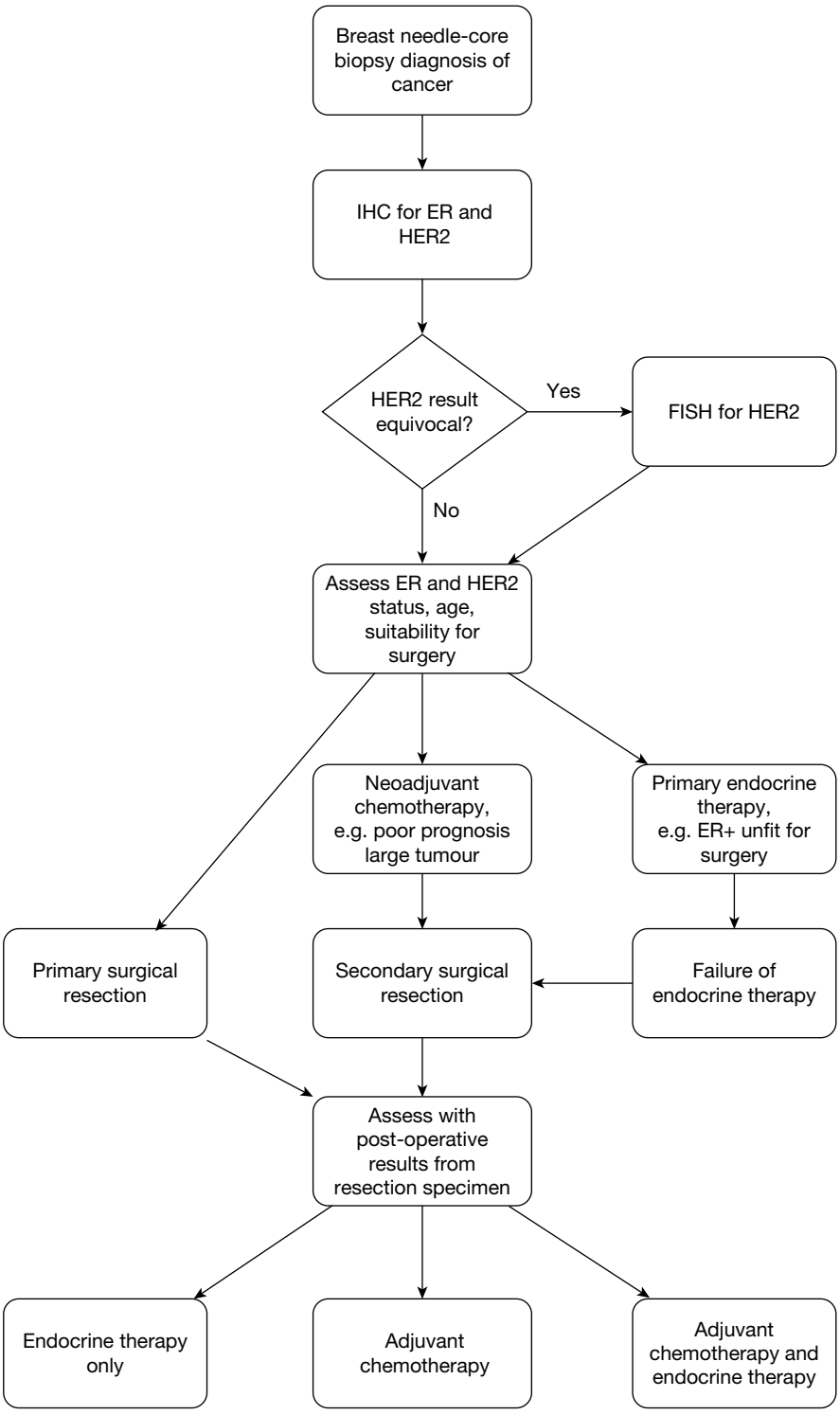


FIGURE 1 Diagnosis and management pathway in breast cancer. FISH, fluorescence in situ hybridisation; IHC, immunohistochemistry.

Current guidelines

Current National Institute for Health and Care Excellence (NICE) clinical guidelines (CG80)[7] indicate that adjuvant therapy should be considered for all patients with early invasive breast cancer after surgery, based on assessment of the prognostic and predictive factors and the potential benefits and side effects of the treatment. These guidelines do not make specific reference to the use of gene expression profiling (GEP) or expanded immunohistochemistry (IHC) tests to aid decision-making. The guidelines do indicate that decisions should be made following discussion of these factors with the patient and recommend consideration of the use of Adjuvant! Online to support estimations of individual prognosis and the absolute benefit of adjuvant treatment for patients with early invasive breast cancer.[7] The NPI is also commonly used as the basis for many local guidelines on the management of chemotherapy for patients with early breast cancer.

Adjuvant! Online

The Adjuvant! Online computer program is designed to provide estimates of the benefits of adjuvant endocrine therapy and chemotherapy. The current version of Adjuvant! Online does not include HER2 status and the potential benefit of trastuzumab. Patient and tumour characteristics are entered into the programme and provide an estimate of the baseline risk of mortality or relapse for patients without adjuvant therapy. Information about the efficacy of different therapy options is derived from Early Breast Cancer Trialists' Collaborative Group (EBCTCG) meta-analyses in order to provide estimates of the reduction in risk at 10 years of breast cancer-related death or relapse for selected treatments. These estimates are then provided on printed sheets in simple graphical and text formats to be used in consultations.

Nottingham Prognostic Index

The NPI is a composite prognostic parameter involving both time-dependent factors and aspects of biological aggressiveness. The NPI score is based on a mix of grade, lymph node involvement and tumour size. The score is calculated as follows: add numerical grade (1, 2 or 3), lymph node score (negative = 1, one to three nodes = 2, more than three nodes = 3) and 0.2 × tumour size in cm. Patients can be divided into three prognostic groups on the basis of the NPI score: a good prognostic group (NPI < 3.4), a moderate prognostic group (3.4 < NPI < 5.4) and a poor prognostic group (NPI > 5.4).

Clinical opinion suggests that there is wide variation in clinical practice between trusts in the UK, with some centres using Adjuvant! Online and/or NPI, in addition to other clinical parameters.

Description of technologies under assessment

Gene expression profiling and expanded IHC tests aim to improve the use of chemotherapy in breast cancer by improving the stratification of patients and identification of those patients who will gain most benefit from chemotherapy. These tests typically report two types of information – breast cancer subtype and/or risk of recurrence. Tests developed to provide information on subtypes can be used either before surgery for informing decisions on neoadjuvant therapy or after primary surgery for informing decisions on adjuvant chemotherapy. 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. These tests are typically indicated for women with ER+ and LN– tumours (and sometime LN+ tumours if number of nodes is small).

In conjunction with other information available about tumour size, grade, etc., test results are likely to be used to guide the decision on which patients should be offered adjuvant

chemotherapy. Tests that require samples to be sent away for central review following surgery may introduce a short delay (of up to 2–3 weeks) before the decision can be taken on whether or not to offer chemotherapy.

Nine tests were identified in the NICE scope[25] and are included in this assessment: six are based on GEP and three on IHC (protein expression profiling) technology.

Gene expression profiling

Gene expression profiling tests assess the identity and number of messenger ribonucleic acid (mRNA) transcripts in a specific tissue sample. As only a fraction of the genes encoded in the genome of a cell are expressed by being transcribed into mRNA, gene expression profiling provides information about the activity of genes that give rise to these mRNA transcripts. Given that mRNA molecules are translated into proteins, changes in mRNA levels are ultimately related to changes in the protein composition of the cells, and consequently to changes in the properties and functions of tissues and cells (both normal and malignant) in the body.

Various assays are used in the management of breast cancer. These assays investigate the expression of specific panels of genes (also known as a gene profile or gene signature). They work by making use of different techniques to measure mRNA levels in breast cancer specimens, including real-time reverse transcription-polymerase chain reaction (RT-PCR) and deoxyribonucleic acid (DNA) microarrays. Many of these assays have been designed to measure the risk of cancer recurrence. Other uses of the assays include breast cancer subtyping (using molecular classification systems), predicting the likely benefit from certain types of therapy (e.g. chemotherapy) and diagnosing breast cancer.

There are various ways of preparing the RNA and different protocols are used to prepare the specimens [e.g. formalin-fixed paraffin-embedded (FFPE), snap-frozen and fresh samples]. Most UK hospitals currently base their pathology services around FFPE tissue and therefore the use of tests requiring fresh samples would raise major service configuration issues. Furthermore, there are varying algorithms that can be used to combine the raw data to obtain a summary measure. All of these factors can affect the reproducibility and reliability of GEP tests.

These tests provide an estimate of the risk of recurrence and/or information about the intrinsic molecular subtype of cancer. The definition of risk group varies between tests, that is, patients classified as high risk by the OncotypeDX test will be at a different level of risk from patients classified as high risk by the Mammostrat test. The definition of subtype is typically based on the classification system first described by Perou *et al.*[26] in 2000 and refined to include five groups – luminal A, luminal B, HER2 amplified, basal-like and unclassified. Subtype information can potentially be used to provide an indication of risk. For instance, cancers identified as luminal A typically have better prognosis than those identified as luminal B and this information may therefore aid in the risk stratification of ER+ tumours.

The six gene expression profiling tests that are included are as follows:

1. The Randox Breast Cancer Array (BCA) (Randox Laboratories, Crumlin, UK) is a complementary DNA (cDNA)-based expression biochip assay that aims to accurately define the clinical subtypes of breast cancer tumours before initiating treatment. The target population is all individuals with diagnosed breast cancer.
2. MammaPrint® (Agendia, Amsterdam, the Netherlands) is based on microarray technology and uses a 70-gene expression profile. MammaPrint is intended as a prognostic test for women of all ages, LN– and LN+ (up to three nodes positive), with a tumour size of ≤ 5.0 cm. MammaPrint is used to determine the risk of distant recurrence of early breast cancer.

Patients are stratified into two distinct groups – low risk (good prognosis) or high risk (poor prognosis) of distant recurrence. It is cleared by the Food and Drug Administration as an in vitro diagnostic multivariate index assay.

3. BluePrint™ (Agendia, Amsterdam, the Netherlands) is used in addition to the MammaPrint test for molecular subtyping. It is an 80-gene microarray with a target population of patients with early-stage (stage I or II), LN– or LN+ (up to three nodes positive), ER+ or ER– breast cancer. BluePrint provides information on breast cancer subtype using three categories: basal-type, luminal-type and *ERBB2*-type cancers.
4. The PAM50 gene expression assay (ARUP Laboratories, Salt Lake City, UT, USA) identifies the major intrinsic biological subtypes of breast cancer. The current version of the test provides classification of breast cancer subtype and quantitative values for (gene/protein) *ESR1*/ER, *PGR*/PR, *ERBB2*/HER2, proliferation score and luminal score (ER pathway). The PAM50 Breast Cancer Intrinsic Classifier test is recommended for all patients diagnosed with invasive breast cancer, regardless of stage or ER status.
5. OncotypeDX™ (Genomic Health Inc., Redwood City, CA, USA) quantifies gene expression for 21 genes in breast cancer tissue using RT-PCR. It predicts the likelihood of recurrence in women of all ages with newly diagnosed stage I or II, ER+, LN– or LN+ (up to three nodes) breast cancer treated with tamoxifen. The test assigns the breast cancer a recurrence score (RS) and a risk category: low (RS < 18), intermediate (18 ≤ RS ≤ 30) or high (RS ≥ 31). The test also reports ER, PR and HER2 status.
6. The Breast Cancer Index (BCI)℠ (bioTheranostics Inc., San Diego, CA, USA) is a RT-PCR assessment of the ratio of expression of two genes, *HOXB13* and *IL17BR*, combined with the five gene Molecular Grade Index (MGI) and gives an indication of recurrence risk. The target population is those with ER+ and LN– early breast cancer. The BCI RS ranges from 0 to 10 and divides patients into three risk groups: low risk is defined as a score < 5, intermediate risk is a score of 5–6.3 and high risk is a score ≥ 6.4.

Key details of the individual GEP tests are provided in *Table 6*.

Expanded immunohistochemistry (protein expression profiling) tests

Immunohistochemistry markers are being developed to provide similar information to that given by the GEP tests. Some of these tests offer the advantage of using existing IHC technology (such as ER and HER2 markers), which is routinely available in all UK pathology departments.

The three included expanded IHC tests for protein expression are:

1. The IHC4 test (academic sponsor: Royal Marsden Hospital and Queen Mary, University of London) assesses the levels of four key proteins in a breast cancer sample: ER, PR, HER2 and Ki-67. The IHC4 score is calculated based on the percentage of cells positive for Ki67 and PR (0–100%); the Histoscore (a measure of the percentage of cells positive multiplied by the intensity, range 0–300) for ER status; and the tumour HER2 status, expressed as a binary measure (positive/negative). The final algorithm for IHC4 calculates a risk score for distant recurrence based on ER, PR, HER2 and Ki-67 in addition to classical clinical and pathological variables (composite risk score IHC4 + clinical score referred to as IHC4 in our report). Of note, an online calculator is expected to be available at the beginning of 2012 (Professor Mitch Dowsett, Royal Marsden Hospital, London, July 2011, personal communication).
2. The Mammostrat test uses five immunohistochemical markers [solute carrier family 7 (amino acid transporter light chain, L system), member 5 (SLC7A5), *Hpa*II tiny fragments locus 9c protein (HTF9C), protein 53 (p53), N-*myc* downstream regulated 1 (NDRG1) and carcinoembryonic antigen-related cell adhesion molecule 5 (CEACAM5)] to stratify patients into risk groups to inform treatment decisions. These markers are independent of

TABLE 6 Gene expression profiling tests

	OncotypeDX	MammaPrint	PAM50	BCI	BluePrint	Randox BCA
Function	Risk of recurrence	Risk of recurrence	Subtyping	Risk of recurrence	Subtyping – to be used after MammaPrint	Subtyping
Technology	RT-PCR (21 genes)	Microarray (70 genes)	Microarray (55 genes)	RT-PCR, *HOXB13*:*IL17BR* ratio and Molecular Grade Index (seven genes)	Microarray (80 genes)	Low-density biochip array
Location of testing	Central testing – USA	Central testing – Amsterdam and Irvine, USA	Central	Central	Central testing – USA	Local – purchase of array processing unit
Type of sample	FFPE	Fresh (use of FFPE to be introduced in 2012)	FFPE	FFPE	Fresh	Fresh
Staining material	Resection/core biopsy	Resection	Resection/core biopsy	Resection	Resection/core biopsy	Resection/core biopsy
Population	ER+, LN–; also LN+ (one to three nodes)	ER+ (or ER–), LN– and LN+ (one to three nodes)	All women	ER +, LN–	All – previously split into risk group by MammaPrint	All women
Key output of test	RS score – point estimate of the 10-year risk of recurrence	Risk of recurrence score – high/low (based on distant recurrence at 5 years)	Five subtypes: luminal A, luminal B, HER2, basal-like and normal-like	BCI RS	Three subtypes: basal-type, luminal-type and *ERBB2*-type cancers	Five subtypes: luminal A, luminal B, HER2, basal-like and normal-like
Presentation of results	RS and risk group (low < 18, intermediate 18–30, high ≥ 31)	Two categories: low and high risk	Subtype and quantitative values for proliferation, luminal gene expression, *ESR1*, *PGR* and *ERBB2*	Risk score: 0–10. Three risk groups: (low ≤ 5, intermediate 5 –6.3, high ≥ 6.4) and 10-year risk of distant recurrence	Subtype	Unknown
Commercially available in the UK	Yes	Yes	Yes	Yes	Yes	No
Cost	£2580	£2675	US$3200	US$3200 (assuming 20% discount)	No additional cost (over and above MammaPrint)	Unknown

one another and do not directly measure either proliferation or hormone receptor status. The test calculates a relative risk of recurrence through the use of a weighted algorithm, which is interpreted in the context of published clinical studies of appropriate patient populations. Patients are classified into three risk categories: prognostic index ≤ 0, defined as the low-risk group; prognostic index > 0 and ≤ 0.7, defined as the moderate-risk group; and prognostic index > 0.7, defined as the high-risk group.

3. NPI plus (NPI+) (University of Nottingham) is a biomarker-based prognostic assay that integrates 10 predictive biomarkers [ER, PR, HER2, cytokeratin s/b (CK5/6), CK7/8, epidermal growth factor receptor (EGFR), HER3, HER4, p53, mucin 1 (MUC1; cell surface associated)] of long-term survival and therapeutic response with existing clinical and molecular pathology knowledge to support individualised clinical decision-making. This test is under development and outputs/presentation are not yet finalised.

Key details of the individual IHC tests are provided in *Table 7*.

TABLE 7 Expanded IHC tests

	IHC4	NPI+	Mammostrat
Function	Risk of recurrence	Subtyping and risk of recurrence	Subtyping and risk of recurrence
Technology	Combines four IHC tests and clinical parameters to derive prognostic score	Uses 10 biomarkers to derive prognostic score (plus others – to be defined)	Uses five biomarkers to derive risk score
Location of testing	Local? (but quality assurance issues need to be addressed)	Not known	Central
Type of sample	FFPE	FFPE	FFPE
Staining material	Resection/core biopsy	Resection/core biopsy	Resection/core biopsy
Population	Postmenopausal, ER+, LN–	All women, age 18–79 years	ER+, LN–, tamoxifen treated
Key output of test	Continuous IHC4 score	Not yet finalised. To include biological class and projected survival	Risk index and risk group
Presentation of results	IHC4 risk score	Not yet finalised. Likely to be similar to Adjuvant! Online	Risk groups: high, moderate and low
Commercially available in the UK	Algorithm available. Quality assurance issues to be addressed	No	No
Cost	Approx. £100–200	Approx. £500	Approx. £1120–1620

Current usage of gene expression profiling and expanded immunohistochemistry tests in the NHS

Use of these tests has been limited within the NHS to date. The OncotypeDX test has been available in the UK since 2007.[27] There are two ongoing clinical trials for OncotypeDX with some UK recruitment. Outside of this the use of OncotypeDX in the NHS appears to be relatively limited, with a small amount of self-funding by NHS patients, occasional primary care trust funding and charitable funding. Private health insurers offer reimbursement on a case-by-case basis. Use of the other GEP and expanded IHC tests appears to be negligible.

Cost of the tests

The cost of each test is included in *Tables 6* and *7*.

Fresh tissue collection is not routine in the NHS and so there will be additional costs associated with tests requiring fresh tissue samples. These costs could be considerable at hospitals where the dissection facilities are already filled to capacity (which is likely to be a significant proportion of hospitals) and where explicit staffing for collection of fresh tissue is not already in place. This is discussed further in *Chapter 3*, *Model inputs: general*.

Description of the decision problem

Background

Since 2002 NICE has recommended that women at intermediate or high risk of recurrence who have not had neoadjuvant chemotherapy should normally be offered a multiagent chemotherapy that includes anthracyclines.[28] Chemotherapy is defined as the use of cytotoxic medications with the intention of preventing cancer recurrence in patients. It should be noted that, for the purposes of this assessment, chemotherapy does not include other forms of systemic therapy such as endocrine treatments or targeted biological therapy (trastuzumab).

Meta-analyses of randomised controlled trials (RCTs) by the EBCTCG have indicated that the use of adjuvant chemotherapy (chemotherapy following surgery) is associated with a reduction in

the risk of relapse and death in women with early-stage breast cancer.[29] Although chemotherapy can reduce the likelihood of cancer recurrence and death for women with breast cancer, it has considerable adverse effects. Short-term and long-term adverse events will affect a proportion of patients receiving chemotherapy, imposing costs and reducing quality of life. Short-term adverse events that occur during chemotherapy are usually temporary and reversible. The most common side effects include nausea, vomiting, mouth soreness, diarrhoea, tiredness, hair loss and temporary lowering of the blood counts. Long-term side effects such as damage to the heart and a small increase in the risk of leukaemia are not reversible. Although chemotherapy may prevent relapse in some, not all women with early-stage breast cancer will benefit and many women remain recurrence free at 10 years without chemotherapy. However, a subset of patients with a 'good' prognosis may still develop recurrence after curative surgery and adjuvant therapy. This presents a great challenge to clinicians in estimating prognosis and making the most appropriate therapeutic decisions relating to whether or not to use adjuvant chemotherapy in women with early-stage breast cancer.

Recommendations about which patients should receive chemotherapy are typically based on estimations of recurrence risk and expected benefit of therapy. Historically, clinicopathological factors, such as patient age, tumour size, nodal involvement, histological grade, ER expression, HER2 overexpression and comorbidities, have been assessed and considered alongside patient preference. In the UK, guidelines based on NPI and Adjuvant! Online have been developed to assist decision-making relating to adjuvant chemotherapy. These guidelines assist clinicians in deciding the benefits of prescribing chemotherapy for a particular patient. NPI provides information about prognosis that is largely based on pathological parameters (e.g. tumour size, grade and lymph node status), with the addition of ER receptor status, age and comorbidity for Adjuvant! Online. However, these clinicopathological tools are imperfect; different guidelines can give different results and it has been suggested that a proportion of women with early-stage breast cancer are over- or undertreated. This may result in unnecessary use of toxic and expensive chemotherapy for women who derive little or no benefit, or avoidable deaths in women for whom chemotherapy was withheld.

Role of new tests

Gene expression profiling and expanded IHC tests aim to improve the targeting of chemotherapy in breast cancer by improving the stratification and identification of patients who will gain most benefit from chemotherapy. The new tests will provide an indication of the risk of recurrence of patients (based on the results of an algorithm to estimate risk of recurrence or indirectly by identifying the cancer subtype). This is based on the knowledge that certain biological features of cancers may indicate an increased likelihood of rapid growth and metastatic potential. The management of these patients, that is, the decision whether or not to prescribe chemotherapy, will be influenced by the test results, and this may result in a change of management of patients compared with current practice (a decision made based on NPI and/or Adjuvant Online). By more accurately guiding the selection of patients to receive adjuvant chemotherapy in early breast cancer management, the use of GEP or expanded IHC tests in patients with early-stage breast cancer may improve health outcomes and quality of life compared with currently used decision-making protocols.

Comparators

The comparator is standard UK practice. This varies between trusts and encompasses the use of Adjuvant! Online and/or guidelines based on NPI to guide decisions on which patients with early breast cancer should be offered adjuvant chemotherapy.

Identification of important subgroups

The NICE scope[25] identifies the population under assessment as people diagnosed with early breast cancer. However, many of the GEP and expanded IHC tests have been developed for use in a specific subpopulation or currently have evidence of efficacy only within a specific subpopulation. For tests providing a risk of recurrence output, the majority of evidence relates to populations with ER+, LN– early breast cancer. Some of these tests also have more limited evidence in LN+ populations (for patients with one to three nodes involved) and in patients with ER– disease.

These tests will have an impact on the health of patients only if they lead to changes in patient management. This is most likely to happen in populations in which the decision on whether or not to offer chemotherapy is currently uncertain. One such group is patients with ER–, LN–, HER2– early breast cancer for whom prognostic factors suggest that they are at intermediate risk. The definition of this 'intermediate group' is not clear-cut. Clinical advice suggests that patients with a NPI score of ≤ 3.4 are typically considered at low risk either using current prognostic tools (except for a few very young women with aggressive early breast cancer) or based on the new tests and are unlikely to receive chemotherapy; therefore, their management is unlikely to change. Few patients with ER–, LN–, HER2– early breast cancer will have a NPI score > 5.4 and therefore those with a NPI score > 3.4 can be considered as being at intermediate risk.

Current treatment protocols indicate that women with HER2+, ER– early breast cancer or with several positive nodes are likely to receive chemotherapy in most centres in England and Wales. Although the use of GEP or expanded IHC tests might be able to spare chemotherapy in a proportion of these patients, the evidence base for the use of these tests in this population is more limited and clinical opinion therefore considered the assessment of these tests in this population to be a lower priority.

Patients with ER+ LN–, HER2– early breast cancer are therefore considered to be an important population in which to assess these tests, given the current evidence base. Within this population those at intermediate risk for whom the decision about whether or not to offer chemotherapy is not clear cut are considered to be an important subgroup.

Outcomes

The clinical effectiveness review will consider the clinical effectiveness of the tests in relation to:

- Analytical validity (i.e. the ability of the test to accurately and reliably measure the expression of mRNA or proteins by breast cancer tumour cells).
- Clinical validity (i.e. the degree to which the test could accurately predict the risk of an outcome such as disease recurrence and discriminate patients with different outcomes). This relates to the prognostic ability of the test.
- Clinical utility (i.e. the ability of the test to discriminate between those who will have more or less benefit from a therapeutic intervention). This includes evidence relating to how the tests will influence decision-making in terms of which patients will be offered chemotherapy and evidence relating to the predictive ability of the test, that is, the extent to which the test identifies those patients who will benefit most in terms of the relative reduction in the risk of recurrence from treatment.

The outcomes of interest for the economic evaluation are the morbidity and mortality associated with invasive breast cancer and its treatment. Outcomes from the model are expressed in terms of cost per quality-adjusted life-year (QALY).

Aim and objectives of the assessment

The overall aim of the assessment is to assess the clinical effectiveness, effect on patient outcomes and cost-effectiveness of the new GEP and expanded IHC tests.

The objectives of the assessment are:

- To conduct a systematic review of the published evidence on the clinical effectiveness and cost-effectiveness of the nine GEP and expanded IHC tests. In relation to clinical effectiveness, evidence relating to the following outcomes will be sought:
  - analytical validity – the ability of the test to accurately and reliably measure the expression of mRNA or proteins by breast cancer tumour cells
  - clinical validity – the degree to which the test can accurately predict the risk of an outcome (typically distant recurrence) and discriminate patients with different outcomes; this relates to the prognostic ability of the test
  - clinical utility – the ability of the test to discriminate between those who will have more or less benefit from a therapeutic intervention.
- To develop a decision model to investigate the benefits, harms and cost-effectiveness of the GEP and expanded IHC tests compared with current prognostic tools to guide the use of adjuvant chemotherapy in early breast cancer. Outcomes from the model are expressed in terms of cost per QALY.

Note

This report contains reference to confidential information provided as part of the NICE appraisal process. This information has been removed from the report and the results, discussions and conclusions of the report do not include the confidential information. These sections are clearly marked in the report.

Chapter 2

Assessment of clinical effectiveness

A systematic review of the evidence on the clinical effectiveness of nine GEP and expanded IHC tests to guide the use of adjuvant chemotherapy in breast cancer management was undertaken according to the general principles recommended in the Centre for Reviews and Dissemination (CRD) guidance for undertaking systematic reviews,[30] the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) statement[31] and the NICE Diagnostic Assessment Programme Interim Methods Statement.[32] The review protocol can be accessed at www.nice.org.uk/nicemedia/live/13283/54425/54425.pdf and is registered as PROSPERO 2011:CRD4201100136, available from www.crd.york.ac.uk/PROSPERO/full_doc.asp?ID=CRD42011001361.

In addition to the systematic review evidence, a separate section summarising supplementary evidence provided by the manufacturers of the tests will be presented within the section relating to each test. This evidence will simply be summarised and will not be subject to the stages of the systematic review as it is not evidence derived as part of the systematic review process.

Methods for reviewing effectiveness

Background context

The present review evaluates nine prognostic tests for guiding chemotherapy treatment decisions in early-stage breast cancer.

For two of the nine tests (OncotypeDX and MammaPrint) the current review updates an existing systematic review of GEP tests for breast cancer. Two previous systematic reviews[33,34] reviewed the literature relating to both OncotypeDX and MammaPrint (one[34] is an update of the other[33]). In the Marchionni *et al.*[33] review the authors conducted an exhaustive literature review of various electronic databases (covering biomedical literature) between 1990 and 2006. Additional sources included the grey literature (conference proceedings), hand searching the reference lists of included studies and pertinent reviews, contacting the manufacturers of the two tests and regulatory authorities and querying experts in the field. In the Smartt review,[34] the authors updated the Marchionni *et al.*[33] review by updating the search strategy to include all relevant available literature between January 2007 and December 2009. Further details are provided in *Overview of existing systematic reviews of the OncotypeDX and MammaPrint tests*.

In the present review, new search strategies were developed for all of the nine tests based on scoping searches (and strategies reported in the two existing systematic reviews for the OncotypeDX and MammaPrint tests).

Identification of studies

Electronic databases

Studies were identified by searching the following electronic databases:

- MEDLINE (via Ovid SP) 1950–May 2011
- MEDLINE In-Process & Other Non-Indexed Citations (via Ovid SP) 1950–May 2011

- EMBASE (via Ovid SP) 1980–May 2011
- Cochrane Central Register of Controlled Trials (CENTRAL) (via Cochrane Library Issue 3, 2011)
- Cochrane Database of Systematic Reviews (CDSR) (via Cochrane Library Issue 8, 2011)
- NHS Database of Abstracts of Reviews of Effects (DARE) (via Cochrane Library Issue 3, 2011)
- Health Technology Assessment (HTA) database (via Cochrane Library Issue 3, 2011)
- BIOSIS previews (via Ovid SP) 1926–May 2011
- Web of Science (includes Science Citation Index and Conference Proceedings Citation Index) (via WOK) 1899–May 2011.

Extensive searches were undertaken to identify all literature relating to the clinical effectiveness of GEP and expanded IHC tests to guide the use of chemotherapy in breast cancer management. Sensitive keyword strategies using free text and, where available, thesaurus terms using Boolean operators and database-specific syntax were developed to search the electronic databases. Synonyms related to the condition (i.e. breast cancer) were combined with synonyms related to the test (i.e. MammaPrint, OncotypeDX, Randox BCA, BluePrint, PAM50, BCI, IHC4, NPI+).

Searches were not restricted by publication type or language; however, all searches were limited by date. For the OncotypeDX and MammaPrint tests, the searches were restricted to January 2009–May 2011 as the search strategies from the existing systematic reviews appear to be of good quality and are clearly reported and, as a result, all studies up to 2009 would have been identified. For the remaining seven tests, the searches were restricted to January 2002–May 2011. The first evidence for the GEP and expanded IHC tests was reported in 2002 for OncotypeDX and MammaPrint. As these are the most established tests and the furthest along the validation pathway, evidence for subsequent tests will not predate this. An example of the MEDLINE search strategy is provided in *Appendix 1*.

Other resources

To identify additional published, unpublished and ongoing studies, the reference lists of all relevant studies (including existing systematic reviews) and information received by the manufacturers were hand searched and key experts in the field were contacted.

All identified citations from the electronic searches and other resources were imported into and managed using the Reference Manager bibliographic software version 12.0 (Thomson ResearchSoft, San Francisco, CA, USA).

Inclusion and exclusion criteria

The inclusion of potentially relevant articles was undertaken using a two-stage process. First, one experienced systematic reviewer screened all titles and abstracts and excluded any citations that clearly did not meet the inclusion criteria. Second, the full manuscripts of all potentially eligible articles were assessed for inclusion by the same reviewer. At each step, articles that did not satisfy the inclusion criteria were excluded. Any uncertainties in the selection process were resolved through discussion with a second reviewer. The relevance of each article for the clinical effectiveness review was assessed according to the following criteria.

Population

All people diagnosed with early invasive breast cancer being treated in the adjuvant setting were included. People diagnosed with early invasive breast cancer being treated in the neoadjuvant setting were excluded.

Index test

The following GEP tests or expanded IHC tests (that guide treatment decisions in early breast cancer management) were included:

- OncotypeDX
- MammaPrint
- BluePrint
- PAM50
- BCI
- Randox BCA
- Mammostrat
- IHC4
- NPI+.

Reference standard

There was no existing reference standard for the index tests.

Comparator

For studies of clinical validity and clinical utility, relevant comparators were those used in current UK clinical practice. Specifically, studies with Adjuvant! Online and/or NPI as comparators to predict risk of recurrence and survival for patients with early breast cancer were sought, although studies including other comparators and those without a comparator were eligible for inclusion. Further details of the comparators are included in *Description of technologies under assessment* (it should be noted that, by definition, no comparator was necessary for studies of analytical validity).

Outcomes

The following outcome measures (where reported) were included:

- Analytical validity – the ability of the test to accurately and reliably measure the expression of mRNA or proteins by breast cancer tumour cells, that is, repeatability and reproducibility.
- Clinical validity – the degree to which the test can accurately predict the risk of an outcome (typically distant recurrence) and discriminate patients with different outcomes. This relates to the prognostic ability of the test – does the test have evidence on clinical validity and has this been externally validated (in an independent data set).
- Clinical utility – the ability of the test to discriminate between those who will have more or less benefit from a therapeutic intervention.

Clinical utility relates to improvements in clinical outcomes such as overall survival (OS), disease-free survival (DFS), chemotherapy toxicity or quality of life. Based on the conclusion of previous reviews it is not anticipated that prospective studies reporting on long-term outcomes such as OS will be available. In the absence of such studies the following outcomes were to be included:

- Reclassification of risk compared with existing prognostic variables (correlations between test score and score on existing measures such as NPI, Adjuvant! Online), that is, how does the test change the classification of risk for patients.
- Impact of the test results on clinical decision-making – how do the tests results translate into changes in decision-making, for example changes in the proportion of patients receiving adjuvant chemotherapy.

- Predictive ability of the test – does the test accurately predict patients who will benefit most from chemotherapy, that is, do patients classified as high risk benefit more in relative terms than patients classified as low risk.
- Quality of life – directly as a result of knowledge of the test score (e.g. reduction in anxiety) or indirectly through changes in the use of chemotherapy (and consequent changes in quality of life).

Study design

All study designs were included. For the outcome of analytical validity, studies incorporating any pathology method were included. For the outcomes of clinical validity and clinical utility, priority was given to prospective RCT data if available. In the absence of these data prospective and retrospective cohort studies and case–control studies with and without a comparator were eligible for inclusion.

Reviews of primary studies were not included in the review of clinical effectiveness but were retained for discussion and identification of additional studies. The following publication types were excluded from the review: animal models, preclinical and biological studies, editorials, opinions, studies applied only to breast cancer biology, studies published only in languages other than English (unless no other comparable data exist) and non-peer-reviewed reports in which insufficient methodological details are reported to allow critical appraisal of the study quality.

Data abstraction strategy

Data abstraction was performed by one reviewer into a standardised data extraction form and independently checked for accuracy by another reviewer. Discrepancies were resolved by discussion and if agreement could not be reached a third reviewer was consulted. When multiple publications of the same study were identified, data were extracted and reported as a single study. Where appropriate, the authors of the studies (or the manufacturer/sponsor of the test) were contacted to provide further details in cases in which information was missing from the articles.

The following information was extracted for all studies when reported: study details [author, year of publication, country, study design, number of eligible patients, number of included patients, follow-up time, evidence type (analytical validity, clinical validity, clinical utility), funding], patient characteristics (age, lymph node status, ER status, tumour size, grade, HER2 status, mean NPI score, and treatment) and results [outcomes/end points, results (in the format presented in the study), authors' conclusions]. Numerical data extracted from the studies were varied and included the following: numbers and percentages of patients having a change in management as a result of the test, association between test score and risk of outcomes [distant recurrence, time to distant recurrence (TTDR)] [*p*-values and associated hazard ratios and 95% confidence intervals (CIs)], correlation between test score and comparator score, differences (*p*-values) between cases and control subjects on test score.

Critical appraisal strategy

There are no validated (or widely agreed) tools for the assessment of prognostic (predictive factor) studies and there is little empirical evidence to support the importance of particular study features affecting the reliability of findings, including the avoidance of bias. Although there are several published quality assessment checklists for assessing prognostic studies in cancer,[35,36] they vary considerably, both in their content and complexity. For this review a generic list of important methodological features recommended by Altman[37] was deemed to be the most appropriate (useful) to assess the internal validity of the included studies. Further details on the methodological assessment tool are provided in *Appendix 2*.

The methodological quality of each included study was assessed by one reviewer and checked by another reviewer using the criteria recommended by Altman.[36] Any discrepancies were resolved by discussion, with involvement of a third reviewer when necessary. Blinding of the quality assessor to author, institution or journal was not considered necessary. The quality assessment items recommended by Altman employed six dimensions relating to the risks of bias of prognostic studies and included the following: sample of participants, follow-up of participants, outcome, prognostic variable, analysis and treatment subsequent to inclusion in cohort. Study quality was assessed with each item scored as 'yes', 'no' or 'unclear'. When a study was reported in more than one publication, its quality was assessed on the basis of the combined data from all relevant publications. Studies were rated as high quality if they received a positive assessment of at least 17 out of 21 methodological quality items.

As the current review updates two existing systematic reviews of GEP tests for breast cancer (OncotypeDX and MammaPrint tests), the methodological quality of these two systematic reviews was assessed using the criteria recommended by Shea *et al.*[38] (assessment of multiple systematic reviews – AMSTAR). The quality assessment checklist for assessing systematic reviews included items on a priori design, data extraction, literature searching, quality assessment, data synthesis, publication bias and conflicts of interest. Further details on the methodological assessment tool together with the details of the assessment of each review are provided in *Appendix 3*.

Methods of data synthesis

Studies that met the entry criteria were eligible for inclusion in meta-analyses if this was appropriate in terms of comparability of the study populations, outcomes and diagnostic thresholds, and if the studies were unlikely to be biased. However, because of the degree of heterogeneity, meta-analysis was not considered appropriate. The presentation of results is therefore limited to a narrative review. The results were grouped in separate sections by test. For each test a summary of the evidence in terms of evidence type, overall quality and key findings was presented in table form at the beginning of the results section. More detailed summaries of the evidence were presented in narrative form in the subsequent sections, arranged by evidence type. Studies relating to analytical validity were detailed first, followed by those relating to clinical validity and then those relating to clinical utility. The studies relating to clinical utility were further divided when possible by those relating to the predictive ability of the test (benefit of chemotherapy), reclassification of risk against existing prognostic variables, changes in treatment recommendations, quality of life and patient anxiety. A summary of the evidence was then presented, again by evidence type.

Results

This section will first provide an overview of the evidence from the two existing systematic reviews of GEP tests (OncotypeDX and MammaPrint) for breast cancer. Second, this section will present the results of the current systematic review of each of the nine tests. Where applicable, supplementary evidence (from the manufacturers and other sources) will also be provided.

Overview of existing systematic reviews of the OncotypeDX and MammaPrint tests

In January 2008, Marchionni *et al.*[33] published a systematic review of the impact of GEP tests on breast cancer outcomes. The objective of the review was to examine the available evidence relating to the analytical and clinical validity of breast cancer GEP in predicting disease recurrence and the clinical utility of these tests in improving chemotherapy choices and patient outcomes. Three gene signatures and their commercially available tests were reviewed: the

OncotypeDX test, the MammaPrint test and the two-gene ratio test (*HOXB13*:*IL17BR*) (not the subject of this review). In 2010, Smartt[34] updated this systematic review and included all relevant evidence from January 2007 to December 2009.

Although a number of other systematic reviews examining GEP tests have been reported, it was felt that the Marchionni *et al.*[33] and Smartt[34] reviews were the most appropriate reviews to update. Other reviews predated those of the Marchionni *et al.*[33] and Smartt[34] reviews were considered to be of lower quality as they did not describe the search strategy and processes of the systematic review in as much detail or did not report the findings in as much detail.

The methodological quality of both systematic reviews was reasonably high (as assessed using the criteria recommended by Shea *et al.*;[38] for further details see *Appendix 3*). Both reviews provided an a priori design, details of a comprehensive literature search and details of conflicts of interest both for the review and for the included studies and combined the findings of the studies in an appropriate way. Marchionni *et al.*[33] provided details of duplicate study selection and detailed that data extraction had been performed by one reviewer and checked by a second, whereas this information was not provided for the Smartt review.[34] Marchionni *et al.*[33] stated that they had searched for and included grey literature as appropriate; however, although Smartt[33] stated that the same procedure had been followed as for Marchionni *et al.*,[33] no specific reference was made to searching or including grey literature. A list of included studies was provided by both reviews; however, a list of excluded studies was provided only for the Marchionni *et al.* review.[33] In both reviews, characteristics tables for included studies were not clearly presented and appeared only in the appendices in the case of the Smartt review.[34] Both reviews presented the methods used for quality assessment, although how this was actually carried out was presented in more detail in Smartt,[33] and both reviews used study quality when formulating conclusions. Neither review assessed publication bias.

In total, 21 studies on the OncotypeDX test and 13 on the MammaPrint test were identified and included by Marchionni *et al.*[33] and Smartt.[34] A summary of the evidence type and overall quality of each study is provided for OncotypeDX and MammaPrint in *Tables 8* and *9* respectively.

Summary of evidence: Marchionni et al.[33]

OncotypeDX

Marchionni *et al.*[33] reported that OncotypeDX was furthest along the validation pathway, with strong retrospective evidence that it predicts distant metastasis and chemotherapy benefit to a clinically relevant extent over standard predictors in a well-defined clinical subgroup with clear treatment implications. A more detailed summary of the main results is provided in *Appendix 4*.

Analytical validity Marchionni *et al.*[33] reported a number of studies on analytical validity and overall success rate of OncotypeDX. They concluded that evidence existed for some of the operational characteristics of this test but that there was limited evidence for the reproducibility of the test in terms of reproducibility across different samples of the same block and across samples from different blocks. No direct evidence was available about the effect of sample preparation. There was indirect evidence that the overall success rate of extracting analysable mRNA was fairly high. Centralisation was considered to be a current strength of OncotypeDX with regard to reproducibility.

Clinical validity Marchionni *et al.*[33] reported fairly strong support for the clinical validity of OncotypeDX over and above that of standard clinical predictors in ER+, LN– and tamoxifen-treated patients, with a clear treatment indication for adjuvant chemotherapy. Paik *et al.*[47] showed

TABLE 8 Existing data from Marchionni *et al*.[32] and Smartt[33] on OncotypeDX

Author (year)	Evidence type	Overall quality[a]
[b]Cronin *et al.* (2004)[39]	Analytical validity	Not reported
[b]Cronin *et al.* (2007)[40]	Analytical validity	Not reported
[b]Habel *et al.* (2006)[41]	Analytical validity/clinical validity (prognosis)	Not reported
[b]Paik *et al.* (2004)[42]	Analytical validity/clinical validity (prognosis)	Not reported
[b]Cobleigh *et al.* (2005)[43]	Analytical validity/clinical validity (prognosis)	Not reported
[b]Esteva *et al.* (2005)[44]	Analytical validity/clinical validity (prognosis)	Not reported
Bryant (2005)[45] (poster)[b]	Clinical validity (prognosis)	Not reported
Hornberger *et al.* (2005)[46] (poster)[b]	Clinical validity (prognosis)	Not reported
Paik *et al.* (2004)[42] (poster)[b]	Clinical validity (prognosis)	Not reported
[b]Oratz *et al.* (2007)[48]	Analytical validity/clinical utility (indirect evidence only)	Not reported
[b]Paik *et al.* (2006)[49]	Analytical validity/clinical utility (indirect evidence only)	Not reported
[c]Goldstein *et al.* (2008)[50]	Clinical validity	Reasonably sound evidence
[c]Wolf *et al.* (2008)[51]	Clinical validity	Low-quality evidence
[c]Asad *et al.* (2008)[52]	Clinical utility (indirect evidence only)	Low-quality evidence
[c]Henry *et al.* (2009)[53]	Clinical utility (indirect evidence only)	Low-quality evidence
[c]Li *et al.* (2009)[54]	Clinical utility (indirect evidence only)	Low-quality evidence
[c]Rayhanabad *et al.* (2008)[55]	Clinical utility (indirect evidence only)	Low-quality evidence
Erb *et al.* (2007)[56] (abstract)[c]	Clinical utility (indirect evidence only)	Not reported
Gold *et al.* (2009)[57] (abstract)[c]	Clinical utility (indirect evidence only)	Not reported
Lo *et al.* (2007)[58] (abstract)[c]	Clinical utility (indirect evidence only)	Not reported
Shak *et al.* (2009)[59] (abstract)[c]	Clinical validity	Not reported

a The quality of the included evidence was documented in narrative form but not categorised in the Marchionni *et al.* review; hence, overall quality was not reported for the studies included in that review. Furthermore, quality assessment for abstracts was not reported in either review.
b Data from Marchionni *et al.*[33]
c Data from Smartt.[34]

TABLE 9 Existing data from Marchionni *et al*.[32] and Smartt[33] on MammaPrint

Author (year)	Evidence type	Overall quality[a]
[b]Ach *et al.* (2007)[60]	Analytical validity	Not reported
[b]Buyse *et al*, (2006)[61]	Analytical validity/clinical validity	Not reported
[b]Glas *et al.* (2006)[62]	Analytical validity/clinical validity	Not reported
[b]Van't Veer *et al.* (2002)[63]	Clinical validity	Not reported
[b]van de Vijver *et al.* (2002)[64]	Clinical validity	Not reported
[c]Mook *et al.* (2009)[65]	Clinical validity	Reasonably sound evidence
[c]Wittner *et al.* (2008)[66]	Clinical validity	Low-quality evidence
[c]Bueno-de-Mesquita *et al.* (2007)[67]	Clinical utility	Reasonably sound evidence
Bender *et al.* (2009)[68] (abstract)[c]	Clinical utility	Not reported
de Snoo *et al.* (2009)[69] (abstract)[c]	Clinical validity	Not reported
Glas *et al.* (2008)[70] (abstract)[c]	Clinical validity	Not reported
Knauer *et al.* (2009)[71] (abstract)[c]	Clinical validity	Not reported
Saghatchian *et al.* (2009)[72] (abstract)[c]	Clinical validity	Not reported

a The quality of the included evidence was documented in narrative form but not categorised in the Marchionni *et al.* review; hence, overall quality was not reported for the studies included in that review. Furthermore, quality assessment for abstracts was not reported in either review.
b Data from Marchionni *et al.*[33]
c Data from Smartt.[34]

that RS was significantly correlated with DFS ($p = < 0.001$) and OS ($p = < 0.001$). RS alone was a better predictor of distant recurrence at 10 years than traditional clinicopathological predictors.

Clinical utility Marchionni *et al.*[33] concluded that the Paik *et al.*[49] study represented the strongest evidence for the clinical utility of the OncotypeDX test. Using data from ER+, LN– patients in the National Surgical Adjuvant Breast and Bowel Project (NSABP) B20 trial, Paik *et al.*[49] compared a group of patients treated with tamoxifen and chemotherapy with a group treated with tamoxifen only. RS was found to be correlated with chemotherapy benefit, defined in terms of 10-year distant recurrence-free survival (DRFS), with a significant benefit from the use of chemotherapy in the high RS group ($p = 0.001$). However, in a multivariate analysis the benefit from chemotherapy was unclear because of large CIs in the low- and intermediate-risk groups. Marchionni *et al.*[33] noted that, although prospective confirmation of these findings was required, this evidence provided reasonable justification in the interim for the use of the test by ER+, LN– women.

MammaPrint

The evidence reported by Marchionni *et al.*[33] for MammaPrint was more limited. A more detailed summary of the main results is provided in *Appendix 5*.

Analytical validity Two technical studies[60,62] provided evidence relating to the analytical validity of MammaPrint. Repeated gene expression measurements over time, within and across individual microarrays and across different laboratories, protocols, instruments and operators provided data on the variability and reproducibility of the test. Buyse *et al.*[61] reported an overall success rate of the assay of 80.9%. Marchionni *et al.*[33] concluded that, although these studies suggested that MammaPrint could be used in a clinical setting, they could not be considered to be direct validations of the assay. The review also noted that evidence underpinning the analytical validity of the test was obtained from a limited number of patients and a moderate number of replications. The only validation study using the MammaPrint assay (rather than the underlying 70-gene signature) showed that only about 80% of fresh-frozen specimens were analysable.

Clinical validity Marchionni *et al.*[33] concluded that, overall, the available published evidence supported MammaPrint as a better predictor of the 5-year risk of distant recurrence than traditional clinical predictors.[61] Buyse *et al.*[61] compared MammaPrint with Adjuvant! Online for prediction of distant metastases within 5 years and for death within 10 years. Similar sensitivities were found for both methods but a higher specificity was demonstrated for MammaPrint. However, the cohorts used were clinically heterogeneous, meaning that generalisations of the findings to a particular patient group are more difficult.

Clinical utility No evidence on the clinical utility of the test was reported.

Summary of evidence: Smartt[34]

The updated systematic review by Smartt[34] found that the additional studies (published between January 2007 and December 2009) on OncotypeDX and MammaPrint addressed some but not all of the outstanding issues relating to the clinical validity and clinical utility of these tests. A summary of the main results is provided in *Appendices 4* and *5*.

OncotypeDX

Analytical validity No further evidence was reported.

Clinical validity Smartt[34] identified two further studies[50,51] on the clinical validity of OncotypeDX. Goldstein *et al.*[50] reported that OncotypeDX was a more accurate predictor of relapse than

standard clinical features for hormone receptor-positive, chemotherapy/hormonal therapy-treated patients and provides complementary information to standard clinicopathological measures. Wolf *et al.*[51] assessed the correlation between standard clinical and pathological breast cancer characteristics and the RS in a cohort of Israeli breast cancer patients and compared the stratification of patients using the RS with that using commonly used clinical guidelines. Neither standard clinicopathological features nor the chosen clinical guidelines/assessment tools could reliably predict the RS among referred breast cancer patients. The clinical utility of these comparisons was not made clear.

Clinical utility Smartt[34] identified four studies[52,55,73,74] on the clinical utility of OncotypeDX. Smartt reported that the studies examined the ability of the test to predict response to treatment or its impact on clinical decision-making. The studies all reported a positive impact of the test on clinical decision-making and generally claimed that there was a reduction in the number of patients who were or would have been considered for chemotherapy. However, the studies generally had methodological weaknesses and were likely to have overestimated the effect/influence of the test and they were not designed to assess the effect of the test on clinical outcomes.

MammaPrint

Analytical validity No further evidence was reported.

Clinical validity Smartt[34] identified two studies[66,75] on the clinical validity of MammaPrint. Mook *et al.*[75] reported that MammaPrint predicted disease outcome better than traditional clinical prognostic factors in patients with one to three positive nodes and was able to accurately identify node-positive patients with an excellent prognosis. The potential clinical utility of MammaPrint was demonstrated in 72 (34%) clinically high-risk patients with a good prognosis signature who had a 10-year breast cancer disease-specific survival of 94% and therefore might be spared chemotherapy. Wittner *et al.*[66] reported a study on LN– patients. MammaPrint had a high negative predictive value (NPV) and provided some information that was additional to that provided by Adjuvant! Online. However, with an extremely low positive predictive value (PPV) and non-significant differences in OS between MammaPrint high- and low-risk patients, the prognostic utility of MammaPrint in this population remained unproven. Moreover, although MammaPrint classified a significant proportion of study patients as high risk, few of these developed metastatic disease.

Clinical utility Smartt[34] identified one study on the clinical utility of MammaPrint. Bueno-de-Mesquita *et al.*[67] reported a prospective study of 427 patients with a MammaPrint profile. The study demonstrated a lack of congruence with well-known clinical guidelines for risk assessment in breast cancer; in approximately one-third of patients there was discordance. The addition of MammaPrint to the standard Dutch clinical assessment of risk (modified by patient preference) increased by 20 the number of patients receiving adjuvant systemic therapy. Follow-up was not long enough to provide evidence of its effect on clinical end points such as distant metastasis-free survival (DMFS) or its utility in predicting treatment benefit.

Key evidence gaps identified by these reviews

OncotypeDX

- Analytical validity – there is limited evidence for the reproducibility of the tests in terms of reproducibility across different samples of the same block and across samples from different blocks. Centralisation was considered to be a current strength of OncotypeDX with regard to reproducibility.

- Clinical validity (prognostic ability of the tests) – there is fairly strong support for OncotypeDX over and above standard clinical predictors, but only in a well-defined population (ER+, LN–). Evidence is required to assess the stability of risk categories in other populations.
- Clinical utility – very few of the studies, particularly in isolation, provided compelling evidence of the test's clinical utility.

MammaPrint

- Analytical validity – there were limited data on variability and reproducibility, with a limited number of patients and a moderate number of replications.
- Clinical validity (prognostic ability of the tests) – evidence was based on retrospective data using clinically heterogeneous cohorts; evidence from RCTs is needed.
- Clinical utility – very limited evidence was available on clinical utility; robust evidence on the prediction of chemotherapy benefit is required.

Marchionni *et al.*[32] concluded (at the time of publication) that for both tests the relationship of predicted to observed risk in different populations needed further study, as did their incremental contribution, optimal implementation and relevance to patients on current therapies. Smartt[34] concluded that the largest volume of evidence related to the OncotypeDX test.

Studies included in the current systematic review

The literature searches identified 5993 potentially relevant citations. Of the titles and abstracts screened, 218 relevant full papers or abstracts were retrieved and assessed for inclusion. A flow chart describing the process of identifying relevant literature is shown in *Figure 2*. A total of 32 citations evaluating the effectiveness of nine prognostic tests (for guiding chemotherapy treatment decisions in early-stage breast cancer) met the inclusion criteria. *Figure 2* also shows the numbers of studies included for each prognostic test. Studies excluded from the review are listed in *Appendix 6* (only those citations that were excluded after a full-text reading for reasons not immediately apparent from the full text).

OncotypeDX

OncotypeDX quantifies gene expression for 21 genes in breast cancer tissue using RT-PCR. It is intended to predict the likelihood of recurrence in women of all ages with newly diagnosed stage I or II, ER+, LN– or LN+ (up to three nodes) breast cancer treated with tamoxifen. The test assigns the breast cancer a RS and a risk category: low (RS ≤ 18), intermediate (18 ≤ RS ≤ 30) or high (RS ≥ 31). The test also reports ER, PR and HER2 status and can provide an indication of how responsive the cancer is likely to be to hormonal therapy. Further details are provided in *Table 6*.

Description of included studies

The present review identified an additional 12 studies (13 citations) for the OncotypeDX test. This included 11 fully published peer-reviewed papers and two meeting abstracts. Of these citations five were related to clinical validity and the remaining eight to clinical utility.

The design and patient characteristics of the 12 included studies are provided in *Tables 10* and *11* respectively. Most of the studies used a retrospective analysis of archived tumour samples together with a database of patient characteristics and prognostic information. Only three studies stated that the design was prospective.[76–78] The majority of participants analysed in the studies were ER+, LN–, and the mean age was around 50–60 years. Most studies included a small number of participants (range 25–367), although three analysed relatively large cohorts [Dowsett *et al.*[79] ($n = 1231$), Mamounas *et al.*[80] and Tang *et al.*[81] (both $n = 1674$ – analyses of the B14 and B20 trials)]. Follow-up was short or not reported for a number of studies; again, the exceptions

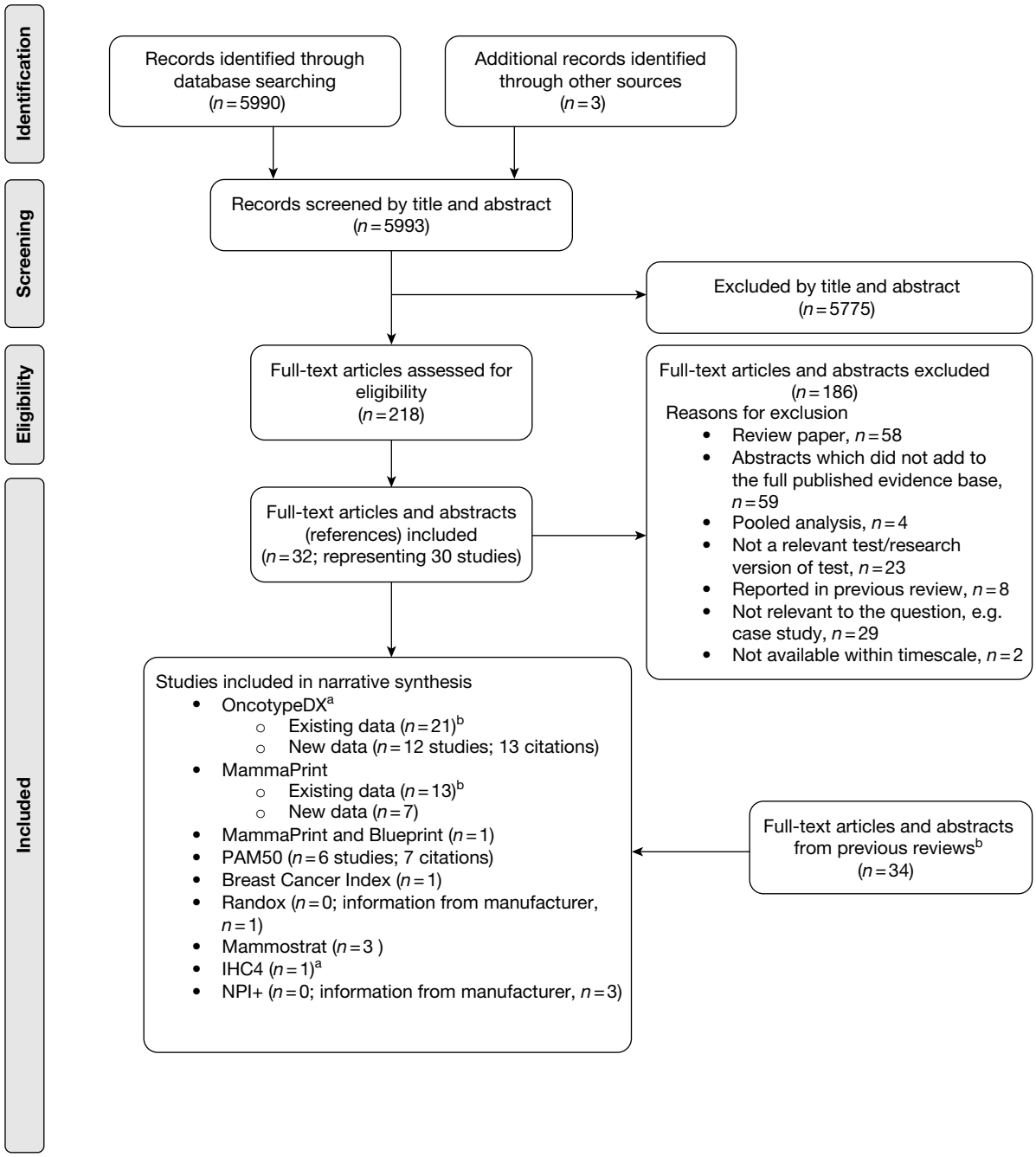


FIGURE 2 Prognosis review: PRISMA (adapted) flow chart. a, one paper is included for both OncotypeDX and IHC4 as it contains data relating to both tests; b, from the systematic reviews of Marchionni *et al*.[33] and Smartt.[34]

were the studies by Dowsett *et al.*[79] (9 years) and Mamounas *et al.*[80] and Tang *et al.*[81] (minimum of 10 years).

Quality of included studies: OncotypeDX

The methodological quality of the 12 included studies[76–88] is summarised in *Figure 3* (further details are provided in *Appendix 7*). Generally, only three studies (four citations) performed well, receiving a positive assessment for at least 17 out of 21 methodological quality items.[80–82,84]

TABLE 10 Study design characteristics of included studies: OncotypeDX test (new data)

Author (year) Country	Study design	Number of patients	Follow-up	Outcomes/end points	Evidence type	Funding
Ademuyiwa *et al.* (2011)[82] USA	Observational, retrospective, consecutive series (2005–9)	Eligible sample: 276 Sample included: 276	NR	Impact on clinical decision-making in terms of recommending chemotherapy	Clinical utility – reclassification against existing prognostic variables – and changes in treatment recommendations	NR
Albain *et al.* (2010)[83] USA, Canada	Retrospective cohort (1989–95) from a Southwest Oncology Group intergroup trial (SWOG-8814, INT 0100)	Eligible sample: 413 Sample included: 367 Samples excluded because of exhaustion of invasive tumour in block, no submission of primary tumour or technical issues	Maximum 13 (median 8.94) years	The degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes	Clinical utility – predictive ability (benefit of chemotherapy)	National Cancer Institute and Genomic Health
Cuzick *et al.*[84] (2011) Multinational including UK	Retrospective cohort from the TransATAC trial (1990–8) FFPE	Eligible sample: 1125 Sample included: 1125	Follow-up: 100-month median follow-up	Distant recurrence (within 10 years), TTDR	Clinical validity	Royal Marsden National Institute for Health Research Biomedical Research Centre, Cancer Research UK, Breakthrough Breast Cancer and AstraZeneca
Dowsett *et al.* (2010)[79] Multinational including UK	Retrospective cohort (dates NR) from TransATAC trial	Eligible sample: 1372 Sample included: 1231 Samples excluded because of unsuccessful RT-PCR analysis ($n=64$) and clinical characteristics ($n=77$)	9 years	Degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes	Clinical validity	Breakthrough Breast Cancer and AstraZeneca
Geffen *et al.* (2009)[77] Israel	Prospective cohort, consecutive patients (2002–6) Subset who received test: NR	Eligible sample: 328 Sample included: 25 Samples excluded as the test was not available for the majority of patients	Subset: NR Whole cohort: NR	Impact on clinical decision-making	Clinical utility – changes in treatment recommendations	Unfunded
Holt *et al.* (2011)[78] (abstract only) UK	Prospective, cohort (dates NR)	Eligible sample: 107 Sample included: 106 One patient excluded because of inadequate sample	NR	Impact on clinical decision-making	Clinical utility – changes in treatment recommendations	NR
Kelly *et al.* (2010)[85] USA	Observational, consecutive patients (2004–8), retrospective analysis of prospective database	Eligible sample: 309 Sample included: 309	NR (states that it was short)	Correlation with Adjuvant! Online and risk prediction	Clinical utility – reclassification against existing prognostic variables	NR

Author (year) Country	Study design	Number of patients	Follow-up	Outcomes/end points	Evidence type	Funding
Lo *et al.* (2010)[76] USA	Observational, consecutive (2005–6), prospective	Eligible sample: 93 Sample included: 89 Four samples excluded as these patients did not complete both pre- and post-RS assay questionnaires	Up to 12 months	Impact of the 21-gene RS assay on clinical decision-making and patient preferences. End points include (1) changes in physician treatment recommendations, (2) physician self-assessed changes in long-term adjuvant treatment, (3) patient anxiety, (4) quality of life, (5) relapse data	Clinical utility – changes in treatment recommendations – and quality of life and patient anxiety	Genomic Health
Tang *et al.* (2011)[81] Mamounas *et al.* (2010)[80] USA	Retrospective, tissue from patients in two trials [NSABP B14 (1982–8) and B20 (1988–93)]	Eligible sample: 1349 Sample included: 1319 Tamoxifen (TAM) treated: $n=895$; TAM + chemotherapy treated: $n=424$ 30 samples excluded because of unsuccessful RT-PCR: B14: $n=11$; B20: $n=19$	Median for distant recurrence-free interval: B14 ($n=668$): 14.3 years; B20 ($n=651$): 10.6 years	The degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes	Clinical validity (Mamounas *et al.*); clinical utility – predictive ability (benefit of chemotherapy) (Tang *et al.*)	National Cancer Institute
Tang *et al.* (2010)[86] (abstract only) USA	Retrospective cohort (dates NR) of patients from the randomised NSABP B20 trial	Eligible sample: NR Sample included: 625 (with RS and ER score ≥ 6.5)	NR	Distant recurrence Value of the integration of RS and clinicopathological factors in the prediction of chemotherapy benefit in reducing risk of recurrence	Clinical utility – predictive ability (benefit of chemotherapy)	NR
Toi *et al.* (2010)[87] Japan	Retrospective cohort (1992–8)	Eligible sample: 325 Sample included: 200 Samples excluded because of LN+ disease (exclusion criteria) ($n=80$); limited or no clinical data ($n=12$); ineligible by pathology evaluation ($n=27$); insufficient RNA ($n=1$); failed RT-PCR ($n=5$)	NR	The degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes	Clinical validity	Japanese Ministry of Health, Labor, and Welfare
Yorozuya *et al.* (2010)[88] Japan	Case–control, retrospective (2000–8)	Eligible sample: 40 Sample included: 40 (10 cases, 30 control subjects)	Cases: 53.4 months; control subjects: 55 months	The degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes	Clinical validity	NR

NR, not reported; ATAC, Arimidex, Tamoxifen, Alone or in Combination trial.

TABLE 11 Patient characteristics of included studies: OncotypeDX test (new data)

Author (year)	Age (years), mean (SD)	LN status	ER status	Tumour size	Grade	HER2 status	Mean NPI score	Treatment
Ademuyiwa *et al.* (2011)[82]	54.8 (range 29–82)	All LN–	All ER+	≤1 cm: 63 (22.8%); 1.1–2 cm: 159 (57.6%); >2 cm: 54 (19.6%) Mean 1.6 cm (range 0.3–4.5 cm); median 1.4 cm	I: 104 (37.7%); II: 139 (50.4%); III: 33 (12.0%)	All HER2–	Excellent or good: 220 (79.7%); moderate: 56 (20.3%)	Chemotherapy: 88 (31.9%); no chemotherapy: 188 (68.1%)
Albain *et al.* (2010)[83]	Overall: 60.4 (7.5) (range 42–81) 30–54: 90 (24.5%); 55–64: 169 (46.0%); ≥65: 108 (29.4%)	All LN+	All ER+	<2 cm: 120 (32.7%); 2–5 cm: 230 (62.7%); >5 cm: 17 (4.6%)	I: 131 (35.7%); II: 194 (52.9%); III: 42 (11.4%)	HER2+: 43 (11.7%)	NR	Tamoxifen alone: 148; chemotherapy, then tamoxifen: 219
Cuzick *et al.*[84] (2011)[a]	G1: NR Median 64 (IQR 57–70)	–/+/unknown 793 (70%)/299 (26%)/44 (4%) Those with unknown nodal status taken to be node negative in analyses	NR (reported to be ER and/or PR positive)	≤1 cm: 177 (16%); 1–2 cm: 574 (51%); >2–3 cm: 272 (24%)	Poor: 206 (18%); moderate: 690 (61%); well differentiated: 229 (21%); unknown: 49 (4%)	HER2+: 116 (10%)	NR	Tamoxifen: 565 (50%); anastrozole: 560 (50%)
Dowsett *et al.* (2010)[79]	64.3 (NR)	Negative: 71%; positive: 25%; unknown: 4.3% (Note 0.7% unaccounted for)	100% hormone receptor positive; does not state if progesterone or oestrogen	≤2 cm: 67%; 2–5 cm: 31%; >5 cm: 1.5%; unknown: 0.3%	Well: 27%; moderate: 52%; poor: 16%; unknown 4.6%	NR	NR	Radiotherapy: 68%; received HRT: 36%; tamoxifen before surgery: 3.9% Tamoxifen: 609/1231; anastrozole: 622/1231
Geffen *et al* (2009)[77]	Subset: NR Whole cohort: <35: 5 (1.5%); 35–55: 107 (32.6%); 56–75: 190 (58.0%); >75: 26 (7.9%)	All LN–	Subset: NR Whole cohort: 288/328 (87.8%)	NR	Subset: NR Whole cohort: low: 70 (21.3%); intermediate: 144 (43.9%); high: 63 (19.2%); undetermined: 51 (15.5%)	Subset: NR Whole cohort: HER2 overexpression: 21 (6.4%)	Subset: NR Whole cohort: NR	Subset: NR Whole cohort: lumpectomy: 297; mastectomy: 31; local therapy: 328; endocrine therapy: 283 (57 had chemotherapy as well); no systemic therapy: 20; chemotherapy: 25

Author (year)	Age (years), mean (SD)	LN status	ER status	Tumour size	Grade	HER2 status	Mean NPI score	Treatment
Holt *et al.* (2011)[78] (Abstract only)	NR	Pathological negative or pathological stage N1 % NR	All ER+	NR	NR but states 'early stage'	NR	NR	Patient choice after OncotypeDX: no chemotherapy: 74/106 (69.8%); chemotherapy: 32/106 (30.2%)
Kelly *et al.* (2010)[85]	Mean NR Median (IQR) at diagnosis: 54 (47–62)	0 nodes: 292 (95%); one to three nodes: 16 (5%); four nodes: 1 (0.3%)	NR but states that all are hormone receptor positive (unclear if this refers to progesterone or oestrogen)	All: grade I and grade III NR Grade II ($n = 191$) median tumour size: 1.3 (IQR 1.0–1.8) cm	I: 45 (15%); II: 191(62%); III: 72 (23%)[a]	HER2: negative: 307 (99%); positive: 2 (0.7%) Median Ki-67: 10 (IQR 5–20)	NR	Adjuvant or neoadjuvant therapy: 84 (27%)
Lo *et al.* (2010)[76]	55 (NR) (range 35–77)	All LN–	All ER+	Mean 1.7 cm (SD NR) (range 0.6–3.5)	Low: 19/89 (21.3%); intermediate: 58/89 (65.2%); high: 12/89 (13.5%)	HER2: negative: 6 (7%)	NPI NR RS: < 18: 38 (42.7%); 18–30: 42 (47.2%); ≥ 31: 9 (10.1%)	Treatment option chosen by patients after test: chemotherapy and hormone therapy: 20; hormone therapy: 65; observation: 3; equipoise: 1
Tang *et al.* (2011)[81] Mamounas *et al.* (2010)[80]	Trial B14: < 50: 194 (29%); 50 to ≤ 60: 173 (26%); ≥ 60: 301 (45%) Trial B20: < 50: 289 (44%); 50 to ≤ 60: 166 (26%); ≥ 60: 196 (30%)	All LN–	All ER+	Trial B14: 0–1.0 cm: 112 (17%); 1.1–2.0 cm: 303 (45%); 2.1–4.0 cm: 218 (33%); ≥ 4.1 cm: 35 (5%) Trial B20: 0–1.0 cm: 83 (13%); 1.1–2.0 cm: 313 (48%); 2.1–4.0 cm: 226 (35%); ≥ 4.1 cm: 29 (4%)	NR	NR	NR	Trial B14: lumpectomy plus irradiation: 393 (38%); mastectomy: 630 (62%) Trial B20: lumpectomy plus irradiation: 277 (43%); mastectomy: 374 (57%)
Tang *et al.* (2010)[86] (abstract only)	NR	All N–	All ER+	NR	NR	NR	NR	Tamoxifen with or without adjuvant chemotherapy

continued

TABLE 11 Patient characteristics of included studies: OncotypeDX test (new data) (*continued*)

Author (year)	Age (years), mean (SD)	LN status	ER status	Tumour size	Grade	HER2 status	Mean NPI score	Treatment
Toi *et al.* (2010)[87]	< 50: 68 (34%); ≥ 50: 132 (66%)	All LN–	All ER+	≤ 2 cm: 92 (46%); > 2 cm: 108 (54%)	I: 30 (15%); II: 80 (40%); III: 36 (18%); unknown: 54 (27%)	NR	NR	Mastectomy: 143 (72%); breast conservation: 57 (29%); tamoxifen: 200/200 (100%)
Yorozuya *et al.* (2010)[88]	Cases: 49.1 (12.9) (range 37–76) Control subjects: 50.9 (11.9) (range 37–78)	All LN–	All ER+	Cases: 18.9 (SD 4.6) (95% CI 15.7 to 22.2) mm Control subjects: 15.8 (SD 6.3) (95% CI 13.4 to 18.2) mm	Cases: I: 1 (10%); II: 1 (10%); III: 8 (80%); unknown 0 (0%) Control subjects: I: 15 (50%); II: 9 (30%); III: 5 (17%); unknown: 1 (3%)	Cases: negative: 7 (70%); positive: 2 (20%); unknown: 1 (10%) Control subjects: negative: 27 (90%); positive: 0 (0%); unknown: 3 (10%)	NR	Cases: mastectomy: 8 (80%); partial mastectomy: 2 (20%); adjuvant hormone therapy: 9 (90%); adjuvant chemotherapy: 1 (10%) Control subjects: mastectomy: 11 (37%); partial mastectomy: 19 (63%); adjuvant hormone therapy: 26 (87%); adjuvant chemotherapy: 4 (13%)

CI, confidence interval; HRT, hormone replacement therapy; IQR, interquartile range; NR, not reported; SD, standard deviation.
a Data is correct (as reported in the original paper).

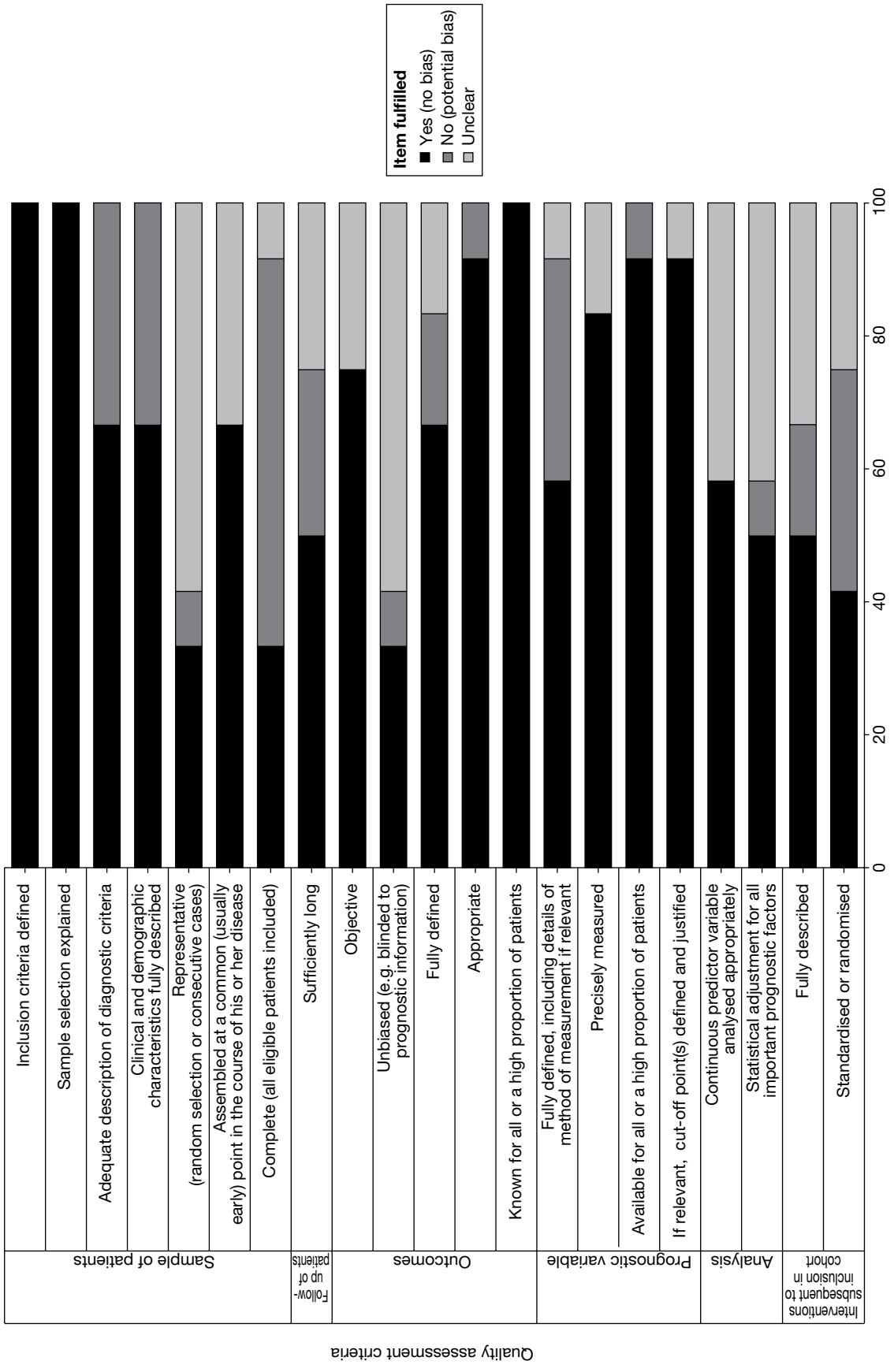


FIGURE 3 OncotypeDX test: methodological quality graph (review authors' judgements about each methodological quality item presented as percentages across all studies).

Although 9 of the 12 studies used a retrospective study design,[79–88] other potential sources of bias were generally related to the following domains: sample of patients (inadequate description of diagnostic criteria, clinical/demographic characteristics not fully described and not including all eligible patients with tumour samples in the study), follow-up of patients, prognostic variables (not fully defined) and interventions subsequent to inclusion in the study (interventions were not described or standardised, thus precluding an unbiased assessment of the prognostic ability of the test).[30]

The assessment of study quality was further hampered by poor reporting of the following methodological items: whether or not sample of patients was representative and assembled at an early point in the course of their disease, whether or not outcomes were fully defined, objective and unbiased and whether or not appropriate statistical analyses were undertaken (continuous predictor variables analysed appropriately and statistical adjustment made for all prognostic factors). Overall, the risk of bias from the 12 included studies was judged to be moderate.

Results: OncotypeDX

In this section a summary of the clinical evidence for OncotypeDX is presented (*Table 12*) followed by a narrative summary of each study. Full data extraction tables are provided in *Appendix 7*.

Analytical validity

No new data examined analytical validity.

Clinical validity

Using 1231 tissue samples from the UK TransATAC (Arimidex, Tamoxifen, Alone or in Combination trial) trial, Dowsett *et al.*[79] assessed postmenopausal, hormone receptor-positive and majority LN– patients. They demonstrated that a 50-point increase in RS in all LN– patients (e.g. RS = 55 vs. RS = 5) was significantly associated with an increased risk of distant recurrence [hazard ratio (HR) 3.92, 95% CI 2.08 to 7.39; $p < 0.001$] when adjusted for the effects of tumour size, local grade (grade derived from case record forms), age and treatment. They also reported that, when local grade was replaced with central grade (assessed using the Elston and Ellis system) in multivariate analysis, adjusted RS was also significantly associated with risk of distant recurrence (HR 5.25, 95% CI 2.84 to 9.73; $p < 0.001$). RS was also significantly associated with TTDR in both node-negative (HR 5.25, 95% CI 2.84 to 9.73; $p < 0.001$) and node-positive patients (HR 3.47, 95% CI 1.64 to 7.38; $p < 0.002$). Correlation between RS-predicted distant recurrence and Adjuvant! Online-predicted recurrence was low but statistically significant by central grade (Spearman rank correlation = 0.23; $p < 0.001$) or local grade (Spearman rank correlation = 0.22; $p < 0.001$). Only approximately 5% of the variability in the estimates of recurrence using either of these scores was explained by the other. The authors concluded that the findings demonstrated that RS is an independent predictor of distant recurrence in LN– and LN+ hormone receptor-positive patients treated with anastrozole, adding value to estimates with standard clinicopathological features. As the patients were recruited as part of a large-scale trial this study benefits from a large sample size of UK-based patients and has a relatively long follow-up time (9 years).

Yorozuya *et al.*[88] reported a very small case–control study (10 cases, 30 control subjects) of ER+, LN– Japanese patients. The cases were those who had metastases after surgery; control subjects did not develop metastases. Significant differences were shown between the groups in terms of mean RS (cases: mean RS = 40.0, 95% CI 21.1 to 58.9; control subjects: mean RS = 17.8, 95% CI 13.8 to 21.9; $p < 0.001$). The study found significant differences between cases and control subjects in the proportions assigned to different risk categories [low: 3 (30%) cases vs. 19 (63%) control subjects; intermediate: 1 (10%) vs. 8 (27%); high: 6 (60%) vs. 3 (10%); $p = 0.005$]. Multivariate

TABLE 12 Summary of evidence for the OncotypeDX test

Author (year)	Evidence type	Overall quality	Key findings
Ademuyiwa *et al.* (2011)[82]	Clinical utility – reclassification against existing prognostic variables – and changes in treatment recommendations	High	276 ER+, LN– patients from two cancer centres in the USA. Impact on clinical decision-making in terms of recommending CT based on clinicopathological characteristics. 37 fewer patients received CT using RS to help decide CT use. 38% of patients had a change in management as a result of the RS. Authors reported a significant association between RS risk group and NPI ($p < 0.001$). Conclusion: RS score had a significant impact on the receipt of adjuvant CT. Limitations: sample size relatively small, use of retrospective chart review
Albain *et al.* (2010)[83]	Clinical utility – predictive ability (benefit of chemotherapy)	Medium	367 postmenopausal ER+ and LN+ US and Canadian patients from the SWOG-9914 trial. RS is prognostic for tamoxifen-treated patients with positive nodes and predicts significant benefit of CAF in tumours with a high RS. Conclusion: a low score identifies women who might not benefit from anthracycline-based chemotherapy, despite positive nodes. Limitations: moderate sample size, time over which tumour samples were collected not reported, therefore they may be differences in diagnostic criteria being applied
Cuzick *et al.*[84] (2011)	Clinical validity	High	1125 patients, majority LN– and hormone receptor positive; multinational including UK. The authors reported the mean change in (likelihood ratio chi-squared) for the addition of the RS to the classical score (higher values indicate more added prognostic information) for TTDR and TR (all recurrences). For TTDR the (likelihood ratio chi-squared) for all patients was 25.3 (95% CI 25.2 to 25.9) and for LN– patients was 20.9 (95% CI 20.7 to 21.6). For TR the LR-$_{\chi}^{2}$ for all patients was 25.6 (95% CI 25.2 to 25.9) and for LN– patients was 25.7 (95% CI 25.4 to 26.4). The authors report that the OncotypeDX RS adds prognostic information to traditional clinicopathological measures. This study has been rated as high quality and benefits from a large sample of patients
Dowsett *et al.* (2010)[79]	Clinical validity	Medium	1231 UK, postmenopausal, hormone receptor-positive, LN– patients. Increase in RS significantly associated with an increased risk of distant recurrence. RS was also significantly associated with TTDR. Correlation between RS-predicted distant recurrence and Adjuvant! Online-predicted recurrence was low but statistically significant. Conclusion: RS is an independent predictor of distant recurrence in LN– and LN+ hormone receptor-positive patients treated with anastrozole, adding value to estimates using standard clinicopathological features. Large sample size, UK-based patients
Geffen *et al.* (2009)[77]	Clinical utility – changes in treatment recommendations	Medium	25 LN– Israeli patients. Each patient had a RS assay. Study reported findings on the impact of the OncotypeDX RS on clinical decision-making. Nine patients (36%) had their treatment recommendations changed based on the score, six from CT to no CT. Limitations: very small sample size
Holt *et al.* (2011)[78] (Abstract only)	Clinical utility – changes in treatment recommendations	Medium	106 UK, ER+ and either LN– or N1 patients. 35 patients (33%) had their initial recommendation changed as a result of the RS; for 71 patients (67%) there was no change. RS added prognostic information beyond that from NPI alone. Conclusion: authors concluded that early results suggest that OncotypeDX is applicable and feasible to perform in the UK setting with a reduction in the use of adjuvant CT. Limitations: although UK based only conducted in one centre, small sample size, abstract data
Kelly *et al.* (2010)[85]	Clinical utility – reclassification against existing prognostic variables	Medium	309 hormone receptor-positive, LN– patients at clinically intermediate risk. Of these, 52% were assigned a low risk on RS, 9% high risk and 39% intermediate risk. Conclusion: findings suggest that OncotypeDX has utility in reclassifying clinically intermediate patients into the three OncotypeDX risk groups. Employed recently diagnosed patients. Limitations: small sample size and a short follow-up time
Lo *et al.* (2010)[76]	Clinical utility – changes in treatment recommendations – and quality of life and patient anxiety	Medium	89 ER+, LN– patients. Prospective US-based study of RS effects on physician and patient adjuvant treatment selection and satisfaction, and quality of life. Changes in physician treatment recommendations for 28 patients (31.5%); 24 patients (27.0%) changed their own treatment decision. Most of the treatment changes were from CHT to HT alone for both physicians and patients. DCS score and state anxiety were significantly reduced across time points (pre and post RS), and the FACT-G score was marginally significantly reduced. Trait anxiety and the FACT-B score were not significantly different. Limitations: small sample size, only 16 physician self-reports

continued

TABLE 12 Summary of evidence for the OncotypeDX test (*continued*)

Author (year)	Evidence type	Overall quality	Key findings
Tang *et al.* (2011)[81] Mamounas *et al.* (2010)[80]	Clinical validity (Mamounas *et al.*); clinical utility – predictive ability (benefit of chemotherapy) (Tang *et al.*)	High	1319 ER+, LN– patients from two large US trials (NSABP B14 and B20). Tang *et al.* – both RS and Adjuvant! Online provided strong independent prognostic information in tamoxifen-treated patients. In the B20 cohort RS was significantly predictive of CT benefit (for DRFI, OS and DFS) but Adjuvant! Online was not. In the larger B20 subcohort, Adjuvant! Online was significantly predictive of CT benefit for OS but not for DRFI or DFS. Conclusion: prognostic estimates can be optimised by combining RS and Adjuvant! Online. RS should be used for estimating relative CT benefit. Mamounas *et al.* – in the tamoxifen-treated patients, RS was a significant predictor of locoregional recurrence. Large sample size with a long follow-up. Limitation: relatively old tumour samples, may be differences in diagnostic criteria applied
Tang *et al.* (2010)[86] (Abstract only)	Clinical utility – predictive ability (benefit of chemotherapy)	Medium	625 ER+, LN–, US patients from the NSABP B20 trial. Examined the value of the SPC (integration of RS and clinicopathological factors) in the prediction of CT benefit in reducing risk of recurrence. Authors concluded that RS used alone remains the best predictor of CT benefit in ER+, LN– breast cancer. Large sample size. Limitations: abstract data, Tang *et al.*[80] and Mamounas *et al.*[79] also used the NSABP cohorts – limitations in using the same data because of risks of double counting in the evidence base as a whole
Toi *et al.* (2010)[87]	Clinical validity	Medium	200 ER+, LN–, Japanese patients. Patients categorised as low risk had a significantly lower risk of distant recurrence than patients in the high-risk category. Continuous RS was significantly associated with the risk of distant recurrence. Conclusion: OncotypeDX has value in providing prognostic information in Asian populations with ER+, LN– breast cancer. Limitation: small sample, Japanese patients so generalisability to UK practice may be limited
Yorozuya *et al.* (2010)[88]	Clinical validity	Medium	40 ER+, LN–, Japanese patients. Compared those who had metastases after surgery with those who did not develop metastases. Significant differences were shown between the groups in terms of mean RS and there were significant differences in the proportions assigned to the different OncotypeDX risk categories. Conclusion: both histological grade and risk score classification were effective in identifying women at risk of developing distant metastases after initial therapy. Limitations: very small sample size, may not be generalisable to the UK setting

CAF, cyclophosphamide, doxorubicin and fluorouracil; CHT, chemotherapy plus hormone therapy; CI, confidence interval; CT, chemotherapy; DCS, Decisional Conflict Scale; DRFI, distant recurrence-free interval; FACT-B, Functional Assessment of Cancer Therapy breast cancer scale; FACT-G, Functional Assessment of Cancer Therapy general scale; HT, hormone therapy; TR, time to recurrence.

logistical regression analysis of age, ER score, PR score, RS, histological grade and lymphatic invasion compared with distant metastases showed that RS was not significant [RS ≥ 50 vs. RS < 50, $p = 0.579$, odds ratio (OR) 2.85, 95% CI 0.07 to 115.552] although the authors conclude that the OR indicates that it has value. The authors concluded that both histological grade and RS classification were effective in identifying women at risk of developing distant metastases after initial therapy for ER+, LN– stage I or IIA breast cancer. There are significant limitations in terms of the generalisation of the findings because of the extremely small sample size used in this study; furthermore, as the study was Japan based, generalisations to the UK setting are limited.

Toi *et al.*[87] examined the prognostic ability of OncotypeDX in 200 ER+, LN– Japanese patients. They showed that patients categorised as low risk had a significantly lower risk of distant recurrence than patients in the high-risk category ($p < 0.001$, log-rank test). No recurrences were identified in the intermediate recurrence group. Continuous RS was significantly associated with the risk of distant recurrence for a 50-point increase in RS (HR 6.20, 95% CI 2.27 to 17.0). In multivariate analyses the continuous RS maintains statistical significance when adjusting for age and clinical tumour size (HR 6.03, 95% CI 2.17 to 16.7). For risk of recurrence the HR was 3.38 (95% CI 1.32 to 8.69), for risk of recurrence or death the HR was 2.09 (95% CI 0.84 to 5.20) and for risk of death the HR was 2.67 (95% CI 0.93 to 7.62). The authors concluded that OncotypeDX has value in providing prognostic information in Asian populations with ER+,

LN– breast cancer. This study had a small sample size and as it was conducted using Japanese patients generalisability to UK practice may be limited; however, the study does benefit from the fact that the tumour samples used were from patients who presented and were treated relatively recently (1992–8).

Cuzick *et al.*[84] reported data that aimed to assess how much of the information in the RS is contained in standard IHC markers (data from this report relating to the IHC4 test is presented in *IHC4 test*). Patients comprised a retrospective cohort from the TransATAC trial (multinational including the UK). The 1125 patients were mainly LN– and hormone receptor positive, and there were a total of 195 recurrences of which 145 were distant recurrences. In LN– women there were 101 recurrences of which 67 were distant recurrences. The authors reported the mean change in likelihood ratio chi-squared for the addition of GHI-RS (Genomic Health Recurrence Score) v to the classical score in the validation halves of 100 random splits of the data (higher values indicate more added prognostic information) for TTDR and time to recurrence (all recurrences). For TTDR the likelihood ratio chi-squared for all patients was 25.3 (95% CI 25.2 to 25.9) and for LN– patients was 20.9 (95% CI 20.7 to 21.6). For time to recurrence the LR-$_{\chi}^{2}$ for all patients was 25.6 (95% CI 25.2 to 25.9), and for LN– patients was 25.7 (95% CI 25.4 to 26.4). The authors report that the OncotypeDX RS adds prognostic information to traditional clinicopathological measures. This study has been rated as high quality and benefits from a large sample of patients.

Mamounas *et al.*[80] (and Tang *et al.*,[81] reported in the following section) undertook a retrospective analysis of ER+, LN– patients who had been recruited into two large US trials (NSABP B14 and B20). They showed a significant association between RS and the proportion of patients with locoregional recurrence at 10 years for 355 placebo-treated patients (NSABP B14), 895 tamoxifen-treated patients (NSABP B14 and B20) and 424 tamoxifen plus chemotherapy-treated patients (NSABP B20). Multivariate Cox regression analysis in the cohort of 895 tamoxifen-treated patients showed that RS was a significant predictor of locoregional recurrence (HR 2.16, 95% CI 1.26 to 3.68; $p < 0.005$). The authors concluded that a significant association exists between RS and risk for locoregional recurrence. This information has biologic consequences and potential clinical implications relative to locoregional therapy decisions for patients with LN– and ER+ breast cancer. These studies appeared to be of reasonable quality and, as the patients were recruited as part of two large-scale trials, the studies benefit from a large sample size with a long follow-up. However, across the two trials tumour samples were collected as long ago as 1982 until 1993. This means that there may be differences in diagnostic criteria applied and this may limit the generalisability of these findings to current practice.

Clinical utility

Predictive ability of the OncotypeDX test (benefit of chemotherapy) Tang *et al.*[81] (and Mamounas *et al.*,[80] as reported in the previous section) undertook a retrospective analysis of the NSABP B14 and B20 trial data on ER+, LN– patients. They compared the prognostic and predictive utility of the OncotypeDX RS and Adjuvant! Online, with an end point of distant recurrence-free interval (DRFI). Cox proportional hazards models were used to compare the prognostic and predictive utility of RS and Adjuvant! Online. Both RS ($p < 0.001$) and Adjuvant! Online ($p = 0.002$) provided strong independent prognostic information in tamoxifen-treated patients. Combining RS and individual clinicopathological characteristics provided greater prognostic discrimination than combining RS and the composite Adjuvant! Online. In the B20 cohort with RS results ($n = 651$), RS was significantly predictive of chemotherapy benefit (interaction $p = 0.031$ for DRFI, $p = 0.011$ for OS, $p = 0.082$ for DFS) but Adjuvant! Online was not. However, in the larger B20 subcohort ($n = 1952$), Adjuvant! Online was significantly predictive of chemotherapy benefit for OS (interaction $p = 0.009$) but not for DRFI or DFS. The authors concluded that prognostic estimates can be optimised by combining RS and Adjuvant! Online. RS should be used for estimating

relative chemotherapy benefit. As stated above, these studies appeared to be of reasonable quality and, as the patients were recruited as part of two large-scale trials, the studies benefit from a large sample size with a long follow-up. However, across the two trials tumour samples were collected as long ago as 1982 until 1993. This means that there may be differences in diagnostic criteria applied and this may limit the generalisability of these findings to current practice.

Albain *et al.*[83] reported findings for 367 postmenopausal ER+ and LN+ US and Canadian patients from the SWOG-9914 trial. They aimed to investigate whether or not RS is prognostic in women treated with tamoxifen alone and whether or not it identified those who might not benefit from anthracycline-based chemotherapy despite higher risks of recurrence. RS was prognostic in the tamoxifen alone group (HR 2.64, 95% CI 1.33 to 5.27; $p = 0.006$) using a 50-point difference in RS as a threshold. There was no benefit of chemotherapy with cyclophosphamide, doxorubicin and 5-fluorouracil (CAF) in patients with a low RS but an improvement in DFS for those with a high RS (score ≥ 31) (HR 0.59, 95% CI 0.35 to 1.01; $p = 0.033$), after adjustment for number of positive nodes. The RS by treatment interaction was significant in the first 5 years ($p = 0.029$), with no additional prediction beyond 5 years, although cumulative benefit remained at 10 years. There were similar findings for OS and breast cancer-specific survival (BCSS). The authors concluded that RS is prognostic for tamoxifen-treated patients with positive nodes and predicts significant benefit of CAF in tumours with a high RS. A low score identifies women who might not benefit from anthracycline-based chemotherapy, despite positive nodes. This study employed a moderate sample size. The authors did not report the length of time over which tumour samples were collected; therefore, it is unclear whether or not it is likely that there were differences in the diagnostic criteria applied.

Tang *et al.*[86] reported a study in abstract form that included 625 ER+, LN– US patients treated with tamoxifen with or without adjuvant chemotherapy from the NSABP B20 trial. They aimed to examine the value of the integration of RS and clinicopathological factors (RSPC) in the prediction of chemotherapy benefit in reducing risk of recurrence. They reported that in 60 of the 625 patients distant recurrence occurred. The RS showed a significant interaction with chemotherapy treatment ($p = 0.037$) with a standardised HR of 0.836. The interaction of RSPC with treatment was not significant ($p = 0.10$) although there was a trend in the same direction as for RS (HR 0.833). The authors concluded that RS used alone remains the best predictor of chemotherapy benefit in ER+, LN– breast cancer. This study benefits from having a large sample size. However, there are significant limitations in making any interpretations from this evidence as it is derived only from an abstract. It has been shown that there may be discrepancies between data made available in abstracts and the reporting of results in subsequently published full-length articles.[89] Because of incomplete reporting the methodological quality of studies cannot be confidently assessed by systematic reviewers. It should also be noted that Tang *et al.*[81] and Mamounas *et al.*[80] also used the NSABP cohorts. There are limitations in using the same data because of the risks of double counting in the evidence base as a whole.

Reclassification of risk against existing prognostic variables Kelly *et al.*[85] considered the correlation between OncotypeDX and Adjuvant! Online in a US population of 309 consecutive patients with hormone receptor-positive, majority LN– early breast cancer of clinically intermediate risk. They demonstrated a low correlation between Adjuvant! Online risk prediction and RS, and between death after 5 years of tamoxifen therapy and RS. Of these patients considered to be of clinically intermediate risk, 52% ($n = 160$) were assigned a low risk on RS, 9% ($n = 27$) a high risk and 39% ($n = 122$) an intermediate risk. The authors concluded that OncotypeDX yielded potentially informative risk assignments in patients who may be considered at indeterminate risk by routine clinical variables. However, 40% of the time they remain as intermediate risk using RS thresholds; this increases to 66% when using thresholds that have been revised for an ongoing trial of OncotypeDX [Trial Assigning Individualized Options for Treatment (TAILORx) – which will be described in *Ongoing trial: the Trial Assigning Individualized Options for Treatment*] (the revised

thresholds are as follows: low risk ≤ 10; intermediate risk 11–25; high risk ≥ 26). These findings suggest that OncotypeDX has utility in reclassifying clinically intermediate patients into the three OncotypeDX risk groups. The study benefits from the fact that all patients had been diagnosed relatively recently (2004–8), although it also has limitations, including a small sample size and a short follow-up time (actual follow-up time was not reported). The authors were not able to report recurrence and survival results because of the short follow-up time.

Ademuyiwa *et al.*[82] reported a study on 276 ER+, LN– patients from two cancer centres in the USA. They reported a significant association between RS risk group and NPI ($p < 0.001$), although there were a number of discordant cases (comparisons are difficult because NPI and RS have two and three risk categories respectively). This was only a brief report and it therefore lacked the detail necessary to make adequate judgements about quality. Furthermore, the sample size was relatively small. Further data on clinical decision-making from this study are reported in the following section.

Changes in treatment recommendations Geffen *et al.*[77] reported findings on the impact of the OncotypeDX RS on clinical decision-making in 25 LN– patients in Israel. Nine patients (36%) had their treatment recommendations changed based on the scores, six of these from chemotherapy to no chemotherapy. The generalisability of these findings is limited, primarily because of the very small sample size. Furthermore, as this study was conducted in Israel, generalisability to the UK is limited.

Lo *et al.*[76] reported a prospective US-based study of 89 ER+, LN– patients to examine whether or not RS affects physicians' and patients' adjuvant treatment selection and satisfaction. They reported changes in physician treatment recommendations for 28 (31.5%) patients; 24 (27.0%) patients changed their own treatment decision. The largest change after RS results was conversion in 20 (22.5%) cases from physicians' pretest recommendation of chemotherapy plus hormone therapy to a post-test recommendation of hormone therapy. Nine (10.1%) patients changed their treatment decision from chemotherapy plus hormone therapy to hormone therapy. The authors concluded that the RS assay impacts significantly on physician and patient adjuvant treatment decision-making. Most of the treatment changes were from a pretreatment recommendation of chemotherapy plus hormone therapy to hormone therapy alone, for both physicians and patients. In addition, Lo *et al.*[75] reported, based on physician self-reports, that RS results have an enduring impact on physician confidence in their treatment recommendation. The generalisability of these findings is limited because of the very small sample size of 89 patients and only 16 physician self-reports.

Ademuyiwa *et al.*[82] investigated the impact on the use of clinicopathological features (based on patient records with oncologists blind to RS) in decision-making for chemotherapy utilisation. The study included 276 ER+, LN– patients from two cancer centres in the USA. In total, 37 fewer patients received chemotherapy using RS to help decide chemotherapy use; 38% of the patients had a change in management as a result of RS. The authors reported a significant association between RS risk group and NPI ($p < 0.001$), although there were a number of discordant cases (comparisons are made difficult because NPI and RS have two and three risk categories respectively). They concluded that the RS had a significant impact on the receipt of adjuvant chemotherapy. This was only a brief report and therefore lacked the detail necessary to make adequate judgements about quality. Furthermore, the sample size was relatively small and there may also be significant limitations from the use of retrospective chart review.

In a conference poster Holt *et al.*[78] reported a study investigating the impact of RS on clinical decision-making in Wales. The 106 patients included in the study were ER+ and either LN– or N1. The authors reported data on change in recommendations pre RS assay to post RS assay.

They demonstrated that 35 patients (33.0%) had their initial recommendation changed as a result of RS [change chemotherapy to no chemotherapy: 25 (23.6%); change no chemotherapy to chemotherapy: 10 (9.4%)] whereas for 71 patients (66.9%) there was no change [no change no chemotherapy: 49 (46.2%); no change chemotherapy: 22 (20.8%)]. The Spearman's rank correlation comparing RS with individual components of the NPI showed that, of size, LN status and grade, only grade was significantly correlated. The authors concluded that early results suggest that OncotypeDX is applicable and feasible to perform in the UK setting, with a reduction in the use of adjuvant chemotherapy consistent with the findings of other studies. RS added prognostic information beyond that from NPI alone. Although the study was UK based it was conducted in only one centre with a very small sample size, making generalisations of the findings difficult. Furthermore, because more chemotherapy was given in the comparator arm, more benefits are likely to be derived from the use of OncotypeDX. In addition, there are significant limitations in making any interpretations from this evidence as it is derived only from an abstract. It has been shown that there may be discrepancies between data made available in abstracts and the reporting of results in subsequently published full-length articles.[89] Because of incomplete reporting the methodological quality of studies cannot be confidently assessed by systematic reviewers.

Quality of life and patient anxiety Lo *et al.*[76] also reported quality of life and patient anxiety data for 89 ER+, LN– patients. Patients were asked to complete standardised measures to assess decisional conflict and personal perceptions of decision-making [Decisional Conflict Scale (DCS)] pre and immediately post RS; anxiety – state anxiety refers to a transitory emotional state or condition and trait anxiety denotes relatively stable individual differences in anxiety proneness [State-Trait Anxiety Inventory (STAI)] pre RS, immediately post RS and 12 months post RS; and quality of life [Functional Assessment of Cancer Therapy (FACT)-B, which is specific to breast cancer, and FACT-G, which is the general scale] pre RS and 12 months post RS. The results showed that DCS score was significantly reduced post RS compared with pre-RS ($p < 0.001$); the STAI demonstrated that state anxiety was significantly reduced across the three time points ($p = 0.007$) whereas trait anxiety was not significantly different across the three time points. For quality of life the FACT-B score pre RS was not significantly different from the score at 12 months post RS; however, the FACT-G score was marginally significantly reduced at 12 months compared with pre RS ($p = 0.49$). The authors concluded that patient anxiety and decisional conflict were significantly lower after RS results. The small sample size used in this study limits the generalisability of the findings and further research in this area is necessary before definitive conclusions on quality of life improvements and reductions in patient anxiety can be formed.

Summary of evidence: OncotypeDX

Analytical validity of OncotypeDX

In the earlier systematic reviews evidence exists on the technical and operational aspects of the test and on assay variability and reproducibility. Studies showed reasonable within-laboratory replicability.

Our findings indicated no new evidence.

Clinical validity (prognostic ability) of OncotypeDX

In earlier systematic reviews the evidence shows that RS was significantly correlated with DFS and OS. RS alone was shown to be a better predictor of distant recurrence at 10 years than traditional clinicopathological predictors.[42] Key gaps relate to the stability of risk categories in populations other then ER+, LN– patients.

Our findings indicate that further larger studies now exist which support the prognostic capability of OncotypeDX. In particular, a large UK study in 1231 postmenopausal women with

hormone receptor-positive, LN– early breast cancer concluded that an increase in risk score was significantly associated with an increased risk of distant recurrence.[90] This study and the Mamounas *et al.*[80] study provide new evidence on the clinical validity of OncotypeDX, which employs cohorts of patients from large-scale RCTs and is rated as high quality. Furthermore, the evidence base has been extended to include the LN+ population[83] and there are the beginnings of an evidence base for the validity of OncotypeDX in different populations such as in Japanese patients.[87,88]

Clinical utility of OncotypeDX

In the earlier systematic reviews, evidence on clinical utility is limited. Paik *et al.*[49] demonstrated a significant benefit from the use of chemotherapy in the OncotypeDX high-risk group, although the review highlighted that the study may have been subject to bias as some patients in the validation data set were also in the training data set. Clinical experts indicated that more effective chemotherapy regimes are currently used in the UK. In total, > 44% of patients were aged < 50 years. The benefit of chemotherapy (reduction in distant recurrence) was greater in this population than in women aged > 50 years. The HR for the benefit of chemotherapy (reduction in distant recurrence) in women aged > 50 years compared with younger women was 2.02 (95% CI 0.75 to 5.47; $p = 0.162$).

Further supporting evidence was needed. Key gaps relate to the extent to which the test added to the management of patients and the proportion of patients who would benefit from the test. The role of the OncotypeDX test in guiding treatment of HER2-positive patients was unclear, as most of these patients were classified in the high-risk RS group in the initial trials. Prospective confirmation of the clinical utility of OncotypeDX was required.

Our findings indicate that there are no prospective studies reporting the impact of OncotypeDX on long-term outcomes such as OS. Four new studies[76–78,82] presented further evidence on the impact of OncotypeDX on clinical decision-making. These indicate that the use of OncotypeDX leads to changes in decision-making for between 31.5% and 38% of patients. However, only one of these studies was UK based, and limitations in relation to study design were identified for this study. Specifically, these data were based on a small sample size ($n = 106$) and were derived from a conference poster,[78] which was lacking the detail necessary to make judgements about the quality of the evidence. Two new studies (with three related citations[81,83,86]) provided evidence supporting the case that OncotypeDX predicts chemotherapy benefit. The Tang *et al.*[81,86] studies were based on ER+, LN– patients and Albain *et al.*[82] reported evidence for ER+, LN+ patients. These studies were based on trial data and the sample sizes were moderate in the case of Albain *et al.*[83] ($n = 367$) and large in the Tang *et al.*[81,86] analyses ($n = 625–1319$). These studies also had long follow-up times. Study quality was judged to be medium[83,86] or high,[81] although as Tang *et al.*[86] was a conference abstract we were unable to access the detail necessary to make adequate judgements about the quality of the evidence.

The first evidence relating to improvements in quality of life and reductions in patient anxiety as a result of using the test have been reported, although generalisations should be made with caution because of the small sample sizes employed. Further research in this area is required.

Key gaps in the evidence remain:

- Few of the studies were considered to be of high quality ($n = 3$). A number of studies in the current review were judged to provide medium-quality (although retrospective) evidence for OncotypeDX ($n = 9$). One of the most characteristic features of the studies was their heterogeneity. The studies varied considerably in their size, study design, patient populations and objectives. A large proportion of the OncotypeDX studies were small and retrospective. Many studies used old archived tumour samples and included the use of retrospective chart

review to elicit treatment recommendations before and after OncotypeDX testing. There was a lack of standardised decision-making tools both within and between studies, and non-standardised methods of patient selection for OncotypeDX testing were used.
- Further direct evidence of the clinical utility of OncotypeDX is still required. This will be addressed by the ongoing TAILORx trial.
- The generalisability of the findings may be limited because of the small number of studies that were conducted in the UK setting and because a number of the studies were funded by the manufacturer, giving rise to possible conflicts of interest and publication bias.

Overall summary

The OncotypeDX evidence is the furthest along the validation pathway compared with other similar tests, and the evidence base, in particular in relation to the prognostic ability of the test, was reasonably sound. This review has identified further studies supporting the prognostic ability (clinical validity) of the test. These are generally of moderate to high quality. Our findings indicate that there are no prospective studies reporting the impact of OncotypeDX on long-term outcomes such as OS. Four additional studies on the impact of OncotypeDX on decision-making indicate that the use of OncotypeDX leads to changes in decision-making for 31.5–38% of patients, but only one of these relates to the UK setting. Two further studies on the predictive benefit of the test were identified, one for LN+ patients. The first evidence relating to improvements in quality of life and reductions in patient anxiety as a result of using the test has been reported, but this is based on small patient numbers and further evidence is required.

Ongoing trial: the Trial Assigning Individualized Options for Treatment

The TAILORx trial commenced in April 2006 and is due to complete primary outcomes in April 2014. It aims to demonstrate that endocrine treatment alone is non-inferior to chemoendocrine treatment in women with an intermediate OncotypeDX score (11–25). Patients aged 18–75 years with ER+ and/or PR+, HER2/neu-negative tumours who are LN– (and who will be treated with tamoxifen) are eligible for inclusion. All patients receive OncotypeDX profiling and are then allocated to risk groups. Those at low risk (≤ 10) will receive endocrine therapy alone and those at high risk (≥ 26) will receive endocrine therapy and adjuvant chemotherapy. Those at intermediate risk (11–25) will receive endocrine therapy and be randomly assigned to chemotherapy or no chemotherapy. The trial is closed for recruitment.[91] Funding for the study is provided by the National Cancer Institute. Further details of this trial are included in *Appendix 8*.

MammaPrint

MammaPrint is based on microarray technology and uses a 70-gene expression profile. MammaPrint is intended as a prognostic test for women of all ages, LN– and LN+ (up to three nodes positive) with a tumour size of ≤ 5.0 cm. MammaPrint is used to determine the risk of distant recurrence of early breast cancer. Patients are stratified into two distinct groups – low risk (good prognosis) or high risk (poor prognosis) of distant recurrence. Further details are provided in *Table 6*.

Description of included studies

The present review identified an additional seven studies for the MammaPrint test. This included six full published peer-reviewed papers and one dissertation chapter.

The design and patient characteristics of the seven included studies are provided in *Tables 13* and *14* respectively. Most of the studies included retrospective analyses of archived tumour samples together with a database of patient characteristics and prognostic information. Only one study stated that the design was prospective.[92] The populations used in the studies were somewhat heterogeneous, with some using only LN– patients and others using a mixture of LN– and LN+ patients. There was a similar pattern relating to ER status. The mean age was around 50 years.

TABLE 13 Study design characteristics of included studies: MammaPrint test (new data)

Author (year) Country	Study design	Number of patients	Follow-up	Outcomes/end points	Evidence type	Funding
Bueno-de-Mesquita *et al.* (2009)[93] Netherlands	Consecutive cohort (1996–9) Fresh frozen tumour samples (pT1–2, LN–)	Eligible sample: NR Sample included: 123 G1: 64 low-risk prognostic signature G2: 59 high-risk prognostic signature	Median: 5.8 (range 0.1–9.0) years	Time from surgery to distant metastasis as first event (counted as failures); OS (defined as time from surgery to death)	Clinical validity; clinical utility – reclassification against existing prognostic variables	NR
Gevensleben *et al.* (2010)[94] Germany	Consecutive cohort (2005–8) Frozen tumour samples (evaluates concordance)	Eligible sample: 170 Sample included: 140 G1: 78 good prognosis signature G2: 62 poor prognosis signature Samples excluded because of inadequate RNA extraction ($n=30$)	NR	Comparison of risk prediction using the MammaPrint test with those of St Gallen criteria[95] and Adjuvant! Online	Clinical utility – reclassification against existing prognostic variables; changes in treatment recommendations	NR
Ishitobi *et al.* (2010)[96] Japan	Retrospective cohort (1998–2001) Frozen tumour samples	Eligible sample: 117 Sample included: 102 G1: 20 low-risk prognostic signature G2: 82 high-risk prognostic signature Samples excluded because of failure of microarray profiling ($n=15$)	Median: 7.1 (range 0.5–9.8) years	DMFS (not defined); correlation between the MammaPrint test risk category and clinicopathological parameters (St Gallen criteria[97,98])	Clinical validity; clinical utility – reclassification against existing prognostic variables	NR
Kok *et al.* (2010)[99] Netherlands	Two datasets: G1: 1985–94; G2: 1982–96 Adjuvant tamoxifen (G1): retrospective, frozen tumour samples No adjuvant systemic treatment (G2): consecutive series from van de Vijver[63] ($n=100$) and Mook *et al.*[64] ($n=51$), FFPE samples	Eligible sample: NR Sample included: 272 G1: 121 (83 low-risk, 38 high-risk prognostic signature) G2: 151 (85 low-risk, 66 high-risk prognostic signature)	Median: G1: 9.6 years; G2: 11.1 years	BCSS (defined as time from surgery to breast cancer-related death)	Clinical validity	NR
Kunz *et al.* (2011)[92] Germany	Prospective cohort (2004–8) Fresh tumour samples (T1–3, NO-3) (evaluates concordance)	Eligible sample: 56 Sample included: 44 Samples excluded because of insufficient sample ($n=6$); lost in transit, ($n=4$); not eligible because of metastases ($n=2$)	NR	Comparison of risk prediction using the MammaPrint test with that of the St Gallen guidelines[97] (2007/9) and Adjuvant! Online	Clinical utility – reclassification against existing prognostic variables	NR

continued

TABLE 13 Study design characteristics of included studies: MammaPrint test (new data) (*continued*)

Author (year) Country	Study design	Number of patients	Follow-up	Outcomes/end points	Evidence type	Funding
Mook *et al.* (2010)[75] Netherlands	Consecutive series (1984 –6) Frozen tumour samples (T1–2, LN–)	Eligible sample: 173 Sample included: 148 G1: 91 good prognosis signature G2: 57 poor prognosis signature Samples excluded because of insufficient sample ($n=22$); poor RNA quality, ($n=3$)	Median: 11.6 years	DMFS (defined as time from surgery to distant metastasis as first event: counted as failures); BCSS (defined as time from surgery to breast cancer-related death); comparison of risk prediction using the MammaPrint test with that of Adjuvant! Online	Clinical validity; clinical utility – reclassification against existing prognostic variables	European Commission Framework Program VI-TRANSBIG; Dutch National Genomics Initiative Cancer Genomics Center; Agendia BV
Na *et al.* (2011)[100] Republic of Korea	Retrospective cohort (2008–9) Fresh tumour samples (T1–2, LN–, M0) (evaluates concordance)	Eligible sample: 48 Sample included: 36 G1: 5 low-risk prognostic signature G2: 31 high-risk prognostic signature Samples excluded because of sampling failure ($n=10$); not eligible because of metastases ($n=2$)	NR	Comparison of risk prediction using the MammaPrint test with those of the St Gallen criteria,[95] National Institutes of Health guideline[101] and Adjuvant! Online	Clinical utility – reclassification against existing prognostic variables	NR

NR, not reported.

Most studies included a small number of participants (range 36–272). Follow-up was either short (< 10 years and in some cases < 5 years) or not reported for a number of studies.

Quality of included studies: MammaPrint

The methodological quality of the seven included studies[75,92–94,96,99,100] is summarised in *Figure 4* (further details are provided in *Appendix 9*). Generally, only two studies[75,93] performed well, receiving a positive assessment for at least 17 out of 21 methodological quality items. Although the majority of the studies (as reported by the authors) used a retrospective study design,[75,96,99,100] other potential sources of bias were generally related to the following domains: prognostic variable (inadequate reporting and justification of cut points used), statistical analysis (lack of statistical adjustment of all prognostic factors and inappropriate analysis of continuous predictor variables, for example categorising of continuous variables leads to loss of statistical power, and data-dependent categorisation leads to overoptimism)[30] and interventions subsequent to inclusion in the study (interventions were not described or standardised). In the majority of studies, the assessment of study quality was hampered by poor reporting of the following methodological items: length of follow-up of patients, whether or not the sample of patients was representative and assembled at an early point in the course of the disease and whether or not outcomes were fully defined and appropriate (including whether or not the outcome assessment was unbiased). Overall, the risk of bias from the seven included studies was judged to be moderate.

Results: MammaPrint

A summary of the clinical evidence on MammaPrint is presented in *Table 15*, followed by a narrative summary of each study. Full data extraction tables are provided in *Appendix 9*.

DOI: 10.3310/hta17440

TABLE 14 Patient characteristics of included studies: MammaPrint test (new data)

Author (year)	Age (years), mean (range)	LN status	ER status	Tumour size	Grade	HER2 status	Mean NPI score	Treatment
Bueno-de-Mesquita *et al.* (2009)[93]	47 (27–55)	All LN–	+/–: G1: 62/2; G2: 32/27	pT1 (≤ 2 cm): G1: 46; G2: 30 pT2 (2.1–5 cm): G1: 18; G2: 29 Mean: 2 cm; range 0.5–5 cm	I: G1: 18; G2: 2 II: G1: 35; G2: 18 III: G1: 11; G2: 39	+/–: G1: 2/61; G2: 7/52	NR	Adjuvant chemotherapy: G1: 1; G2: 17 Adjuvant endocrine therapy: G1: 9; G2: 5 Both: G1: 9; G2: 4 None: G1: 45; G2: 33
Gevensleben *et al.* (2010)[94]	NR	+/–: G1: 24/54; G2: 22/40	+/–: G1: 77/1; G2: 39/23	≤ 1 cm: G1: 4; G2: 3 > 1 to ≤ 2 cm: G1: 39; G2: 18 > 2 to ≤ 5 cm: G1: 34; G2: 38 > 5 cm: G1: 1; G2: 3	NR	+/–/unknown: G1: 3/74/1; G2: 6/55/1	NR	Adjuvant systemic therapy: 134 (chemotherapy: 23; endocrine therapy: 59; both: 52); no treatment: 2; unknown: 4
Ishitobi *et al.* (2010)[96]	NR (but < 70 years)	All LN–	+/–: G1: 19/1; G2: 33/38[a]	≤ 2 cm: G1: 9;[a] G2: 40 > 2 cm: G1: 9;[a] G2: 42	I: G1: 11; G2: 9[a] II: G1: 6; G2: 23[a] III: G1: 3; G2: 49[a]	NR	NR	Chemotherapy: G1: 2; G2: 27 Hormone therapy: G1: 17; G2: 57
Kok *et al.* (2010)[99]	NR	G1: LN–: 20; N1–3: 74; N > 3: 20; unknown: 7 G2: LN–: 138; N1–3: 10; N > 3: 3	All ER+	≤ 2 cm: G1: 55; G2: 96 > 2 cm: G1: 65; G2: 55 Unknown: G1: 1	I: G1: 36; G2: 52 II: G1: 63; G2: 54 III: G1: 18; G2: 45 Unknown: G1: 1	NR	NR	G1: adjuvant tamoxifen monotherapy (about 70% for at least 2 years) and no neoadjuvant therapy G2: no adjuvant systemic treatment
Kunz *et al.* (2011)[92]	44 (32–56)	LN–: 27 N1: 16 N2/N3: 3	+/–: 36/8 (intermediate: 2)	pT1: 26 (pT1b: 2; pT1c: 24) pT2: 19 pT3: 1 Mean: 2.06 cm	I: 18 II: 18 III: 10	+/–: 5/41	NR	Chemotherapy: 32; no chemotherapy: 14; peritumoural invasion: 14; no invasion: 32
Mook *et al.* (2010)[75]	NR (55–70)	All LN–	+/–: G1: 88/3; G2: 28/29	pT1 (≤ 2 cm): G1: 59; G2: 24 pT2 (> 2 to 5 cm): G1: 32; G2: 33	I: G1: 52; G2: 3 II: G1: 28; G2: 15 III: G1: 11; G2: 39	NR	NR	Adjuvant endocrine (tamoxifen) therapy: G1: 17; G2: 10 Note: inclusion criteria specified no adjuvant chemotherapy
Na *et al.* (2011)[100]	47 (23–68)	All LN–	+/–: G1: 4/1; G2: 25/6	T1 (≤ 2 cm): G1: 3; G2: 20 T2 (> 2 to 5 cm): G1: 2; G2: 11 Mean: 2.0 cm	I: G1: 3; G2: 4 II: G1: 2; G2: 15 III: G1: 0; G2: 12	+/–: G1: 1/4; G2: 7/24	NR	No neoadjuvant treatment. No other details provided

NR, not reported.
a Group with missing data.

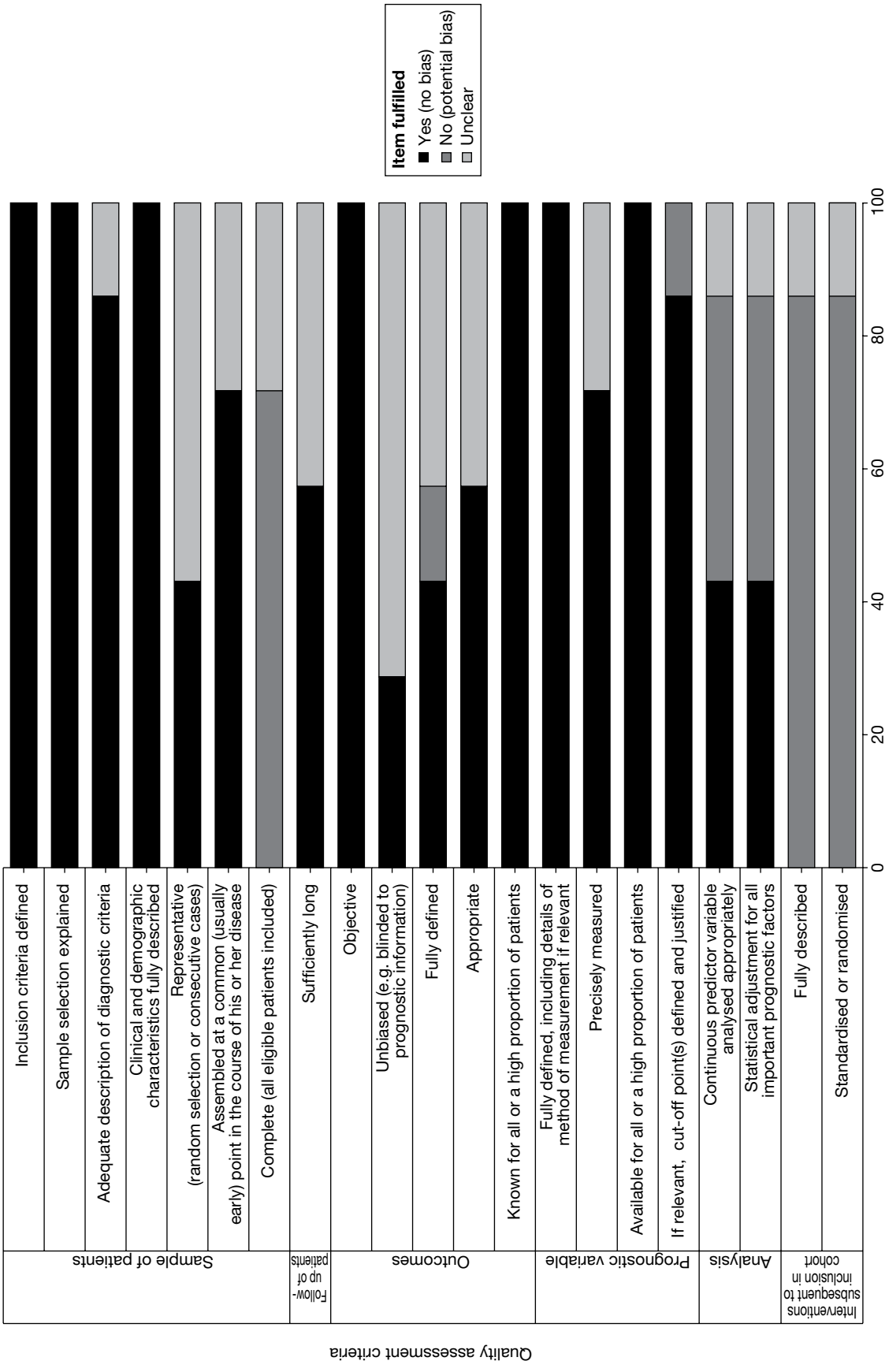


FIGURE 4 MammaPrint test: methodological quality graph (review authors' judgements about each methodological quality item presented as percentages across all studies).

TABLE 15 Summary of evidence for the MammaPrint test

Author (year)	Evidence type	Overall quality	Key findings
Bueno-de-Mesquita *et al.* (2009)[93]	Clinical validity; clinical utility – reclassification against existing prognostic variables	High	123 LN–, majority ER+ patients from the Netherlands. The rates of discordance between MammaPrint and four different standard clinicopathological measures were relatively high (38%, 41%, 26%, 30%). OS probability was 97% for good and 82% for poor prognosis signature patients with an estimated HR of 3.4 (95% CI 1.2 to 9.6; $p=0.021$). The probability of remaining free of distant metastasis (as first event) was 98% for good and 78% for poor prognosis signature patients with an estimated HR of 5.7 (95% CI 1.6 to 20; $p=0.007$). MammaPrint was shown to be a strong independent prognostic factor in multivariate analyses, outperforming the clinicopathological risk indexes. Limitations: small sample size, follow-up limited to 5 years
Gevensleben *et al.* (2010)[94]	Clinical utility – reclassification against existing prognostic variables; changes in treatment recommendations	Moderate	140 LN– and ER+ patients from Germany. MammaPrint and Adjuvant! Online were concordant in 83 cases and discordant in 57 cases (41%). A retrospective analysis of treatment given (where available) compared with treatment indication by MammaPrint was performed showing that, according to MammaPrint, 40% of patients had been either undertreated or overtreated. Limitation: small sample size
Ishitobi *et al.* (2010)[96]	Clinical validity; clinical utility – reclassification against existing prognostic variables	Moderate	102 LN–, majority ER+ patients from Japan. NPV for time to distant metastasis was high (100%), indicating that all patients were correctly classified, whereas PPV was low (9.8%), indicating that many of the cases classified as high risk were incorrectly classified. The relatively young patient population and 5-year follow-up may also explain why the probability of DMFS was also very high for the high-risk group. Limitations: small sample size, particularly in the low-risk group, findings may not be generalisable to the UK setting
Kok *et al.* (2010)[99]	Clinical validity	Moderate	272, all ER+ patients from the Netherlands. Inpatients treated with adjuvant tamoxifen (mainlyLN+), both MammaPrint and the endocrine response categories were associated with BCSS at 10 years. Inpatients treated with tamoxifen, combined analysis of MammaPrint and ER/PR revealed additional value. Inpatients who did not receive tamoxifen, only MammaPrint was associated with outcome. Limitation: small sample size
Kunz *et al.* (2011)[92]	Clinical utility – reclassification against existing prognostic variables	Moderate	44 LN– and majority ER+ patients from Germany. Comparison of numbers of patients classified into risk groups using St Gallen,[97] Adjuvant! Online and MammaPrint. The authors concluded that gene expression analysis as an additional tool can accurately separate patients with an intermediate clinical risk into low- and high-risk groups. Limitation: very small sample size
Mook *et al.* (2010)[75]	Clinical validity; clinical utility – reclassification against existing prognostic variables	High	148 ER+, LN–, postmenopausal patients from the Netherlands. Distant metastasis-free survival at 5 years was 93% in the low-risk group and 72% in the high-risk group ($p=0.07$) with an associated HR of 4.6 (95% CI 1.8 to 12.0; $p=0.001$). At 10 years the difference was not significant. BCSS at 5 years was 99% in the low-risk group and 80% in the high-risk group ($p=0.036$) with an associated HR of 19.1 (95% CI 2.5 to 148; $p=0.005$). At 10 years the difference was not significant. MammaPrint and Adjuvant! Online were concordant in 107 cases and discordant in 41 cases. The authors concluded that the MammaPrint signature can accurately select postmenopausal patients at low risk of breast cancer in terms of related death within 5 years of diagnosis and can be of clinical use in selecting postmenopausal women for adjuvant chemotherapy. Limitations: small sample size, assessed only postmenopausal women
Na *et al.* (2011)[100]	Clinical utility – reclassification against existing prognostic variables	Moderate	36 LN–, majority ER+ patients from Republic of Korea. Clinical risk concordant with the prognostic signature for 29 (81%) patients according to the St Gallen guidelines;[95] 30 (83%) patients according to the National Institutes of Health guidelines and 23 (64%) patients according to Adjuvant! Online. Limitations: very small sample size, no follow-up data, may not be generalisable to the UK setting

Analytical validity

Our searches did not reveal any studies that examined analytical validity.

Clinical validity (prognostic ability)

Kok *et al.*[99] assessed whether or not analysing both MammaPrint score and hormone receptors provides superior prediction of outcome than hormone receptors alone in 272 Dutch patients. One group comprised LN+, ER+, tamoxifen-treated patients and a second group comprised LN–, ER+ patients who had received no adjuvant systemic treatment. Hormone receptors were evaluated using the St Gallen consensus recommendations[102] (highly endocrine responsive: ER and PR ≥ 50%; incompletely endocrine responsive: ER and/or PR low or with either one absent). In patients treated with adjuvant tamoxifen (mainly LN+), both MammaPrint (adjusted for endocrine response categories, HR 2.78; 95% CI 1.30 to 5.94) and the endocrine response categories (adjusted for MammaPrint, HR 7.22; 95% CI 2.17 to 24.0) were associated with BCSS at 10 years. Also, in patients treated with tamoxifen for metastatic disease, combined analysis of MammaPrint and ER/PR revealed additional value (multivariate Cox regression; $p = 0.013$). In patients who did not receive tamoxifen, only MammaPrint was associated with outcome. The authors concluded that both methods provide independent information on outcome after tamoxifen for LN+ breast cancer. There are a number of limitations to this study: the second patient group comprised patients included in two previously reported evaluations, the overall sample size was small and tumour samples had been collected over a number of years (1982–97), which has implications for changes in diagnosis, treatment and standards of care. The study did benefit from an adequate follow-up time of 10 years.

Ishitobi *et al.*[96] examined risk classification using MammaPrint and disease outcome for 102 LN–, majority ER+, relatively young breast cancer patients in Japan. The results relating to clinical validity are presented here and the results relating to clinical utility are presented in the relevant section below. Among the 102 patients, 20 (20%) were classified as low risk and 82 (80%) as high risk. The authors reported that the probability of DMFS at 5 years was 100% for the low-risk group and 94% for the high-risk group. They did not report a HR. The NPV for time to distant metastasis was high (100%, 20/20), whereas the PPV was quite low (9.8%, 8/82). The NPV indicates the proportion of patients classified as low risk who were correctly classified using MammaPrint, whereas the small PPV indicates that many of the cases classified as high risk were incorrectly classified. The authors concluded that the 70-gene prognosis signature accurately identified Japanese breast cancer patients as being at low risk of developing recurrences, as 100% of the individuals in the low-risk group remained metastasis free for the duration of the observation period. The authors suggest that the low number of individuals in the low-risk group is consistent with previous findings on patient groups of ≤ 54 years. However, these low numbers make any generalisations of the findings limited. The young patient population may also explain why the probability of DMFS was also very high for the high-risk group, together with the fact that this was assessed at only 5 years, given that the majority of distant recurrences and deaths from breast cancer occur > 5 years after diagnosis. This study employed a very small sample size and, furthermore, as this study was performed in a Japanese population any generalisations to the UK population are significantly limited.

In a Netherlands-based study, Bueno-de-Mesquita *et al.*[93] assessed 123 LN–, majority ER+ patients who had been assigned MammaPrint risk categories. They reported risk classification and probability of disease outcome (time from surgery to distant metastasis and OS). OS probability was 97% (±2%) for good and 82% (±5%) for poor prognosis signature patients (p-value not reported) with an estimated HR of 3.4 (95% CI 1.2 to 9.6; $p = 0.021$). The probability of remaining free of distant metastasis (as first event) was 98% (±2%) for good and 78% (±6%) for poor prognosis signature patients (p-value not reported) with an estimated HR of 5.7 (95% CI 1.6 to 20; $p = 0.007$). In multivariate analysis, the authors demonstrated that MammaPrint

was a strong independent prognostic factor, outperforming the clinicopathological risk indexes. They concluded that the 70-gene prognosis signature is also an independent prognostic factor in LN– breast cancer patients for women diagnosed in recent years. Again, as this study used a small sample size and the follow-up assessment was limited to 5 years, generalisations of the findings are limited. This study also reported reclassification findings, which are detailed in the relevant section below.

Mook *et al.*[75] examined 148 LN– and majority ER+, specifically postmenopausal patients. The study, conducted in the Netherlands, investigated disease outcome (DMFS and BCSS at 5 years), and prediction of early breast cancer-specific death (BCSD) using MammaPrint risk categories. The authors also assessed reclassification and these findings will be presented below. DMFS at 5 years was 93% in the low-risk group and 72% in the high-risk group ($p = 0.07$) with an associated HR of 4.6 (95% CI 1.8 to 12.0; $p = 0.001$). At 10 years it was 80% in the low-risk group and 67% in the high-risk group (HR not reported, *p*-value not reported). Over the entire follow-up period the HR was 1.8 (95% CI 0.9 to 3.5; $p = 0.07$). BCSS at 5 years was 99% in the low-risk group and 80% in the high-risk group ($p = 0.036$) with an associated HR of 19.1 (95% CI 2.5 to 148, $p = 0.005$). At 10 years it was 90% for the low-risk group and 69% for the high-risk group (HR not reported, *p*-value not reported). Over the entire follow-up period the HR was 2.0 (95% CI 1.0 to 4.0; $p = 0.04$). In terms of the prediction of early BCSD, the HR for BCSS at 5 years was 14.4 (95% CI 1.7 to 122.2; $p = 0.01$) and at 10 years was 4.4 (95% CI 1.4 to 13.6; $p = 0.01$). Subgroup analyses showed that the HR for BCSS in hormonal therapy-naive patients (untreated) at 5 years was 10.8 (95% CI 1.2 to 94.7; $p = 0.03$). The authors concluded that the MammaPrint signature can accurately select postmenopausal patients at low risk of breast cancer in terms of related death within 5 years of diagnosis, although not at 10 years, and can be of clinical use in selecting postmenopausal women for adjuvant chemotherapy. Again this study employed a very small sample size and was based on postmenopausal women, limiting the applicability of the findings.

Clinical utility

Reclassification against existing prognostic variables Kunz *et al.*[92] compared the MammaPrint result with St Gallen criteria[97] and Adjuvant! Online and conducted risk assessment using MammaPrint according to nodal status in 44 women in Germany. The majority of patients were LN– and ER+. MammaPrint classified 29 patients as low risk and 15 patients as high risk. St Gallen criteria classified four patients as low risk, 34 patients as intermediate risk and six patients as high risk. In the group of women with intermediate risk according to St Gallen, MammaPrint assigned 23 patients to low risk and 11 to high risk. Adjuvant! Online classified 19 patients as low risk and 25 patients as high risk (for Adjuvant! Online, patients were classified as having low clinical risk when the 10-year OS rate as predicted by Adjuvant! Online was > 88% for ER+ tumours and > 92% for ER– tumours). MammaPrint classified 13 patients with LN+ disease as low risk and five as high risk. For those with LN– disease, 17 were classified as low risk and nine as high risk. The authors concluded that, by using gene expression analysis as an additional tool, patients with an intermediate clinical risk can be accurately separated into low- and high-risk groups. Follow-up data would be required to verify the authors' conclusions that the gene expression analysis provides more accurate information on recurrence risk than conventional clinicopathological criteria. This was a reasonable quality study with a prospective design although the interpretation of the findings is limited because of the very small sample size. Studies on larger sample sizes would be required to verify these conclusions.

Na *et al.*[100] compared MammaPrint with the St Gallen criteria,[95] the National Institutes of Health (NIH) guidelines[101] and Adjuvant! Online in 36 LN– and majority ER+ Korean patients. The 70-gene prognosis signature identified 5 (14%) patients with a low-risk prognosis signature

and 31 (86%) patients with a high-risk prognosis signature. Clinical risk was concordant with the prognostic signature for 29 (81%) patients according to the St Gallen guidelines, 30 (83%) patients according to the NIH guidelines and 23 (64%) patients according to Adjuvant! Online. The authors concluded that the 70-gene prognostic signature gave somewhat different results in Korean patients with breast cancer from those in European patients. They suggested that further studies should assess whether or not a gene disparity between Asians and Europeans influenced the results. Further large-scale studies with a follow-up evaluation are required to assess whether or not the use of the 70-gene prognostic signature can predict the prognosis of Korean patients with breast cancer. (Note that as St Gallen has three risk categories and MammaPrint has two, a calculation of concordance is not possible.) This study had a very small sample size and, as the results could have been influenced by a gene disparity between European and Asian patients, any generalisations to the UK population are significantly limited. Furthermore, as there was no follow-up evaluation, no conclusions regarding the prognostic value of MammaPrint compared with clinicopathological guidelines can be made.

Ishitobi *et al.*[96] examined risk classification using MammaPrint and disease outcome for breast cancer in 102 LN–, majority ER+ patients in Japan. The results relating to clinical validity have been presented in the previous section. Among the 102 patients, 20 (20%) were classified as low risk and 82 (80%) as high genomic risk. Based on the 1998 St Gallen criteria,[103] all patients were classified as intermediate or high risk. The 2009 St Gallen criteria[97] use more refined criteria to define the low-risk group and they classified 7 (of 100) patients as low risk compared with 20 (of 102) patients identified as low risk by MammaPrint ($p = 0.009$). The authors concluded that the 70-gene prognosis signature accurately identified Japanese breast cancer patients at low risk of developing recurrences, as 100% of the individuals in the low-risk group remained metastasis free for the duration of the observation period. Overall, this study employed a very small sample size; furthermore, as this study was performed using a Japanese population, any generalisations to the UK population are significantly limited.

Bueno-de-Mesquita *et al.*[93] made a comparison between MammaPrint risk categories and risk assessment based on Adjuvant Online!, St Gallen guidelines,[103,104] NPI and Dutch Institute for Healthcare Improvement (CBO) guidelines (2004)[105,106] in 123 LN– and majority ER+ Dutch patients. Discordance between the measures was 38% (47/123), 41% (50/123), 26% (32/123) and 30% (37/123) respectively. These rates of discordance appear relatively high although we do not know which predictor is classifying correctly. Again, as this study used a small sample size and the follow-up assessment was limited to 5 years, generalisations of the findings are limited.

Mook *et al.*[75] examined 148 LN– and majority ER+, specifically postmenopausal patients. The study, conducted in the Netherlands, investigated classification using MammaPrint and disease outcome (the results of the latter are presented in the earlier section). MammaPrint risk categories of high and low were compared with Adjuvant! Online categories of high and low (low clinical risk: predicted 10-year BCSS > 88% for ER+ tumours and > 92% for ER– tumours). MammaPrint and Adjuvant! Online were concordant in 107 cases and discordant in 41 cases, although again we cannot make any assertions regarding which indicator is more accurate. The authors concluded that the 70-gene prognosis signature can accurately select postmenopausal patients at low risk of breast cancer. Again, these findings were based on a very small sample size and assessed only postmenopausal women, limiting the applicability of the findings to younger women.

A German-based study[95] investigating 140 majority LN– and ER+ patients was reported by Gevensleben *et al.*[94] The authors made a comparison between risk prediction using the 70-gene prognostic signature and risk prediction using the St Gallen criteria[95] and Adjuvant! Online (Adjuvant! Online risk classification according to Ravdin *et al.*[107]). MammaPrint and Adjuvant!

Online were concordant in 83 cases and discordant in 57 cases (41%). The authors concluded that MammaPrint provides improved prediction of recurrence risk compared with currently used guidelines. The generalisability of the findings is limited because of the small sample size employed.

Changes in treatment recommendations Gevensleben *et al.*[94] in their study investigating 140 LN– and ER+ German patients, also examined treatment advice. For 59 patients (out of 62) with poor prognosis identified by the 70-gene prognosis signature, the clinical treatment was recorded. In total, 19 (32%) of these patients did not receive adjuvant systemic treatment other than endocrine therapy and were potentially undertreated. In contrast, 35 (out of 77) patients who were classified as having a good prognosis by the 70-gene prognosis signature, and for whom treatment was known, received chemotherapy and were potentially overtreated. As a result, the authors concluded that the 70-gene prognosis signature would have resulted in altered treatment advice in 40% of the patients, based on the assumption that all high-risk patients would receive chemotherapy and all low-risk patients would not. There are limitations to this study, including that it was based on a small sample of patients and that the data relating to changes in treatment recommendations are retrospective and relate only to potential and not actual changes.

Supplementary evidence: MammaPrint

Four further citations[108–111] were excluded from the review as they did not meet the inclusion criteria on the basis that they are pooled analyses of existing data. This evidence was not in the form of meta-analyses of the separate studies. This approach is methodologically inappropriate. However, because of the paucity of data on the clinical utility of MammaPrint, a number of these studies have been used to inform the economic model and therefore they will be summarised here for completeness.

These studies suffer from several major limitations. First, they are based on pooled analyses and it is unclear how individual patient data have been combined from the various primary studies incorporated. Furthermore, there is likely to be significant heterogeneity in the chemotherapy used, standards of care and diagnosis as patients were recruited over a period of >20 years (1984–2006). This makes any generalisability of the conclusions to current practice difficult.

Knauer *et al.*[112] evaluated the predictive value of the 70-gene prognostic signature for response to chemotherapy. They created a pooled database of patients from six previous studies, including 541 women with unilateral stage T1–3, N-1, M0 invasive breast cancer diagnosed between 1984 and 2006. Each tumour had been classified as having a high- or low-risk signature using the MammaPrint test: 252 (47%) as low risk and 289 (53%) as high risk. Median follow-up was 7.1 years, but all analyses were censored at 5 years. The signature was prognostic: women with a low-risk tumour signature had a 5-year BCSS of 97% and a 5-year DMFS of 95% whereas women with a high-risk tumour signature had a 5-year BCSS of 87% and a 5-year DMFS of 82%. However, women in both risk categories appeared to benefit from chemotherapy, although the estimates were not statistically significant in the low-risk group. For BCSS the unadjusted HR for chemotherapy was 0.58 (95% CI 0.07 to 5.0) in the low-risk group and 0.21 (95% CI 0.07 to 0.59) in the high-risk group. The p-value for interaction between use of chemotherapy and the risk signature was not statistically significant ($p = 0.45$). For DMFS the unadjusted HR for chemotherapy was 0.26 (95% CI 0.03 to 2.0) in the low-risk group and 0.35 (95% CI 0.17 to 0.71) in the high-risk group. The p-value for the interaction was not reported, but in this case the trend was towards greater relative benefit from chemotherapy in the low-risk group. This study, however, has some major statistical flaws. For instance, data were truncated arbitrarily at 5 years, despite that fact that the median follow-up was 7.1 years. Censoring the follow-up at 5 years biased the results in favour of the utility of the prognostic signature because the association between the 70-gene signature and recurrent disease falls quickly after 5 years of follow-up.[113] As

the majority of distant recurrences and deaths from breast cancer occur > 5 years after diagnosis, this is a significant limitation.[113]

Knauer *et al.*[110] investigated whether or not MammaPrint identifies HER2-positive patients with a favourable outcome. A total of 168 T1–3, N-1, HER2-positive patients were identified from a pooled database, classified by the MammaPrint test as having a good or a poor prognosis, and correlated with long-term outcome. A total of 89 of these patients did not receive adjuvant chemotherapy. In these patients, after a median follow-up of 7.4 years, 35 (39%) distant recurrences and 29 (33%) BCSDs occurred. The test classified 20 (22%) patients as having a good prognosis, with 10-year distant disease-free survival (DDFS) of 84%, compared with 69 (78%) poor prognosis patients with a 10-year DDFS of 55%. The estimated HRs were 4.5 (95% CI 1.1 to 18.7, $p = 0.04$) and 3.8 (95% CI 0.9 to 15.8, $p = 0.07$) for DDFS and BCSS respectively. In multivariate analysis adjusted for known prognostic factors and hormone therapy, HRs were 5.8 (95% CI 1.3 to 26.7, $p = 0.03$) and 4.7 (95% CI 1.0 to 21.7, $p = 0.05$) for DDFS and BCSS respectively. The authors concluded that the test is an independent prognostic indicator that identified a subgroup of HER2-positive early breast cancer patients with a favourable long-term outcome.

Mook *et al.*[108] aimed to evaluate the accuracy of MammaPrint in T1 breast cancer. They selected 964 patients with pT1 tumours (≤ 2 cm) from a pooled database. The samples had been previously analysed and classified as having good or poor prognosis. The median follow-up was 7.1 years. The 10-year DMFS and BCSS probabilities were 87% [standard error (SE) 2%] and 91% (SE 2%), respectively, for the good prognosis group ($n = 525$) and 72% (SE 3%) and 72% (SE 3%), respectively, for the poor prognosis group ($n = 439$). The signature was an independent prognostic factor for BCSS at 10 years (multivariate HR 3.25; 95% CI 1.92 to 5.51; $p < 0.001$]). Moreover, the test predicted DMFS at 10 years for 139 patients with pT1ab cancers (HR 3.45; 95% CI 1.04 to 11.50; $p = 0.04$). The authors concluded that the test is an independent prognostic factor in patients with pT1 tumours and can help to individualise adjuvant treatment recommendations.

Bueno-de-Mesquita *et al.*[109] evaluated the additional prognostic value of MammaPrint compared with a combination of established prognostic guidelines. A total of 701 patients from a pooled database with an existing MammaPrint result were evaluated. Only 6% (10/156) of ER– tumours had a good prognosis signature. The signature was not useful for ER+ tumours with a concordant high Adjuvant! Online, high St Gallen and/or high NPI risk ($n = 139$). The 10-year OS estimate for the good prognosis group with these characteristics was < 80% and adjuvant systemic treatment (AST) would therefore be appropriate irrespective of the signature result. In contrast, for patients with a concordant low Adjuvant! Online, low St Gallen and/or low NPI risk and in discordant clinical risk patients, the signature identified low-risk patients in whom AST could be safely withheld (10-year OS < 90%). The authors concluded that the MammaPrint signature provides additional prognostic information, especially in ER+, LN– breast cancer patients with a predominantly low or discordant clinical risk on the basis of Adjuvant! Online, the St Gallen guidelines and/or NPI.

Summary of evidence: MammaPrint

Analytical validity of MammaPrint

In the earlier systematic reviews limited data are available on variability and reproducibility, with a limited number of patients and a moderate number of replications. The only validation study using the MammaPrint assay (rather than the underlying 70-gene signature) showed that only about 80% of fresh-frozen specimens were analysable.

Our review identified no new evidence.

Clinical validity (prognostic ability) of MammaPrint

Earlier systematic reviews identified a range of studies providing evidence on the prognostic ability of the test in heterogeneous populations. The evidence relating to the clinical validity of MammaPrint was not always conclusive nor supportive of the prognostic value of the test. Four studies suggested that the test could predict prognosis, one study failed to verify the prognostic utility of the test and in another the methods and results were at variance with those of other studies.

The current review identified four additional studies that contain data on clinical validity. Of these, two were rated as high quality and two as moderate quality. These studies demonstrated that the MammaPrint score is a strong independent prognostic factor and may provide additional value to standard clinicopathological measures. The majority of the evidence suggests that the test is reliable at predicting outcome at 5 years.[74] However, the population in all of these studies was relatively small, ranging between 102 and 272 patients. One of the studies was conducted in a Japanese population, making generalisation to UK practice difficult. Follow-up was limited to only 5 years in two of the studies.

Clinical utility of MammaPrint

Earlier systematic reviews identified one study on clinical utility, which demonstrated that MammaPrint had an impact on clinical decision-making. The addition of MammaPrint to the standard Dutch clinical assessment of risk (modified by patient preference) in a cohort of 427 patients increased the number of patients receiving adjuvant systemic therapy by 20. However, the follow-up was not long enough to provide evidence of its effect on clinical end points such as DMFS or its utility in predicting treatment benefit.

The current review identified six studies that contained data on the clinical utility of MammaPrint. Of these, two were rated as high quality and four as moderate quality. Five of the six studies reported on how the MammaPrint test reclassifies patients into high- and low-risk groups compared with the risk assigned in current practice. None of these was based in a UK setting. These studies reported that there was a high level of discordance between MammaPrint and current practice, although the studies did not demonstrate how this would impact on treatment decisions. One study reported that the use of MammaPrint would result in altered treatment advice for 40% of patients, but this was based on the assumption that all patients classified as high risk would receive chemotherapy and no patients classified as low risk would receive chemotherapy rather than by providing evidence of actual changes in practice. A study on the benefits of chemotherapy by MammaPrint risk group was identified but omitted from the systematic review because it was based on a pooled analysis of six primary studies (which were included in the review in their own right). This study reports findings on chemotherapy benefit for MammaPrint high-and low-risk groups but the findings are not considered to be robust as the authors do not reanalyse the tumour samples and it is unclear how individual patient data were combined. All of the studies on clinical utility were based on small sample sizes.

Key gaps in the evidence base remain:

- Robust evidence of clinical utility is needed. It is not yet clear whether or not the use of the MammaPrint test will change the management of patients, for example reduce the number of patients unnecessarily treated with chemotherapy or improve patient outcomes through increases in DFS and OS. The ongoing Microarray In Node-negative Disease may Avoid ChemoTherapy (MINDACT) trial, which is summarised in *Ongoing randomised trial: the Microarray In Node-negative Disease may Avoid ChemoTherapy trial* (see *Appendix 8* for further detail of the MINDACT trial), will provide this evidence.

- Only two studies were considered to be of high quality. The rest of the studies in the current review were judged to provide moderate-quality (although retrospective) evidence for MammaPrint. All of the included studies employed very small sample sizes. One of the most characteristic features of the studies was their heterogeneity, particularly with respect to patient populations. All but one[92] of the MammaPrint studies were retrospective, and many used old archived tumour samples and non-standardised methods of patient selection. In addition to patient heterogeneity, there is likely to be significant heterogeneity in the chemotherapy treatment as patients were diagnosed with breast cancer over a period of >20 years (1984–2006) and the standards of care have changed considerably.
- Further issues that may limit the extent to which we can generalise the findings include publication bias and the fact that no studies were conducted in the UK setting.

Overall summary

The evidence base, in particular in relation to the prognostic ability of the test, is developing but is based on small sample sizes (≤272). None of the studies used UK-based patients and the data were all based on cohort studies. The test appears to be prognostic at 5 years although the validity of the test to predict longer-term outcomes does not seem to have been established. Robust evidence of clinical utility is needed as it is not yet clear to what extent the use of the MammaPrint test will change the management of patients and to what extent chemotherapy would be offered to patients classified as having a good or a poor prognosis with MammaPrint. It is also unclear to what extent MammaPrint risk groups are predictive of chemotherapy benefit or how the use of MammaPrint will improve patient outcomes through increases in DFS and OS. The evidence for MammaPrint to date is mainly derived from premenopausal women, but younger women are more likely to be classified as having a poor prognosis using MammaPrint, which might overestimate the benefit of the test.

Ongoing randomised trial: the Microarray In Node-negative Disease may Avoid ChemoTherapy trial

The MINDACT trial started recruiting patients in 2006 and has an estimated completion date of 2019. It is a partially randomised, open-label, prospective, multicentre clinical trial that aims to assess the value of the 70-gene prognostic signature in predicting which patients would benefit from chemotherapy compared with Adjuvant! Online, which is based on clinical characteristics. Women >18 years (the upper age limit of 70 years was recently removed) with histologically confirmed unilateral invasive breast cancer, T1–3 operable disease, up to three positive lymph nodes and no distant metastases are eligible for enrolment. The target for enrolment recently increased to 6600 from 6000. This is predicted to result in 55% of patients assessed as high risk by both methods, who will go on to have chemotherapy, and 13% of patients assessed as low risk by both methods, who will go on to have no chemotherapy. The 32% who are assessed as high risk by one method and low risk by the other will then be randomised to follow the treatment indicated by MammaPrint or the treatment indicated by Adjuvant! Online. Two further objectives of the trial relating to the efficacy of different chemotherapy agents and endocrine treatment strategies are addressed by two further stages of randomisation. Regardless of previous randomisation and risk categorisation, patients who are to receive chemotherapy are randomised to docetaxel or capecitabine regimens and patients who are hormone receptor-positive are randomised to a single-agent upfront aromatase inhibitor (letrozole) for 7 years or tamoxifen for 2 years followed by letrozole for 5 years. The primary outcome measures are DMFS at 5 years and DFS, followed up for a minimum of 15 years. As of October 2012, the study had enrolled 6700 patients. Further details of this trial are included in *Appendix 8*.

MammaPrint and BluePrint

BluePrint is used in addition to MammaPrint for molecular subtyping. It is an 80-gene microarray and the target population is patients with early-stage (stage I or II), LN– or LN+

(up to three nodes), ER+ or ER– breast cancer. BluePrint provides information on breast cancer subtype using three categories: basal-type, luminal-type and *ERBB2*-type cancers.

Description of included studies

The searches did not identify any full peer-reviewed papers relating to the BluePrint test; however, one meeting abstract by Stork-Sloots *et al.*[114] met the inclusion criteria. This study related to clinical validity, the design was retrospective and the sample size was moderate ($n = 469$). No further details of the design or the study populations were reported.

Quality of the included study: MammaPrint and Blueprint

Although the assessment of study quality was hindered by poor reporting in the domains of outcome, prognostic variable, analysis and interventions subsequent to inclusion in the cohort, the overall methodological quality of the included study[114] was judged to be low, indicating a high risk of bias (the study received a positive assessment of at least two out of 21 methodological quality items).

Results: MammaPrint and Blueprint

Full data extraction tables are provided in *Appendix 10*.

Analytical validity

No available evidence.

Clinical validity

Stork-Sloots *et al.*[114] compared BluePrint directly with the intrinsic subtyping using the original intrinsic gene set as developed by Perou *et al.*,[26] from which the PAM50 gene set has originated. Using 469 independent samples and two publicly available data sets ($n = 215$, $n = 159$), the authors reported 5-year survival estimates for the subtypes and for the groups further separated by high- and low-risk MammaPrint categories. They reported that samples with a *ERBB2*-like or basal-like gene profile showed equally poor 5-year survival rates of ~65%. However, the *ERBB2*-like subset of MammaPrint low-risk patients (15%) showed an 89% (95% CI 71% to 100%) survival rate without trastuzumab treatment. When the luminal-like subtype was separated into high and low risk by MammaPrint the survival rate was 56% (95% CI 46% to 68%) for high-risk luminal-like samples and 94% (95% CI 90% to 99%) for low-risk luminal-like samples. The authors concluded that the developed multigene profile can classify breast tumours into luminal-, *ERBB2*- and basal-like subgroups. By combining this molecular subtyping with MammaPrint risk classification, specific groups of patients can be recognised that are at high risk of recurrence. The low-risk patients within the luminal- and *ERBB2*-like subclasses have a very low risk of recurrence. There are significant limitations in making any interpretations from this evidence as it is derived only from an abstract. It has been shown that there may be discrepancies between data made available in abstracts and the reporting of results in subsequently published full-length articles.[89] Because of incomplete reporting the methodological quality of studies cannot be confidently assessed by systematic reviewers.

Clinical utility

No available evidence.

Summary of evidence: MammaPrint and BluePrint

Because of the limited available data (reported in abstract form only), no firm conclusions can be drawn about the clinical validity (prognostic ability) of the MammaPrint and BluePrint test, although it does appear that the test has been validated in an independent cohort. No published evidence is yet available on the clinical utility of the test in combination with MammaPrint. Further evidence for this test is required.

PAM50 test

The PAM50 gene expression assay identifies the major intrinsic biological subtypes of breast cancer. The current version of the test provides classification of breast cancer subtype and quantitative values for (gene/protein) *ESR1*/ER, *PGR*/PR, *ERBB2*/HER2, proliferation score and luminal score (ER pathway). The current version does not provide a risk of recurrence score or category. The PAM50 Breast Cancer Intrinsic Classifier test is recommended for all patients diagnosed with invasive breast cancer, regardless of stage or ER status. Further details are provided in *Table 6*.

Description of included studies

The searches identified six studies for the PAM50 test. This included two full peer-reviewed papers,[115,116] three meeting abstracts[117–120] and an unpublished manuscript provided by the manufacturer.[121]

The design and patient characteristics of the six included studies are provided in *Tables 16* and *17* respectively. All of the reported studies had a retrospective design analysing archived tumour samples. The populations used in the studies were somewhat heterogeneous, although in most studies more patients were LN+ and ER+ than LN– and ER–. The ages of the patients varied across the studies, from a median age of 47.5 years in one paper[121] to a median of 67 years in another.[116] Most studies included a moderate number of tumour samples. Follow-up was not reported for a number of studies but was around a median of 10 years in those that did report a follow-up time.

Quality of included studies: PAM50

The methodological quality of the six included studies (seven citations)[115–121] is summarised in *Figure 5* (further details are provided in *Appendix 11*). Of these, two studies performed well,[117,121] receiving a positive assessment for at least 17 out of 21 methodological quality items. Although all studies used a retrospective study design, other potential sources of bias were generally related to the following domains: sample of patients (all eligible patients were not included), outcomes (not fully defined) and interventions subsequent to inclusion in the cohort (interventions were not fully described or standardised). Overall, the risk of bias from these studies was judged to be moderate.

Results: PAM50

A summary of the clinical evidence for PAM50 is presented in *Table 18* followed by a narrative summary of each study. Full data extraction tables are provided in *Appendix 11*.

Analytical validity

Ebbert *et al.*[118] reported an analytical validity study using 171 tumour samples from US patients. They reported that within-platform cross-validation of the clinical subtype predictor showed 91.6% concordance. There was 100% reproducibility in subtype predictions across 46 runs testing different subtypes. Subtype predictions across platforms showed 88.1% concordance. Dilution experiments, introducing 'normal' breast tissue RNA into breast cancer RNA, showed a systematic switch towards the 'normal' signature, with luminal A and luminal B subtypes being most susceptible. The authors concluded that the PAM50 Breast Cancer Intrinsic Classifier is highly reproducible within and across platforms and that the clinical test has utility in the management of ER+ and ER– invasive breast cancer of all stages. The quality of this report was judged as low. Furthermore, the study was based on a small number of tumour samples and full details of the patient characteristics were not provided. In addition, there are significant limitations in making any interpretations from this evidence as it is derived only from an abstract. It has been shown that there may be discrepancies between data made available in

TABLE 16 Study design characteristics of included studies: PAM50

Author (year) Country	Study design	Number of patients	Follow-up (years)	Outcomes/end points	Evidence type	Funding
Bernard *et al.* (2011)[119] (abstract) Additional data from Martin *et al.*[120] (abstract) Country NR	Retrospective cohort (dates not reported) from randomised prospective GEICAM 9906 trial FFPE blocks with invasive breast tissues	Eligible sample: NR Sample included: 793 Tumour samples analysed using PAM50 by RT-qPCR method	8.7	Analytical outcomes including accuracy, reproducibility, etc.	Analytical validity; clinical utility – predictive ability (benefit of chemotherapy) Comparison of PAM50 with IHC by RT-qPCR and identification of potential predictive markers of taxane clinical benefit	NR
Cheang *et al.* (2011)[117] (abstract) Additional data from unpublished manuscript Canada	Retrospective cohort (dates not reported) from randomised NCIC.CTG MA.5 trial	Eligible sample: 476 Sample included: 476 Tumour samples analysed using PAM50	NR	Responsiveness of intrinsic subtypes to adjuvant anthracyclines vs. non-anthracyclines – tumour classification and correlation with RFS and OS	Clinical validity; clinical utility – predictive ability (benefit of chemotherapy)	NR
Chia *et al.* (2011)[121] Canada, USA	[Academic-in-confidence (AIC) information has been removed]	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	Clinical validity; clinical utility – predictive ability (benefit of chemotherapy) (AIC information has been removed)	Canadian Breast Cancer Foundation and National Cancer Institute
Ebbert *et al.* (2011)[118] (abstract) USA	Retrospective (dates not reported) FFPE blocks with invasive ($n=155$) and normal ($n=16$) breast tissues (training set)	Eligible sample: NR Sample included: 171 Tumour samples analysed using PAM50 by RT-qPCR method	NR	Analytical outcomes including accuracy, reproducibility, etc.	Analytical validity	NR
Nielsen *et al.* (2010)[116] Unpublished supplementary data from the manufacturer Canada	Retrospective cohort (1986–1992) FFPE, minimum two tumour cores extracted	Eligible sample: 991 Sample included: 786 (archived tumour samples) Samples excluded because of insufficient yield ($n=180$), technically insufficient, ($n=25$)	Median: 11.7	Numbers assigned to each intrinsic subtype, risk of recurrence score Comparator: clinical, IHC (ER, PR, HER2, Ki-67) Adjuvant! Online used to generate breast cancer RFS and DSS estimates for each patient	Clinical validity – comparison of clinical, IHC and GEP models of prognosis	NR
Parker *et al.* (2009)[115] Canada, USA	Retrospective cohort (dates not reported) Fresh-frozen and FFPE tissues Independent publicly available data sets including data sets previously reported by Loi, Wang, and Ivs.hima	Eligible sample: NR Sample included: 950 G1: 189 (training set for development of the subtype predictor); G2: 761	NR	Distribution of intrinsic subtypes compared with clinical marker status risk of relapse models for prognosis in LN– breast cancer	Clinical validity	NR

DSS, disease-specific survival; NR, not reported; RFS, recurrence-free survival; RT-qPCR, reverse transcription-quantitative polymerase chain reaction.

DOI: 10.3310/hta17440

TABLE 17 Patient characteristics of included studies: PAM50 test

Author (year)	Age (years), mean (SD)	LN status	ER status	Tumour size	Grade	HER2 status	Mean NPI score	Treatment
Bernard *et al.* (2011)[119] (abstract) Additional data from Martin *et al.*[120] (abstract)	NR	All LN+	NR	NR	NR	NR	NR	NR
Cheang *et al.* (2011)[117] Additional data from unpublished manuscript	NR (AIC information has been removed)	All LN+	+/–/missing: (AIC information has been removed)	NR	I: (AIC information has been removed); II: (AIC information has been removed); III: (AIC information has been removed); unknown: (AIC information has been removed)	NR	NR	Adjuvant cyclophosphamide, methotrexate and 5-fluorouracil (CMF) or adjuvant cyclophosphamide, epirubicin and 5-fluorouracil (CEF) chemotherapy after randomisation
Chia *et al.* (2011)[121]	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)
Ebbert *et al.* (2011)[118] (abstract)	NR	NR	NR	NR	NR	NR	NR	NR
Nielsen *et al.* (2010)[116]	Median 67 (range NR)	+/–/unknown: 511/222/53	+/–/missing: 768/9/9	Median 2.1 cm	I: 34; II: 338; III: 370; unknown: 44	+/–/ unknown: 75/696/15	NR	All patients had adjuvant tamoxifen
Parker *et al.* (2009)[115]	G1: 58 (15); G2: 53 (13)	+/–: G1: 96/100;[a] G2: 35/710[a]	+/–: G1: 114/77;[a] G2: 544/195[a]	≤/> 2 cm: G1: 63/136;[a] G2: 408/339[a]	Low/medium/high: G1: 12/56/127;[a] G2: 133/218/286[a]	+/–: G1: NA; G2: 66/192[a]	NR	G1: training set G2: no adjuvant systemic therapy

NA, not applicable; NR, not reported; SD, standard deviation.
a Group with missing data.

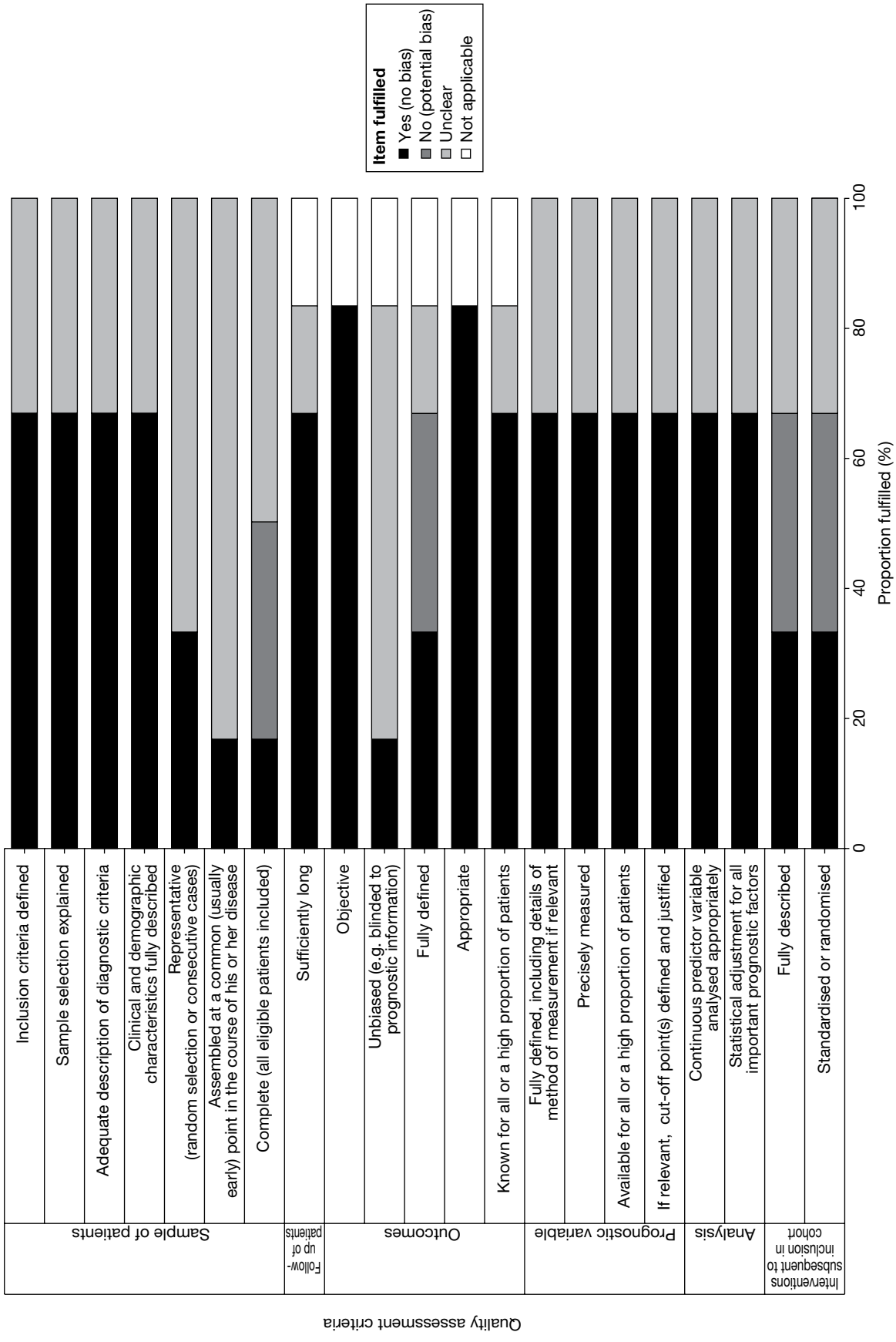


FIGURE 5 PAM50 test: methodological quality graph (review authors' judgements about each methodological quality item presented as percentages across all studies).

TABLE 18 Summary of evidence for the PAM50 test

Author (year)	Evidence type	Overall quality	Key findings
Bernard *et al.* (2011)[119] (abstract) and Martin *et al.*[120] (abstract)	Analytical validity; clinical utility – predictive ability (benefit of chemotherapy) Comparison of PAM50 by RT-qPCR with IHC and identification of potential predictive markers of taxane clinical benefit	Low	793 LN+ patients from the GEICAM 9906 randomised trial. Bernard *et al.* reported agreement between RT-qPCR gene expression and IHC scoring for clinical markers. They showed that there was good agreement between (gene/protein) *ESR1*/ER, *PGR*/PR and *ERBB2*/HER2. The accuracy was significantly lower for *MKI67*/Ki-67, *EGFR*/EGFR and *KRT5*/CK5/6. The authors concluded that calling cut-points based on RT-qPCR expression across subtypes is reproducible across data sets and has good agreement with expression by IHC for clinically used biomarkers. Martin *et al.* reported (AIC information has been removed). The authors concluded that (AIC information has been removed). Limitations: abstract data
Cheang *et al.* (2011)[117] (AIC information has been removed)	Clinical validity; clinical utility – predictive ability (benefit of chemotherapy) (AIC information has been removed)	High	(AIC information has been removed)
Chia *et al.* (2011)[121]	Clinical validity; clinical utility – predictive ability (benefit of tamoxifen) (AIC information has been removed)	High	(AIC information has been removed)
Ebbert *et al.* (2011)[118] (abstract)	Analytical validity	Low	171 US patients. Reported that within-platform cross-validation of the clinical subtype predictor showed 91.6% concordance. There was 100% reproducibility of subtype predictions across 46 runs testing different subtypes. Subtype predictions across platforms showed 88.1% concordance. The authors concluded that the PAM50 Breast Cancer Intrinsic Classifier is highly reproducible within and across platforms. Limitations: small sample size, abstract data
Nielsen *et al.* (2010)[116] (additional data from unpublished version of the manuscript)	Clinical validity Comparison of clinical, IHC and GEP models of prognosis	Moderate	786 LN+ or higher-risk LN–, ER+ Canadian patients. Assessed numbers of patients assigned to each intrinsic subtype and risk of relapse score against IHC (ER, PR, HER2, Ki-67). Reported that the included patients were considered to be high risk with overall 10-year RFS of 62% and DSS of 72%. Those assigned to luminal A tumours had significantly better outcomes (10-year RFS 74%; DSS 83%) than those assigned to luminal B, HER2-enriched and basal-like tumours. The authors concluded that IHC approaches do work and provide significant prognostic information; however, PAM50 is superior to these in terms of adding significant additional information and in its capacity to identify a particularly low-risk group. Incorporated a relatively large sample
Parker *et al.* (2009)[115]	Clinical validity	Moderate	950 majority LN–, ER+ Canadian and US patients. Reported that the intrinsic subtypes showed prognostic significance (for RFS) in untreated (no systemic therapy) patients ($p < 0.0001$) and remained significant in multivariable analyses that incorporated clinical covariates (ER status, histological grade, tumour size and LN status) ($p < 0.0001$). The authors concluded that the intrinsic subtype and risk predictors based on the PAM50 gene set add significant prognostic and predictive value to pathological staging, histological grade and standard clinical molecular markers

DSS, disease-specific survival; RFS, recurrence-free survival; RT-qPCR, reverse transcription-quantitative polymerase chain reaction.

abstracts and the reporting of results in subsequently published full-length articles.[89] Because of incomplete reporting the methodological quality of studies cannot be confidently assessed by systematic reviewers.

Bernard *et al.*,[119] using a cohort of 793 LN+ tumour samples from the GEICAM 9906 randomised trial, reported agreement between reverse transcription-quantitative polymerase chain reaction (RT-qPCR) gene expression and IHC scoring for clinical markers. They showed that there was good agreement between (gene/protein) *ESR1*/ER, *PGR*/PR and *ERBB2*/HER2. The accuracy was significantly lower for *MKI67*/Ki-67, *EGFR*/EGFR and *KRT5*/CK5/6. Discrepancies between the Hercep test and chromogenic in situ hybridisation (CISH) procedure for 2+ and 3+ staining-intensity samples showed that RT-qPCR agreed better with the Herceptest [area under the curve (AUC): 0.95 vs. 0.93). The authors concluded that calling cut-points based on RT-qPCR expression across subtypes is reproducible across data sets and has good agreement with expression by IHC for clinically used biomarkers. The quality of this report was judged as low. Although the study benefits from a relatively large sample size, as these data were reported in abstract form only there are significant limitations in using these results to make generalisations.

Clinical validity

Parker *et al.*[115] investigated the distribution of intrinsic subtypes in comparison with clinical marker status and risk of relapse models for prognosis in a cohort of 950 Canadian and US majority LN–, ER+ breast cancer patients. They reported that the intrinsic subtypes showed prognostic significance (for recurrence-free survival, RFS) in untreated (no systemic therapy) patients ($p < 0.0001$) and remained significant in multivariable analyses that incorporated clinical covariates (ER status, histological grade, tumour size and node status) ($p < 0.0001$). The authors concluded that the intrinsic subtype and risk predictors based on the PAM50 gene set add significant prognostic and predictive value to pathological staging, histological grade and standard clinical molecular markers. The quality of this study was judged to be moderate and the sample size was relatively large.

Nielsen *et al.*[116] used a cohort of 786 LN+ or higher-risk LN–, ER+ Canadian patients with tumours collected between 1986 and 1992 to assess the numbers of patients assigned to each intrinsic subtype and risk of relapse score against IHC (ER, PR, HER2, Ki67). Adjuvant! Online was used to generate breast cancer recurrence-free survival and disease-specific survival estimates for each patient. They reported that the included patients were considered to be high risk, with overall 10-year RFS of 62% and disease-specific survival of 72%. Those assigned to luminal A tumours had significantly better outcomes (10-year RFS 74%; disease-specific survival 83%) than those assigned to luminal B, HER2-enriched and basal-like tumours. In Cox models incorporating standard prognostic variables, HRs for breast cancer disease-specific survival over the first 5 years of follow-up, relative to the most common luminal subtype, were 1.99 (95% CI 1.09 to 3.64) for the luminal B subtype, 3.65 (95% CI 1.64 to 8.16) for the HER2-enriched subtype and 17.71 (95% CI 1.71 to 183.33) for the basal-like subtype ($p = 0.0018$). The authors concluded that IHC approaches do work and provide significant prognostic information; however, PAM50 is superior to these in terms of adding significant additional information and in its capacity to identify a particularly low-risk group. This study was judged to be of moderate quality and incorporated a relatively large sample size.

Chia *et al.*[121] (AIC information has been removed).

Cheang *et al.*[117] (AIC information has been removed).

Clinical utility

In addition to the clinical validity data reported above, Chia *et al.*[121] (AIC information has been removed).

(AIC information has been removed.) Using the same data reported by Bernard *et al.*[119] for a cohort of 793 LN+ tumour samples, Martin *et al.*[120] (AIC information has been removed). The

quality of this report was judged as low. Although the study benefits from a relatively large sample size, as these data were reported in abstract form only there are significant limitations in using these results to make generalisations.

Cheang *et al.*[117] (AIC information has been removed).

Summary of evidence: PAM50

Analytical validity of PAM50

Two abstracts[118,119] reported data on analytical validity, both rated as low quality. One[118] employed a relatively small sample ($n = 171$) whereas the other[119] was based on a much larger sample ($n = 793$). These studies provide a comparison of PAM50 against standard IHC measurements. There are significant limitations in making any interpretations from this evidence as it is derived only from abstracts. It has been shown that there may be discrepancies between data made available in abstracts and the reporting of results in subsequently published full-length articles.[89] Because of incomplete reporting the methodological quality of studies cannot be confidently assessed by systematic reviewers.

Clinical validity (prognostic ability) of PAM50

Four studies,[115–117,121] two rated as high quality and two rated as low quality, were identified that contain data on clinical validity. Two of these are as yet unpublished. These studies demonstrate that the intrinsic subtypes are significantly associated with outcome, provide additional information to IHC approaches and standard clinicopathological measures and can identify a particularly low-risk group. They demonstrate that prognostic ability has been validated in external cohorts. However, the population in most of the studies was LN+, with the exception of that by Parker *et al.*,[115] who assessed LN– patients; therefore, the generalisability of these findings to LN–, ER+ patients is limited. Furthermore, no studies were UK based, limiting the generalisation of the findings to UK practice. Finally, as two of these studies were unpublished there are significant limitations in drawing conclusions from this evidence.

Clinical utility of PAM50

(AIC information has been removed).

Key gaps in the evidence base remain and are summarised below:

- The evidence base to date is still immature as the majority of the evidence presented here is in abstract or unpublished form only. Only two studies were considered to be of high quality (these presented clinical validity and clinical utility data), two were of moderate quality and two were of low quality, although it should be noted that only two studies were full peer-reviewed papers.
- Further evidence around analytical validity is required as the current evidence is based on abstracts only.
- Robust evidence of clinical utility is needed. It is not yet clear whether or not the use of the PAM50 test will change the management of patients, for example reduce the number of patients unnecessarily treated with chemotherapy or improve patient outcomes through increases in DFS and OS.
- The fact that no studies were conducted in a UK setting may limit the extent to which we can generalise the findings.

Overall summary

The evidence base, in particular in relation to the prognostic ability of the test, is developing.

None of the studies used UK-based patients and the data were all based on cohort studies. Most of the evidence is in LN+ patients. The main limitation is that currently most of the evidence is unpublished or is in abstract form only. Robust evidence of clinical utility is needed as it is not yet clear to what extent the use of the PAM50 test will change the management of patients, to what extent PAM50 subtypes are predictive of chemotherapy benefit or how the use of PAM50 will improve patient outcomes through increases in DFS and OS.

Breast Cancer Index

The BCI is a RT-PCR assessment of the ratio of expression of two genes, *HOXB13* and *IL17BR*, combined with the MGI and gives an indication of recurrence risk. The target population is those with ER+ and LN– early breast cancer. The BCI RS ranges from 0 to 10 and divides patients into three risk groups. Low risk is defined as a score < 5; intermediate risk as a score of 5–6.3; and high risk as a score ≥ 6.4. Further details are included in *Table 6*.

Description of included studies

The searches identified one full peer-reviewed study relating to the BCI.[122] The design and patient characteristics of the included study are provided in *Tables 19* and *20*.

Quality of included studies: Breast Cancer Index

Although the assessment of study quality was hindered by poor reporting of whether or not outcomes were unbiased, the overall methodological quality of the included study was judged to be high, indicating a low risk of bias (received a positive assessment for at least 19 out of 21 methodological quality items).

TABLE 19 Study design characteristics of the included study: BCI

Author (year) Country	Study design	Number of patients	Follow-up (years)	Outcomes/end points	Evidence type	Funding
Jerevall *et al.* (2011)[122] Sweden	Retrospective cohort (1976–1990) from the randomised, prospective Stockholm trial FFPE RT-PCR	Eligible sample: 808 Sample included: 588 G1: 314; G2: 274 Samples excluded because of insufficient tumour (n = 37); failure of RT-PCR, (n = 2); ER– cases (n = 181) (exclusion criterion of study)	Mean: 17	Time to distant metastasis; DMFS; BCSD	Clinical validity	Swedish Cancer Society, Swedish Research Council, King Gustaf V Jubilee Fund, National Cancer Institute, Avon Foundation and bioTheranostics

TABLE 20 Patient characteristics of the included study: BCI

Author (year)	Age (years), mean (SD)	LN status	ER status	Tumour size	Grade	HER2 status	Mean NPI score	Treatment
Jerevall *et al.* (2011)[122]	NR (all postmenopausal)	All LN–	All ER+	G1: ≤ 2 cm: 256 (82%); > 2 cm: 55 (18%); unknown: 3 (1%) G2: ≤ 2 cm: 223 (81%); > 2 cm: 49 (18%); unknown: 2 (1%).	G1: I: 67 (21%); II: 209 (67%); III: 38 (12%) G2: I: 67 (24%); II: 172 (63%); III: 35 (13%)	G1: +/–/unknown: 14 (4%)/272 (87%)/28 (9%) G2: +/–/unknown: 13 (5%)/238 (87%)/23 (8%)	NR	G1: tamoxifen G2: systemically untreated

NR, not reported; SD, standard deviation.

Results: Breast Cancer Index

Full data extraction tables are presented in *Appendix 12*.

Analytical validity

No available evidence.

Clinical validity (prognostic ability)

Jerevall *et al.*[122] reported a retrospective analysis of (a subcohort of) a randomised prospective trial cohort. The 588 patients were all postmenopausal, LN– and ER+. The authors reported the development and testing of *HOXB13:IL17BR* plus MGI as a continuous index (BCI) in a training set (G1) and a test set (G2) of the same trial data. In the training set (G1) BCI classified 59% of patients as low risk. The rate of distant recurrence for the low-risk group was 7.1% (95% CI 0% to 3.5%) and the rate of death was 1.1% (95% CI 0% to 2.6%). In total, 22% were classified as intermediate risk, with a rate of distant recurrence of 17.8% (95% CI 7.6% to 26.8%) and a rate of death of 14.5% (95% CI 5.2% to 22.9%), and 18.4% were classified as high risk, with a rate of distant recurrence of 20.0% (95% CI 8.7% to 30.0%) and a rate of death of 14.7% (95% CI 4.7% to 23.6%). In the test set (G2) 53% of patients were classified as low risk (rate of distant recurrence 8.3%, 95% CI 4.7% to 14.4%; rate of death 5.1%, 95% CI 1.3% to 8.7%), 27% were intermediate risk (rate of distant recurrence 22.9%, 95% CI 14.5% to 35.2%; rate of death 19.8%, 95% CI 10.0% to 28.6%) and 20% were high risk (rate of distant recurrence 28.5%, 95% CI 17.9% to 43.6%; rate of death 28.8%, 95% CI 15.3% to 40.2%). The authors concluded that the BCI was a strong prognostic factor for distant recurrence and BCSD, independent of tumour size, grade, HER2 status and PR status (although tumour size did contribute prognostic value to distant recurrence). The prognostic utility of the BCI was also assessed compared with Adjuvant! Online for the test set (G2). Both BCI and Adjuvant Online were significant predictors of BCSD (BCI: HR 2.3, 95% CI 1.5 to 3.7, $p < 0.001$; Adjuvant Online: HR 1.4, 95% CI 1.0 to 1.8, $p = 0.04$) and distant recurrence (BCI: HR 2.0, 95% CI 1.3 to 3.1, $p = 0.001$; Adjuvant! Online: HR 1.4, 95% CI 1.0 to 1.8, p 0.03). The authors concluded that the retrospective analysis of this randomised, prospective trial cohort validated the prognostic utility of *HOXB13:IL17BR* plus MGI and it was used to develop and test a continuous risk model that enables prediction of distant recurrence risk at the patient level. The study had a long mean follow-up time and a moderate sample size. The study used tumour samples that dated back to 1976, introducing differences in the diagnostic criteria applied to patients.

Clinical utility

No available evidence.

Summary of evidence: Breast Cancer Index

Based on the limited available data, no firm conclusions can be made about the BCI. Further evidence on analytical validity, clinical validity and clinical utility is required. The test has also not been validated in an external cohort.

Randox Breast Cancer Array

The Randox BCA is a cDNA-based expression biochip assay that aims to accurately define the clinical subtype of breast cancer tumour before initiating treatment. The target population is all individuals with diagnosed breast cancer.

Description of included studies

The searches did not identify any relevant full peer-reviewed papers or meeting abstracts relating to the Randox assay. Supplementary evidence was provided by the manufacturer of the test.

Results: the Randox assay

Supplementary evidence

The manufacturer submitted a description of the data gathered on the Randox assay up to August 2011.[123] A summary of this information is provided here.

The objective of the study was to improve the discrimination, to include basal and unclassified (triple-negative) types, subdivide luminal groups into A and B and assign samples to an *ERBB2* group. The main indicator of correct typing is whether the samples are correctly typed as ER+ or ER–. A total of 150 archived tumour samples were collected and used on the Randox biochip array. Exclusion criteria included a lack of ER status information or one or both of the housekeeping genes failing to reach adequate expression levels, preventing normalisation of the remaining gene set on the chip. The sample set included information on the following: DFS, OS ER status, PR status and other standard clinical measurements. However, HER2 (*ERBB2*) status was not available for any patients; thus, hormonal status was either luminal positive or negative. The initial summary, using a patient cohort of 78 individuals, shows an overall agreement of 79% between hormonal status (primarily ER) and the multiplex biochip assay. The authors concluded that these findings were encouraging.

Methodological detail was lacking for the summarised study and the sample size was very small. Quality assessment could not be undertaken because of the limited detail available. It should also be noted that, although Randox separates luminal A and luminal B groups, the overall reported agreement of 79% was based on luminal A and B combined and hormonal status, and did not differentiate between the subtypes.

Summary of evidence: the Randox assay

No conclusions can be drawn from this evidence. Further evidence is required.

Mammostrat test

The Mammostrat test uses five IHC markers (SLC7A5, HTF9C, p53, NDRG1 and CEACAM5) to stratify patients into risk groups to inform treatment decisions. These markers are independent of one another and do not directly measure either proliferation or hormone receptor status. The current version of the test provides classification into one of five breast cancer subtypes, and quantitative values for (gene/protein) *ESR1*/ER, *PGR*/PR, *ERBB2*/HER2, proliferation, and luminal score (ER pathway), along with a RS and category (low, moderate and high). For further information see *Table 7*.

Description of included studies

The searches identified three full peer-reviewed studies relating to the Mammostrat test.[124–126] All studies contained data relating to clinical validity. Ross *et al.*[126] also reported on clinical utility in terms of the predictive ability of the test by risk group. The studies are described in *Tables 21* and *22*.

Quality of included studies: Mammostrat

The methodological quality of the three included studies[124–126] is summarised in *Figure 6* (further details are provided in *Appendix 13*). Of these, two studies performed well,[124,126] receiving a positive assessment for at least 17 out of 21 methodological quality items. Although all studies used a retrospective study design, other potential sources of bias were generally related to the following domains: sample of patients (clinical/demographic characteristics not fully described) and interventions subsequent to inclusion in the cohort (interventions were not fully described or standardised). The assessment of study quality was generally hampered by poor reporting of the

TABLE 21 Study design characteristics of included studies: Mammostrat test

Author (year) Country	Study design	Number of patients	Follow-up (years)	Outcomes/end points	Evidence type	Funding
Bartlett *et al.* (2010)[124] UK	Retrospective, consecutive sample series (1981–98) Microarray	Eligible sample: 1540 Sample included: 1540 G1 (all ER+): 1189; G2 (ER+, tamoxifen only): 831; G3 (ER+, N–, tamoxifen only): 657	Minimum 9	DRFS; RFS; OS	Clinical validity	Sarah Percy Endowment Fund
Ring *et al.* (2006)[125] [Commercial-in-confidence (CIC) information has been removed] USA	Retrospective cohort (G1: 1989–2002; G2: 1995–6; G3: 1974–95)	Eligible sample: NR Sample included: 1109 G1: 466 (also training cohort); G2: 299; G3: 344	G1: NR; G2: 5; G3: mean 11.7	DFS at 5 years	Clinical validity	NR
Ross *et al.* (2008)[126] NR	Retrospective cohort (dates not specified) from the randomised, prospective NSABP B14 and B20 trials	Eligible sample: NR Sample included: 1267 Placebo: 287; B14 tamoxifen: 550; B20 tamoxifen: 161; B20 chemotherapy: 269	Minimum 10	Recurrence-free interval; DRFI; BCSD	Clinical validity; clinical utility – predictive ability (benefit of chemotherapy)	NR

NR, not reported.

following methodological items: whether or not all eligible patients were included, representative and assembled at an early point in the course of their disease and whether or not outcomes were unbiased and known for all or a high proportion of patients. Overall, the risk of bias from these studies was judged to be low.

Results: Mammostrat

A summary of the clinical evidence for the Mammostrat test is presented in *Table 23* followed by a narrative summary of each study. Full data extraction tables are provided in *Appendix 13*.

Analytical validity

No available evidence.

Clinical validity

Bartlett *et al.*[124] investigated assignment to risk groups using Mammostrat for 1540 UK patients with LN– tumours. They demonstrated that significantly more cases were assigned to high-risk groups for ER– than for ER+ cases (45% vs. 16%; $p<0.001$), but there were no differences between other groups. They also looked at associations between risk scores and DRFS, RFS and OS across the different groups: all cases, all ER+ cases, all ER+ cases treated with tamoxifen only and ER+, LN– cases treated with tamoxifen only. For all groups there were significant associations between risk score and RFS, DRFS and OS (with the exception of the ER+, LN– treated with tamoxifen only group for which there was a trend only for OS). Multivariate analyses for each of the four groups showed that risk score was a significant independent predictor of RFS, DRFS and OS, along with clinicopathological predictors, for all cases and all ER+ cases. Risk score was a significant independent predictor of DRFS and OS with a trend for RFS for ER+ cases treated with tamoxifen only, and there was a trend towards significance for Mammostrat score to predict RFS and DRFS in ER+, LN– cases treated with tamoxifen only. The authors concluded that Mammostrat can act as an independent prognostic tool for ER+, tamoxifen-treated breast cancer and that there is a possible association with outcome regardless of LN status and ER– tumours. This study was rated as being of high quality on the basis of the quality assessment

TABLE 22 Patient characteristics of included studies: Mammostrat test

Author (year)	Age (years)	LN status	ER status	Tumour size	Grade	HER2 status	Mean NPI score	Treatment
Bartlett *et al.* (2010)[124]	<50/>50/missing All cases: 660 (42.8%)/879 (57.1%)/1 (0.1%) G1: 505 (42.5%)/683 (57.4%)/1 (0.1%) G2: 284 (34.2%)/547 (65.8%)/0 G3: 243 (37.0%)/414 (63.0%)/0	LN–/1–3/4–9/10+/ missing All cases: 1164 (75.6%)/321 (20.8%)/44 (2.9%)/9 (0.8%)/38 (2.5%) G1: 889 (74.8%)/264 (22.2%)/29 (2.4%)/5 (0.4%)/2 (0.2%) G2: 657 (79.1%)/154 (18.5%)/17 (2.0%)/3 (0.4%) G3: 657 (100%)	Allred score <2/3–5/6–8/ missing All cases: 278 (18.1%)/347 (22.5%)/823 (53.4%)/92 (6.0%) G1: 19 (1.6%)/347 (29.2%)/823 (69.2%) G2: 15 (1.8%)/213 (25.6%)/603 (72.6%) G3: 14 (2.1%)/174 (26.5%)/469 (71.4%)	<2 cm/>2 cm/missing All cases: 1150 (74.7%)/314 (20.4%)/76 (4.9%) G1: 903 (75.9%)/224 (18.8%)/62 (5.2%) G2: 648 (78.0%)/148 (17.8%)/35 (4.2%) G3: 531 (80.8%)/99 (15.1%)/27 (4.1%)	I/II/III/missing All cases: 411 (26.7%)/710 (46.1%)/381 (24.7%)/38 (2.5%) G1: 359 (30.2%)/581 (48.9%)/233 (19.6%)/16 (1.3%) G2: 269 (32.4%)/416 (50.1%)/135 (16.2%)/11 (1.3%) G3: 215 (32.7%)/323 (49.2%)/109 (16.6%)/10 (1.5%)	NR	NR	Breast-conservation surgery, axillary node sampling or clearance, and whole-breast radiotherapy 1102 treated with adjuvant tamoxifen without chemotherapy, 92 with other hormone therapy and 197 no adjuvant hormone therapy or chemotherapy
Ring *et al.* (2006)[125]	<50/≥50 G1: 135 (29%)/327 (71%) G2: 74 (25%)/225 (75%) G3: not available	LN–/N1/N2/N3/unknown G1: 264 (58%)/184 (40%)/8 (2%)/0/10 G2: 170 (62%)/68 (25%)/24 (9%)/12 (4%)/25 G3: 200 (66%)/0/103 (34%)/0/0	+/+LN–/– G1: 325 (70%)/195 (42%)/141 (30%) G2: 232 (78%)/137 (46%)/67 (22%) G3: 273 (79%)/159 (46%)/71 (21%)	T1/T2/T3/T4/unknown G1: 242 (54%)/173 (39%)/21 (5%)/12 (3%)/18 G2:167 (57%)/96 (33%)/19 (6%)/13 (4%)/4 G3: not available	I/II/III/unknown G1: 59 (16%)/168 (45%)/148 (39%)/91 G2: 45 (18%)/120 (48%)/83 (33%)/51 G3: not available	+/–/unknown (+ = 2–3 nodes; – = 0–1 nodes) G1: 378 (85%)/68 (15%)/20 G2: 253 (88%)/34 (12%)/12 G3: not available	Good/moderate/ poor/unknown G1: 120 (36%)/156 (47%)/55 (17%)/135 G2: not available G3: 69 (38%)/164 (58%)/31 (11%)/80	NR
Ross *et al.* (2008)[126]	NR (% of cohort only) <50/50–59/>60: 32/27/41	NR	NR (% of cohort only – fmol/mg) 10–49/50–99/>100: 37/22/41	NR (% of cohort only) <2 cm/2.1–4.0 cm/>4.1 cm: 59/35/5	NR	NR	NR	Placebo: 287; tamoxifen treated (B14 and B20): 711; chemotherapy (B20): 269

+LN–, ER positive, node negative; NR, not reported.

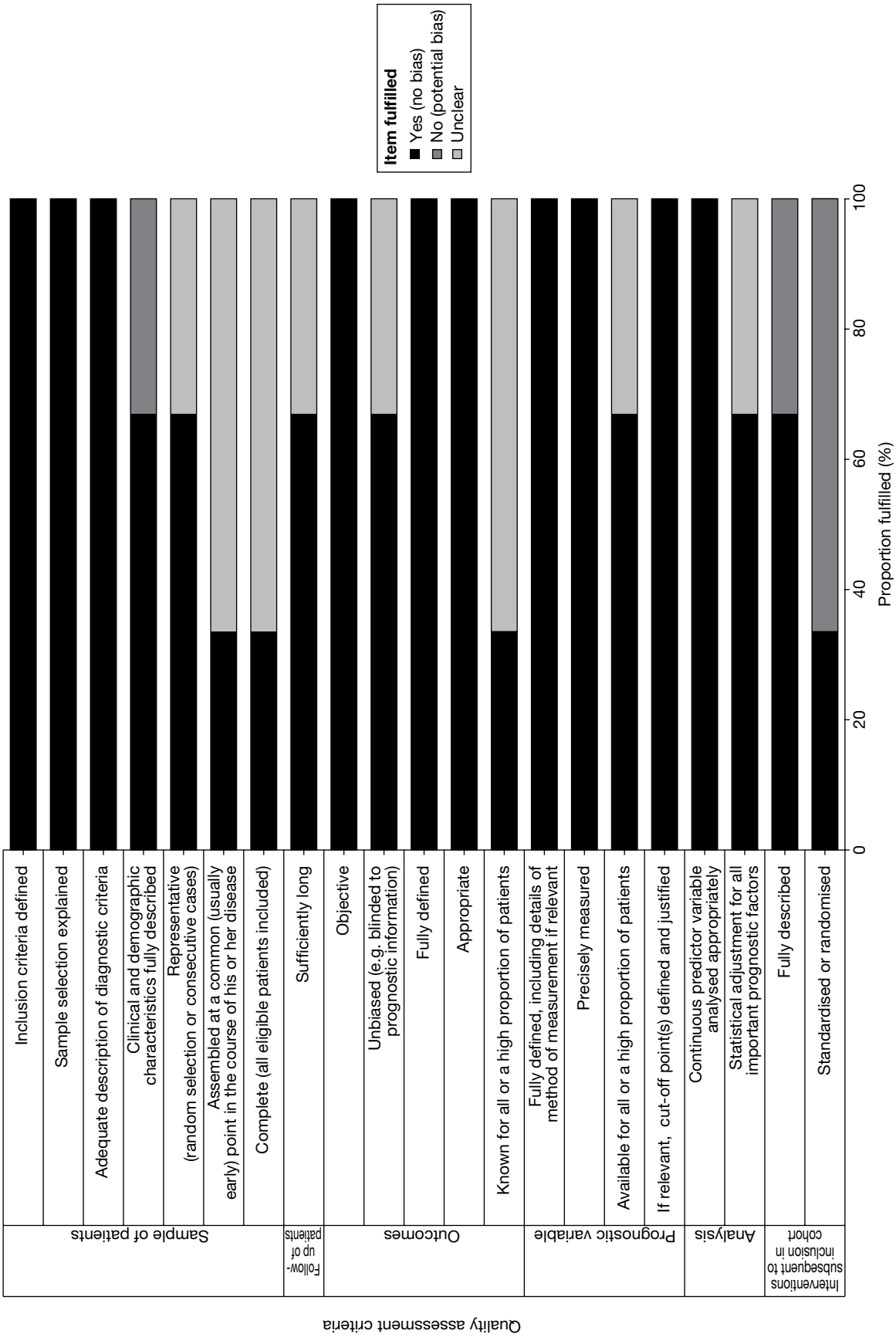


FIGURE 6 Mammostrat test: methodological quality graph (review authors' judgements about each methodological quality item presented as percentages across all studies).

TABLE 23 Summary of evidence for the Mammostrat test

Author (year)	Evidence type	Overall quality	Key findings
Bartlett *et al.* (2010)[124]	Clinical validity	High	1540 LN– UK patients. Also looked at associations between risk score and outcomes across the different groups: all cases, all ER+, all ER+ treated with tamoxifen only and ER+, LN– treated with tamoxifen only. For all cases and across the three groups there were significant associations between risk score and RFS, DRFS and OS (with the exception of the ER+, LN– treated with tamoxifen only group for which there was only a trend for OS). The authors concluded that Mammostrat can act as an independent prognostic tool for ER+, tamoxifen-treated breast cancer and that there is a possible association with outcome regardless of node status and ER– tumours. This study had a large sample size overall, a relatively long follow-up of 9 years and employed UK patients. Limitations: subsets analysed had relatively small numbers of patients within each; the samples used in these analyses were relatively old, dating back to 1981, and therefore there may be differences between the patients in terms of stage at presentation and diagnosis
Ring *et al.* (2006)[125]	Clinical validity	Moderate	1109 majority LN–, ER+ US patients. Cox model was able to identify patients classified in different risk categories based on outcomes. In both independent cohorts the model was independent of stage, grade and LN status. The authors concluded that the test can significantly improve on traditional prognosticators in predicting outcome for ER+ breast cancer patients. Large sample of patients. Validated in an external cohort. Limitation: tumour samples used in these analyses dated back as far as 1974 and therefore there may be differences between the patients in terms of stage at presentation and diagnosis
Ross *et al.* (2008)[126]	Clinical validity; clinical utility – predictive ability (benefit of chemotherapy)	High	711 ER+, LN–, tamoxifen-treated patients taken from the NSABP B14 and B20 trials. In multivariate analyses the Mammostrat test had significant prognostic power independent of age and tumour size (HR 1.3; 95% CI 1.1 to 1.6; $p=0.007$). Concluded that the risk index was significantly associated with clinical outcome among the ER+, LN–, tamoxifen-treated patients. Clinical utility: in terms of recurrence-free interval, patients in the low-risk group improved by 5% from 86% to 91% (HR 0.4; 95% CI 0.2 to 0.8) and patients in the high-risk group improved by 21% from 64% to 85% (HR 0.4; 95% CI 0.2 to 0.9), showing that these groups benefited from chemotherapy, whereas the patients in the intermediate risk group did not. Limitation: data from two trials were used but it was unclear how the data were combined

employed here. It also benefits from a large sample size overall, although it should be noted that the subsets analysed had relatively small numbers of patients within each. It had a relatively long follow-up of 9 years, and as it employed UK patients the findings should be applicable to UK practice. The samples used in these analyses were relatively old, dating back to 1981; therefore, there may be differences between the patients in terms of stage at presentation and diagnosis.

In a US-based study, Ring *et al.*[125] assessed 1109 majority LN–, ER+ patients. Using a training cohort the authors demonstrated that a Cox model identified a group of patients as having either poor or moderate prognosis, with a 5-year DFS rate of approximately 75%, as opposed to patients classified as having good prognosis, who had a 5-year DFS rate of approximately 95% ($p<0.001$). In the first independent cohort the model identified poor prognosis patients with a 5-year DFS rate of 50%, compared with approximately 70% for patients classified as having moderate prognosis and 87% for patients classified as having good prognosis ($p=0.008$). In the second independent cohort the model distinguished ER+ patients classified as having poor prognosis with OS rates of 55%, compared with 75% for patients classified as having moderate prognosis and 90% for patients classified as having good prognosis ($p=0.0039$). In both independent cohorts the model was independent of stage, grade and LN status. In the combined independent cohort, for patients with poor or good prognosis (82%), sensitivity for poor prognosis in predicting disease progression was 38%, whereas specificity was 88%. The PPV of poor prognosis was 38% (95% CI 32% to 44%), whereas the NPV was 88% (95% CI 84% to 92%). The authors concluded that the test can significantly improve on traditional prognosticators in predicting outcome for ER+ breast cancer patients. The quality assessment indicated that this study was of moderate quality and it used a large sample of patients. The test was also validated in an external cohort. However, there are limitations in that the tumour samples used in these analyses date

back as far as 1974; therefore, there may be differences between the patients in terms of stage at presentation and diagnosis.

Ross *et al.*[126] examined the association between the clinical outcomes recurrence-free interval (RFI), DRFI and BCSD and stratification by the Mammostrat test in 711 ER+, LN– tamoxifen-treated patients taken from the NSABP B14 and B20 trials. Of this group approximately 58% were classified as low risk, 21% as moderate risk and 21% as high risk. There was a significant association between patients stratified by the Mammostrat test and RFI (HR 1.3; 95% CI 1.1 to 1.6; $p = 0.006$). This was not significant in the low-risk group compared with the moderate-risk group (log-rank, $p = 0.05$) but was significant in the low-risk group compared with the high-risk group (HR 1.8; 95% CI 1.2 to 2.6). The authors reported a significant association between patients stratified by the test and DRFI (HR 1.4, 95% CI 1.1 to 1.7; $p = 0.001$). In the low-risk group compared with the moderate-risk group this was not significant whereas in the high-risk group compared with the low-risk group it was significant (HR 2.1, 95% CI 1.4 to 3.1; $p = 0.0004$). They also reported a significant association between patients stratified by the test and BCSD (HR 1.5, 95% CI 1.2 to 1.9; $p = 0.0003$). In the low-risk group compared with the moderate-risk group this was not significant whereas in the high-risk group compared with the low-risk group it was significant (HR 2.3; 95% CI 1.5 to 3.5; $p < 0.0001$). The Kaplan–Meier estimate of the proportion of patients recurrence free after 10 years was 82% (95% CI 79% to 85%) for the group overall, 85% (95% CI 81% to 88%) for the low-risk group, 85% (95% CI 80% to 91%) for the moderate-risk group and 73% (95% CI 65% to 80%) for the high-risk group. In multivariate analyses the Mammostrat test had significant prognostic power independent of age and tumour size (HR 1.3; 95% CI 1.1 to 1.6; $p = 0.007$). The authors concluded that the risk index was significantly associated with clinical outcome among the ER+, LN– tamoxifen-treated patients. Ross *et al.*[126] also reported data relating to clinical utility, which are detailed in the following section.

Clinical utility

Predictive ability (benefit of chemotherapy) Ross *et al.*[126] also presented evidence on the ability of the test to identify patients who have greater absolute benefit from adjuvant chemotherapy compared with unstratified patient populations. These analyses were based on the tamoxifen- and cytotoxic chemotherapy-treated patients ($n = 269$) and the B20 tamoxifen only-treated patients ($n = 161$) from the trial data. In terms of RFI patients in the low-risk group improved by 5% from 86% to 91% (HR 0.4; 95% CI 0.2 to 0.8) and patients in the high-risk group improved by 21% from 64% to 85% (HR 0.4; 95% CI 0.2 to 0.9), showing that these groups benefited from chemotherapy, whereas the patients in the intermediate-risk group did not. This study was rated as high quality although data from two different trials were used and it was unclear how the data were combined.

Supplementary evidence

(CIC information has been removed.)

Summary of evidence: Mammostrat

Analytical validity of Mammostrat

No evidence was found on the analytical validity of the test.

Clinical validity (prognostic ability) of Mammostrat

The three studies identified suggest that Mammostrat can act as an independent prognostic tool for ER+, tamoxifen-treated breast cancer. Two of the studies were rated as high quality and one as moderate quality. The test has been validated in an external cohort. Although the evidence base

for Mammostrat is relatively immature, these initial studies include a large sample size, appear to be of reasonable quality and, in the case of Bartlett *et al.*,[127] use a UK-based population.

Clinical utility of Mammostrat

One study reported on clinical utility and was rated as high quality. Initial evidence suggests that low- and high-risk groups benefited from chemotherapy, with high-risk patients benefitting more than low-risk patients. The moderate-risk group did not appear to benefit. There was no published evidence on reclassification of risk groups compared with conventional risk classifiers, and no evidence on the impact of the test on decision-making. Further evidence is required.

Overall summary

The evidence base, in particular in relation to the prognostic ability of the test, was of reasonably high quality. Further evidence of analytical validity and clinical utility is required.

IHC4 test

IHC4 assesses the levels of four key proteins (ER, PR, HER2 and Ki-67) in a breast cancer sample. The IHC4 score is calculated based on the percentage of cells positive for Ki-67 and PR (0–100%); the Histoscore for ER status (a measure of the percentage of cells positive multiplied by the intensity, range 0–300); and the tumour HER2 status, expressed as a binary measure (positive/negative). The final algorithm for IHC4 calculates a risk score for distant recurrence based on ER, PR, HER2 and Ki-67 in addition to classical clinical and pathological variables (composite risk score IHC4 + clinical). No risk category is given. Further details are included in *Table 7*.

Description of included study

The searches did not identify any relevant full peer-reviewed papers relating to IHC4. One relevant meeting abstract was identified, but had been superseded by a full paper.[83] The study design and patient characteristics are detailed in *Tables 24* and *25* respectively. The investigators also provided further information on the test (Professor Mitch Dowsett, Royal Marsden Hospital, London, September 2011, personal communication) and this information is detailed in the supplementary evidence section.

Quality of included studies: IHC4

Although the assessment of study quality was hindered by poor reporting of whether or not outcomes were fully defined and unbiased, the overall methodological quality of the included study was judged to be high, indicating a low risk of bias (received a positive assessment for at least 19 out of 21 methodological quality items).

Results: IHC4

Full data extraction tables are presented in *Appendix 14*.

Clinical validity

Cuzick *et al.*[83] reported a study assessing the prognostic value of IHC4. The IHC4 score was created and validated in one cohort (G1) and further validated in an independent cohort (G2). G1 was a retrospective cohort comprising patients from the TransATAC trial (a multinational trial, including the UK). The majority of the 1125 patients in G1 were LN– and hormone receptor positive. In this cohort there was a total of 195 recurrences of which 145 were distant recurrences. In LN– women there were 101 recurrences of which 67 were distant recurrences. The authors determined the value of each of the four IHC markers in three ways: univariately, as an addition to a model containing the classical variables, and when added to a model containing the classical variables and the other three IHC markers; this was carried out for all women and separately for LN– women only. They found that each of the four variables added a significant amount of

TABLE 24 Study design characteristics of the included study: IHC4 test

Author (year) Country	Study design	Number of patients	Follow-up (months)	Outcomes/end points	Evidence type	Funding
Cuzick *et al.* (2011)[84] G1: multinational, including UK G2: UK	G1: retrospective, cohort study of TransATAC trial G2: cohort study IHC4 assessed in a group of hormone receptor-positive patients with a GHI-RS score FFPE	Eligible sample: 1911 Sample included: 1911 G1: 1125 hormone receptor-positive patients who did not receive adjuvant chemotherapy, had a GHI-RS and adequate tissue G2: 786 patients. A predictive model using classical variables and the four IHC markers (IHC4 score) was created and assessed in a separate cohort	Median: 100	Distant recurrence (within 10 years); TTDR	Clinical validity To determine how much of the information contained in the GHI-RS is contained in standard IHC markers	Royal Marsden NIHR Biomedical Research Centre, Cancer Research UK, Breakthrough Breast Cancer and AstraZeneca

GHI-RS, Genomic Health Recurrence Score ; NIHR, National Institute for Health Research.

TABLE 25 Patient characteristics of included studies: IHC4 test

Author (year)	Age (years), median (IQR)	LN status	ER status	Tumour size	Grade	HER2 status	Mean NPI score	Treatment
Cuzick *et al.* (2011)[84]	G1: 64 (57–70) G2: 55 (48–63)	G1: –/+/unknown: 793 (70%)/288 (26%)/44 (4%) Those with unknown nodal status taken to be node negative in analyses G2: –/+: 487 (62%)/299 (38%)	NR (reported to be ER+ and/or PR+) Reported to be ER+ and/or PR+	G1: ≤1 cm: 177 (16%); 1–2 cm: 574 (51%); >2–3 cm: 272 (24%) G2: ≤1 cm: 105 (13%); 1–2 cm: 415 (53%); >2–3 cm: 190 (24%)	G1: poor: 206 (18%); moderate: 690 (61%); well differentiated: 229 (21%) G2: poor: 178 (23%); moderate: 336 (43%); well differentiated: 272 (34%)	G1: +: 116 (10%) G2: +: 41 (5%)	NR	G1: tamoxifen: 565 (50%); anastrozole: 560 (50%) G2: no endocrine treatment: 410 (52%); tamoxifen: 376 (48%)

IQR, interquartile range; NR, not reported.

information. Ki-67 was the most powerful univariately, but not in multivariate analyses because of its correlation with grade. For the multivariate models PR was most prognostic overall, but less so in LN– patients, in whom ER, HER2 and Ki-67 had similar values. The overall contribution of the IHC measurements for distant recurrence was highly significant [χ^2 (4 degrees of freedom, df) = 39.1; $p < 0.0001$], and it was reported that the median IHC4 score for all patients was –4.2 and the interquartile range (IQR) –29.9 to 29.9. The HR for a change from the 25th to the 75th percentile of the IHC4 score for all patients was 5.7 (95% CI 3.4 to 9.7) in univariate analysis and 3.9 (95% CI 2.4 to 6.7) when added to clinical score. In a second validation cohort of 786 ER+ younger patients treated in the UK (G2), the authors demonstrated that IHC4 score was highly significantly predictive of outcome (HR 4.8; 95% CI 2.2 to 10.2) for a change from the 25th to the 75th percentile in univariate analysis, and gave similar results when added to clinical score (HR 4.4; 95% CI, 2.0 to 9.3; $p < 0.0001$). The authors concluded that they have created a prognostic model that integrates IHC information with classical clinical and pathological variables and may prove helpful in managing early ER+ breast cancer in postmenopausal patients, but that additional studies are needed to determine the general applicability of the IHC4 score. This study was rated as high quality on the basis of the quality assessment checklist. It has employed a large sample size and the test has been validated in an external cohort of UK patients.

Supplementary information

Supplementary data were provided by the co-investigators of this study[84] after request by the External Assessment Group (EAG) (Professor Mitch Dowsett, Royal Marsden Hospital, London, September 2011, personal communication). Two analyses were conducted in the TransATAC trial – among women with LN–, ER+, HER2– early breast cancer ($n = 707$) and among all women ($n = 1117$).

Discussion with the co-investigators of the study (Professor Mitch Dowsett, Royal Marsden Hospital, London, July 2011, personal communication) indicated that the test was meant to be used in conjunction with clinicopathological parameters and therefore data for the final algorithm (using age, grade and tumour size) were used in the economic model. They provided data on the reclassification and risk of distant recurrence of patients using IHC4 plus clinical score (for simplicity the term IHC4 will be used in the report but the data refer to the use of the test in conjunction with clinicopathological parameters). Although cut-offs are not available for IHC4, for the purpose of the economic assessment investigators provided risk classification evidence of IHC4 based on low, intermediate and high risk of distant recurrence. Cut-offs were defined using a similar approach to the classification with OncotypeDX (< 10%, 10–20% and > 20% risk of distant recurrence). The cut-offs used for IHC4 are, however, exploratory and were defined only to populate the economic model. More details are available in *Chapter 3*.

Among the 707 women with LN–, ER+, HER2– early breast cancer, 85.3% of patients ($n = 603$) were classified as having a low risk of distant recurrence using IHC4. The proportions of patients classified as having an intermediate and a high risk of distant recurrence were 9.9% ($n = 70$) and 4.8% ($n = 34$). The risk of distant recurrence for patients classified as low, intermediate or high risk of distant recurrence by IHC4 is shown in *Figure 7*.

The reclassification of the three IHC4 risk groups against the two NPI risk groups used in the economic model (NPI ≤ 3.4 and NPI > 3.4) is presented in *Table 26*.

Summary of evidence: IHC4

Analytical validity of IHC4

We found no evidence on analytical validity. Although the use of IHC4 may be extended for use in other laboratories, the included paper suggests that there may be a lack of reproducibility

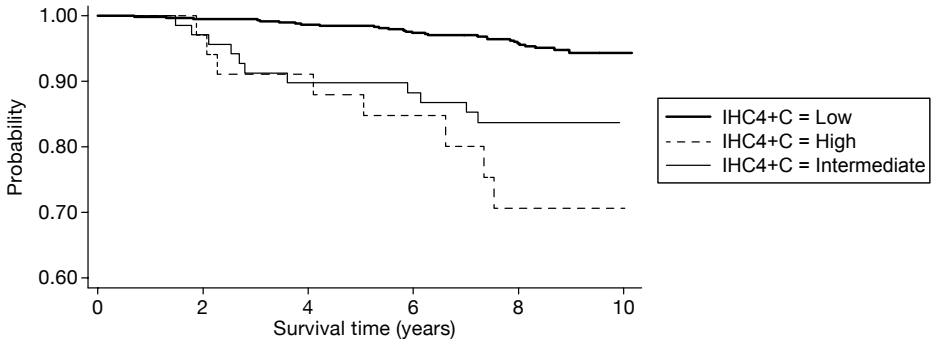


FIGURE 7 Time to distant recurrence by IHC4 + clinical score group for patients who were LN–, ER+, HER2–.

TABLE 26 Reclassification of ER+, LN–, HER2– patients from the TransATAC trial by IHC4 + clinical score and NPI group (Professor Mitch Dowsett, Royal Marsden Hospital, London, UK, September 2011, personal communication)

	Low risk IHC4, *n* (%)	Intermediate risk IHC4, *n* (%)	High risk IHC4, *n* (%)	Total, *n* (%)
NPI ≤ 3.4	437 (97.3)	12 (2.7)	0	449 (100)
NPI > 3.4	166 (64.6)	58 (22.6)	33 (12.8)	257 (100)
Total	603 (85.4)	70 (9.9)	33 (4.7)	706 (100)

of the test in relation to Ki-67. The authors suggest that, because IHC4 offers the possibility of carrying out the test in local laboratories, full validation would require evaluation of the IHC4 score when carried out in a range of local laboratories. Reproducibility of the test would need to be confirmed and quality assurance programmes put in place.

Clinical validity (prognostic ability) of IHC4

One study on the clinical validity of IHC4 was available, which claims that the IHC4 score is a highly significant predictor of distant recurrence. This initial study included a large sample size and detailed the development of the test in one cohort and the external validation of the test in an independent cohort. The study has been rated as high quality on the basis of the quality assessment employed.

Clinical utility of IHC4

There is currently no evidence on the clinical utility of IHC4 in terms of its ability to change treatment decisions or its ability to predict chemotherapy benefit. Although there are no published data on clinical utility, unpublished data were obtained to populate the economic model.

Overall summary

The evidence base for IHC4 is currently limited to clinical validity (prognostic ability), although the evidence for clinical validity is relatively strong given that the test has been developed using a large cohort of patients and has been validated in an external cohort. Further evidence is required on the analytical validity and clinical utility of IHC4.

Nottingham Prognostic Index plus

NPI+ is a biomarker-based prognostic assay that integrates 10 predictive biomarkers of long-term survival and therapeutic response with existing clinical and molecular pathology knowledge to support individualised clinical decision-making. This test is under development and outputs/presentation are not yet finalised. Further details are provided in *Table 7*.

Description of included studies

The searches did not identify any relevant full peer-reviewed papers or meeting abstracts relating to NPI+. Supplementary evidence was provided by the manufacturer of the test.

Supplementary evidence

The manufacturers submitted two draft full papers based on the same data and one draft abstract (of a full paper). The study design and patient characteristics included in these documents are presented in *Tables 27* and *28*.

Quality of included studies: Nottingham Prognostic Index plus

The overall methodological quality of the two (unpublished) included studies is provided in *Appendix 15*. Both studies were deemed to be of (AIC information has been removed).

Results: Nottingham Prognostic Index plus

A summary of the evidence provided is present in *Table 29* followed by a narrative summary. Full data extraction tables are presented in *Appendix 15*.

TABLE 27 Study design characteristics of included studies: NPI+ test (information submitted by the manufacturer)

Author (year) Country	Study design	Number of patients	Follow-up (years)	Outcomes/ end points	Evidence type	Funding
Green *et al.* (unpublished) and Nottingham Prognostics (2011)[128] (AIC information has been removed)	(AIC information has been removed)	Eligible sample: (AIC information has been removed) Sample included: (AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)
Nottingham Prognostics (2011) (abstract)[128] (AIC information has been removed) (AIC information has been removed)	(AIC information has been removed)	Eligible sample: (AIC information has been removed) Sample included: (AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)

TABLE 28 Patient characteristics of included studies: NPI+ test (information submitted by manufacturer)

Author (year)	Age (years), mean (SD)	LN status	ER status	Tumour size	Grade	HER2 status	Mean NPI score	Treatment
(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)
(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)	(AIC information has been removed)

SD, standard deviation.

TABLE 29 Summary of evidence for the NPI+ test

Author (year)	Evidence type	Overall quality	Key findings
(AIC information has been removed)	(AIC information has been removed)	Moderate	(AIC information has been removed)
(AIC information has been removed)	(AIC information has been removed)	Low	(AIC information has been removed)

(AIC information has been removed.)

Summary of evidence: Nottingham Prognostic Index plus

The evidence base for NPI+ is currently insufficient to draw any firm conclusions regarding the analytic and clinical validity of the test, and as yet there is no available evidence on the clinical utility of the test. Further evidence on the prognostic ability of the test is required. According to the unpublished abstract from the manufacturers of the test, validation in an external cohort is ongoing but as yet results are not available.

Chapter 3

Economic analysis

A systematic review of existing cost-effectiveness evidence is reported in the following section. This is followed by reviews of the economic evaluations submitted by two of the manufacturers/sponsors of the tests in response to the request for information issued by NICE at the start of the assessment process. The relevance of existing cost-effectiveness evidence for NICE decision-making is then summarised. This is followed by a description of the independent economic model and its results, and a comparison of the independent economic model with the evaluations from the two manufacturers/sponsors. Finally, the independent economic model results are discussed.

Systematic review of existing cost-effectiveness evidence

This section of the report describes a review of the existing published evidence on the cost-effectiveness of GEP and expanded IHC (or protein expression profiling) tests to guide the use of adjuvant chemotherapy in breast cancer management.

Methods

A systematic search of the existing literature evaluating the cost-effectiveness of the nine GEP and expanded IHC tests identified by NICE (OncotypeDX, MammaPrint, Mammostrat, PAM50, BluePrint in combination with MammaPrint, IHC4, Randox BCA, BCI and NPI+) to guide adjuvant chemotherapy treatment decision-making in the management of early breast cancer was undertaken. Only full economic evaluations published in English addressing the cost-effectiveness of those tests compared with NPI, Adjuvant! Online or any adaptations of these tools in clinical practice were included in the review. Cost-effectiveness studies that used St Gallen, the National Comprehensive Cancer Network (NCCN)[129] and NIH guidelines[101] were excluded from the review because of time and resource constraints as these comparators are not directly relevant to the UK context, but such studies were scanned by the reviewers to inform the model development.

The following databases were searched for relevant published literature: MEDLINE, MEDLINE In-Process & Other Non-Indexed Citations (Ovid), CINAHL, EMBASE, NHS Economic Evaluation Database and HTA (via the Cochrane Library), Web of Science (which includes the Science Citation Index) and BIOSIS. In addition, literature searches were undertaken for the clinical effectiveness review (see *Chapter 2, Methods for reviewing effectiveness*) and relevant cost papers were identified from these searches. In addition, the reference lists of relevant articles were hand searched. Full details of the search strategies used in MEDLINE are presented in *Appendix 1* (these have been adapted for use in other databases). Searches were not restricted by language.

Studies were selected for inclusion through a two-stage process. Titles and abstracts were examined for inclusion by one reviewer. Full manuscripts of selected citations were then retrieved and assessed by the same reviewer. The quality of the cost-effectiveness studies was assessed using a critical appraisal checklist adapted from the Drummond and Jefferson[130] and Eddy[131] checklists.

The aim of the review was to identify published economic evaluations and summarise the main limitations of the existing models. Because of time constraints it was not possible to provide a detailed direct comparison of the models.

Results

Identified studies

The search retrieved 72 citations relating to cost-effectiveness (*Figure 8*) and two additional references were known by the authors. Fifty-six articles were excluded at title stage and four articles were excluded at abstract level. Thirteen studies (corresponding to 14 references) were examined at full-text level and four studies (corresponding to five references) were identified as meeting the inclusion criteria of the systematic review of economic evaluations.[132–136] This included an economic evaluation developed as part of the Ontario Health Technology Assessment (OHTA) described in a report[134] and PowerPoint presentation slides.[136]

Nine articles were excluded after retrieving the full text because there were insufficient details to assess the validity of assumptions,[137] the economic evaluation was available only in abstract form,[138–140] the study used a different comparator[141] or for other reasons.[143–145] Klang *et al.*[141] was excluded from the review as the exact nature of the comparator defined as clinical practice in Israel was unclear.

Of the four identified economic studies (corresponding to five references), two compared MammaPrint against Adjuvant! Online[132,133] and two compared OncotypeDX against Adjuvant! Online.[134–136] None of the four published economic evaluations was conducted in a UK setting.

Description of published cost-effectiveness studies evaluating the use of MammaPrint

Two economic evaluations compared treatment guided using the MammaPrint test with treatment guided using Adjuvant! Online and used a health-care payer perspective.[132,133] A Markov approach was employed in both economic evaluations but the populations considered in the models differed slightly. Retel *et al.*[133] addressed the cost-effectiveness of MammaPrint in

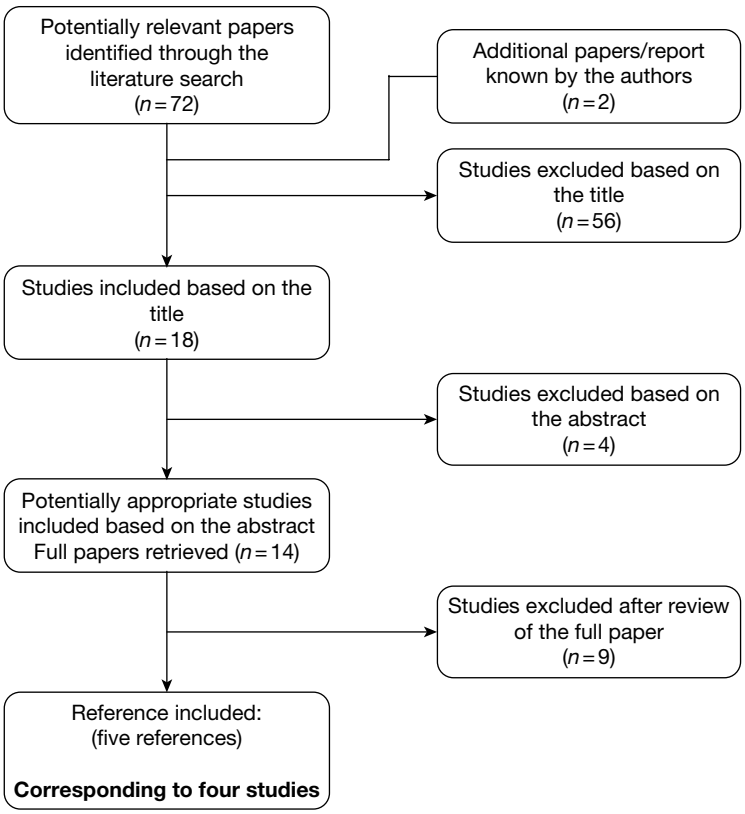


FIGURE 8 Flow diagram of economic evaluation selection/exclusion.

women with LN–, ER+, HER2+/– early breast cancer. Chen *et al.*[132] included ER– early breast cancer but excluded HER2+ early breast cancer.

A tabulated summary of the key features and data sources and a quality assessment for the two studies included in the cost-effectiveness review of MammaPrint are presented in *Appendices 16* and *17* respectively. It was not possible for the EAG to check the economic models as only the publications were available in the public domain.

A narrative description and assessment of each economic evaluation is presented in the following sections.

Description and critique of Retel et al.*:[146] cost-effectiveness of the 70-gene signature compared with St Gallen guidelines and Adjuvant! Online for early breast cancer*

Brief overview The aim of the study was to estimate the cost-effectiveness of MammaPrint compared with that of the St Gallen guidelines and Adjuvant! Online to guide adjuvant treatment decisions in Dutch women with LN–, ER+, HER2 +/– early breast cancer.

The model used a Markov approach and followed women over 20 years in four possible health states: disease free, relapse (local, regional or contralateral relapse), distant metastasis and death. The study adopted the perspective of the Dutch health-care payer and costs and QALYs were discounted at 4.0% and 1.5% per annum respectively. The mean age of patients entering the model was 50 years.

Summary of effectiveness data For the strategy using MammaPrint and Adjuvant! Online, the sensitivity and specificity were calculated and patients were classified into four risk groups of developing distant metastasis: true high, true low, false high and false low. The sensitivity and specificity were calculated from a pooled analysis of three validation studies[60,63,93] using 10-year BCSS as a final outcome (thus, patients were classified as low or high risk according to the probability of survival rather than the probability of developing distant metastasis).

The study classified patients as low risk with Adjuvant! Online if the predicted 10-year survival was >88%. MammaPrint classified patients into two categories: low (good prognosis) and high (poor prognosis) risk of developing distant recurrence. Patients classified as high risk either by MammaPrint or by Adjuvant! Online were assumed to receive chemotherapy in addition to endocrine therapy. Low-risk patients were assumed to receive endocrine therapy alone.

Patients classified as true low and false high had a zero probability of relapse and distant metastasis. The probability of relapse and distant metastasis for true high-risk patients was based on an analysis conducted in a sample of 20,624 Swedish breast cancer patients derived from Lidgren *et al.*,[147] with a constant risk within three time periods: 1–5, 5–10 and 10–20 years. Patients classified as false low were assumed to have a 100% probability of developing distant metastases.

Patients could have only one relapse (and then possibly move to distant metastasis). In total, 10% of high-risk patients were assumed to be HER2+, with a risk twice as high as observed for HER2– patients. A relative risk reduction with a HR of 0.64 (95% CI 0.54 to 0.76) was applied for patients treated with trastuzumab. Adverse events associated with chemotherapy were included for chronic heart failure only.

Finally, utilities (see *Appendix 16*) were measured using the European Quality of Life-5 Dimensions (EQ-5D).[148]

Summary of resource utilisation and cost data The costs of the health states (see *Appendix 16*) were based on Lidgren *et al.*[148] Drug costs for chemotherapy and hormonal therapies were based on Dutch sources. Chemotherapy costs included drug costs, day-care costs (including administration), laboratory and diagnostic imaging costs (mammography, tumour markers) and prevention. The cost of MammaPrint was assumed to be €2675. Costs were expressed in 2005 euros.

Summary of cost-effectiveness The results for the base case are presented in *Table 30*. Treatment guided using MammaPrint was associated with a cost per QALY gained of €4614 compared with Adjuvant! Online.

The impact of key model parameters was examined in one-way univariate sensitivity analysis and probabilistic sensitivity analysis (PSA) and showed that the results were sensitive to data used to calculate the sensitivity and specificity of the tests (5-year risk of distant metastasis instead of 10-year risk of BCSS, and using data for each individual validation study) and the cost of chemotherapy.

Comments Based on the description of the model, this appears to be a reasonably well-conducted cost-effectiveness analysis, although it has a number of limitations. The generalisability of the results from this study to the UK context is limited.

The study used sensitivity/specificity of the tests to reclassify patients into risk group categories. Patients were classified according to their risk of developing distant metastasis, but the sensitivity/specificity were calculated using the 10-year BCSS risk in the base-case analysis. Using the 5-year risk of distant metastasis instead of the 10-year of risk of BCSS was tested in sensitivity analysis. An assumption has also been made when calculating the sensitivity/specificity of the tests that low-risk patients cannot die from breast cancer. Although low-risk patients are less likely to die from breast cancer, they could still die from their cancer. There are also some limitations associated with the use of the sensitivity/specificity for tests providing a continuous risk score (especially for Adjuvant! Online).

The evaluation assumes that the decision to receive chemotherapy will be based on the test results for MammaPrint and Adjuvant! Online alone. However, it is likely that MammaPrint would be used in conjunction with other clinical parameters to inform the treatment recommendation. The assumption that the prognostic test results and treatment guidelines would be followed in all cases (patients and physicians are 100% compliant) is simplistic. Furthermore, it is unclear if the cut-off of ≤ 88% used to identify high-risk patients with Adjuvant! Online reflects actual clinical practice. Discussions with clinical experts indicated that the risk score estimated using Adjuvant! Online on its own is less informative than the complete output, which includes estimates of reduction in risk at 10 years of breast cancer-related death or relapse for selected treatments.

TABLE 30 Base-case results for the cost-effectiveness of MammaPrint compared with Adjuvant! Online estimated by Retel *et al.*[146a]

	Life-years	QALYs	Cost (€)	Cost/QALY gained (€)	Cost/life-year gained (€)
Adjuvant! Online	15.68	12.20	26,915		
MammaPrint	15.88	12.44	28,045	4614	5736

a Results for the cost-effectiveness of MammaPrint compared with the St Gallen guidelines are not presented here but are available in the original paper.[146]

The test was assumed to be administered to women with both HER2+ and HER2– early breast cancer; however, UK clinical opinion indicated that the vast majority of patients with HER2+ early breast cancer are typically offered chemotherapy and the MammaPrint test may therefore be considered unnecessary.

The authors also assumed that patients classified as low risk (true low or false high) have a zero probability of having a relapse or distant metastasis. This seems to be a very simplistic assumption. Furthermore, the authors modelled only one relapse but acknowledge that about 30% of patients develop more than one relapse.

Many assumptions have also been made about the probability of moving between health states, and the impact of chemotherapy is unclear. The risk of recurrence for patients treated with endocrine therapy was extracted from studies of patients receiving tamoxifen only; however, more effective agents are now available, potentially reducing the risk of recurrence. The starting age of the cohort was low (50 years) given that the majority of breast cancers are diagnosed in women >50 years of age. Finally, the use of fresh tissue samples for MammaPrint will have service configuration implications for UK pathology laboratories.

Description and critique of Chen et al.:[132] cost-effectiveness of the 70-gene MammaPrint signature in node-negative breast cancer

Overview The aim of the study was to estimate the cost-effectiveness of MammaPrint compared with Adjuvant! Online to guide the adjuvant treatment decision in US patients aged ≤60 years with ER+/–, T1 or T2, LN–, HER2– breast cancer. The model used a Markov approach and followed patients over their lifetime in three possible health states: disease free, death from cancer and death from other causes. The study adopted the perspective of the US payer, and costs and QALYs were discounted at 3.0% per annum.

Summary of effectiveness data Two separate models were constructed using effectiveness data from a validation study[61] and Surveillance, Epidemiology and End Results (SEER) data[149] to reflect US clinical practice (the alternative model). This was carried out as no low-risk patients with ER– tumours were included in the Buyse *et al.*[61] study.

In the base-case model, the risk classification and 10-year OS were extracted from Buyse *et al.*[61] In the alternative model, the risk reclassification was adapted from SEER and Buyse *et al.*,[61] assuming the same rate of cross-classification between high- and low-risk patients as observed in the Buyse *et al.* study, as data for MammaPrint were unavailable. A range of assumptions was necessary to use the SEER data.

Patients with ER+ early breast cancer were assumed to receive endocrine therapy (tamoxifen) whereas ER– patients were not; chemotherapy was given to patients classified as high risk only. The benefit of chemotherapy was extracted from a meta-analysis of RCTs (EBCTCG 1998),[150] applying a reduction in all-cause deaths of 26% in ER+ and 32% in ER– patients.

Utilities used to calculate quality of life were extracted from the published literature.[46,151] A utility of 0.70 was applied for patients undergoing chemotherapy for 6 months and 0.98 for patients after completion of chemotherapy or disease free. The authors did not report the valuation method or the quality of life instrument used to estimate the utility values.

Summary of resource utilisation and cost data Costs are presented in 2007 US dollars. Costs included the costs of endocrine therapy, chemotherapy, administration, treatment-related toxic effects and breast cancer surveillance. The cost of recurrence and terminal care (with cancer) was

included for women dying from cancer. A cost of terminal care was included for women dying from other causes.

The cost of chemotherapy was derived from insurance claims[152] and included the costs of chemotherapy medication, hospitalisation and emergency room for chemotherapy adverse events or all causes, ambulatory encounters and prescription. The study included patients receiving alkylating agents (58%), anthracyclines (51%), taxanes (25%) and antimetabolites (18%). The cost of the MammaPrint test was $4200.

Summary of cost-effectiveness Results for the base-case and alternative model are presented in *Table 31*. The incremental cost-effectiveness ratio (ICER) was also presented by ER status subgroup – US$5908 per QALY gained (US$6167 per life-year) for ER+ patients. MammaPrint was dominated in ER– patients in the base-case model.

The impact of the main model parameters was examined in one-way univariate sensitivity analysis, which showed that the results were mostly sensitive to the proportion of ER+ patients classified as high risk by MammaPrint, the estimate of OS, the cost of MammaPrint and the cost of chemotherapy.

Comments Based on the description of the model, this appears to be a reasonable cost-effectiveness analysis. The generalisability of the results from this study is limited given that it is based on the US health-care system.

The model is simplistic – patients either stay alive or die. The impact of recurrence is incorporated only in terms of the cost for patients dying from breast cancer. This ignores the health effect. Furthermore, a proportion of patients will have a relapse but not die from breast cancer. The authors also did not discuss the selected cut-off for Adjuvant! Online. It was unclear how the benefit of chemotherapy was applied to breast cancer deaths.

The benefit of chemotherapy was extracted from a meta-analysis and was assumed to be the same whether patients were classified as low or high risk with MammaPrint or Adjuvant! Online.

Furthermore, there were limitations in the data used. As highlighted by the authors, no low-risk patients with ER– early breast cancer were included in the Buyse *et al.* study.[61] Therefore, an alternative model was constructed using SEER data. However, a series of assumptions were necessary in order to make use of the SEER data, increasing the uncertainty relating to these results.

The authors stated that patients can experience local, regional or distant recurrence before death. It is unclear what the relative contribution of each of the types of relapse was on breast

TABLE 31 Base-case results for the cost-effectiveness of MammaPrint compared with Adjuvant! Online estimated by Chen *et al.*[132]

	Life-years	QALYs	Cost ($)	Cost/QALY gained ($)	Cost/life-year gained ($)
Base-case model					
Adjuvant! Online	21.596	21.065	162,140		
MammaPrint	21.739	21.218	163,580	9428	10,059
Alternative model					
Adjuvant! Online	20.659	21.191	163,108		
MammaPrint	21.230	21.751	163,509	702	716

cancer survival. This can have implications in terms of costs and health effects if included in the economic model. Likewise, the authors report neither the valuation method nor the quality of life instrument used to estimate utilities. No PSA was conducted.

Furthermore, the risk of recurrence for patients treated with endocrine therapy was extracted from patients receiving tamoxifen only; however, more effective agents are now available, reducing the risk of recurrence. The use of fresh tissue samples associated with MammaPrint will have service configuration and cost implications for UK pathology laboratories.

The health-state utility value for the recurrence-free health state was high (0.98). Evidence indicates that the utility in the general population for a similar age cohort would be lower.[153] Less gain would be accrued in the model by preventing a recurrence if a lower value was used.

The mean age of patients entering the model is unclear. The economic evaluation considered only patients aged ≤ 60 years. MammaPrint is now licensed for both younger and older women with breast cancer; however, the cost-effectiveness of the test in an older population is not known.

Description of published cost-effectiveness studies evaluating the use of OncotypeDX

Two economic evaluations[134–136] compared treatment guided using OncotypeDX with that guided using Adjuvant! Online and used a health-care perspective. The same model structure was used in both studies, with the model developed by Tsoi *et al.*[135] being made available to the OHTA and adapted.[134,136] Both studies addressed the cost-effectiveness of OncotypeDX in Canada in women with LN–, ER+, HER2– early breast cancer. The mean age of women entering the model was 50 years.

A tabulated summary of the key features and data sources and a quality assessment for the two studies included in the cost-effectiveness review of OncotypeDX are presented in *Appendix 18*. It was not possible for the EAG to check the economic models as only the publications were available in the public domain.

A narrative description and assessment of each economic evaluation is presented in the following sections.

Description and critique of Tsoi et al.*:[135] cost-effectiveness analysis of recurrence score-guided treatment using a 21-gene assay in early breast cancer*

Overview The aim of the study was to estimate the cost-effectiveness of OncotypeDX compared with Adjuvant! Online to guide the adjuvant treatment decision in Canadian patients with LN–, ER+, HER2 early breast cancer. The model used a Markov approach using a monthly cycle and followed patients over their lifetime in four possible health states: chemotherapy, recurrence free, distant recurrence and death. The study adopted the perspective of Canadian health care, and costs and QALYs were discounted at 5.0% per annum. The mean age of patients entering the model was 50 years.

Summary of effectiveness data The probability of reclassification was based on Bryant *et al.*[45] High and intermediate risks defined by OncotypeDX were grouped together. High-risk patients according to Adjuvant! Online were defined as patients with a 10-year mortality ≤ 91%. Patients were first classified according to Adjuvant! Online (low vs. high). For the strategy including OncotypeDX, patients classified as low or high risk using Adjuvant! Online were further reclassified into low and intermediate/high risk using OncotypeDX. Patients were assumed

to receive chemotherapy if they were considered at intermediate/high risk and entered the chemotherapy state for 6 months during which they might experience toxicity. The probability of developing toxicity (major and minor) was obtained from the literature.[154,155] Patients in the recurrence-free state received tamoxifen for 5 years. Patients could develop distant metastases, remain disease free or die. Death from other causes than breast cancer was included.

The 10-year risk of recurrence was obtained from Paik *et al.*[49] for each risk group category (for both the Adjuvant! Online and OncotypeDX arms). A relative risk reduction of 30%, taken from a meta-analysis conducted by the EBCTCG,[29] was applied for patients classified in the high-risk group to represent the effect of chemotherapy. The median survival after distant metastasis was assumed to be 21 months. The probabilities were assumed to follow an exponential distribution.

Utility values were extracted from the published literature and were estimated using different approaches, including standard gamble and a visual analogue scale.

Summary of resource utilisation and cost data Costs were reported in 2008 Canadian dollars. The cost of chemotherapy was obtained from the Sunnybrook Odette Cancer Centre pharmacy, Toronto, Ontario. In the base case, patients were assumed to receive four cycles of doxorubicin and cyclophosphamide (AC). Other chemotherapy regimes were considered in sensitivity analysis [four cycles of docetaxel and cyclophosphamide (TC) and six cycles of 5-flurouracil, epirubicin and cyclophosphamide-docetaxel (FEC-D)]. The costs of chemotherapy included the costs of the chemotherapeutic agent, supportive medications, laboratory evaluation and human resources.

No costs were assumed for minor toxicities as it was assumed that they were already included in the cost of supportive medication. The cost of major toxicities included the cost for the management of febrile neutropenic complications and growth factor support. The model included the cost of fatal toxicities. The cost of hormonal treatment was applied to all patients for 5 years or until death. In addition to the costs of the health state (recurrence free and recurrence), the model included the cost for terminal care.

The cost of OncotypeDX was assumed to be C$4404.

Summary of cost-effectiveness Results for the base-case analysis are presented in *Table 32*.

The impact of changes in the main model parameters was examined in one-way univariate sensitivity analysis, which showed that the results were sensitive to the reclassification probabilities, recurrence rates used, discounting, baseline age and cost of OncotypeDX.

Comments Based on the description of the model, this appears to be a reasonably well-conducted cost-effectiveness analysis. The generalisability of the results from this study to the UK context is, however, limited.

TABLE 32 Base-case results for the cost-effectiveness of OncotypeDX compared with Adjuvant! Online estimated by Tsoi *et al.*[135]

	Life-years	QALYs	Cost (C$)	Cost/QALY gained (C$)	Cost/life-year gained (C$)
Adjuvant! Online	13.933	13.573	15,645		
OncotypeDX + Adjuvant! Online	13.997	13.638	19,747	63,064 (approx. £39,917[a])	63,911 (approx. £40,466[a])

a 1C$ = £0.632967 (www.xe.com, accessed 22 September 2011).

The study assumed that the decision to receive chemotherapy will be based on OncotypeDX or Adjuvant! Online alone; however, it is likely that both tools will be used in clinical practice to inform the treatment recommendation. It is also assumed that the prognostic test results and treatment guidelines would be followed in all cases (patients and physician are 100% compliant). This is unlikely to be the case in clinical practice.

High- and intermediate-risk group patients identified by OncotypeDX were grouped together and assumed to receive chemotherapy. However, it is unclear from existing studies whether or not patients classified in the intermediate-risk group would benefit from chemotherapy. Furthermore, the benefit of chemotherapy was assumed to be the same irrespective of OncotypeDX risk score. There is some evidence to suggest that high-risk patients gain a greater proportionate benefit, although this evidence has a number of weaknesses.

There was also an issue with the definition of risk groups. Bryant *et al.*[45] rank order outputs from Adjuvant! Online so that a similar proportion of cases would be categorised as low risk (50%) as for OncotypeDX. This is arbitrary and therefore may introduce biases into the analysis.

Local and regional recurrences were not included in the model. No long-term adverse events were included.

The risk of recurrence for patients treated with endocrine therapy was extracted from data on patients receiving tamoxifen only; however, more effective agents are now available, reducing the risk of local and systemic recurrence.

Finally, utilities were extracted from a variety of sources using different valuation methods. This might bias the cost-effectiveness results. The starting age of the cohort was low (50 years) compared with the average age of patients presenting with early breast cancer in the UK.

Description and critique of OHTA analysis:[134,136] gene expression profiling for guiding adjuvant chemotherapy decisions in women with breast cancer

Overview The aim of the study was to estimate the cost-effectiveness of Adjuvant! Online in combination with OncotypeDX compared with Adjuvant! Online alone to guide the adjuvant treatment decision in Canadian patients (Ontario) with LN–, ER+, HER2– early breast cancer. The analysis was built on the economic model developed by Tsoi *et al.*[135]

Compared with the original model,[135] the OHTA analysis classified patients into low, intermediate and high risk using OncotypeDX or Adjuvant! Online, analysed all possible combination to give OncotypeDX to specific group of patients according to the Adjuvant! Online score (all patients, low, intermediate or high only, intermediate and high), modelled different chemotherapy regimens and conducted a PSA.

The majority of costs have been inflated from Tsoi *et al.*[135] to reflect 2010 prices. The cost of OncotypeDX was updated to C$4191. The authors also stated that the cost of chemotherapy was updated but this was not reported. The benefit of chemotherapy was assumed to be different between risk group categories, based on evidence from the Paik *et al.* study.[48] As in the original analysis, the risk reclassification and the probability of distant recurrence were derived from Bryant *et al.*[45] and Paik *et al.*[49]

Summary of cost-effectiveness Incremental analysis was conducted comparing the most effective strategy with the next most effective strategy (*Table 33*). Assuming that OncotypeDX was

TABLE 33 Base-case results for the cost-effectiveness of OncotypeDX compared with Adjuvant! Online estimated by the OHTA analysis (2010)[134,136]

	Cost (C$)	QALYs	ICER (C$)
No patients	13,298	13.34	
Adjuvant! Online high risk	13,660	14.04	518 (approx. £328[a])
Adjuvant! Online intermediate/high risk	13,961	14.42	795 (approx. £503[a])
All patients	17,466	14.64	23,983 (approx. £15,179[a])

a 1C$ = £0.632967 (www.xe.com, accessed 22 September 2011).

provided only to high-risk patients classified by Adjuvant! Online resulted in an estimated ICER of C$518 per QALY gained compared with not using OncotypeDX. The ICER was C$795 per QALY gained if OncotypeDX was given to intermediate- and high-risk patients compared with high-risk patients only classified by Adjuvant! Online. Finally, giving OncotypeDX to all patients is associated with higher benefit and costs. The ICER comparing this strategy (OncotypeDX to all patients) with OncotypeDX given only to patients classified as high and intermediate risk by Adjuvant! Online was C$23,983 per QALY gained.

Univariate sensitivity analysis was performed. PSA was also performed using Monte Carlo simulation. The PSA indicated that, at the willingness-to-pay threshold of $75,000 per QALY gained, the probability that OncotypeDX is cost-effective is 83.5% for patients identified as Adjuvant! Online low risk, 99.8% for patients identified as Adjuvant! Online intermediate risk and 65.8% for patients identified as Adjuvant! Online high risk.

Comments Based on the description of the model, this appears to be a reasonably well-conducted cost-effectiveness analysis. The generalisability of the results from this study to the UK context are, however, limited.

The description of the model and its assumptions is minimal in the report, but this is explained by the fact that this is an adaptation of a previously published cost-effectiveness evaluation. The authors were also contacted and a greater description of the model and results are due for publication soon. Despite the adaptations, key limitations remain regarding the data used to reclassify patients and for utility estimates, the probability of distant metastases, the type of relapse modelled, long-term adverse events after chemotherapy and the benefit of chemotherapy.

Assessment of the economic evaluation submitted by Genomic Health

An economic evaluation was submitted by Genomic Health[27] comparing the use of OncotypeDX with current clinical practice in the UK and included a full report and an electronic model submitted in Microsoft Excel (Microsoft Corporation, Redmond, WA, USA). The economic model was reviewed to check that the parameters presented in the report corresponded to those used in the economic model and assessed using a critical appraisal checklist adapted from the Drummond and Jefferson[130] and Eddy[131] checklists (*Table 34*).

TABLE 34 Critical appraisal of the economic model submitted by Genomic Health[27]

Modelling assessments should include:		Economic evaluation submitted by Genomic Health[26]
1	A statement of the problem	Yes
2	A discussion of the need for modelling vs. alternative methodologies	Yes
3	A description of the relevant factors and outcomes	Yes
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 submodel	Yes
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	Yes
6	A list of assumptions pertaining to the structure of the model (e.g. factors included, relationships and distributions) and the data	Yes
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 which will be used in a sensitivity analysis	Yes
8	The results derived from applying the model for the base case	Yes
9	The results of the sensitivity analyses: unidimensional, best/worst case, multidimensional (Monte Carlo/ parametric), threshold	Yes
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	Yes
11	A description of the validation undertaken including concurrence of experts, internal consistency, external consistency, predictive validity	Unclear
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	Unclear
13	A description of research in progress that could yield new data that could alter the results of the analysis	Unclear

Description of the economic model submitted by Genomic Health

Overview

The economic model submitted by Genomic Health[27] used a Markov approach with individuals moving between three possible health states: recurrence free, distant recurrence and death (from breast cancer or other causes) (*Figure 9*). The model compared the cost-effectiveness of the addition of OncotypeDX to clinical and pathological parameters (using NPI and Adjuvant! Online; termed usual care in the economic model submitted by Genomic Health) with that of clinical and pathological parameters alone in women with ER+ and LN– or single node-positive early breast cancer in the UK. The starting age in the model was 60.55 years and patients were followed up for 30 years. The study adopted the perspective of the UK NHS, with costs and QALYs discounted at 3.5%. A tabulated summary of the key features and data sources of the economic model submitted by Genomic Health is presented in *Table 35*.

The structure was based on an original model by Hornberger *et al.*[46] Patients with ER+ and LN– or single node-positive [pN1(mic)] early breast cancer with no contraindications for adjuvant chemotherapy are assigned adjuvant therapy based on:

- clinical and pathological parameters alone (using NPI and Adjuvant! Online) or
- the addition of OncotypeDX RS to usual care [terminology used in the Sponsor submission (SS)].

Patients are categorised as low, intermediate or high risk according to the OncotypeDX classification. Among each risk group category patients are further divided according to the treatment they received (either hormonal therapy alone or hormonal therapy in addition to chemotherapy). In each cycle of the model, the risk of recurrence was evaluated for each

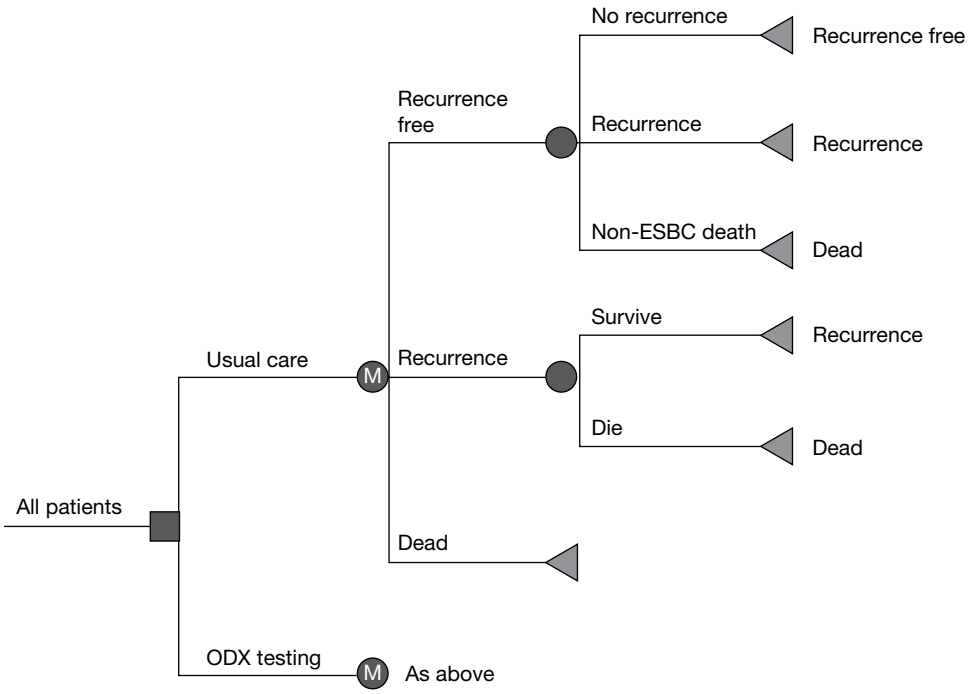


FIGURE 9 Overview of the OncotypeDX cost-effectiveness model structure submitted by Genomic Health (reproduction of figure 6–2 in the report submitted by Genomic Health).[27]

simulated patient based on their RS-defined category of low, intermediate or high risk as reported for the NSABP B20 cohort.[49] The risk was then adjusted for patients who received chemotherapy, based on whether or not patients received chemotherapy as per the initial recommendation (in the arm termed usual care) and based on the recommendation following the additional information provided by the OncotypeDX RS. The benefit of chemotherapy varied by OncotypeDX risk group, based on Paik *et al.*[49] For PSA, recurrence risks and relative risk reductions for chemotherapy were sampled from normal distributions, with the assumed variance derived from published data. Non-breast cancer death was captured as a competing risk in the model, based on UK life tables for women in 2007–9.[156] For patients experiencing distant recurrence, the median survival was assumed to be 3.3 years.[157]

Summary of effectiveness data

The impact of OncotypeDX on treatment recommendations was obtained from the preliminary results from a Welsh cohort study by Holt *et al.*[78] reporting on the first 107 patients.[27] The study considered the treatment recommendations made based on usual care (chemotherapy or no chemotherapy) and then the treatment recommendations made following the additional knowledge of the OncotypeDX test result. In this study, 33% of patients had their initial treatment recommendations changed following OncotypeDX testing.[27,78] Not all treatment decisions were directly influenced by the high- or low-risk category from the report, that is, some high-risk patients did not receive chemotherapy and vice versa (*Table 36*).

The risk of recurrence for patients in the low-, intermediate- and high-risk groups assessed by OncotypeDX and the impact of chemotherapy on risk of recurrence by risk group was taken from Paik *et al.*[49] The risk of recurrence was assumed to be constant over time, modelled by an exponential distribution. All patients within each OncotypeDX risk category were assumed to have the average risk of recurrence for that group.

TABLE 35 Tabulated summary of the key features and data of the economic model submitted by Genomic Health[27]

Parameter	Key features/data
Country	UK
Perspective (costs)	NHS and PSS
Comparators (NPI, Adjuvant Online!)	Usual care (NPI and Adjuvant Online!)
Starting age in the model	60.55 years
Population	ER+, LN– (or single node-positive) early breast cancer
Model structure (type, health states)	Markov model with three health states (recurrence free, distant recurrence and death)
Definition of relapse	Distant recurrence only
Time horizon	30 years
Endocrine therapy regime	Mixed – tamoxifen and aromatase inhibitors – according to NICE guidance (see text)
Chemotherapy regime	Six cycles of 5-fluorouracil, epirubicin and cyclophosphamide (FEC75)
Benefit of chemotherapy by RS risk group	Low-risk group: no benefit (assumed); intermediate group: 39%; high risk: 74%
Adverse events	Short-term adverse events included in the cost and disutility associated with chemotherapy
Other assumptions	The probability of dying after distant metastases was the same irrespective of risk classification. Transitional probabilities are assumed to be exponential
Definition of high risk	In the usual care arm patients were offered chemotherapy based on the treatment decision taken using NPI and Adjuvant Online! In the OncotypeDX arm patients were offered chemotherapy based on the treatment decision taken using NPI and Adjuvant Online! and knowledge of the OncotypeDX test result. Note that the results of the test were not always followed
Quality of life	Different sources, valuation methods. Recurrence free = 0.78; decrement from chemotherapy = 0.07; distant recurrence = 0.6
Costs and resources used	2010 UK pounds
	OncotypeDX test: £2580; chemotherapy (all cycles): £3194 (chemotherapy) + £4534 (adverse events associated with chemotherapy and use of G-CSF); recurrence free (yearly): £0; endocrine therapy (mixed): £853 (first 5 years) + £40 (adverse events first 5 years) + £123 (adverse events 6–8 years); DM (3.3 years): £916 monthly; terminal care (last 3 months): £0
Discounting (per annum)	3.5% for both costs and benefits
Uncertainty	One-way and probabilistic sensitivity analysis
% of HER2+	Unclear from the submission
Cost per QALY	£6232

DM, distant metastasis; G-CSF, granulocyte colony-stimulating factor; PSS, Personal Social Services.

The probability of dying from distant metastasis was derived from Thomas *et al.*,[157] assuming a median life expectancy of 3.3 years. Again, this was assumed to be the same for all risk groups.

Utilities were extracted from the published literature. The quality of life associated with recurrence (0.60) was taken from Milne *et al.*,[158] who reported an analysis in New Zealand women with advanced breast cancer and assumed treatment with endocrine therapy. The disutility associated with chemotherapy (–0.07) was taken from Peasgood *et al.*[159] The health utility associated with 1 year in the recurrence-free state (0.78) was assumed to be the same during and after endocrine therapy.[160]

Summary of resource utilisation and cost data

All drug costs were taken from the *British National Formulary* (BNF).[161] Five endocrine therapy regimes were considered in line with NICE guidelines: (1) tamoxifen for 5 years, (2) anastrozole for 5 years, (3) letrozole for 5 years, (4) tamoxifen for 2 years plus exemestane for the final 3 years

TABLE 36 Proportions of patients in the preliminary analysis of the Holt *et al.*[78] study receiving chemotherapy before and after OncotypeDX testing (by RS)[27]

Initial recommendation			Post ODX	
RS Group	HT	CT	RS Group	HT
Low	30.5%	23.8%	Low	30.5%
Int.	16.2%	10.5%	Int.	16.2%
High	8.6%	10.5%	High	8.6%
All	55.24%	44.76%	All	55.24%

CT, chemotherapy; HT, hormonal therapy; int., intermediate.

and (5) tamoxifen for 5 years followed by extended therapy with letrozole for a further 3 years. The probability that a patient was treated with each regime was taken from NICE TA112.[162] The annual cost over the first 5 years was £669.03 and the annual cost over the following 3 years was £108.40. All patients were assumed to be 100% compliant with endocrine therapy. Adverse event probabilities and costs of adverse events associated with endocrine therapies were derived from Hind *et al.*[163] and inflated to 2010 prices.

Patients treated with chemotherapy were assumed to receive six cycles of FEC75 (5-fluorouracil, epirubicin and cyclophosphamide) based on the regime description of FEC75 given by Avon, Somerset and Wiltshire Cancer Services in its chemotherapy protocol documents.[27] Administration costs were taken from the National Schedule of Reference Costs 2009–10[164] for NHS trusts on an outpatient basis. The costs of adverse events and the probability of their occurrence was taken from Wolowacz *et al.*[165] for patients treated with CAF in the absence of sufficient evidence for FEC. The model did not include long-term adverse events associated with chemotherapy.

Finally, it was assumed that all patients were treated with granulocyte colony-stimulating factor (G-CSF) in order to prevent neutropenia.

Summary of cost-effectiveness

The base-case analysis is presented in *Table 37*. The addition of OncotypeDX to current practice (clinical and pathological parameters) resulted in an ICER of £6231.91 per QALY gained.

One-way sensitivity analysis showed that the base-case outcomes were most sensitive to variation in patient age, the cost of OncotypeDX testing and the change in chemotherapy recommendations for low-risk patients. The PSA results indicated that, at a willingness-to-pay threshold of £20,000 per QALY, there was a 99.6% probability that OncotypeDX would be cost-effective compared with current clinical practice (*Figure 10*).

A scenario analysis was also presented for node-positive patients using data from treatment decisions in node-positive patients in the German setting.[166] (Commercial in-confidence information has been removed.)

TABLE 37 Base-case result for the cost-effectiveness of OncotypeDX compared with current clinical practice in the UK estimated by Genomic Health (reproduction of table 6–13 in the report submitted by Genomic Health)[27]

	Usual care	OncotypeDX testing	Difference
Cost (£)	11,847.24	12,734.93	887.69
QALYs	11.39	11.54	0.14
Life expectancy (years)	14.73	14.89	0.16
ICER (£/QALY gained)	6231.91		
ICER (£/life-year gained)	5633.30		

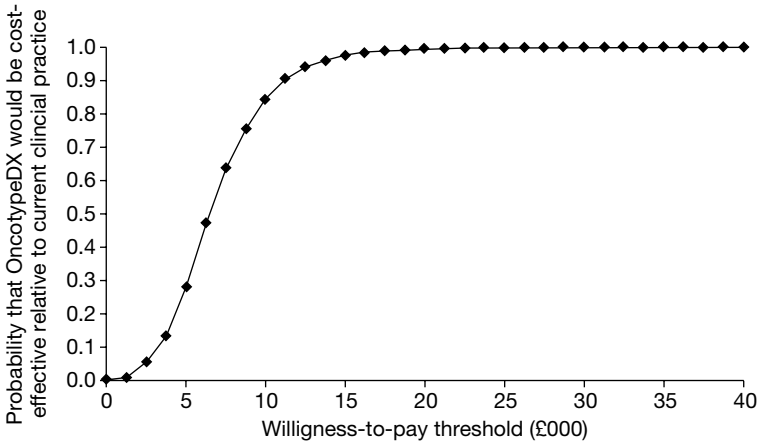


FIGURE 10 Cost-effectiveness acceptability curve for the probability of OncotypeDX being cost-effective at various willingness-to-pay thresholds estimated by Genomic Health.[27]

Critique of the economic evaluation submitted by Genomic Health

The EAG has reviewed the economic model and report submitted by Genomic Health. A detailed critique is presented below. In summary, the model was considered to be of a good standard given the evidence available; however, there are a number of limitations with the model structure, assumptions and data inputs that need further consideration.

Impact of OncotypeDX on chemotherapy decision-making in the UK

The economic model used data from Holt *et al.*[78] to reflect current practice in England and Wales and the impact of OncotypeDX on treatment recommendations.[27] The study was conducted in a Welsh cohort and is the only identified evidence of the impact of the test on UK decision-making.

The EAG have several concerns regarding the use of the Holt *et al.* study. These concerns have been discussed in *Chapter* 3, *Results: OncotypeDX* and are further detailed in *Model inputs: test-specific parameters*. Briefly:

- Data used to populate the economic model were taken from a preliminary analysis conducted in a small sample of 106 patients.
- The study was conducted in Wales in two centres and it is unclear to what extent results are generalisable to the rest of England and Wales.
- It is unclear how the decision to recommend chemotherapy was made.
- There are concerns that patients may not be representative of patients seen in clinical practice in England and Wales. In *Model inputs: test-specific parameters*, the NPI distribution of patients included in the Holt *et al.* study[78] is compared with the NPI distribution of patients

from two registries [Eastern Cancer Registration and Information Centre (ECRIC) and West Midland Cancer Intelligence Unit (WMCIU)], which shows that patients included in the Holt *et al.* study were more severe – with larger tumours and a higher proportion of grade II and III tumours (analysis conducted by the EAG).
- The proportion of patients recommended for chemotherapy under current practice in the Holt *et al.* study appears to be overestimated when compared with the proportion of patients who are actually offered chemotherapy, based on data from two cancer registries in England and Wales (see *Model inputs: test-specific parameters*). If this is the case, use of these data in the model may increase the predicted benefits derived from the use of OncotypeDX, resulting in a potential overestimation of the ICER.

Because of these limitations, the EAG did not consider the Holt *et al.* study to be an appropriate study to reflect current practice in England and Wales.

Risk of recurrence

In the absence of follow-up in the Holt *et al.* study,[27,78] a separate data source was used to estimate the risk of recurrence for patients classified as being of low, intermediate or high risk of distant recurrence with OncotypeDX. The 10-year risk of distant recurrence used in the model for patients on endocrine therapy (tamoxifen) was 3.2% for the low RS group, 9.1% for the intermediate RS group and 39.5% for the high RS group.[49] The EAG expresses three main concerns with these data:

- Data were taken from a US cohort of women prior to 2006 and therefore the results might not be transferable to current treatment practice for women in England and Wales.
- The Paik *et al.* study[49] is based on pre- and postmenopausal women who received tamoxifen only; however, a mixture of different endocrine therapies is now used in the UK.
- Biases could have been introduced because two separate sources of data were used for the risk classification and risk of recurrence. Although the EAG acknowledges the rationale of the manufacturer to use two separate data sources, the EAG considers this approach to be inappropriate because of the high correlation between the two parameters. This is particularly important as the ICER is likely to be sensitive to these assumptions. Data from a previous US study indicated a 10-year risk of distant recurrence of 6.8%, 14.3% and 30.5% for the low, intermediate and high RS groups, respectively, in women treated with tamoxifen only. The TransATAC trial conducted in the UK[79] showed a 10-year risk of recurrence of 4.0%, 12.0% and 25.0% for women classified in the low, intermediate and high RS groups, respectively, based on postmenopausal, LN–, HER2+/– women treated with tamoxifen and anastrozole. In sensitivity analyses, the manufacturer shows that using data for the risk of recurrence from the TransATAC trial increases the ICER from £6232 to about £9160 per QALY gained.

In addition, the manufacturer assumed the risk of recurrence to be constant over time. Evidence shows that the hazard of distant recurrence declines with time, with a plateau after approximately 15 years.[167] Assuming a reduction in recurrences over time, and certainly beyond 15 years, would have resulted in a higher ICER as the use of OncotypeDX would prevent fewer recurrences.

Finally, the risk of recurrence was applied according to the OncotypeDX classification, that is, patients recommended chemotherapy or not using current clinical practice have the same risk of recurrence. However, it seems likely that patients who are recommended chemotherapy within current clinical practice (based on the use of the NPI and/or Adjuvant Online) have a higher risk of recurrence than patients who are not recommended chemotherapy (even within the same RS group). For example, as shown in *Model inputs: test-specific parameters*, the risk of distant recurrence for patients classified using OncotypeDX is different from the risk of distant

recurrence for patients classified by NPI. Ignoring the prognostic value of the treatment decision using clinicopathological parameters is likely to produce a more favourable ICER.

Patients who are offered the test

The model assumes that OncotypeDX is given to all women with ER+ and LN– or single node-positive early breast cancer. However, clinical opinion indicates that in the UK only a subgroup of patients might be offered OncotypeDX – those patients at intermediate risk for whom the decision for adjuvant treatment is uncertain. Assuming that all women receive the test is considered to be a conservative assumption and the ICER is likely to be more favourable if only selected patients receive the test.

Benefit of chemotherapy

Data from the Paik *et al.* study[49] were used to determine the benefit of chemotherapy. Although the study showed a consistent benefit in women classified as being of intermediate or high risk of distant recurrence with OncotypeDX, there are some concerns with the study design. Indeed, data from the training set (used to develop the test) were used to estimate the benefit of chemotherapy. This is likely to positively bias the observed effect of chemotherapy. More discussion is available in *Chapter 2*, *OncotypeDX test*. Given that the Paik *et al.* study is based on patients who received tamoxifen only, it is not clear how this evidence relates to women in the UK who currently receive a mixture of different endocrine therapies. This impact is not, however, expected to be large as the use of different endocrine therapies does not generate large differences in OS. In addition, the study included women with HER2+ early breast cancer. Those women are likely to have a high risk of distant recurrence (and are more likely to be classified with a high RS) and derive a greater benefit from chemotherapy.

The chemotherapy regimen used to define the impact of chemotherapy by risk group in Paik *et al.*[49] was CMF (cyclophosphamide, methotrexate and 5-fluorouracil) or MF (methotrexate and 5-fluorouracil). Discussion with clinical experts indicated that newer and more effective regimens are used in the UK. The economic model assumes the use of FEC75 in the UK. The impact of this assumption (on both efficacy and cost) has not been discussed. It is not known how this will influence the impact of chemotherapy on distant recurrence.

Time between recurrence and death

The economic model submitted by Genomic Health assumed that the time between distant recurrence and death was the same irrespective of the risk group. Discussion with clinical experts indicated that it is likely that the time between recurrence and death may be shorter for patients at high risk. This has not been discussed by the manufacturer and it is unclear how this would affect the ICER.

Exclusion of local and regional recurrences

The economic model submitted by Genomic Health included only distant metastases. The omission of local and regional recurrences is likely to produce a less favourable ICER as additional benefits might be accrued by the use of the new test with no additional cost.

Cost and utility associated with recurrence

The cost of recurrence was taken from the study by Thomas *et al.*,[157] which was conducted in a mixture of patients, some with ER–, HER2+ and LN– early breast cancer. The manufacturer discussed the limitations of using data from this study.

There are concerns that the cost of recurrence is applied as a one-off cost, which has implication when discounting costs. It is further assumed that patients remain in the recurrence health state for 3.3 years, whether they are aged 60 years or 90 years. However, in reality, older women are

likely to spend less time in the recurrence health state. It is unclear how this would affect the ICER as it potentially results in an overestimation of the cost of distant recurrence but also of the QALYs gained whilst in the recurrence health state.

Adverse events

Long-term adverse events associated with chemotherapy, such as cardiotoxicity and secondary cancers, are not captured in the model. This is likely to produce a less favourable ICER if the use of OncotypeDX reduces the proportion of patients receiving chemotherapy, as it does in the Genomic Health model.

Short-term adverse events were included in the model. Costs relating to the use of G-CSF to prevent neutropenia were considered to be overestimated as it was assumed that all patients receive G-CSF for each of the six cycles. This is a concern given the high cost of G-CSF in the model. The cost of G-CSF accounts for £4118 (53%) of the total cost of chemotherapy (drug, administration, monitoring, adverse events) in the model (£7728). Discussion with clinical experts indicated that in the UK G-CSF is typically used for the secondary prevention of febrile neutropenia (i.e. after an event or following a dose delay due to neutropenia); it is given only to a proportion of patients (approximately 25%) as secondary prophylaxis for all subsequent cycles (five or fewer) following an episode of febrile neutropenia or dose delay. This assumption used in the Genomic Health model is likely to produce a lower (more favourable) ICER.

Cost of endocrine therapy and chemotherapy

Finally, the economic model submitted by Genomic Health assumed no wastage and a dosage per body surface area (BSA) of 1.8 mg/m^2. A UK study reported that the mean dosage per BSA for women with breast cancer in the UK was 1.75 mg/m^2.[168] This is likely to overestimate the drug cost based on the BSA, such as the cost of chemotherapy or endocrine therapy, and therefore produce a more favourable ICER.

Probabilistic sensitivity analysis

Finally, the EAG had some concerns about the PSA conducted by the manufacturer. The benefit of chemotherapy for patients classified as low risk with OncotypeDX was not varied in the PSA. Although no benefit (reduction in distant recurrence) was observed for this group of patients (HR 1.31), the CI was wide enough (95% CI 0.46 to 3.78) that a benefit is not impossible.[49] Costs were varied in the PSA assuming a normal distribution and an arbitrary SE of 10% around the mean. Neither the proportion of patients receiving chemotherapy under clinical practice nor the classification of patients was varied in the PSA. The change in treatment allocation after knowledge of the OncotypeDX result was, however, varied using a normal distribution using an arbitrary SE of 10%. The EAG did not consider this approach appropriate as this ignores the correlation between the risk reclassification used for both the comparator and the intervention arms and the change in treatment allocation for the intervention arm.

Assessment of the economic evaluation submitted by Clarient

An economic evaluation for the use of Mammostrat in the UK (report only) was submitted by Clarient late in the appraisal process shortly before the finalisation of the EAG report.[169] Because of its direct relevance to the UK, the EAG felt it useful to report the method and main finding. Because of the late submission and time constraints, only a brief description and critique is reported thereafter.

The submission[169] included a full report only and therefore it was not possible for the EAG to check the economic model.

Brief description of the economic model submitted by Clarient

Description of the method and data inputs

The model compared the cost-effectiveness of treatment guided using Mammostrat with that of treatment guided using the NPI in women with ER+, LN– early breast cancer in the UK, including both pre- and postmenopausal women. Patients were followed up for 10 years and the study adopted the perspective of the UK NHS, with costs and QALYs discounted at 3.5%. The economic model used a Markov approach with individuals moving between three possible health states: recurrence free, recurrence (all recurrences) and death (from breast cancer or other causes).

(CIC information has been removed.)

Summary of results

(CIC information has been removed.)

(CIC information has been removed.)

Critique of the economic evaluation submitted by Clarient

It was not possible for the EAG to check the economic model as only the report was provided by the manufacturer. Based on the description of the model alone the robustness of the model cannot be verified. A large number of assumptions were made to link the evidence available and it is not possible to fully assess the impact of this. Therefore, the results from this study should be interpreted with caution.

Because of time and resource constraints and the absence of the Microsoft Excel model, it was not possible to provide a detailed critique of the economic evaluation submitted by Clarient; however, the main limitations/concerns are highlighted below:

- The model uses a 10-year time horizon. This is believed to be very short given that recurrences can usually occur after 10 years.
- (CIC information has been removed.)

(CIC information has been removed.)

Relevance of existing cost-effectiveness evidence for NICE decision-making

The existing cost-effectiveness evidence has limited relevance for the UK setting. Only two of the nine tests (OncotypeDX and MammaPrint) have any published cost-effectiveness evidence

TABLE 38 Base-case results for the cost-effectiveness of Mammostrat compared with the NPI estimated by Clarient (reproduction of table 17 in the report submitted by Clarient[169])

(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)
(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)
(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)
(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)
(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)
(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)

to date[132,134–136,146] and each presented a number of limitations (see *Systematic review of existing cost-effectiveness evidence*).

Genomic Health and Clarient each submitted an economic evaluation considering the cost-effectiveness of OncotypeDX[26] and Mammostrat[169] in the UK, respectively, and the submitted economic evaluations are therefore potentially more relevant for UK decision-making. However, there were a number of issues in the evaluations that require further consideration:

- the assumption about the baseline level of chemotherapy in clinical practice in England and Wales
- the assumption about the risk of distant recurrence in a UK population
- the assumption about the proportion of patients who would be offered chemotherapy after reclassification with the new test in England and Wales
- the assumption about who would be offered the test in England and Wales
- the assumptions about the cost of chemotherapy and the treatment of adverse events generated by the chemotherapy in England and Wales.

Independent economic model: methods

This section of the report describes the development of the de novo economic model. The following sections describe the population under assessment, the interventions to be modelled, the comparators and the subgroups of interest. The economic model is described in *Description of the de novo economic model*. This gives an overview of the model and a more detailed description of the model structure, followed by a description of the model input parameters – first, those common to all models (including costs and utilities) and, second, those that are test specific (clinical parameters).

The key objective of the economic assessment is to address the cost-effectiveness of the use of GEP and expanded IHC tests compared with current practice to guide adjuvant chemotherapy decision-making in women with early breast cancer in England and Wales. Only two of the tests, OncotypeDX and MammaPrint, have published evidence about their economic value[134–136,146] but these evaluations are not UK specific. Two UK economic evaluations were submitted by Genomic Health (OncotypeDX)[27] and Clarient (Mammostrat)[169] as part of the NICE request for additional information to the manufacturers. The review of the published cost-effectiveness evidence and the critique of the economic evaluations submitted by the manufacturers for this appraisal revealed a number of limitations that need to be addressed.

Therefore, a de novo economic model was constructed to address these limitations where possible and to estimate the cost-effectiveness of a wider range of GEP and expanded IHC tests. Notably, the EAG economic assessment uses UK-specific data and addresses limitations over the proportion of patients who currently receive chemotherapy in England and Wales and the risk of distant recurrence in a UK population; carries out a subgroup analysis offering the test to patients who are considered the most likely to benefit from the test; and seek to undertake a more accurate estimation of the cost of chemotherapy in England and Wales.

The economic model considers the selection of patients for chemotherapy using the new tests (intervention arm) compared with the selection of patients for chemotherapy using current prognostic tools (comparator arm). Patients who receive chemotherapy are assumed to experience a reduction in the risk of recurrence (and subsequent deaths) compared with those patients who receive endocrine therapy only. The costs of chemotherapy, along with the

costs and the reduction in quality of life resulting from the adverse events associated with the chemotherapy, are taken into account within the model.

Population under assessment

The NICE scope[25] identifies the population under assessment as people diagnosed with early breast cancer. However, most of the GEP and expanded IHC tests have been developed for use in a specific subpopulation or have evidence of efficacy only within a specific subpopulation (see *Chapter 2*, *Results*). The economic assessment focuses on women with ER+, LN–, HER2– early breast cancer. This subgroup was selected after review of the evidence available (see *Chapter 2*, *Results*) and the indications of the tests (see *Tables 6* and *7*), discussion with clinical experts and the perceived likelihood of the use of the test resulting in a change in current clinical practice.[170] This was considered to be the population for which the new tests had the most robust evidence base and the population in which the tests were most likely to be used in the first instance in England and Wales. Patients with HER2+ early breast cancer or with positive nodes were not considered in this assessment because of time and resource constraints and lack of evidence, but they should be the subject of future research. Of particular note, the role and cost-effectiveness of GEP and expanded IHC tests in LN+ women may be explored as part of the planned Optimal Personalised Treatment of breast cancer using Multi-parameter Analysis (OPTIMA) trial, although funding for this trial is not yet confirmed. The proposed aim of the OPTIMA trial is to identify an effective method, using multiparameter analysis, to target women with ER+, HER2 normal primary breast cancer who are likely to benefit or not from chemotherapy. A health economic evaluation is planned as part of the study (Dr Peter Hall, Clinical Research Fellow, University of Leeds, July 2011, personal communication).

Interventions

Nine tests were identified by NICE in the scope[25] (OncotypeDX, MammaPrint, Mammostrat, IHC4, PAM50, BCI, Randox BCA, NPI+ and BluePrint). These tests are described in detail in *Chapter 1*, *Description of technologies under assessment*. Our systematic review of the evidence (see *Chapter 2*, *Results*) indicated considerable differences in the level, quality and reporting of evidence between tests. Although some of the included tests have a relatively well-developed evidence base, some tests are still under development or have a relatively immature evidence base (e.g. NPI+, Randox BCA). Furthermore, there are differences in the output of the tests. Many of the tests predict the likelihood of distant recurrence, providing either a risk of recurrence score (as a continuous scale) or a risk category (e.g. high/low), but three of the tests (Randox BCA, current version of the PAM50 test and BluePrint) provide information about subtyping alone. The impact of information on subtype on the management of patients with early breast cancer is not yet clearly understood. No evidence on the impact of subtype information on clinical decision-making in early breast cancer in England and Wales was identified. This makes the potential comparison between tests particularly difficult.

To allow a sensible comparison between tests based on the available evidence, and given the time and resource constraints of the project, the EAG defined four minimum criteria that a test had to fulfil to be included in the economic evaluation. These criteria have been defined after consideration of the NICE scope,[25] discussion with clinical experts and consideration of the review of the existing cost-effectiveness evidence:

1. The test has been validated in an external cohort (clinical validity).
2. There is evidence about risk reclassification against one of the comparators defined by NICE within the scope (i.e. NPI, Adjuvant! Online or clinical practice in the UK).[25] In other words, there is evidence on how the new test reclassifies patients into risk groups relative to their initial risk group as defined by current practice.

3. The test provides an estimate of risk of recurrence in the form of a risk score or risk category. Following discussion with clinical experts tests that provide only information about subtyping were excluded as it is not yet clear how this knowledge will impact on the treatment decision-making process.
4. The outputs of the test, which will be used to inform the decision about whether or not to offer chemotherapy, are well defined.

A summary of these criteria for each of the nine tests considered for this appraisal is presented in *Table 39*.

Overall, only a subset of these tests met the criteria for inclusion in the economic evaluation defined by the EAG: OncotypeDX, MammaPrint, IHC4 and Mammostrat.

Although the PAM50 test has evidence about risk reclassification against OncotypeDX and Adjuvant! Online,[171] this test was excluded for the following reasons:

- PAM50 was not available in the UK at the time of writing of the report. Furthermore, the current version of the commercialised test (not available in the UK) does not provide a risk score but only information about subtyping. Following discussion with clinical experts it remained unclear how subtyping would be used to inform treatment decisions. An in vitro diagnostic version of the test is expected to be commercialised and this version will calculate a risk score; however, this is still under development.
- The evidence for risk reclassification was derived from a cohort in which the majority of women had positive nodes and therefore fall outside the subgroup of interest for this assessment.

The evidence base for NPI+ is developing. External validation of the test in an independent cohort is currently underway but has not yet been published. At the time of writing this report there was no published evidence on risk reclassification with NPI+.

There is no published evidence on risk reclassification against any of the comparators defined in the scope for BluePrint, Randox BCA and BCI. Furthermore, BluePrint and Randox BCA provide subtyping only.

TABLE 39 Summary of the inclusion/exclusion criteria for the economic evaluation

	Test	External validation of the test in an independent cohort	Evidence about risk reclassification	Final version of the test provides risk of recurrence	Clear use of the test	Other comments
Included	OncotypeDX	✓	✓ Adjuvant! Online, NPI, clinical practice	✓	✓	
	MammaPrint	✓	✓ Adjuvant! Online, NPI	✓	✓	
	Mammostrat	✓	✓ NPI	✓	✓	
	IHC4	✓	✓ NPI (OncotypeDX)	✓	✓	
	PAM50	✓	✓ OncotypeDX, Adjuvant! Online	✗ In vitro diagnostic version in development	✓	Reclassification evidence in a mix of LN+/–
Excluded	BluePrint	✗	✗	✗	✓	
	NPI+	✗ Nearing completion	✓ Unpublished	✓	✗	
	Randox CA	✗	✗	✗	✗	
	BCI	✗	✗	✓	✓	

Interventions to be assessed in the economic evaluation

Four tests were evaluated: OncotypeDX, MammaPrint, IHC4 and Mammostrat. These tests are described in detail in *Chapter 1*, *Description of technologies under assessment*. However, as indicated in the systematic review of the literature conducted as part of this project, the level and quality of evidence for these tests varies considerably.

The primary analysis evaluated the cost-effectiveness of adjuvant chemotherapy guided using OncotypeDX and IHC4. The systematic review of the evidence indicated that OncotypeDX is the furthest along the validation pathway compared with other similar tests, and the evidence base, in particular in relation to the prognostic ability of the test, was considered to be reasonably sound. The evidence for IHC4 is less developed; however, there is evidence relating to the performance of IHC4 compared with OncotypeDX and this allowed both tests to be modelled within the same model structure. A number of additional assumptions were, however, necessary to evaluate the cost-effectiveness of IHC4.

The final algorithm for IHC4 calculates a risk score for distant recurrence based on ER, PR, HER2 and Ki-67 in addition to classical clinical and pathological variables (composite risk score IHC4 + clinical score). This version of the algorithm was considered in the economic analysis (the term IHC4 will be used for simplicity but it refers to the composite risk score IHC4 + clinical score). Of note, an online calculator is expected to be made available (Professor Mitch Dowsett, Royal Marsden Hospital, London, July 2011, personal communication).

Analyses were performed for MammaPrint and Mammostrat but these were considered to be exploratory as there were significant gaps and/or limitations in the evidence base available for both tests (see *Model inputs: test-specific parameters*).

Comparators

Description of potentially relevant comparators

NICE CG80[7] indicates that adjuvant therapy should be considered for all patients with early invasive breast cancer after surgery, based on assessment of the prognostic and predictive factors alongside the potential benefits and side effects of the treatment. The guidelines recommend consideration of the use of Adjuvant! Online to support estimation of individual prognosis and the absolute benefit of adjuvant treatment for patients with early invasive breast cancer.[7] In addition, guidelines based on NPI are widely used in England and Wales. Clinical opinion suggests that there is wide variation in clinical practice between centres in the UK, with some centres using Adjuvant! Online, some using NPI-based guidelines and some using a combination of the two.

Adjuvant! Online

Despite NICE recommendations to use Adjuvant! Online,[7] clinical experts indicated that it is not comprehensively used in the UK for a number of reasons:

- It is based on a US population and there are some difficulties in applying the Adjuvant! Online data to the UK population.
- Although it is a useful aid for discussing risk of recurrence and benefits of chemotherapy with patients, it is viewed by some as complex to use and interpret for decision-making purposes.
- It cannot be used by all NHS trusts as access is blocked by some trusts for information technology security reasons.

The published evidence reports outcomes based only on the risk of recurrence estimated using Adjuvant! Online. However, both the risk of recurrence and predicted impact of adjuvant treatments would be used to inform treatment decisions.

Nottingham Prognostic Index

Nottingham Prognostic Index-based guidelines are widely used in some parts of the UK to inform decisions about adjuvant chemotherapy. The NPI forms part of the National Cancer Dataset for breast cancer so the NPI score should be given in the report of every invasive breast cancer case in the UK. It is simple to use although it may be considered to be less informative and therefore potentially less useful than Adjuvant! Online, particularly when discussing prognosis and potential treatments with patients.

Comparator used in the economic model

The comparator used in the model was current clinical practice. Clinical opinion indicated that, although NPI and Adjuvant Online! are used to aid the decision-making process, the decision whether or not to offer adjuvant chemotherapy to a specific patient is complex and includes other demographic and pathological parameters. Consequently, the EAG economic assessment used cancer registry data to reflect current clinical practice in England and Wales in terms of the proportion of women who currently receive chemotherapy. Summary data from ECRIC and WMCIU were obtained to populate the economic model (West Midland Cancer Intelligence Unit, July 2011, personal communication; Eastern Cancer Registration and Information Centre, July 2011, personal communication). The use of registry data reflects decision-making based on actual clinical practice, using NPI and/or Adjuvant! Online or other prognostic information.

The ECRIC registers all malignant tumours and some precancerous lesions occurring in people resident in the East of England at the time of diagnosis. Analyses for this assessment were constrained to women with ER+, LN–, HER2– early breast cancer (stage I or II) aged <75 years at diagnosis. An age cut-off was applied to reflect the fact that older women are likely to benefit less from the test (with a high proportion ineligible or unwilling to undergo chemotherapy because of frailty, comorbidities, etc.). It is acknowledged that there is no specific age cut-off but in practice the proportion of women receiving chemotherapy falls significantly for women aged ≥ 70 years and is very low for women aged ≥ 75 years.[172] Overall, 4475 patients were included in the analysis from 2007 onwards. Of these, around 800 had unknown HER2 status. The mean (median) age of included patients was 58.3 (60.0) years. The mean (median) tumour size of included patients was 16.9 (14.0) mm and 23.7% had grade I breast cancer, 56.0% grade II breast cancer and 20.2% grade III breast cancer.

The WMCIU registers all malignant tumours and some precancerous lesions occurring in people resident in the West Midlands. Again, analyses were constrained to women with ER+, LN–, HER2– invasive breast cancer and who were aged <75 years at diagnosis. The WMCIU had incomplete information on stage; therefore, early breast cancer was defined as women with no metastases and having had surgery (mastectomy or breast-conserving surgery). Data for the years 2007 and 2008 were available but data from 2007 only were used in the economic model as this was believed to be more accurate as the data were supplemented by national audit data (West Midland Cancer Intelligence Unit, July 2011, personal communication). Overall, 1214 patients with ER+, LN–, HER2– early breast cancer, who were diagnosed in 2007, were included. The mean (median) age of included patients was 58.0 (60.0) years. The mean (median) tumour size of included patients was 17 (15) mm and 26.6% had grade I breast cancer, 56.5% grade II breast cancer and 16.5% grade III breast cancer.

Cancer registry data from ECRIC and WMCIU were combined by the EAG and used in the base-case analysis to reflect the current levels of chemotherapy in England and Wales for women

with ER+, LN–, HER2– early breast cancer. Data were obtained for patients with a NPI score ≤3.4 and patients with a NPI score >3.4 to allow a subgroup analysis to be performed and to take account of the prognostic value of the current treatment decision based on clinicopathological parameters. Registry data reflect how both NPI or Adjuvant! Online are used currently in the decision-making process; however, it is not known which particular tools/guidelines (e.g. NPI or Adjuvant! Online, both, other tools) were used to inform adjuvant treatment decisions in the trusts within these cancer registry areas. For the purposes of the economic model it is assumed that data from these two areas are representative of all trusts in England and Wales. The term 'clinical practice' is used to define the comparator selected for this appraisal (i.e. current levels of adjuvant chemotherapy, based on the use of current prognostic tools, such as NPI and Adjuvant! Online).

Subgroups for whom the new tests are most likely to be used

Previous economic evaluations have typically assumed that the new tests will be offered to all women with ER+, LN–, HER2– early breast cancer. However, after discussion with clinical experts, it seems likely that, in England and Wales, the new tests may be targeted at a subgroup of this population – those at intermediate risk (and typically those aged <75 years) for whom the decision about whether or not to give chemotherapy is most uncertain. The definition of this 'intermediate group' is not clear-cut (see *Chapter 1, Identification of important subgroups*) but clinical advice suggested that usually typically patients with a NPI score of ≤3.4 are unlikely to receive chemotherapy (except for a few very young women with aggressive early breast cancer).

Consequently, two analyses are presented:

- The new test is given to all women with ER+, LN–, HER– breast cancer.
- The new test is given only to women with a NPI score >3.4 (based on the assumption that the vast majority of women with a NPI score ≤3.4 would not be considered for chemotherapy).

Of note, this subgroup is a proxy for the intermediate-risk group that might benefit the most from the test, but this may subgroup also include patients at the top end of the NPI distribution for whom the decision of chemotherapy is more certain. This subgroup was used as it was not possible, because of data restrictions, to create an intermediate-only group by separating out the high NPI risk group. However, because our population is ER+, LN–, HER2– it is rare to have a patient with a NPI score >5.4 and therefore the number of high-risk patients is expected to be low.

Finally, the EAG acknowledges that the cut-off is arbitrary and, although NPI is used in clinical practice to guide treatment decisions in some centres in England and Wales, treatment decision will not be based on NPI alone.

Description of the de novo economic model

Overview

A probabilistic decision-analytic model was developed to estimate the costs and QALYs of adjuvant chemotherapy guided by GEP and expanded IHC tests compared with current clinical practice (using cancer registry data) in England and Wales. The economic model was programmed using Microsoft Excel software (2011) and used a 6-monthly cycle length and followed patients over a lifetime horizon (100 years as the upper age limit) in the base case. Shorter time horizons were examined in sensitivity analyses. In accordance with NICE's interim methods guide for diagnostics,[32] the economic model adopted the perspective of the UK NHS and Personal Social Services (PSS) with costs and benefits discounted at 3.5% per annum.

No prospective studies that follow patients from initial diagnosis through to final health outcomes have been identified for any of the tests. Two prospective studies, MINDACT (MammaPrint) and TAILORx (OncotypeDX), are ongoing but not due to report for several years (see *Chapter 2*, *Results*). The economic model therefore needed to combine clinical data from several different sources in order to model how the results from the new tests translated into final outcomes in the form of QALYs.

Four tests were selected for the economic evaluation (OncotypeDX, IHC4, MammaPrint and Mammostrat). It is envisaged that these tests will be used as an addition to existing prognostic tools. As indicated in the systematic review, there are differences in the level and quality of evidence supporting each of the tests. Three separate analyses were performed using the best direct sources of data available for each test and these should not be directly compared. This was carried out because the EAG considered that combining evidence from different studies, based on different methodologies and with different patient characteristics (see *Chapter 2*, *Results*), limited the conclusions that could be drawn from the analyses and, in particular, the comparisons that could be made between the analyses.

The primary analysis compared current clinical practice with the adjuvant treatment decision based on the addition of OncotypeDX to current clinical practice and the addition of IHC4 to current clinical practice. Two exploratory analyses were undertaken to compare current clinical practice with the addition of MammaPrint and Mammostrat to current clinical practice. These analyses were considered to be exploratory only because of significant limitations in the evidence base.

Model structure

The key objective of the economic assessment is to address the cost-effectiveness of the use of GEP and expanded IHC tests to guide adjuvant chemotherapy decisions in women with early breast cancer in England and Wales. The model takes into account the reduction in the risk of relapses (and subsequent deaths) associated with the use of adjuvant chemotherapy. It also takes into account the costs and reduction in quality of life resulting from the adverse events associated with the chemotherapy.

All patients in the model are assumed to receive endocrine therapy. A proportion of patients in the comparator arm (current practice) received chemotherapy, based on cancer registry data. In the intervention arm (addition of new test) patients were assigned into a risk category using the new test and this additional information influenced the decision to prescribe chemotherapy.

The economic model comprises three key components:

1. Patients were assigned to risk categories according to the assigned risk score/group using the new test.
2. Women who would receive chemotherapy, as well as endocrine therapy, were identified, using the additional knowledge of the assigned risk group.
3. The natural history of breast cancer for patients treated with endocrine therapy alone or with the addition of chemotherapy was then simulated using a state transition model.

These three components are described in detail in the following sections.

Assignment of patients into different risk groups

OncotypeDX, MammaPrint and Mammostrat assign women into risk groups – high, intermediate and low risk (OncotypeDX and Mammostrat) or good and poor prognosis (MammaPrint). The IHC4 test provides a risk score only; however, patients have been allocated

into risk groups, similar to the OncotypeDX risk groups, for the purposes of this assessment (see *Model inputs: test-specific parameters* for more details).

In the economic model, women were first stratified into two NPI groups (women with a NPI score ≤ 3.4 and women with a NPI score > 3.4). This was carried out to allow the use of the test with different subgroups of patients to be explored and to allow adjustment of non-UK clinical evidence to reflect the NPI distribution in the UK population. This also takes into account the prognostic value of the treatment decision using clinicopathological parameters. Indeed, within the current treatment decision-making process based on clinicopathological parameters, it is possible to identify patients who are at a higher risk of distant recurrence. Within these two NPI groups, patients were further reclassified into low, intermediate or high risk (or low and high risk in the case of MammaPrint) of recurrence according to the outputs of the new tests.

In simple terms, patients are assigned into different boxes, each with a different prognosis. Patients are assigned to the same boxes for the comparator (current practice) arm or the intervention arm (GEP and expanded IHC tests) as the diagnostic tool does not affect the prognosis of those patients if there is no change in the adjuvant treatment.

Identification of women receiving adjuvant chemotherapy on the basis of the test results

Once women have been assigned into the different boxes (with different prognosis) the next step is to identify which women would receive adjuvant chemotherapy.

The aim of categorising patients into risk groups based on distant recurrence with the GEP and expanded IHC tests is to identify patients who have a greater chance of developing a distant recurrence/recurrence. The risk groups identified by the new tests are therefore expected to influence the targeting of chemotherapy. However, other factors will also influence the decision regarding chemotherapy, including clinical and pathological factors, along with patient choice. In clinical practice a proportion of women classified as low risk of distant recurrence using GEP and expanded IHC tests may still receive chemotherapy; similarly, a proportion of women considered to be at high risk may not receive chemotherapy, as shown in Spain[173] or in the USA[174] for OncotypeDX.

In the intervention arm of the economic model the proportion of patients who would receive chemotherapy is based on the expected interpretation of the test, for example women categorised as high risk of recurrence are more likely to receive chemotherapy than women categorised as low risk. Some previous analyses have assumed that chemotherapy is received based on the risk group only. For instance, all women defined as high risk receive chemotherapy. However, in clinical practice other issues are likely to impact on this decision (clinicopathological factors, age of patient, patient choice, etc.) and it is unlikely that 100% of high-risk patients will receive chemotherapy. An adjustment for such factors was therefore used in the model.

In the comparator (current practice) arm, the proportion of women receiving chemotherapy is based on cancer registry data. Two subgroups are considered: women with a NPI score ≤ 3.4 and women with a NPI score > 3.4. Because the model categorised women into boxes (defined by the new test a posteriori) and the oncologist is blind to the results of the new test, we assumed that the probability of receiving chemotherapy was the same in the current practice arm whether patients were reclassified as low, intermediate or high using GEP and expanded IHC tests. However, it is likely that patients who are classified as high risk by the new test are more likely to have been identified as high risk under current practice and, therefore, are more likely to have received chemotherapy than those patients classified as low or intermediate risk by the new test. To further explore this assumption, data from the Holt *et al.* study[78] were analysed by the EAG

to approximate the proportion of patients who were recommended chemotherapy by RS group before knowledge of the OncotypeDX score (analysis conducted by the EAG using individual patient-level data submitted by Genomic Health).

Overall, 30.43%, 30.30% and 68.42% of patients with a low, intermediate and high RS score were recommended chemotherapy before knowledge of the OncotypeDX test results. Preliminary analyses suggested that the proportion is likely to be higher for patients with a high RS, but the sample size was too small (69 in low RS, 33 in intermediate RS and 19 in high RS) to draw any definitive conclusion. While preliminary, this analysis suggested that our assumption (that the probability of receiving chemotherapy in the comparator is constant irrespective of the RS group) might be conservative as current practice using clinicopathological parameters does appear to add some prognostic value.

Natural history of breast cancer

The final part of the model was a Markov model. Patients were able to move between five possible health states: recurrence free (A), distant recurrence (B), local recurrence (C), long-term adverse events after chemotherapy (D) and death (from breast cancer, long-term adverse events or general causes – E).

As shown in *Figure 11*, patients enter the model in the recurrence-free survival health state (A) and remain in that health state until they develop a distant recurrence (B), have an adverse event after chemotherapy (D) or die from breast cancer or general causes or from their adverse event (E). After a distant recurrence (B), patients remain in this health state until they die from either breast cancer or general causes (E) or develop an adverse event for women treated with chemotherapy (D). Patients developing an adverse event after chemotherapy can remain in that health state, die from their adverse event or die from general causes (E). The estimation of long-term adverse events is simplistic. No distinction was made between patients developing long-term adverse events after a recurrence (B) and patients developing long-term adverse events in the recurrence-free health state (A). Furthermore, patients with a long-term adverse event were assumed to remain in that health state (D) until death (E) and were not allowed to move to other health states.

Local/regional recurrences have been modelled by considering the costs and quality of life decrements (disutility) assuming that a proportion of patients entering the distant recurrence state (B) have previously experienced a local recurrence (C). No transition probabilities were used between this health state and death or adverse events. This is simplistic but justified by the fact that the risk categories (used in the economic model) defined by the new tests (OncotypeDX,

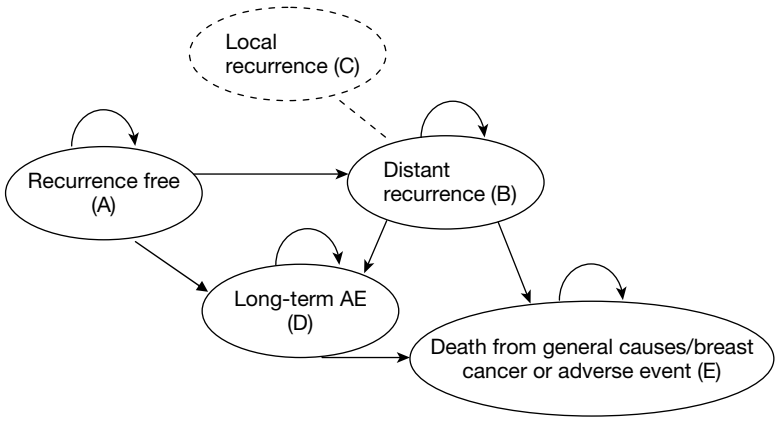


FIGURE 11 Schematic of the model structure.

MammaPrint, IHC4) have been defined according to the risk of developing distant recurrence and there is no robust evidence to accurately model the development of local recurrence and the different transitions between health states for patients reclassified as low, intermediate or high risk with GEP and expanded IHC tests.

Model inputs: general

Model inputs that were common to the assessment of each of the four tests are described below. Model inputs that are test specific, such as clinical parameters, are described in *Model inputs: test-specific parameters*.

Mean age of patients entering the model

The EAG economic assessment focuses on women with ER+, LN–, HER2– who are aged ≤75 years. Patients were assumed to enter the economic model at a mean age of 58.3 years based on the average age in the ECRIC dataset of women with ER+, LN–, HER2– aged <75 years (Eastern Cancer Registration and Information Centre, July 2011, personal communication). Although patient age is not used to determine treatment selection, experts suggested that women aged >70–75 years are much less likely to be offered chemotherapy because of issues of frailty and comorbidities. Sensitivity analysis was conducted varying the mean baseline age.

Of note, the model does not separate pre- and postmenopausal women but most of the evidence was taken from postmenopausal women. In addition, it was not possible to explore different age thresholds as we did not have access to patient-level data.

Baseline Nottingham Prognostic Index distribution in England and Wales

The economic assessment separates women with a NPI score ≤3.4 and women with a NPI score >3.4. The baseline NPI distribution was extracted from the combined (EAG analysis) ECRIC (2007 onwards) and WMCIU (2007 only) data (West Midland Cancer Intelligence Unit, July 2011, personal communication; Eastern Cancer Registration and Information Centre, July 2011, personal communication). Approximately two-thirds of patients had a NPI score ≤3.4 (*Table 40*).

For the scenario assuming that the test is given to all women with ER+, LN–, HER2– early breast cancer, we modelled patients with a NPI score of ≤3.4 and patients with a NPI score >3.4 separately to account for the prognostic value of the treatment decision using clinicopathological parameters.

Additional sources of evidence for the baseline distribution of NPI were considered in sensitivity analysis (*Table 41*). These included women with ER+, LN–, HER2– early breast cancer from the TransATAC trial (Professor Mitch Dowsett, Royal Marsden Hospital, London, September 2011, personal communication) and the ER+, LN–, HER2– population from the Holt *et al.* study[78] (analysis conducted by the EAG using individual patient-level data submitted by Genomic Health).

TABLE 40 Distribution of the NPI score observed in the ECRIC and WMCIU data and used in the economic model

Cohort	NPI ≤3.4, *n* (%)	NPI >3.4, *n* (%)	Total, *n*
ECRIC[a] (2007 onwards)	2602 (65.6)	1365 (34.4)	3967
WMCIU (2007 only)	819 (68.2)	382 (31.8)	1201
Combined data used in the economic model	3421 (66.2)	1747 (33.8)	5168

a Includes HER2 equivocal and unknown.

TABLE 41 Distribution of the NPI score from the TransATAC trial (Professor Mitch Dowsett, Royal Marsden Hospital, London, September 2011, personal communication) and the Holt cohort (EAG analysis) in women with ER+, LN–, HER2– early breast cancer used in sensitivity analysis

Source	NPI ≤ 3.4, *n* (%)	NPI > 3.4, *n* (%)
TransATAC	449 (63.6)	257 (36.4)
Holt (EAG analysis)	70 (57.8%)	51 (42.2%)

Proportion of patients receiving chemotherapy in current clinical practice in England and Wales

As described in *Comparator used in the economic model*, cancer registry data were used to reflect the proportion of women with ER+, LN–, HER2– early breast cancer who currently receive chemotherapy in England and Wales. Data from ECRIC (2007 onwards; Eastern Cancer Registration and Information Centre, July 2011, personal communication) and WMCIU (2007 only) (West Midland Cancer Intelligence Unit, July 2011, personal communication; Eastern Cancer Registration and Information Centre, July 2011, personal communication) were combined (EAG analysis) and showed that overall about 14.4% of women aged < 75 years with ER+, LN–, HER2– early breast cancer received chemotherapy. When separating patients with a NPI score of ≤ 3.4 and a NPI score > 3.4, 4.6% and 33.6% of women received chemotherapy respectively (*Table 42*).

Death from breast cancer causes after a distant recurrence

In the base case, the hazard rate of death after a distant recurrence was taken from Thomas *et al.*[157] Thomas *et al.*[157] analysed the time to death from relapse among 77 relapsed breast cancer patients. The first site of relapse was distant in 51 patients (66%) with the remaining patients having locoregional recurrences. The study included a mix of patients with regard to ER status (55% ER+), nodal involvement (66% LN–) and HER2 status (75% HER2–). The study reported a median survival of 40.1 months (equating to an annual hazard rate of about 0.30) and this value was used in the base case. Sensitivity analyses were conducted varying the time spent in the distant recurrence health state within the reported CI in this study.[155]

In the base case we assumed that the risk of death after a recurrence was independent of the prognosis of the patient, because of the lack of more informative data. Discussion with clinical experts indicated that it is likely that low- and high-risk patients will spend a different amount of time in the distant recurrence health state. High-risk patients are likely to have more aggressive cancer and are likely to spend less time in recurrence before death. A scenario analysis was therefore explored assuming different times in recurrence between risk groups. Because there are no published data to our knowledge on survival after distant metastasis for patients with different prognosis, assumptions were made to examine the impact of this assumption on the ICER.

Costs

All costs are in 2010 prices.

Costs of the different tests No costs were assumed for treatment guided by current clinical practice as pathological parameters that are currently used to guide the adjuvant treatment decision will continue to be collected after the introduction of the new tests. The new tests are considered to be add-ons to the existing tests.

The costs of performing the OncotypeDX, MammaPrint and Mammostrat tests in the UK were assumed to be £2580, £2675 and £1120–1620 (£1135 was used in the economic model) respectively (data received from the respective manufacturers) (*Table 43*). The IHC4 algorithm

TABLE 42 Proportion of patients currently receiving chemotherapy in ECRIC and WMCIU (women with ER+, LN–, HER2– early breast cancer aged < 75 years)

Cohort	Entire cohort (%)	NPI ≤ 3.4 (%)	NPI > 3.4 (%)
ECRIC[a] (2007 onwards)	13.97	4.23	32.53
WMCIU (2007)	15.90	5.86	37.43
Combined data used in the economic model	14.42	4.62	33.60

a Includes HER2 equivocal and unknown.

TABLE 43 Costs of the new tests used in the economic model

	Cost of the new test (£)	Cost of the additional NHS time (£)	Cost of handling fresh tissue (£)
OncotypeDX	2580		
IHC4		100–200	
Mammostrat	1135		
MammaPrint	2675		250

is free; therefore, the only costs will be the additional time required for analysis of Ki-67 and PR (measured in some centres) and quantitative H scoring (a immunohistochemical approach used in the assessment of markers for breast cancer prognosis, by assessing the intensity and distribution of positive staining) of ER. The investigators of IHC4 were contacted to provide an estimate of the likely additional cost to the NHS. Although no formal costings have been made it was estimated that the additional cost of IHC4 would range between £100 and £200, including the pathologist's time for scoring (Professor Mitch Dowsett, Royal Marsden Hospital, London, July 2011, personal communication). In the base case we assumed that it cost an additional £150 to the NHS to run an IHC4 test (see *Table 43*). This figure is varied in the sensitivity analysis.

The MammaPrint test can be performed on fresh tissue preserved in RNA*Retain*® (Asuragen, Austin, TX) or fresh frozen tissue (note: use of FFPE to be introduced in 2012). In addition to the cost of the test, we assumed an additional cost of £250 per patient for MammaPrint for the cost of handling fresh tissue. A sensitivity analysis was conducted assuming no additional cost. Fresh tissue collection is not routine in the NHS (only a few research centres currently have this working arrangement) so there will be additional costs, which would be considerable at hospitals where the dissection facilities are running at capacity (which is likely to be a significant proportion of hospitals) and where explicit staffing for collection of fresh tissue is not in place. Discussion with local clinicians indicates that capital costs could be at least £75,000 per hospital if new dissection tables are required, which is likely to be the case in many hospitals where routine fresh tissue sampling is not in place, and additional staff costs for biomedical scientists and histopathologists will be incurred. If a full fresh tissue service is required and needed to cover all theatre time then additional staff costs could be £20,000–50,000 per year (Simon Cross, Reader and Honorary Consultant, Royal Hallamshire Hospital, Sheffield, July 2011, personal communication). Experts indicated that a charge of about £250 per sample would be necessary to take a fresh tissue sample for a research study because of the time-critical staff-intensive work required. However, this assumes that a fresh tissue sample is collected only in a small number of patients under the current service configuration. A reconfiguration of the entire pathology service would be necessary if fresh tissue samples had to be collected routinely for all patients, which would incur additional costs.

We did not incorporate the additional cost associated with the failure of a test. This was considered to be minimal as contact with the manufacturers indicated that another sample could be sent for free or a refund issued in case of failure of the test.

Endocrine therapy costs

The economic model considers only women with ER+ early breast cancer and assumes that all patients receive endocrine therapy. Five endocrine therapy regimens were considered as per NICE Technology Appraisal 112:[162] (1) tamoxifen for 5 years, (2) anastrozole for 5 years, (3) letrozole for 5 years, (4) tamoxifen for 2 years plus exemestane for the final 3 years and (5) tamoxifen for 5 years followed by extended therapy with letrozole for a further 3 years.

Drug costs were taken from BNF 61.[161] It was assumed that each endocrine therapy was given once daily at a 20-mg, 1-mg, 2.5-mg and 25-mg dosage for tamoxifen, anastrozole, letrozole and exemestane respectively. The annual cost for each drug is presented in *Table 44*.

The probability that a patient will be treated with each regimen was taken from the costing template accompanying TA112.[162] It was assumed that 40% of patients received tamoxifen for 5 years, 20% anastrozole for 5 years, 20% letrozole for 5 years and 20% tamoxifen for 2 years plus exemestane for the final 3 years. It was further assumed that 10% of patients received tamoxifen for 5 years followed by extended therapy with letrozole for a further 3 years. After weighting, the annual drug cost was calculated to be £668.90 for the first 5 years and £110.70 for the remaining 3 years.

In addition to drug costs, monitoring cost were included. We assumed that patients treated with endocrine therapy have two follow-up appointments in the first year and one follow-up appointment every subsequent year (£129 based on NHS reference costs 2009/10,[175] 370 Medical Oncology). We further assumed that patients had one mammogram every year (£46.37 based on Campbell *et al.*[176]) for a maximum of 5 years.

Chemotherapy costs

It was assumed that all patients received FEC75 as clinical opinion indicated that this is the most commonly used chemotherapy regime for this population (ER+, LN–, HER2–). Sensitivity analysis was carried out varying the cost to explore the sensitivity of the results to this assumption. Note that the choice of chemotherapy (FEC75) in the economic model impacts on cost only as the effect of chemotherapy was taken from a separate source of data[49] that uses CMF/MF regimes. No effectiveness data were available for FEC75 for this group of patients.

The cost of the chemotherapy drugs was calculated according to the regime description of FEC75 given by Avon, Somerset and Wiltshire Cancer Services in their chemotherapy protocol documents.[177] All drug costs are taken from BNF 61.[161] We also assumed a dosage per BSA of 1.75 mg/m^2 based on the value reported by Sacco *et al.*,[168] estimated in women with breast cancer in the UK.

The chemotherapy drug cost per cycle is summarised in *Table 45*. No drug wastage was assumed.

TABLE 44 Costs of endocrine therapy drugs used in the economic model

	Dose (mg)	Tablets per pack	Price per pack (£)	Annual cost (£)
Tamoxifen	20	30	2.09	25.45
Anastrozole	1	28	68.56	894.34
Letrozole	2.5	28	84.86	1106.97
Exemestane	25	30	88.80	1081.14

Discussion with clinical experts indicated that FEC75 is usually given for six cycles. The number of cycles was varied in sensitivity analysis. In addition to drug costs, we assumed an additional pharmacy cost of £38[162] per cycle to account for the chemopharmacy/aseptic costs.

Furthermore, administration costs were assumed to be £270.60 for the first cycle of treatment (NHS reference costs 2009/210,[175] S13Z: Deliver more complex Parenteral Chemotherapy at first attendance) and £284.50 for the remaining cycles (NHS reference costs 2009/10,[175] SB15Z: Deliver subsequent elements of a Chemotherapy cycle). Patients were also assumed to have a separate outpatient appointment before administration of each cycle of the chemotherapy (NHS reference costs 2009/10,[175] 370 Medical Oncology – £129).

Patients were also assumed to be monitored and to receive one liver function test (£12.68), a test of urea and electrolytes (£12.30) and a full blood count (£5.81) at each cycle based on 2008/9 data from the Sheffield Teaching Hospital (Sheffield Teaching Hospital Trust, 2007–8, personal communication), uplifted to 2010 prices.[178] Finally, it was further assumed that 25% of patients would have an echocardiogram before starting chemotherapy (NHS reference costs 2009/10,[175] RA60Z – £59).

The total cost of chemotherapy including drug acquisition cost, administration and monitoring was calculated to be £4099 for the entire course of chemotherapy (six cycles of treatment).

Costs of short-term adverse events associated with chemotherapy

Short-term adverse events were included for patients receiving chemotherapy.

The probability of short-term grade 3 and grade 4 adverse events was extracted from the PACS-01 clinical trial for patients treated with FEC100 as no data were available for FEC75.[179] This included anaemia, thrombocytopenia, neutropenic infection, nausea/vomiting and stomatitis. Short-term grade 3 and grade 4 adverse events were costed using the NHS reference costs where appropriate.[175] The total cost of treating short-term adverse events was estimated to be £275.61 per patient (*Table 46*). This excluded the cost associated with the secondary prevention of febrile neutropenia using G-CSF prophylaxis, included separately (see next section).

Costs of the secondary prevention of febrile neutropenia

We assumed that G-CSF prophylaxis was given for the secondary prevention of neutropenia to women receiving chemotherapy only. In the base case it was assumed that about 25% of patients would receive G-CSF prophylaxis (Dr Matthew Winter, Consultant in Medical Oncology, Sheffield Teaching Hospitals NHS Foundation Trust, September 2011, personal communication) and that G-CSF would be given for an average of three cycles. Patients were assumed to receive filgrastim at a dose of 500,000 units/kg daily for six days after each cycle of chemotherapy (maximum three cycles). A mean weight of 66 kg was assumed and the drug cost per injection of 30 million units was assumed to be £59.[161] Filgrastim was assumed to be administrated by a district nurse (£39 per injection) using the cost from the Personal Social Services Research Unit (PSSRU).[178]

TABLE 45 Costs of chemotherapy drugs used in the economic model

	Dose (mg)	Total dose per cycle (mg)	Cost per cycle (£)
Fluorouracil	600	1050	16.00
Epirubicin	75	131	168.78
Cyclophosphamide	600	1050	16.32
Total			201.10

TABLE 46 Costs of short-term adverse events from chemotherapy used in the economic model

	Proportion of grade 3/4 adverse events (%)	HRG code[175]	HRG	Cost of HRG (£)	Cost per patient (£)
Anaemia	1.4	SA09F	Other Red Blood Cell Disorders without CC	1529	21.41
Thrombocytopenia	0.3	SA12F	Thrombocytopenia without CC	1355	4.06
Neutropenic infection	1.6			2286	36.58
Nausea/vomiting	24.2	FZ48C	Malignant General Abdominal Disorders with length of stay 1 day or less	588	142.34
Stomatitis	4.0	CZ24Q	Complex/Major Head, Neck and Ear Disorders without CC	1781	71.22
Cost of adverse event per patient					275.61

CC, Complications and Comorbidities ; HRG, Healthcare Resource Group.

Overall, the cost associated with the secondary prevention of adverse event for patients receiving chemotherapy was estimated to be £485.30 per patient.

Costs associated with long-term adverse events

Potential long-term adverse events include secondary malignancies and congestive heart failure (CHF). Although CHF is more common than secondary malignancies, the development of cancer is likely to have more serious consequences and to be associated with a higher impact on health-care resources than the management of CHF.

The base-case economic model included acute myeloid leukaemia (AML) as a long-term adverse event after chemotherapy. The probability of developing AML was based on the 8-year cumulative probability of developing AML in women treated with epiribucin extracted from a meta-analysis of 19 trials conducted in early breast cancer.[180]

The meta-analysis showed that the 8-year cumulative probability of AML was 0.37% (95% CI 0.13 to 0.61%) in women receiving a cumulative dose of cyclophosphamide $< 6300\ mg/m^2$ and a cumulative dose of epirubicin $< 720\ mg/m^2$ ($n = 4760$).[180]

We further assumed that patients spend 8 months on average in the AML health state at a mean cost of £11,500 based on the approximate mean life-years and mean costs estimated by the manufacturer for the NICE technology appraisal of azacitidine for myelodysplastic syndromes.[181] These assumptions were varied in sensitivity analysis.

Costs associated with the management of distant recurrence

The costs associated with distant recurrence were derived from Thomas *et al.*[157] using a sample of 77 patients with relapsed breast cancer. Costs included active supportive care and end-of-life care. Costs specifically associated with terminal care were removed to avoid double counting as these were included separately in the economic model. (Note that only cost items described as supportive/terminal care were removed.)

After removing cost items that were specific to terminal care, we estimated the 6-monthly cost to be approximately £4082 (uplifted to 2010 prices).[178] We assumed that the cost was constant over time. This is very simplistic as evidence shows that the cost is higher in the first 2 years and decreases thereafter.[182] However, the impact of this assumption is likely to be minimal because this affects both the comparator arm and the intervention arm in the model.

Costs associated with the management of local recurrence

The cost of local recurrence was taken from Karnon *et al.*[182] and was estimated at £14,132 (uplifted to 2010 prices).[178]

Costs associated with the management of death from breast cancer

Finally, the cost associated with terminal care/end of life was taken from Campbell *et al.*[176] and was assumed to be about £4038. This cost was applied as a one-off in the economic model, immediately before death from breast cancer.

Health-state utilities

Quality of life utility scores were identified from a recent systematic review of utility values in breast cancer.[159] The utility values used in the model are given in *Table 47*.

Utility values for patients in the recurrence-free and distant recurrence health states were extracted from Lidgren *et al.*[148] These were EQ-5D values and using this study allowed values for recurrence free (0.824) and distant recurrence (0.685) to be taken from the same study for consistency (see *Table 47*). The study followed 361 breast cancer patients attending the breast cancer outpatient clinic at Karolinska University Hospital Solna between April and May 2005. The decrement in utility per patient experiencing a local recurrence was taken from Campbell *et al.*[176] and assumed to be –0.108 in the base case. We assumed that patients with AML have a utility value of 0.26 based on the value used in a previous economic evaluation conducted in Canada.[183] We further assumed that patients receiving chemotherapy have a disutility of 0.038, taken from Campbell *et al.*[176] for women treated with E-CMF (epirubicin, cyclophosphamide, methotrexate and 5-fluorouracil)/FEC60 in the first year. This is believed to capture the decrement in utility associated with the administration of chemotherapy and related adverse events. Finally, a decrement in utility was applied for patients dying from breast cancer, derived from Campbell *et al.*[176] and Lidgren *et al.*[148] Utility values were varied in sensitivity analysis.

Death from causes other than breast cancer

The mortality rate from causes other than breast cancer was extracted from UK life tables (2007–9) for women[184] after adjustment to remove death attributable to breast cancer.

Proportion of patients with distant recurrence who have previously experienced a local recurrence

In the base case we assumed that 10.5% of patients entering the distant recurrence state have previously experienced a local recurrence. This is based on an analysis conducted in 3601 women with early breast cancer enrolled in previous European Organisation for Research and Treatment in Cancer (EORTC) trials (10801, 10854 and 10902), which showed that the presence of locoregional recurrence was a significant prognostic risk factor for the occurrence of distant

TABLE 47 Utility values used in the model

	Mean utility score	Duration	Source
Recurrence free	0.824 (95% CI 0.785 to 0.857)	1 year	Lidgren *et al.*[148]
Distant recurrence	0.685 (95% CI 0.62 to 0.735)	1 year	Lidgren *et al.*[148]
Local recurrence (decrement per patient)	–0.108	NA	Campbell *et al.*[176]
AML	0.26	1 year	Younis *et al.*[183]
Chemotherapy (decrement per patient)	–0.038	NA	Campbell *et al.*[176]
Utility for patients dying from breast cancer (final 3 months of life)	0.159 (SE 0.04)	3 months	Campbell *et al.*[176]

NA, not applicable.

recurrence.[185] The analysis showed that, among the 1224 patients who developed a distant recurrence, 129 patients had a distant recurrence after a locoregional recurrence.

We did not make any assumptions about the time spent in the local recurrence health state. Local/regional recurrences have been modelled by considering the cost and quality of life decrements (disutility), assuming that a proportion of patients entering the distant recurrence state have previously experienced a local recurrence.

Model inputs: test-specific parameters

Clinical parameters specific to each test are described below. For each test clinical parameters relating to the three main components of the model are described in turn.

Clinical parameters: OncotypeDX and IHC4

The systematic review of evidence indicated that the OncotypeDX test is the furthest along the validation pathway compared with other similar tests, and the evidence base, in particular in relation to the prognostic ability of the test, was reasonably sound. The evidence base for IHC4 is less developed but there is direct evidence relating to the performance of IHC4 compared with that of OncotypeDX and so the clinical evidence relating to these two tests is described together.

The primary analysis compared current clinical practice with treatment guided using OncotypeDX or IHC4 in addition to current practice. This was carried out using evidence that directly compared the test results for OncotypeDX and IHC4. An overview of this evidence was presented in Cuzick *et al.*[84] However, the specific data used in the economic model for IHC4 were unpublished and were made available to the EAG for the purpose of this assessment (Professor Mitch Dowsett, Royal Marsden Hospital, London, September 2011, personal communication).

Assignment of patients into risk category: OncotypeDX and IHC4 Most of the evidence on the ability of OncotypeDX to classify patients into the low, intermediate or high risk group is derived from US studies (see *Chapter 2*, *OncotypeDX test*). The systematic review did, however, identify two UK studies[78,79] that presented classification evidence. Data from one of these studies, the TransATAC trial (Professor Mitch Dowsett, Royal Marsden Hospital, London, September 2011, personal communication), were used in the base-case analysis. The second study, the Holt *et al.* study[78] included a small sample of patients recruited in Wales and reported how the test classified patients by OncotypeDX RS and how this influenced clinical decision-making. However, the systematic review of evidence showed that there are limitations with using data from this study (see *Chapter 2*, *OncotypeDX test results*). Likewise, discussion with clinical experts indicated some concerns that the patients included might not be representative of patients seen in clinical practice. To further explore this point, we compared the baseline characteristics of women with ER+, LN–, HER2– early breast cancer included in the final analysis of the Holt *et al.* study ($n = 121$) (EAG analysis) with the baseline characteristics of cohorts of patients from the cancer registry data provided by ECRIC (2007 onwards) and WMCIU (2007 only) (West Midland Cancer Intelligence Unit, July 2011, personal communication; Eastern Cancer Registration and Information Centre, July 2011, personal communication) (*Table 48*). Overall, the Holt cohort was generally more severe with a higher distribution of Grade 2 and 3 tumours and larger tumour size (20.4 mm vs 17.0 mm). There was also a higher proportion of patients classified as intermediate or high risk according to NPI (NPI > 3.4) in the Holt study (42.15%) compared with patients included in the ECRIC (34.4%) or WMCIU (31.8%) cohort.

The impact of using data from the Holt *et al.* study was explored in a scenario analysis using data used in the economic model submitted by Genomic Health (see *Comparison of assumptions and results with the economic models submitted by Genomic Health and Clarient*).

TABLE 48 Comparison of patients included in the Holt *et al.* study[78] (EAG analysis) and cohorts of patients from the ECRIC and WMCIU cancer registry data (West Midland Cancer Intelligence Unit, July 2011, personal communication; Eastern Cancer Registration and Information Centre, July 2011, personal communication)

Baseline characteristic	Holt *et al.*[78] (*n* = 121)[a]	ECRIC cohort (*n* = 3245)	WMCIU cohort (*n* = 1214)
Age (years)			
Mean	55.88	58.30	58
First interquartile	50.00	51.00	51
Third interquartile	63.00	66.00	66
Median	55.00	60.00	60
Grade distribution (%)			
I	19.01	23.7	26.6
II	63.64	56.0	56.5
III	17.36	20.2	16.5
Tumour size (mm)			
Mean	20.39	16.90	17
First interquartile	13.00	10.00	10
Third interquartile	23.00	20.00	20
Median	18.00	14.00	15
NPI score (%)			
Low (≤ 3.4)	57.85	65.59	68.19
Intermediate/high (> 3.4)	42.15	34.41	31.81

a Analysis conducted by the EAG.

The TransATAC trial evaluated the efficacy and safety of 5 years of anastrozole, tamoxifen or the combination of both treatments in postmenopausal women with localised breast cancer in the UK.[79,84] The study included a much larger sample size. Furthermore, data on both the risk of distant recurrence and risk classification were available in the TransATAC trial from the same cohort for patients using OncotypeDX and IHC4, making it the most robust source to use to populate the economic model. Clinical experts supported the view that patients included in this study were more likely to be representative of patients seen in clinical practice. However, the inherent limitations of the generalisability of such trial data, such as the fact that women with comorbidities would have been excluded from the trial, need to be taken into consideration.

Data from the TransATAC trial were reanalysed by the investigators of the trial to exclude women with HER2+ cancer, to stratify patients by NPI score (≤ 3.4 and > 3.4) and to provide additional data relating to IHC4 (Professor Mitch Dowsett, Royal Marsden Hospital, London, September 2011, personal communication). Reclassification using the OncotypeDX and IHC4 tests compared with NPI group for women with ER+, LN–, HER2– is presented in *Tables 49* and *50* respectively. The IHC4 + clinical score test provides a continuous risk score. Selected cut-offs for three IHC4 risk groups have been defined specifically to populate the economic model (using the same methodology as for OncotypeDX) and therefore this might not reflect how the test will be used in clinical practice.

Among patients with a NPI score ≤ 3.4 ($n = 449$), significantly more patients were reclassified as intermediate/high using OncotypeDX ($n = 126$) than with IHC4 ($n = 12$). Among patients with a NPI score > 3.4 ($n = 257$), more patients were reclassified as having a low risk using IHC4 ($n = 166$) than with OncotypeDX ($n = 133$).

TABLE 49 Reclassification of ER+, LN–, HER2– patients from the TransATAC trial by OncotypeDX and NPI group (Professor Mitch Dowsett, Royal Marsden Hospital, London, September 2011, personal communication)

	Low RS, *n* (%)	Intermediate RS, *n* (%)	High RS, *n* (%)	Total, *n* (%)
NPI ≤ 3.4	323 (71.94)	109 (24.28)	17 (3.79)	449 (100)
NPI > 3.4	133 (51.75)	76 (29.57)	48 (18.68)	257 (100)
Total	456 (64.59)	185 (26.20)	65 (9.21)	706 (100)

TABLE 50 Reclassification of ER+, LN–, HER2– patients from the TransATAC trial by IHC4 + clinical score and NPI group (Professor Mitch Dowsett, Royal Marsden Hospital, London, September 2011, personal communication)

	Low IHC4, *n* (%)	Intermediate IHC4, *n* (%)	High IHC4, *n* (%)	Total, *n* (%)
NPI ≤ 3.4	437 (97.33)	12 (2.67)	0	449 (100)
NPI > 3.4	166 (64.59)	58 (22.57)	33 (12.84)	257 (100)
Total	603 (85.41)	70 (9.92)	33 (4.67)	706 (100)

For the purposes of the economic assessment, patients classified using OncotypeDX were reclassified according to IHC4. Reclassification data for patients with a NPI score ≤ 3.4 and a NPI score > 3.4 used in the EAG economic model are presented in *Tables 51* and *52* respectively.

The impact of using data from Holt[27] was examined in *Comparison with the economic model submitted by Genomic Health*.

Note that the approach of classifying patients into risk categories has some limitations:

- It was assumed that the new tests categorised patients into risk categories; however, both OncotypeDX and IHC4 provide a continuous risk score.
- Although cut-offs are available for OncotypeDX to identify patients at low, intermediate or high risk of distant recurrence, these are informative and not definitive. In the economic model, patients were classified according to the original cut-offs defined by the manufacturer of the technology: low – RS < 18; intermediate – RS between 18 and 30; and high – RS ≥ 31. However, the definition of low, high and mid-range RS was modified in the TAILORx trial for OncotypeDX. A cut-off of ≤ 10 was used instead of the original < 18 to define patients at low risk of distant recurrence. The cut-off for high risk of distant recurrence was modified from ≥ 31 to ≥ 26.
- The IHC4 test does not present cut-offs. The test is intended to be used as a continuous risk score and interpretation is at the discretion of the physician. For the purpose of the economic assessment, investigators provided risk classification evidence of IHC4 based on low, intermediate and high risk of distant recurrence. Cut-offs were defined using a similar approach to that for OncotypeDX (< 10%, 10–20% and > 20% predicted risk of distant recurrence). The cut-offs used for IHC4 are therefore exploratory and were defined only to populate the economic model.

Risk of distant recurrence for patients receiving endocrine therapy only: OncotypeDX and IHC4 The TransATAC trial reported the risk of distant recurrence in patients treated with anastrozole or tamoxifen. The advantage of using data from this trial is that it was possible to extract the risk of distant recurrence for patients classified using OncotypeDX and further reclassified using IHC4 from the same source of data as the reclassification data used above (see *Tables 51* and *52*). The proportions of patients without distant recurrence at 10 years used in the economic model for patients with a NPI score ≤ 3.4 and a NPI score > 3.4 are presented in *Tables 53* and *54* respectively (Professor Mitch Dowsett, Royal Marsden Hospital, London, September 2011,

TABLE 51 Risk classification using OncotypeDX followed by reclassification using IHC4 + clinical score for patients with a NPI score ≤ 3.4 ($n = 449$)

	Low RS, *n* (%)	Intermediate RS, *n* (%)	High RS, *n* (%)	Total, *n* (%)
Low IHC4	321 (71.49)	103 (22.94)	13 (2.90)	437 (97.33)
Intermediate IHC4	2 (0.45)	6 (1.34)	4 (0.89)	12 (2.67)
High IHC4	0	0	0	0
Total	323 (71.94)	109 (24.28)	17 (3.79)	449 (100)

TABLE 52 Risk classification using OncotypeDX followed by reclassification using IHC4 + clinical score for patients with a NPI score > 3.4 ($n = 257$)

	Low RS, *n* (%)	Intermediate RS, *n* (%)	High RS, *n* (%)	Total, *n* (%)
Low IHC4	111 (43.19)	39 (15.18)	16 (6.23)	166 (64.59)
Intermediate IHC4	13 (5.06)	28 (10.89)	17 (6.61)	58 (22.57)
High IHC4	9 (3.50)	9 (3.50)	15 (5.84)	33 (12.84)
Total	133 (51.75)	76 (29.57)	48 (18.68)	257 (100)

TABLE 53 Proportion of patients free of distant recurrence at 10 years in patients with a NPI score ≤ 3.4 – classified using OncotypeDX and then reclassified using IHC4 + clinical score ($n = 449$)

	Low RS (%)	Intermediate RS (%)	High RS (%)
Low IHC4	98	92	91
Intermediate IHC4	**100**	**100**	**100**
High IHC4	–	–	–

TABLE 54 Proportion of patients free of distant recurrence at 10 years in patients with a NPI score > 3.4 – classified using OncotypeDX and then reclassified using IHC4 + clinical score ($n = 257$)

	Low RS (%)	Intermediate RS (%)	High RS (%)
Low IHC4	92	89	93
Intermediate IHC4	100	75	76
High IHC4	**63**	**89**	60

personal communication). Note that inconsistencies can be observed because of the small sample size of patients within each box (in bold for sample size < 10).

We assumed that the risk of distant recurrence was constant over the first 10 years, using an exponential distribution. This was done in the absence of the Kaplan–Meier data for each subgroup and is acknowledged as a limitation as the risk of recurrence may vary over time.

We also assumed that the risk was reduced by half after 10 years as clinical experts indicated that patients who had not experienced a distant recurrence before 10 years have a lower risk of distant recurrence beyond 10 years. We further assumed that no recurrences would occur after 15 years. Discussion with clinical experts indicated that only a minority of recurrences are likely to occur beyond this date. This assumption was tested in sensitivity analysis.

Patients offered chemotherapy based on the result of the test: OncotypeDX and IHC4 The risk groups identified by the new tests are expected to influence the targeting of chemotherapy. However, other factors will also influence the decision regarding chemotherapy, including clinical

and pathological factors, along with patient choice. In clinical practice, a proportion of women classified as having a low risk of distant recurrence may receive chemotherapy; similarly, a proportion of women classified at high risk may not receive chemotherapy, as shown in previous studies for OncotypeDX undertaken in Spain[173] or in the USA.[174]

To populate our economic model, data from the only identified UK study was used for OncotypeDX in the base case.[78] The Holt *et al.* study reported the OncotypeDX RS and the chemotherapy decision based on knowledge of the RS combined with traditional clinical and pathological parameters. Individual patient-level data were made available to the EAG and were reanalysed by NPI group. Results are presented in *Table 55* (analysis conducted by the EAG using individual patient-level data submitted by Genomic Health). This study, however, has a number of limitations, discussed previously, and these assumptions were tested in sensitivity analysis.

No data exist on the proportion of women who are given chemotherapy according to the results of the IHC4 test. Discussion with clinical experts indicated that interpretation of the OncotypeDX and IHC4 results is likely to be similar. We therefore assumed the same proportions for IHC4 as for OncotypeDX. This is, however, a limitation of the analysis and the assumption was tested in sensitivity analysis.

Benefit of chemotherapy (reduction in the risk of distant recurrence): OncotypeDX and IHC4 The systematic review of evidence identified one study, the Paik *et al.* study, evaluating the benefit of chemotherapy for LN– patients receiving chemotherapy in addition to tamoxifen compared with tamoxifen alone and classified using OncotypeDX.[49,86] In the overall population, the addition of chemotherapy compared with tamoxifen alone was estimated to reduce the risk of distant recurrence by 44% (HR 0.56, 95% CI 0.34 to 0.91). No chemotherapy benefit (reduction in distant recurrence) was found for women classified as low risk of distant recurrence with OncotypeDX (HR 1.31, 95% CI 0.46 to 3.78, $p = 0.61$). A reduction of 39% (HR 0.61, 95% CI 0.24 to 1.59, $p = 0.39$) and 74% (HR 0.26, 95% CI 0.13 to 0.53, $p < 0.001$) was found for the risk of distant recurrence for women receiving chemotherapy in addition to tamoxifen compared with tamoxifen alone and classified as intermediate and high risk of distant recurrence with OncotypeDX respectively. The limitations of this study are described in *Chapter 2, Results: OncotypeDX test*.

The base-case analysis used data from the Paik *et al.* study,[49] assuming the test to be predictive of the benefit of chemotherapy. This may be an optimistic assumption as the effect of chemotherapy might be lower than that reported (as the Paik *et al.* study included younger women and women with HER2+ early breast cancer). However, this study is based on a less effective regimen than is currently used and it is unclear what the overall impact of these factors would be. Univariate sensitivity analyses were conducted varying the benefit of chemotherapy. In addition, because of the limitations of this study (see *Chapter 2*, *Results: OncotypeDX test*), the EAG explored a scenario assuming that all women receiving chemotherapy derive the same benefit in terms of reduction in distant recurrence (i.e. that the test is prognostic only). However, not all clinical

TABLE 55 Proportion of patients receiving chemotherapy after knowledge of the OncotypeDX results:[78] final analysis including 121 patients with ER+, LN–, HER2– breast cancer (EAG analysis)

	Entire cohort	NPI ≤ 3.4	NPI > 3.4
Low RS	1.45%	0.00%	4.55%
Moderate RS	42.42%	38.10%	50.00%
High RS	89.47%	50.00%	94.12%

experts supported this assumption and recommended the use of the predictive evidence available despite the limitations noted above.

No studies were identified for the benefit of chemotherapy in women reclassified using IHC4; therefore, we used indirect evidence. Patients classified as low, intermediate or high risk with OncotypeDX were assumed to derive the same benefit from chemotherapy irrespective of their further reclassification as low, intermediate or high with IHC4. The benefit of chemotherapy for a particular risk group for IHC4 was therefore derived from the known mix of patients with OncotypeDX RS of low, intermediate and high within the IHC4 group. In simple terms, because we know the RS classification of patients reclassified by IHC4, it is possible to apply the benefit of chemotherapy from the RS risk group. This assumes that the reclassification with IHC4 does not provide any additional benefit in terms of identifying patients who would benefit the most from chemotherapy. This may be a conservative assumption; however, it is not possible to ascertain the potential bias of such an assumption.

Clinical parameters: Mammostrat

Compared with OncotypeDX, the evidence base for Mammostrat was less developed and some gaps were identified by the systematic review of the literature (see *Chapter 2*, *Quality of included studies: Mammostrat test*). Most of the evidence available for Mammostrat related to the clinical validity (prognostic ability) of the test. No published analyses of reclassification against Adjuvant! Online or NPI or the impact of the test on decision-making were identified.

Because of the gaps and uncertainties in the data available, an exploratory analysis was carried out to assess the cost-effectiveness of Mammostrat in addition to current clinical practice compared with current clinical practice in England and Wales.

Clinical parameters relating to the three main components of the model are described in turn.

Assignment of patients into risk category and risk of (all) recurrence: Mammostrat Unpublished data from a subset of women with ER+, LN– breast cancer included in the Ring *et al.* study[169] were used to populate the economic model for the risk reclassification against NPI (*Table 56*). (CIC information has been removed.)

The number of recurrences in the same subset of patients ($n = 245$) was also obtained for patients classified by Mammostrat and NPI (*Table 57*).[169]

The fraction of patients who recur was calculated by the EAG for patients classified using Mammostrat and NPI to provide an indication of the risk of recurrence (mean follow-up of 11.7 years) for each prognostic group (*Table 58*).

TABLE 56 Risk reclassification by Mammostrat and by NPI group for a subset of 245 patients included in the Ring *et al.* study[169]

(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)
(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)
(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)
(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)

(CIC information has been removed.)

Data from this subset of patients were used in the economic model in the absence of other evidence.

Furthermore, because this study was conducted outside the UK, the risk was adjusted by calculating the ratio of the risk between patients with a NPI score < 3.4 and for patients with a NPI score > 3.4 for patients classified as low, intermediate or high using Mammostrat. The ratio was then applied to the risk of recurrence (DFS) estimated from the TransATAC trial for patients with a NPI score ≤ 3.4 and patients with a NPI score > 3.4 with ER+, LN–, HER2– breast cancer. The estimated adjusted 10-year risk of being free of distant recurrence used in the economic model is presented in *Table 59*. A sensitivity analysis was conducted using the direct data without adjustment.

Patients offered chemotherapy based on the results of the test: Mammostrat There is no published evidence on how Mammostrat will influence treatment decisions in the UK, or elsewhere. In the base case we assumed that the interpretation of the Mammostrat test would be the same as for patients categorised as low, intermediate or high risk of recurrence with OncotypeDX, using data from the Holt *et al.* study[78] (see *Table 55*). A sensitivity analysis was conducted assuming that no patients classified as low risk, 50% of patients classified as intermediate risk and 100% of patients classified as high risk received chemotherapy.

Benefit of chemotherapy (reduction in the risk of all recurrence): Mammostrat One study has been identified by the systematic review reporting the effect of chemotherapy for patients reclassified

TABLE 57 Number of recurrences by Mammostrat and by NPI group for a subset of 245 patients included in the Ring *et al.* study[169]

	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)
(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)
(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)

TABLE 58 Estimated proportions of patients free of recurrence (EAG calculation)

	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)
(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)
(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)

TABLE 59 Proportions of patients free of recurrence at 10 years used in the economic model for the Mammostrat analysis

	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)
(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)
(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)	(CIC information has been removed)

as low, intermediate or high risk using Mammostrat.[126] This study showed that patients with a low Mammostrat risk (HR 0.4, 95% CI 0.2 to 0.8) and a high Mammostrat risk (HR 0.4, 95% CI 0.2 to 0.9) benefited from chemotherapy whereas patients in the intermediate-risk group did not. The limitations of this study are described in *Chapter 2*, *Quality of included studies: Mammostrat test*.

Data from this study by Ross *et al.*[126] were used in the base case. Sensitivity analyses were undertaken using the 95% CIs of this study.

Clinical parameters: MammaPrint

Compared with OncotypeDX, the evidence base for MammaPrint is less well developed and gaps were identified by the systematic review of the literature (see *Chapter 2*, *Results: MammaPrint test*). The systematic review of the literature indicated that the evidence base, in particular in relation to the prognostic ability of the test, is developing but is based on cohort studies with small sample sizes. None of the studies is based on UK patients and studies are mainly derived from premenopausal women and so are not representative of the population of women with ER+, LN–, HER2– early breast cancer. No robust evidence on clinical utility was identified. The one study identified that considered chemotherapy benefit by MammaPrint risk group had significant limitations.[111] No UK studies of the impact of the test on clinical decision-making were identified.

Because of the gaps and uncertainties in the data available, an exploratory analysis was carried out looking at the cost-effectiveness of adjuvant chemotherapy guided using MammaPrint in addition to current clinical practice compared with current clinical practice in England and Wales.

Clinical parameters relating to the three main components of the model are described in turn.

Assignment of patients into risk category and risk of recurrence: MammaPrint Risk classification data from Bueno de Mesquita *et al.*[109] were used in the economic model as this study contained a relatively large sample size (*Table 60*). The study included Dutch women only. In this study, data from previous cohorts were pooled and reanalysed. Women included in this study were mainly premenopausal and included ER– women. Premenopausal women and ER– women are more likely to be classified as poor prognosis with MammaPrint and this raises concerns relating to the generalisability of the reclassification data to the population considered in the economic model.

The systematic review of evidence did not identify any studies that presented the risk of recurrence in a UK population for patients classified using MammaPrint. The risk of distant recurrence for patients receiving endocrine therapy was therefore derived using data from the same non-UK study as the risk reclassification data but from a different subset of patients.[105] This study reported the risk of recurrence (distant metastases as first event) for patients stratified by NPI, Adjuvant! Online and MammaPrint. This evidence relates to premenopausal women in a non-UK cohort and is based on patients receiving tamoxifen only. The study separated patients into those with a low NPI/low Adjuvant! Online score, an intermediate NPI/high Adjuvant! Online score or discordant results, rather than just low and intermediate NPI only (*Table 61*).

TABLE 60 Risk reclassification of patients using MammaPrint by NPI group

	Good prognosis using MammaPrint, *n* (%)	Poor prognosis using MammaPrint, *n* (%)	All, *n* (%)
NPI ≤ 3.4	259 (71.75)	102 (28.25)	361 (100)
NPI > 3.4	84 (24.71)	256 (75.29)	340 (100)
All	343 (48.93)	358 (51.07)	701 (100)

This study also showed, from a different subset of patients, that 94% of the discordances between NPI and Adjuvant! Online occurred in patients with a NPI ≤ 3.4. Consequently, the EAG pooled data for patients with a low NPI and low Adjuvant! Online score and patients with a discordant result between the two prognostic tools.

This non-UK study presented the risk of distant recurrence as first event (rather than time to any distant recurrence) in predominantly premenopausal women. We therefore adjusted the risk using data from the TransATAC trial to more closely reflect the expected risk within a UK population. The ratio of the risk between the good and poor prognosis groups was calculated for patients with a NPI score ≤ 3.4 and a NPI score > 3.4. This ratio was then applied to the 10-year risk of distant recurrence for patients with a NPI score ≤ 3.4 and a NPI score > 3.4 estimated from the TransATAC trial (Professor Mitch Dowsett, Royal Marsden Hospital, London, September 2011, personal communication) in a UK population with ER+, LN–, HER2– early breast cancer. This is acknowledged to be a limitation of the analysis.

The estimated adjusted 10-year risk of being free of distant recurrence used in the economic model is presented in *Table 62*.

Patients offered chemotherapy based on the results of the test: MammaPrint No UK studies were identified on the proportion of patients who receive chemotherapy based on the MammaPrint test results (see *Chapter 2, Results: MammaPrint test*). A Dutch study (the MircoarRAy PrognoSTics in Breast CancER study or RASTER)[186] was identified that provided information on the number of patients who have been recommended and/or offered chemotherapy based on poor or good outcome classification according to the CBO[105] and subsequently following knowledge of the MammaPrint test result.

The study showed that, overall, out of the 273 and 208 patients classified as having good and poor prognosis by MammaPrint, respectively, chemotherapy advice was given to 13.55% and 87.50% respectively. Among patients considered to be at low risk according to the CBO,[105] 1.80% and 68.42% were classified as having good and poor prognosis using MammaPrint, respectively, and were recommended chemotherapy. The figures for patients considered to be at high risk according to the CBO were 32.08% and 98.48% respectively.

Data from this study were used in the base-case analysis in the absence of other evidence, assuming that the NPI category was a proxy for the risk group defined using the CBO

TABLE 61 Proportion of patients free of recurrence at 10 years (distant metastases as first event) for patients reclassified using MammaPrint and NPI/Adjuvant! Online

	Low NPI/ low Adjuvant! Online score (%)	Intermediate NPI/ high Adjuvant! Online score (%)	Discordant results (%)
Good prognosis using MammaPrint	87	77	92
Poor prognosis using MammaPrint	69	45	59
All	82	53	83

TABLE 62 Proportions of patients free of recurrence at 10 years used in the economic model for the MammaPrint analysis

	Good prognosis using MammaPrint (%)	Poor prognosis using MammaPrint (%)
NPI ≤ 3.4	97.7	91.3
NPI > 3.4	94.3	83.7

guidelines.[105] This is a limitation. Furthermore, the study included ER– patients and the MammaPrint test classified most ER– patients as having poor prognosis. Furthermore, there were some concerns with the study design as there were some amendments to the protocol and patients from different cohorts had been pooled. In total, 84% of patients were aged < 55 years and the population included women with ER+ and ER– early breast cancer. It was unclear how representative patients were compared with those seen in clinical practice in the UK and how this study would relate to UK clinical practice.

Benefit of chemotherapy (reduction in the risk of distant recurrence): MammaPrint A study was identified providing data on the benefit of chemotherapy for ER+, LN– patients reclassified as having a good or a poor prognosis with MammaPrint.[112] This study showed that patients with both a poor and a good prognosis benefited from chemotherapy (compared with no chemotherapy) in terms of a reduction in distant metastasis, but the benefit was not significant for patients with a good prognosis. The limitations of this study are described in *Chapter 2, Results: MammaPrint test*.

Summary of inputs used

To summarise, we assign patients into different boxes according to the risk group predicted by the new test (*Figure 12*). A proportion of these patients are assumed to receive chemotherapy. In the current practice arm, the proportion is informed by registry data and is assumed to be independent of the risk group (as oncologists are blind to the results of the new test). For the intervention arm, the proportion is linked to the risk group assigned by the new test. Patients are then at risk of developing a recurrence, and die from breast cancer or other causes.

A summary of the main inputs is presented in *Table 63*.

The main assumptions used in the base-case economic model are summarised below:

- The population assessed was all women with ER+, LN–, HER2– early breast cancer.
- A subgroup analysis was carried out in women with ER+, LN–, HER2– early breast cancer with a NPI score > 3.4 only.
- Interventions assessed were OncotypeDX, MammaPrint, IHC4 and Mammostrat.
- The comparator was current clinical practice in England and Wales. We used registry data from the ECRIC and WMCIU to reflect current clinical practice. We assumed that data from these two registries were representative of all trusts in England and Wales.
- Women were separated by NPI score (≤ 3.4 and > 3.4) to conduct subgroup analysis but also to capture the prognostic value of the treatment decision using clinicopathological parameters based on NPI, that is, women with a low NPI score are likely to have a lower risk than women with a NPI score > 3.4 but are also less likely to receive chemotherapy under current clinical practice.
- The NPI score distribution was taken from data from two registries (ECRIC and WMCIU).
- The starting age of the cohort was 58.3 years.
- We assumed that the new test will not be considered in women aged > 75 years as the decision for chemotherapy might be limited because of frailty and comorbidities.
- The model used a lifetime horizon.
- The model used a 6-monthly cycle length.
- The model adopted the perspective of the UK NHS and PSS.
- Costs and benefits were discounted at 3.5% annually.
- A primary analysis was carried out comparing current clinical practice alone, OncotypeDX in addition to current clinical practice and IHC4 in addition to current clinical practice as this presented the most robust evidence to populate the economic model (mainly derived from UK sources).

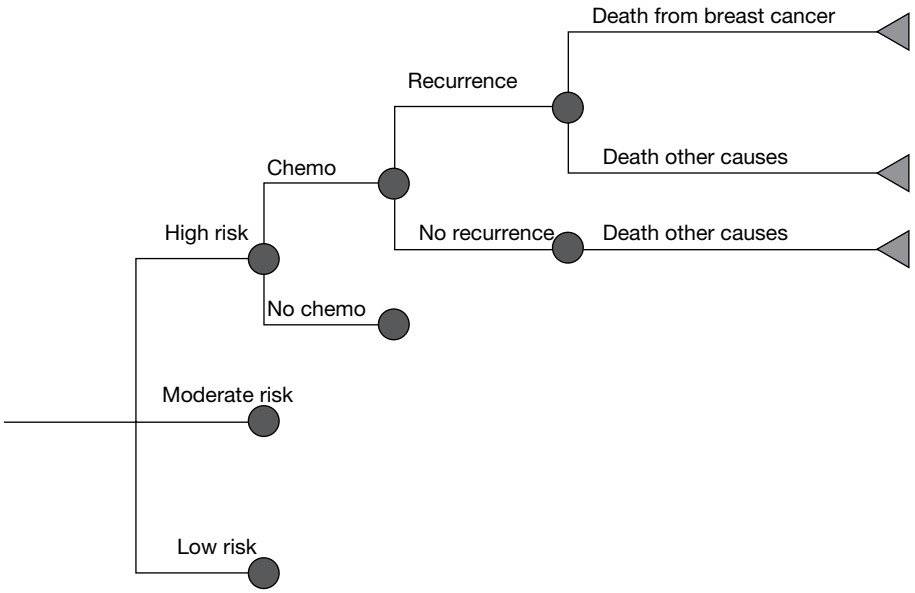


FIGURE 12 Schematic diagram showing how the inputs were assembled.

TABLE 63 Summary of inputs

Parameter	Base-case value	Distribution	Source
Baseline age (years)	58.3		ECRIC (West Midland Cancer Intelligence Unit, July 2011, personal communication; Eastern Cancer Registration and Information Centre, July 2011, personal communication)
Dosage per BSA (mg/m^2)	1.75	Normal (SE 0.01)	Sacco *et al.*[168]
Cost of the tests			
OncotypeDX	£2580		
IHC4	£150		Personal communication and assumption (Professor Mitch Dowsett, Royal Marsden Hospital, London, September 2011)
MammaPrint	£2675		
Mammostrat	£1135		
Handling fresh tissue	£250		Personal communication (Simon Cross, Reader and Honorary Consultant, Royal Hallamshire Hospital, Sheffield, July 2011)
Endocrine therapy cost (6-monthly cost)			
First 5 years	£334	Normal	BNF[161] plus NICE guidance[162] and assumption
Remaining 3 years	£65		
Cost of monitoring in recurrence-free state (6-monthly cost) (£)			
First year	£151	Normal	NHS reference costs 2010/11[175] and assumption
Remaining 4 years	£87		
Chemotherapy cost (one-off)			
Drug, administration and monitoring	£4099	Normal	BNF,[161] NHS reference costs 2010/11[175] and assumption
Short-term adverse events	£276		NHS reference costs 2010/11[175] and assumption
G-CSF	£485		
Long-term adverse events after chemotherapy			
8-year probability of AML	0.37%	Beta (129,1095)	Praga *et al.*[180]
Time spent in AML health state	8 months		Assumption based on NICE STA18[181]
Lifetime cost	£11,500		Assumption based on NICE STA18[181]

TABLE 63 Summary of inputs (*continued*)

Parameter	Base-case value	Distribution	Source
Recurrence cost (6-monthly)	£4082	Normal	Derived from Thomas *et al.*[157]
End-of-life cost (one-off) – after death from breast cancer	£4038	Normal (SE £454)	Campbell *et al.*[176]
Local recurrence cost – one-off	£14,132	Normal (SE £1853)	Derived from Karnon *et al.*[182] and PSSRU[178]
Health state utilities			
Recurrence-free state	0.824	Beta (353,75)	Lidgren *et al.*[148]
Distant recurrence	0.685	Beta (171,79)	Lidgren *et al.*[148]
Local recurrence (decrement per patient)	–0.108	Normal (SE 0.04)	Campbell *et al.*[176]
AML	0.26		Younis *et al.*[183]
Chemotherapy (decrement per patient)	–0.038		Campbell *et al.*[176]
Terminal care cost (final 3 months)	0.159	Normal (SE 0.04)	Campbell *et al.*[176]
Baseline NPI score distribution			
≤ 3.4	66.2%	Beta (2602,819)	ECRIC and WMCIU (West Midland Cancer Intelligence Unit, July 2011, personal communication; Eastern Cancer Registration and Information Centre, July 2011, personal communication)
> 3.4	33.8%		
Proportion of patients receiving chemotherapy under current clinical practice			
NPI ≤ 3.4	4.6%	Beta (3263,158)	ECRIC and WMCIU (West Midland Cancer Intelligence Unit, July 2011, personal communication; Eastern Cancer Registration and Information Centre, July 2011, personal communication)
NPI > 3.4	33.6%	Beta (1160, 587)	ECRIC and WMCIU (West Midland Cancer Intelligence Unit, July 2011, personal communication; Eastern Cancer Registration and Information Centre, July 2011, personal communication)
Proportion of patients receiving chemotherapy after knowledge of the result of the new test (Holt *et al.* study[78])	See *Table 55*		EAG analysis
Risk reclassification			
OncotypeDX and IHC4	See *Tables 51* and *52*	Beta	TransATAC trial (personal communication)
MammaPrint	See *Table 60*	Beta	Bueno-de-Mesquita *et al.*[109]
Mammostrat	See *Table 56*	Beta	Subset of the Ring *et al.* study[125] (Clarient[169])
Risk of recurrence			
OncotypeDX and IHC4	See *Tables 53* and *54*		TransATAC trial (Professor Mitch Dowsett, Royal Marsden Hospital, London, September 2011, personal communication)
MammaPrint	See *Table 62*		Derived from Bueno-de-Mesquita *et al.*[109] and the TransATAC trial (Professor Mitch Dowsett, Royal Marsden Hospital, London, September 2011, personal communication)
Mammostrat	See *Table 59*		Derived Professor Mitch Dowsett, Royal Marsden Hospital, London, September 2011, from a subset of the Ring *et al.* study[125] (Clarient[169]) and the TransATAC trial (Professor Mitch Dowsett, Royal Marsden Hospital, London, September 2011, personal communication)

continued

TABLE 63 Summary of inputs (*continued*)

Parameter	Base-case value	Distribution	Source
Benefit of chemotherapy (HR)			
OncotypeDX		Log-normal	Paik *et al.*[49]
	Low RS: 1.31	Low RS: 95% CI 0.46 to 3.78	
	Intermediate RS: 0.61	Intermediate RS: 95% CI 0.24 to 1.59	
	High RS: 0.26	High RS: 95% CI 0.13 to 0.53	
MammaPrint		Log-normal	Knauer *et al.*[110]
	Good prognosis: 0.26	Good prognosis: 95% CI 0.03 to 2.02	
	Poor prognosis: 0.35	Poor prognosis: 95% CI 0.17 to 0.71	
Mammostrat		Log-normal	Ross *et al.*[126]
	Low: 0.4	Low: 95% CI 0.2 to 0.8	
	Intermediate: 1	Intermediate: 95% CI NA	
	High: 0.4	High: 95% CI 0.2 to 0.9	

NA, not applicable; STA, Single Technology Appraisal.

- Two exploratory analyses were carried out comparing current clinical practice alone against MammaPrint or Mammostrat in addition to current clinical practice.
- We assumed that the diagnostic tool does not affect the prognosis of patients if there is no change in the adjuvant treatment.
- In the comparator arm (current clinical practice), the probability of receiving chemotherapy was taken from data from two registries (ECRIC and WMCIU). It was assumed that the probability of receiving chemotherapy was the same irrespective of the reclassification of patients with the new test in the low-, intermediate- or high-risk group. This is likely to be a conservative assumption.
- The probability of receiving chemotherapy in the intervention arm was based on the assigned risk group, with a greater likelihood of receiving chemotherapy for patients classified as high risk of recurrence/distant recurrence than for patients classified as low risk.
- The economic model used distant recurrence as the primary outcome.
- Only locoregional recurrence that led to a distant recurrence was included, by considering the cost and quality of life decrement associated with locoregional recurrence. We assumed that 10.5% of distant recurrences were preceded by a local recurrence.
- We assumed that patients treated with chemotherapy can develop long-term adverse events. The model included the development of AML only.
- The risk of recurrence was assumed to be constant in the first 10 years and to be halved between 10 and 15 years; no recurrence was assumed after 15 years.
- Median survival after a distant recurrence was assumed to be 40.1 months.
- We assumed that all patients receive endocrine therapy. Five regimens were considered: tamoxifen for 5 years, anastrozole for 5 years, letrozole for 5 years, tamoxifen for 2 years plus exemestane for the final 3 years and tamoxifen (or other endocrine therapy regimens) for 5 years followed by extended therapy with letrozole for a further 3 years.
- Patients were assumed to have two follow-up appointments in the first year and one follow-up appointment every subsequent year. Patients were assumed to have one mammogram every year for a maximum of 5 years.

- Patients treated with chemotherapy receive six cycles of FEC75. We assumed that patients have a separate outpatient appointment before drug administration. Monitoring includes a liver function test, urea and electrolytes tests and a full blood count. It was further assumed that 25% of patients have an echocardiogram before starting chemotherapy.
- Short-term adverse events after chemotherapy were included based on the probability of patients treated with FEC100 suffering from short-term adverse events, as no data were available for FEC75.
- We assumed that 25% of patients treated with chemotherapy receive G-CSF (maximum of three cycles) for the secondary prevention of febrile neutropenia.
- A decrement in utility was assumed for patients receiving chemotherapy and for patients dying from breast cancer-specific causes.

Assumptions specific to each test are described in *Table 64*.

Independent economic model: results

The primary analysis compared current clinical practice with treatment guided using OncotypeDX and IHC4. This involved using data from a direct comparison between OncotypeDX and IHC (Professor Mitch Dowsett, Royal Marsden Hospital, London, September 2011, personal communication).

In addition to the primary analysis, exploratory analyses were undertaken for Mammostrat and MammaPrint. These analyses were exploratory because of the significant limitations in the evidence base, and results need to be interpreted with consideration of the assumptions made and the robustness of the evidence used.

All analyses assumed that the new tests were used in addition to current prognostic tools. Base-case analyses assumed the tests to have predictive ability, that is, patients in the high-risk group benefit relatively more from reduction in recurrences following chemotherapy than patients in the lower-risk groups; this assumption was tested in sensitivity analysis.

Cost-effectiveness of treatment guided using OncotypeDX and IHC4

Two analyses are presented:

1. The tests were given to all women with ER+, LN–, HER2– early breast cancer.
2. The tests were given only to women with ER+, LN–, HER2– early breast cancer with a NPI score >3.4. This subgroup analysis was undertaken to explore the impact of targeting the tests at patients with intermediate risk. It is considered likely that the majority of women with a NPI score ≤ 3.4 would be considered low risk and would not receive chemotherapy under current practice or using the new tests and therefore the tests would have a limited impact on the management of these women.

The new tests were offered to all women with ER+, LN–, HER2– early breast cancer

Deterministic results

Assuming that the tests were offered to all women with ER+, LN–, HER2– early breast cancer, treatment guided using OncotypeDX was predicted to lead to an increase in the proportion of patients receiving chemotherapy compared with current clinical practice under our base-case assumptions (19.11% vs. 14.42%). On the contrary, treatment guided using IHC4 was predicted

TABLE 64 List of assumptions specific to each test

OncotypeDX	IHC4	Mammostrat	MammaPrint
■ OncotypeDX used as a discrete variable (low, intermediate, high RS) instead of continuous RS ■ Original RS groups are used: low (RS < 18), intermediate (RS between 18 and 30), high (RS ≥ 31) ■ Reclassification evidence from the TransATAC trial (UK population) by NPI, by OncotypeDX and by IHC4 ■ Risk of recurrence from the TransATAC trial (UK population) by NPI, by OncotypeDX and by IHC4 ■ Holt *et al.* study (UK population)[78] used to inform the proportion of patients classified as low, intermediate or high (receiving chemotherapy) ■ The benefit of chemotherapy in terms of reduction of distant recurrence was taken from the Paik *et al.* study[49] of a US cohort treated with CMF/CM. The study included a large proportion of premenopausal women and some women with HER2+. No benefit was assumed for women with a low OncotypeDX RS	■ Composite score of IHC4 in addition to clinicopathological parameters ■ The test was assumed to be reproducible (notably Ki-67 element) ■ IHC4 used as a discrete variable (low, intermediate, high) instead of continuous risk of distant recurrence ■ Patients with low, intermediate and high risk of distant recurrence were defined as patients with a predicted risk of distant recurrence of < 10%, 10–20% and > 20% respectively (same approach as for OncotypeDX) ■ Reclassification evidence from the TransATAC trial (UK population) by NPI, by OncotypeDX and by IHC4 ■ Risk of recurrence from the TransATAC trial (UK population) by NPI, by OncotypeDX and by IHC4 ■ We assumed that physicians interpret OncotypeDX and IHC4 in the same way. The Holt *et al.* study (UK population)[78] was used to inform the proportion of patients receiving chemotherapy classified as low, intermediate or high ■ The benefit of chemotherapy in terms of reduction of distant recurrence was applied from the RS risk group (using OncotypeDX – Paik *et al.* study[49]) as no specific data for IHC4. This assumes that reclassification with IHC4 does not provide any additional benefit (possibly conservative assumption) ■ We assumed that it cost an additional £100–200 to perform the IHC4 test	■ Reclassification evidence from a subset of patients from the Ring *et al.* study (USA).[125] We assumed that the reclassification holds for the UK ■ The risk of recurrence was derived from the same subset of the Ring *et al.* study (USA)[125] with adjustment for the UK population using data from the TransATAC trial ■ We assumed that physicians interpret OncotypeDX and Mammostrat in the same way. The Holt study (UK population)[78] was used to inform the proportion of patients receiving chemotherapy classified as low, intermediate or high ■ Health and cost outcomes after a recurrence were assumed to be the same as for distant recurrence as the EAG used distant recurrence as a primary outcome ■ The benefit of chemotherapy in terms of reduction of recurrence (any) was taken from the Ross *et al.* study[126] of a US cohort treated with CMF/CM. No benefit was assumed for women with a moderate Mammostrat risk	■ Reclassification from Bueno-de-Mesquita *et al.*[109] from a pooled analysis of Dutch women (mainly premenopausal and including ER–). We assumed that the reclassification hold for the UK ■ The risk of recurrence was derived from a different subset of patients from the Bueno-de-Mesquita *et al.* study[109] of Dutch women (mainly premenopausal and including ER–), with adjustment for the UK population using data from the TransATAC trial ■ We used data from the RASTER study[186] (Dutch women) to estimate the probability of chemotherapy for patients classified as having good or poor prognosis with MammaPrint. The study separated patients using the CBO guidelines[102] instead of NPI (we therefore assumed that CBO is a proxy for NPI). The study also included a mix of ER+/–, HER2 +/– women. Data from this study were assumed to hold for the UK in the absence of other evidence ■ The benefit of chemotherapy in terms of reduction of distant recurrence was taken from the Knauer *et al.* study[110] of a Dutch cohort treated with CMF or anthracycline +/– taxane regimens (pooled analysis of previous cohort). The study had several flaws. We assumed that women with a good prognosis had a greater benefit (although non-significant) than women with a poor prognosis ■ MammaPrint offers three optional tests for no additional cost. This was not considered

to lead to a reduction in the proportion of patients receiving chemotherapy compared with current clinical practice under our base-case assumptions (9.57% vs. 14.42%). More women were classified as high or intermediate risk with OncotypeDX than with IHC4, and were therefore more likely to receive chemotherapy.

For a cohort of 1000 women with ER+, LN–, HER2– early breast cancer, we predicted that 76 distant recurrences would occur under current clinical practice. Treatment guided using

OncotypeDX and IHC4 was predicted to reduce the number of distant recurrences to 64 and 71, respectively, under the assumptions used for the base-case analysis.

The mean discounted cost of treatment guided using current clinical practice, OncotypeDX and IHC4 was estimated to be £6519, £9094 and £6340 per patient respectively. The breakdown of costs by category is presented in *Table 65*. Treatment guided using OncotypeDX was estimated to reduce the costs associated with the management of recurrences (distant and local recurrences, terminal care) but incurred additional costs to perform the test (£2580) compared with current clinical practice. IHC4 also reduced the costs associated with recurrences, but also reduced the costs associated with chemotherapy, for an additional test cost of £150 per patient compared with current clinical practice.

The mean discounted QALYs were 13.44, 13.54 and 13.49 for current clinical practice, OncotypeDX and IHC4 respectively.

Compared with current clinical practice, the incremental cost for treatment guided using OncotypeDX was estimated to be £26,940 per QALY gained. Chemotherapy treatment guided using IHC4 was dominant (i.e. provided more QALYs at a lower cost) compared with current clinical practice. These results are based on the assumption that OncotypeDX has predictive ability. This assumption was tested in sensitivity analysis.

Treatment guided using OncotypeDX, IHC4 and current clinical practice was also compared using incremental analysis, that is, the least effective strategy was compared with the next least effective strategy that was neither dominated nor extendedly dominated. The base-case costs and QALYs are shown in *Table 66*. Treatment guided using OncotypeDX provided the most benefit (13.54 QALYs) at the highest cost (£9094). The ICER for treatment guided using OncotypeDX compared with treatment guided using IHC4 was £55,406 per QALY gained.

TABLE 65 Breakdown of costs for the primary analysis: current clinical practice compared with OncotypeDX and IHC4

Cost categories	Current clinical practice (£)	OncotypeDX (£)	IHC4 (£)
Recurrence free	926	928	927
Distant recurrence	1277	1081	1199
Terminal care	222	188	209
Local recurrence	92	78	87
Endocrine therapy	3298	3307	3302
Chemotherapy	591	783	392
Short-term adverse events	110	145	73
Long-term adverse events	3	4	2
Cost of test	0	2580	150
Total cost	6519	9094	6340

TABLE 66 Deterministic ICER for the primary analysis comparing OncotypeDX and IHC4 with current clinical practice assuming that the test is given to all women with ER+, LN–, HER2– early breast cancer in England and Wales

	Mean cost (£)	Mean QALYs	ICER (£)	Incremental analysis (£)
OncotypeDX	9094	13.54	26,940	55,406
IHC4	6340	13.49	**Cost saving**	
Current clinical practice	6519	13.44		

Probabilistic results

The results of PSA using 2500 iterations are shown in *Table 67*. Treatment guided using IHC4 remained dominant (i.e. provided more QALYs at a lower cost) compared with current clinical practice. The incremental cost for treatment guided using OncotypeDX was £29,503 per QALY gained compared with current clinical practice and £64,111 per QALY gained compared with IHC4 (see *Table 67*).

Figure 13 shows the cost-effectiveness acceptability curve (CEAC) using results generated over a lifetime horizon. The curve shows the probability of each test being cost-effective for different monetary values that the decision-maker may be willing to pay for an additional QALY. The CEAC shows that treatment guided by IHC4 was the most cost-effective strategy compared with current clinical practice and OncotypeDX when using a willingness-to-pay threshold of £20,000 per QALY gained (in 99.48% of cases). The probability that treatment guided using OncotypeDX was cost-effective at a £20,000 threshold was 0.40% in the incremental analysis and 12.44% compared with current clinical practice alone.

The new tests were offered only to women with ER+, LN–, HER2– early breast cancer with a Nottingham Prognostic Index score > 3.4

Deterministic results

Assuming that the tests were offered only to women with ER+, LN–, HER2– early breast cancer with a NPI score > 3.4, a greater proportion of patients were predicted to receive chemotherapy when using OncotypeDX (under our base-case assumptions) and a lower proportion were

TABLE 67 Probabilistic ICER (2500 iterations) for the primary analysis comparing OncotypeDX and IHC4 with current clinical practice assuming that the test is given to all women with ER+, LN–, HER2– early breast cancer in England and Wales

	Mean cost (£)	Mean QALYs	ICER (£)	Incremental analysis (£)
OncotypeDX	9100	13.52	29,503	64,111
IHC4	6332	13.48	**Cost saving**	
Current clinical practice	6507	13.44		

FIGURE 13 Cost-effectiveness acceptability curve for the primary analysis comparing OncotypeDX and IHC4 with current clinical practice assuming that the test is given to all women with ER+, LN–, HER2– early breast cancer in England and Wales.

predicted to receive chemotherapy when using IHC4 compared with current clinical practice (34.72% vs. 26.31% vs. 33.60% respectively).

Treatment guided using IHC4 was dominant (i.e. provided more QALYs at a lower cost) compared with current clinical practice. The incremental cost for treatment guided using OncotypeDX was £9007 per QALY gained compared with current clinical practice and £26,859 per QALY gained compared with IHC4 (*Table 68*). This is based on the assumption that OncotypeDX has predictive ability. This assumption was tested in sensitivity analysis.

Probabilistic results

The results of PSA using 2500 iterations are shown in *Table 69*. Treatment guided using IHC4 remained dominant (i.e. provided more QALYs at a lower cost) compared with current clinical practice. The incremental cost for treatment guided using OncotypeDX was £9774 per QALY gained compared with current clinical practice and £31,125 per QALY gained compared with IHC4 (see *Table 69*).

The CEAC (*Figure 14*) shows that treatment guided by IHC4 was the most cost-effective strategy when using a willingness-to-pay threshold of £20,000 per QALY gained (in 81.24% of cases). The probability that treatment guided using OncotypeDX was cost-effective at a £20,000 threshold was 18.60% in the incremental analysis and 91.56% compared with current clinical practice alone.

Univariate sensitivity analyses: parameters

A range of univariate sensitivity analyses were carried out to explore the impact of varying the main model parameters. Results of the univariate sensitivity analysis assuming that the tests were offered to all women with ER+, LN–, HER2– early breast cancer are presented in *Table 70*. Results for the univariate sensitivity analysis assuming that the tests were offered only to women with a NPI score >3.4 are presented in *Table 71*.

OncotypeDX

The main model parameters were varied within reasonable ranges. The ICER for OncotypeDX compared with current clinical practice was mainly sensitive (defined as changes in the ICER by ≥±10%) to the assumptions about the time horizon, the starting age of the cohort, the risk of recurrence, the proportion of patients receiving chemotherapy after reclassification with the new test, the benefit of chemotherapy and the NPI score distribution (see *Tables 70* and *71*).

TABLE 68 Deterministic ICER for the primary analysis comparing OncotypeDX and IHC4 with current clinical practice assuming that the test is given only to women with a NPI score >3.4

	Mean cost (£)	Mean QALYs	ICER (£)	Incremental analysis (£)
OncotypeDX	10,911	13.06	9007	26,859
IHC4	8318	12.97	**Cost saving**	
Current clinical practice	8816	12.83		

TABLE 69 Probabilistic ICER for the primary analysis comparing OncotypeDX and IHC4 with current clinical practice assuming the test to be given to women with a NPI >3.4 only

	Mean cost (£)	Mean QALYs	ICER (£)	Incremental analysis (£)
OncotypeDX	10,924	13.05	9774	31,125
IHC4	8305	12.96	**Cost saving**	
Current clinical practice	8797	12.83		

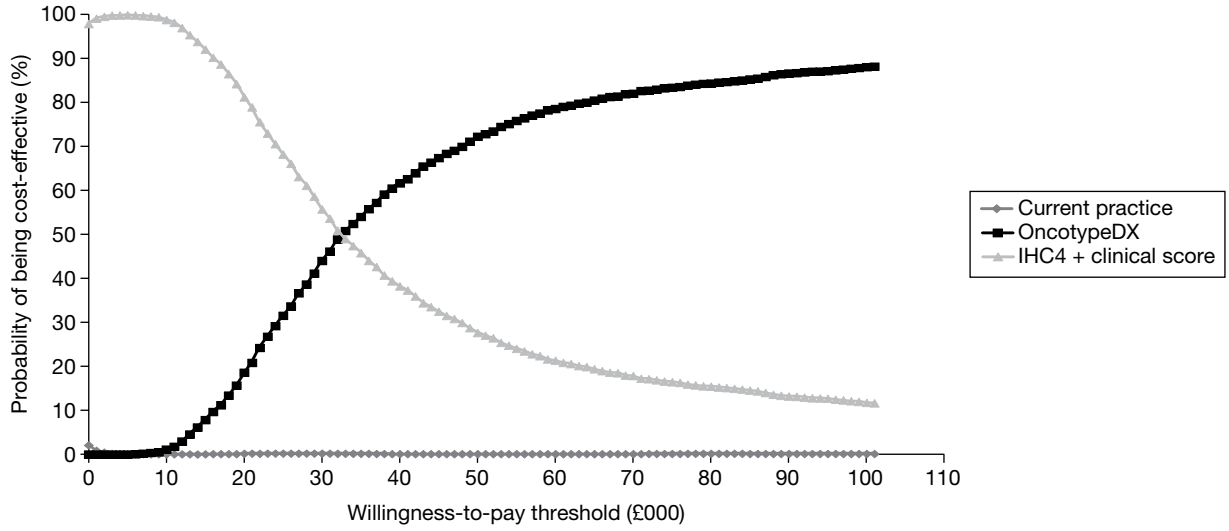


FIGURE 14 Cost-effectiveness acceptability curve for the primary analysis comparing OncotypeDX and IHC4 with current clinical practice assuming that the test is given only to women with a NPI score >3.4.

The ICER was sensitive to the assumed benefit of chemotherapy. The ICER increased (less favourable to OncotypeDX) assuming a lower benefit of chemotherapy (–20% to –40%). The ICER also deteriorated (less favourable to OncotypeDX) significantly assuming that the test was prognostic only, that is, the same relative reduction in the risk of distant recurrence following chemotherapy was applied whether patients were classified as low, intermediate or high risk according to the OncotypeDX RS classification.

The ICER increased (less favourable to OncotypeDX) as the time horizon decreased or the age increased, given that less benefit can be accrued over time.

A reduction in the risk of distant recurrence increased the ICER (less favourable to OncotypeDX) whereas an increase in the risk of distant recurrence improved the ICER in favour of OncotypeDX. Given that more recurrences can be avoided if there is an increase in the risk of distant recurrence, more of the cost of the test can be offset.

Furthermore, the ICER was sensitive to the assumptions about the proportion of patients who received chemotherapy depending on the results of the test (interpretation of the test). If we assumed that chemotherapy was guided solely by the test results, so no women classified as low risk, 50% of women classified as intermediate risk and 100% of women classified as high risk with OncotypeDX receive chemotherapy (for women with both a NPI score ≤3.4 and a NPI score >3.4), the ICER improved (more favourable to OncotypeDX) because chemotherapy is targeted to patients who, according to the test, are likely to benefit the most from it. In addition, the ICER was very sensitive to the assumption about the probability of chemotherapy in patients classified as intermediate risk with OncotypeDX. The ICER ranged from £22,812 to £35,629 if the test was given to all women and from £8371 to £10,022 if the test was given only to women with a NPI score >3.4, assuming that the probability of receiving chemotherapy was the same as for patients classified as low and high risk respectively.

The ICER improved (more favourable to OncotypeDX) when using the NPI distribution from the Holt *et al.* study[77] in the model as more patients were classified with a NPI score >3.4. This group of patients was shown to derive a greater benefit from the new test.

TABLE 70 Results of the univariate sensitivity analysis for the primary analysis assuming that the new tests are offered to all women with ER+. LN–, HER2– early breast cancer in England and Wales

	OncotypeDX		IHC4		Clinical practice		ICER (£)	
	QALYs	Cost (£)	QALYs	Cost (£)	QALYs	Cost (£)	OncotypeDX vs. clinical practice	IHC4 vs. clinical practice
Base case	13.54	9094	13.49	6340	13.44	6519	26,940	**Cost saving**
Time horizon = 5 years	4.11	7947	4.11	5093	4.10	5209	632,318	**Cost saving**
Time horizon = 10 years	7.05	8680	7.04	5883	7.03	6036	120,123	**Cost saving**
Time horizon = 20 years	11.08	9080	11.04	6324	11.01	6502	39,368	**Cost saving**
Starting age = 50 years	15.51	9166	15.45	6416	15.39	6597	21,632	**Cost saving**
Starting age = 60 years	12.97	9066	12.93	6310	12.88	6487	28,932	**Cost saving**
Starting age = 70 years	9.83	8792	9.80	6017	9.77	6184	47,796	**Cost saving**
Reduction in the risk of recurrence by 10%	13.59	8975	13.54	6207	13.50	6379	29,960	**Cost saving**
Reduction in the risk of recurrence by 20%	13.64	8853	13.60	6073	13.56	6236	33,784	**Cost saving**
Increase in the risk of recurrence by 10%	13.49	9212	13.43	6471	13.38	6656	24,494	**Cost saving**
Increase in the risk of recurrence by 20%	13.44	9328	13.37	6599	13.32	6791	22,473	**Cost saving**
Recurrence up to 20 years	13.50	9229	13.45	6488	13.40	6673	25,298	**Cost saving**
No changed in the risk of distant recurrence between 10 and 15 years	13.48	9277	13.42	6541	13.37	6728	24,342	**Cost saving**
Reduction in the risk of distant recurrence by 75% between 10 and 15 years	13.57	9002	13.52	6238	13.48	6411	28,520	**Cost saving**
Proportion of local recurrence before a distant recurrence = 5%	13.54	9053	13.49	6294	13.44	6470	27,034	**Cost saving**
Proportion of local recurrence before a distant recurrence = 20%	13.54	9165	13.49	6418	13.44	6602	26,780	**Cost saving**
Proportion of local recurrence before a distant recurrence = 30%	13.54	9239	13.49	6500	13.44	6689	26,611	**Cost saving**
Time in distant recurrence health state different by prognosis group	13.56	9359	13.51	6648	13.46	6772	26,793	**Cost saving**
Time in distant recurrence – LCI	13.51	8809	13.46	6023	13.41	6181	26,248	**Cost saving**
Time in distant recurrence – UCI	13.56	9354	13.51	6628	13.47	6825	27,629	**Cost saving**
Proportion of patients receiving chemotherapy under current clinical practice = ECRIC only	13.54	9094	13.49	6340	13.44	6493	26,830	**Cost saving**
Proportion of patients receiving chemotherapy under current clinical practice = WMCIU only	13.54	9094	13.49	6340	13.45	6605	27,428	**Cost saving**
Proportion of patients receiving chemotherapy under current clinical practice = Holt *et al.* study[77]	13.54	9094	13.49	6340	13.48	7230	35,951	**Cost saving**

continued

TABLE 70 Results of the univariate sensitivity analysis for the primary analysis assuming that the new tests are offered to all women with ER+. LN–, HER2– early breast cancer in England and Wales (*continued*)

	OncotypeDX		IHC4		Clinical practice		ICER (£)	
	QALYs	Cost (£)	QALYs	Cost (£)	QALYs	Cost (£)	OncotypeDX vs. clinical practice	IHC4 vs. clinical practice
Increase by 20% in the proportion of patients receiving chemotherapy under current clinical practice	13.54	9094	13.49	6340	13.45	6632	28,346	**Cost saving**
Increase by 30% in the proportion of patients receiving chemotherapy under current clinical practice	13.54	9094	13.49	6340	13.45	6688	29,182	**Cost saving**
Reduction by 20% in the proportion of patients receiving chemotherapy under current clinical practice	13.54	9094	13.49	6340	13.43	6405	25,810	**Cost saving**
Reduction by 30% in the proportion of patients receiving chemotherapy under current clinical practice	13.54	9094	13.49	6340	13.43	6348	25,326	**Cost saving**
Assumption about who would receive chemotherapy with the new tests	(CiC information has been removed)	(CiC information has been removed)	13.49	6308	13.44	6519	23,765	**Cost saving**
Reduction in the benefit of chemotherapy by 10%	(CiC information has been removed)	(CiC information has been removed)	13.48	6350	13.44	6525	28,416	**Cost saving**
Reduction in the benefit of chemotherapy by 20%	13.52	9128	13.48	6359	13.43	6532	29,945	**Cost saving**
Reduction in the benefit of chemotherapy by 30%	13.51	9144	13.47	6368	13.43	6538	31,529	**Cost saving**
Reduction in the benefit of chemotherapy by 40%	13.51	9158	13.47	6376	13.43	6543	33,169	**Cost saving**
Assuming the same benefit of chemotherapy for everyone = 40%	13.49	9200	13.47	6374	13.45	6498	64,940	**Cost saving**
Assuming the same benefit of chemotherapy for everyone = 30%	13.46	9264	13.45	6421	13.43	6535	91,274	**Cost saving**
NPI distribution from the Holt *et al.* study[77]	13.48	9323	13.42	6589	13.36	6808	22,281	**Cost saving**
NPI distribution from the TransATAC trial	13.52	9169	13.47	6421	13.42	6613	25,251	**Cost saving**
Cost for IHC4 = £100	13.54	9094	13.49	6290	13.44	6519	26,940	**Cost saving**
Cost for IHC4 = £200	13.54	9094	13.49	6390	13.44	6519	26,940	**Cost saving**
Cost for IHC4 = £300	13.54	9094	13.49	6490	13.44	6519	26,940	**Cost saving**
Cost for IHC4 = £400	13.54	9094	13.49	6590	13.44	6519	26,940	1557
G-CSF is given to all patients receiving chemotherapy	13.54	9373	13.49	6479	13.44	6728	27,655	**Cost saving**
Five cycles of G-CSF (instead of three)	13.54	9156	13.49	6371	13.44	6565	27,099	**Cost saving**
Five cycles of chemotherapy (instead of three)	13.54	8964	13.49	6275	13.44	6420	26,604	**Cost saving**
100% echocardiogram (instead of 25%)	13.54	9103	13.49	6344	13.44	6525	26,961	**Cost saving**

TABLE 70 Results of the univariate sensitivity analysis for the primary analysis assuming that the new tests are offered to all women with ER+. LN–, HER2– early breast cancer in England and Wales (*continued*)

	OncotypeDX		IHC4		Clinical practice		ICER (£)	
	QALYs	Cost (£)	QALYs	Cost (£)	QALYs	Cost (£)	OncotypeDX vs. clinical practice	IHC4 vs. clinical practice
Increase in the cost of chemotherapy by 25%	13.54	9290	13.49	6438	13.44	6666	27,443	**Cost saving**
Reduction in the cost of chemotherapy by 25%	13.54	8899	13.49	6242	13.44	6371	26,437	**Cost saving**
Increase in the cost of endocrine therapy by 25%	13.54	9921	13.49	7166	13.44	7343	26,964	**Cost saving**
Reduction in the cost of endocrine therapy by 25%	13.54	8268	13.49	5515	13.44	5694	26,916	**Cost saving**
Increase in the cost of distant metastases by 25%	13.54	9365	13.49	6640	13.44	6838	26,427	**Cost saving**
Reduction in the cost of distant metastases by 25%	13.54	8824	13.49	6040	13.44	6199	27,453	**Cost saving**
Cost of local recurrence – LCI	13.54	9070	13.49	6313	13.44	6490	26,986	**Cost saving**
Cost of local recurrence – UCI	13.54	9123	13.49	6372	13.44	6552	26,886	**Cost saving**
Terminal care cost – LCI	13.54	9053	13.49	6294	13.44	6470	27,018	**Cost saving**
Terminal care cost – UCI	13.54	9136	13.49	6386	13.44	6568	26,861	**Cost saving**
Utility values – LCI	12.89	9094	12.84	6340	12.80	6519	28,061	**Cost saving**
Utility values – UCI	14.08	9094	14.03	6340	13.98	6519	26,034	**Cost saving**
Increase of 25% in the decrement in utility for patients dying from breast cancer	13.53	9094	13.48	6340	13.44	6519	26,862	**Cost saving**
Decrease of 25% in the decrement in utility for patients dying from breast cancer	13.54	9094	13.49	6340	13.44	6519	27,018	**Cost saving**
Increase of 25% in the decrement in utility for patients receiving chemotherapy	13.53	9094	13.49	6340	13.44	6519	27,066	**Cost saving**
Decrease of 25% in the decrement in utility for patients receiving chemotherapy	13.54	9094	13.49	6340	13.44	6519	26,815	**Cost saving**
Utility for patients with AML = 0.5	13.54	9094	13.49	6340	13.44	6519	26,936	**Cost saving**
Utility for patients with AML = 0.6	13.54	9094	13.49	6340	13.44	6519	26,934	**Cost saving**
Risk of long-term adverse events multiplied by 2	13.52	9096	13.48	6340	13.43	6519	27,954	**Cost saving**
Risk of long-term adverse events multiplied by 3	13.51	9097	13.47	6340	13.42	6520	29,034	**Cost saving**
Proportion of patients classified as intermediate with the new test undergoing chemotherapy = low-risk value	13.50	8683	13.46	6215	13.44	6519	35,629	**Cost saving**
Proportion of patients classified as intermediate with the new test undergoing chemotherapy = high-risk value	13.56	9323	13.52	6434	13.44	6519	22,812	**Cost saving**

LCI, lower confidence interval; UCI, upper confidence interval.

TABLE 71 Results of the univariate sensitivity analysis for the primary analysis assuming that the new tests are offered only to women with a NPI score >3.4

	OncotypeDX		IHC4		Clinical practice		ICER (£)	
	QALYs	Cost (£)	QALYs	Cost (£)	QALYs	Cost (£)	OncotypeDX vs. clinical practice	IHC4 vs. clinical practice
Base case	13.06	10,911	12.97	8318	12.83	8816	9007	**Cost saving**
Time horizon = 5 years	4.07	9012	4.07	6245	4.06	6544	170,573	**Cost saving**
Time horizon = 10 years	6.94	10,170	6.92	7501	6.88	7918	39,573	**Cost saving**
Time horizon = 20 years	10.75	10,884	10.69	8289	10.59	8784	13,108	**Cost saving**
Starting age = 50 years	14.93	11,019	14.81	8434	14.64	8941	7203	**Cost saving**
Starting age = 60 years	12.53	10,868	12.44	8272	12.32	8765	9686	**Cost saving**
Starting age = 70 years	9.56	10,456	9.50	7828	9.42	8286	16,152	**Cost saving**
Reduction in the risk of recurrence by 10%	13.16	10,700	13.07	8085	12.94	8559	10,087	**Cost saving**
Reduction in the risk of recurrence by 20%	13.25	10,485	13.17	7846	13.06	8296	11,440	**Cost saving**
Increase in the risk of recurrence by 10%	12.98	11,117	12.87	8547	12.72	9066	8125	**Cost saving**
Increase in the risk of recurrence by 20%	12.89	11,319	12.77	8770	12.62	9311	7392	**Cost saving**
Recurrence up to 20 years	13.00	11,146	12.90	8577	12.76	9095	8389	**Cost saving**
No changed in the risk of distant recurrence between 10 and 15 years	12.96	11,230	12.85	8670	12.71	9195	8055	**Cost saving**
Reduction in the risk of distant recurrence by 75% between 10 and 15 years	13.12	10,747	13.03	8138	12.90	8619	9585	**Cost saving**
Proportion of local recurrence before a distant recurrence = 5%	13.07	10,836	12.97	8235	12.83	8723	9086	**Cost saving**
Proportion of local recurrence before a distant recurrence = 20%	13.06	11,039	12.97	8459	12.83	8973	8872	**Cost saving**
Proportion of local recurrence before a distant recurrence = 30%	13.06	11,174	12.97	8609	12.83	9139	8729	**Cost saving**
Time in distant recurrence health state different by prognosis group	13.08	11,118	12.99	8587	12.84	8932	9091	**Cost saving**
Time in distant recurrence – LCI	13.02	10,391	12.92	7743	12.78	8175	9116	**Cost saving**
Time in distant recurrence – UCI	13.11	11,383	13.01	8840	12.88	9396	8900	**Cost saving**
Proportion of patients receiving chemotherapy under current clinical practice = ECRIC only	13.06	10,911	12.97	8318	12.83	8775	9019	**Cost saving**
Proportion of patients receiving chemotherapy under current clinical practice = WMCIU only	13.06	10,911	12.97	8318	12.85	8959	8967	**Cost saving**
Proportion of patients receiving chemotherapy under current clinical practice = Holt *et al.* study[77]	13.06	10,911	12.97	8318	12.97	10,037	8818	**Cost saving**
Increase by 20% in the proportion of patients receiving chemotherapy under current clinical practice	13.06	10,911	12.97	8318	12.86	9066	8938	**Cost saving**
Increase by 30% in the proportion of patients receiving chemotherapy under current clinical practice	13.06	10,911	12.97	8318	12.87	9191	8907	**Cost saving**
Reduction by 20% in the proportion of patients receiving chemotherapy under current clinical practice	13.06	10,911	12.97	8318	12.81	8564	9082	**Cost saving**

TABLE 71 Results of the univariate sensitivity analysis for the primary analysis assuming that the new tests are offered only to women with a NPI score >3.4 (*continued*)

	OncotypeDX		IHC4		Clinical practice		ICER (£)	
	QALYs	Cost (£)	QALYs	Cost (£)	QALYs	Cost (£)	OncotypeDX vs. clinical practice	IHC4 vs. clinical practice
Reduction by 30% in the proportion of patients receiving chemotherapy under current clinical practice	13.06	10,911	12.97	8318	12.79	8438	9121	**Cost saving**
Assumption about who would receive chemotherapy with the new tests	13.09	10,799	12.98	8192	12.83	8816	7761	**Cost saving**
Reduction in the benefit of chemotherapy by 10%	13.05	10,950	12.96	8347	12.82	8834	9459	**Cost saving**
Reduction in the benefit of chemotherapy by 20%	13.03	10,987	12.94	8374	12.82	8851	9923	**Cost saving**
Reduction in the benefit of chemotherapy by 30%	13.02	11,022	12.93	8399	12.81	8867	10,400	**Cost saving**
Reduction in the benefit of chemotherapy by 40%	13.00	11,055	12.92	8423	12.80	8882	10,890	**Cost saving**
Assuming the same benefit of chemotherapy for everyone = 40%	(CiC information has been removed)	(CiC information has been removed)	12.93	8419	12.85	8766	28,833	**Cost saving**
Assuming the same benefit of chemotherapy for everyone = 30%	(CiC information has been removed)	(CiC information has been removed)	12.87	8557	12.81	8867	39,579	**Cost saving**
NPI distribution from the Holt *et al.* study[77]	13.06	10,911	12.97	8318	12.83	8816	9007	**Cost saving**
NPI distribution from the TransATAC trial	13.06	10,911	12.97	8318	12.83	8816	9007	**Cost saving**
Cost for IHC4 = £100	13.06	10,911	12.97	8268	12.83	8816	9007	**Cost saving**
Cost for IHC4 = £200	13.06	10,911	12.97	8368	12.83	8816	9007	**Cost saving**
Cost for IHC4 = £300	13.06	10,911	12.97	8468	12.83	8816	9007	**Cost saving**
Cost for IHC4 = £400	13.06	10,911	12.97	8568	12.83	8816	9007	**Cost saving**
G-CSF is given to all patients receiving chemotherapy	13.06	11,416	12.97	8701	12.83	9305	9077	**Cost saving**
Five cycles of G-CSF (instead of three)	13.06	11,023	12.97	8403	12.83	8924	9023	**Cost saving**
Five cycles of chemotherapy (instead of three)	13.06	10,674	12.97	8139	12.83	8586	8974	**Cost saving**
100% echocardiogram (instead of 25%)	13.06	10,926	12.97	8330	12.83	8830	9009	**Cost saving**
Increase in the cost of chemotherapy by 25%	13.06	11,267	12.97	8588	12.83	9160	9056	**Cost saving**
Reduction in the cost of chemotherapy by 25%	13.06	10,555	12.97	8049	12.83	8471	8958	**Cost saving**
Increase in the cost of endocrine therapy by 25%	13.06	11,727	12.97	9132	12.83	9627	9031	**Cost saving**
Reduction in the cost of endocrine therapy by 25%	13.06	10,094	12.97	7504	12.83	8005	8984	**Cost saving**
Increase in the cost of distant metastases by 25%	13.06	11,403	12.97	8862	12.83	9421	8519	**Cost saving**
Reduction in the cost of distant metastases by 25%	13.06	10,419	12.97	7774	12.83	8210	9495	**Cost saving**

continued

TABLE 71 Results of the univariate sensitivity analysis for the primary analysis assuming that the new tests are offered only to women with a NPI score >3.4 (*continued*)

	OncotypeDX		IHC4		Clinical practice		ICER (£)	
	QALYs	Cost (£)	QALYs	Cost (£)	QALYs	Cost (£)	OncotypeDX vs. clinical practice	IHC4 vs. clinical practice
Cost of local recurrence – LCI	13.06	10,867	12.97	8269	12.83	8761	9051	**Cost saving**
Cost of local recurrence – UCI	13.06	10,963	12.97	8375	12.83	8879	8956	**Cost saving**
Terminal care cost – LCI	13.06	10,835	12.97	8235	12.83	8723	9082	**Cost saving**
Terminal care cost – UCI	13.06	10,986	12.97	8402	12.83	8908	8932	**Cost saving**
Utility values – LCI	12.44	10,911	12.35	8318	12.21	8816	9378	**Cost saving**
Utility values – UCI	13.59	10,911	13.49	8318	13.35	8816	8707	**Cost saving**
Increase of 25% in the decrement in utility for patients dying from breast cancer	13.06	10,911	12.97	8318	12.83	8816	8982	**Cost saving**
Decrease of 25% in the decrement in utility for patients dying from breast cancer	13.07	10,911	12.97	8318	12.84	8816	9032	**Cost saving**
Increase of 25% in the decrement in utility for patients receiving chemotherapy	13.06	10,911	12.97	8318	12.83	8816	9011	**Cost saving**
Decrease of 25% in the decrement in utility for patients receiving chemotherapy	13.07	10,911	12.97	8318	12.84	8816	9003	**Cost saving**
Utility for patients with AML = 0.5	13.06	10,911	12.97	8318	12.83	8816	9007	**Cost saving**
Utility for patients with AML = 0.6	13.07	10,911	12.97	8318	12.83	8816	9007	**Cost saving**
Risk of long-term adverse events multiplied by 2	13.04	10,913	12.95	8318	12.81	8817	9061	**Cost saving**
Risk of long-term adverse events multiplied by 3	13.02	10,915	12.93	8318	12.79	8818	9112	**Cost saving**
Proportion of patients classified as intermediate with the new test undergoing chemotherapy = low risk value	13.00	10,450	12.88	8045	12.83	8816	10,022	**Cost saving**
Proportion of patients classified as intermediate with the new test undergoing chemotherapy = high risk value	13.14	11,351	13.06	8565	12.83	8816	8371	**Cost saving**

LCI, lower confidence interval; UCI, upper confidence interval.

IHC4

The ICER for IHC4 was sensitive to a greater number of assumptions than the ICER for OncotypeDX, such as the time spent in the distant recurrence health state, the proportion of patients receiving chemotherapy under clinical practice and the cost of chemotherapy, but remained dominant compared with current clinical practice (i.e. provided more QALYs at a lower cost) except when the cost of IHC4 was raised to £400 (ICER of £1557 per QALY gained compared with current clinical practice).

Univariate sensitivity analyses: structural assumptions

In addition to input parameter values, we also examined the impact of two structural assumptions; the exclusion of IHC4 from the model and the impact of modelling patients as a single group (instead of as two separate subgroups: NPI score ≤3.4 and NPI score >3.4)

Assuming no further reclassification using IHC4 (exclusion of IHC4)

In this scenario we used data for OncotypeDX only (*Table 49* presents data for the risk classification) from the TransATAC trial, assuming no further reclassification with IHC4. Therefore, we calculated the ICER only for OncotypeDX compared with current clinical practice. Patients were split into six possible risk categories (by NPI and OncotypeDX RS) compared with 18 risk categories in the base-case model (by NPI, RS and IHC4).

The impact on the ICER was minimal: a reduction in the ICER from £26,940 (base case) to £25,574 per QALY gained assuming that the test is given to all women with ER+, LN–, HER2– early breast cancer or an increase in the ICER from £9007 (base case) to £10,218 per QALY gained assuming that the test is given only to women with a NPI score > 3.4.

The results of this scenario analysis suggested that our base-case ICER for OncotypeDX was minimally affected by our choice of model structure to accommodate the evaluation of IHC4.

Assuming no further reclassification using IHC4 (exclusion of IHC4) and modelling the entire cohort as a single group (not split by Nottingham Prognostic Index score)

A second structural assumption was tested, to examine to what extent not separating patients into two subgroups (by NPI score) affected the ICER. This assumes that patients with a NPI score ≤ 3.4 and patients with a NPI score > 3.4 within the same risk group (defined by the new test) have the same prognosis.

Again, in this scenario analysis we used data for OncotypeDX only (*Table 49* present data for the risk classification) from the TransATAC trial, assuming no further reclassification with IHC4. Therefore, we calculated the ICER only for OncotypeDX compared with current clinical practice. In this scenario analysis, the model separated patients into three possible risk categories (by RS only) compared with 18 risk categories in the base-case model (by NPI, RS and IHC4).

As expected, this assumption had a positive impact on the ICER (more favourable to OncotypeDX), with the ICER decreasing from £26,940 (base case) to £18,859 per QALY gained assuming that the test is given to all woman with ER+, LN–, HER2– early breast cancer.

By modelling the entire cohort as a single group, the prognostic value of current decision-making using clinicopathological parameters is ignored (i.e. that patients with a low NPI score have a lower risk of recurrence but are also less likely to receive chemotherapy compared with patients with a NPI score > 3.4 under current clinical practice). This is more favourable to OncotypeDX. This scenario assumes that patients within the defined RS risk group are homogeneous; however, it seems more likely that patients with a low RS and low NPI score would have a better prognosis than patients with a low RS and high NPI score.

Exploratory analysis: cost-effectiveness of treatment guided using Mammostrat

An exploratory analysis was carried out to assess the cost-effectiveness of Mammostrat compared with current clinical practice in England and Wales. The evidence base for Mammostrat is less well developed and a number of gaps were identified by the systematic review of the literature.

The EAG economic model was repopulated using the evidence for Mammostrat on reclassification (unpublished) and on the benefit of chemotherapy by risk group, but many assumptions were necessary because of limitations in the evidence available, especially on the impact of the test on decision-making and the extent to which the reclassification data used in the model were generalisable to the UK population. Further uncertainty is introduced given that

the EAG economic model uses distant recurrence as an outcome whereas most of the evidence for Mammostrat was drawn from analyses of DFS and therefore included all recurrences.

Data from a subset of the Ring *et al.* study[125] were used in the economic model; however, (CIC information has been removed). There is major uncertainty regarding the robustness of the reclassification data from the subset of the Ring *et al.* study.[125]

Because of the limitations of the evidence base, any conclusions drawn from this analysis are subject to significant uncertainty.

Deterministic results

The proportion of women receiving chemotherapy was estimated to increase slightly with the use of Mammostrat compared with current clinical practice under our base-case assumptions (21.16% vs. 14.42% in all women; 34.27% vs. 33.60% in women with a NPI score > 3.4 with ER+, LN–, HER– early breast cancer).

Compared with current clinical practice, the incremental cost for treatment guided using Mammostrat was £26,598 per QALY gained under our base-case assumptions, assuming that the test is given to all women with ER+, LN–, HER– early breast cancer (*Table 72*).

If Mammostrat was given only to women with a NPI score > 3.4, the Mammostrat test was dominated (i.e. provided less benefits for a higher cost). (CIC information has been removed.)

Probabilistic sensitivity analysis

The results of the PSA using 2500 iterations are shown in *Table 73*. Treatment guided using Mammostrat had a cost per QALY gained of £27,731 compared with current clinical practice if the test was offered to all women with ER+, LN–, HER2– early breast cancer. If the test was

TABLE 72 Deterministic ICER for the exploratory analysis comparing Mammostrat with current clinical practice in women with ER+, LN–, HER– early breast cancer in England and Wales

	Mean cost (£)	Mean QALYs	ICER (£)
All patients			
Mammostrat	9040	12.91	26,598
Current clinical practice	7699	12.86	
Patients with a NPI score > 3.4			
Mammostrat	10,985	12.29	**Dominated**
Current clinical practice	9717	12.34	

TABLE 73 Probabilistic ICER (2500 iterations) for the exploratory analysis comparing Mammostrat with current clinical practice in women with ER+, LN–, HER2– early breast cancer in England and Wales

	Mean cost (£)	Mean QALYs	ICER (£)
All patients			
Mammostrat	9028	12.90	27,731
Current clinical practice	7683	12.85	
Patients with a NPI score > 3.4			
Mammostrat	10,958	12.29	**Dominated**
Current clinical practice	9685	12.34	

offered only to women with a NPI score > 3.4, Mammostrat was dominated (i.e. provided less QALYs at a higher cost).

The CEAC shows that treatment guided by Mammostrat score is a cost-effective strategy in 36.0% of cases when using a willingness-to-pay threshold of £20,000 per QALY gained (*Figure 15*) if the test were to be given to all women with ER+, LN–, HER2- early breast caner. The probability of treatment guided using Mammostrat being cost-effective at a £20,000 threshold was 18.0% if the test were to be offered only to women with a NPI score > 3.4 (*Figure 16*).

Univariate sensitivity analysis

The impact of key parameters was tested in univariate sensitivity analysis (*Tables 74* and *75*).

The ICER was very sensitive to the assumption about the proportion of patients who would receive chemotherapy based on the result of the new test if the test was offered to all women with ER+, LN–, HER2– early breast cancer (see *Table 74*). Assuming that no patients classified as low risk, 50% of patients classified as intermediate risk and 100% of patients classified as high risk would receive chemotherapy improved the ICER (more favourable to Mammostrat). Furthermore, the ICER was very sensitive to the assumption about the probability of chemotherapy in patients classified as intermediate risk with Mammostrat. The ICER ranged

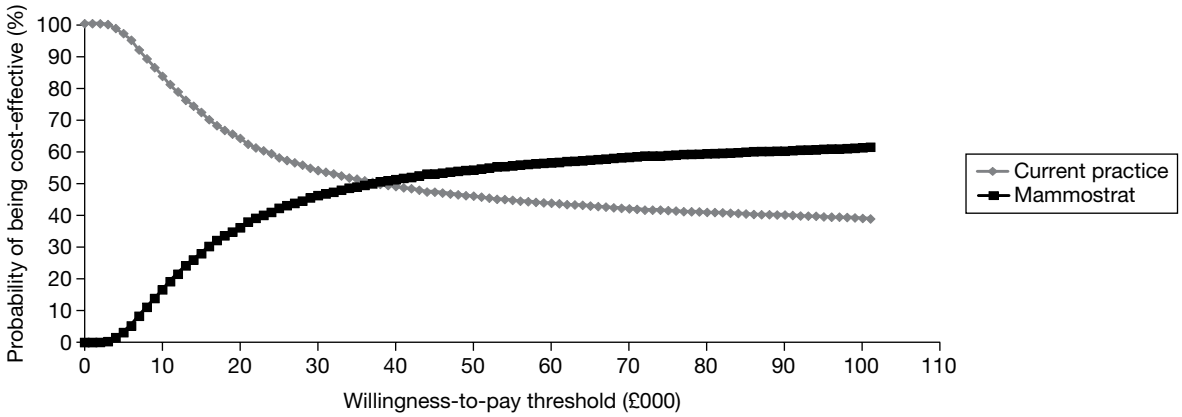


FIGURE 15 Cost-effectiveness acceptability curve for the exploratory analysis comparing Mammostrat with current clinical practice assuming that the test is given to all women with ER+, LN–, HER2– early breast cancer in England and Wales.

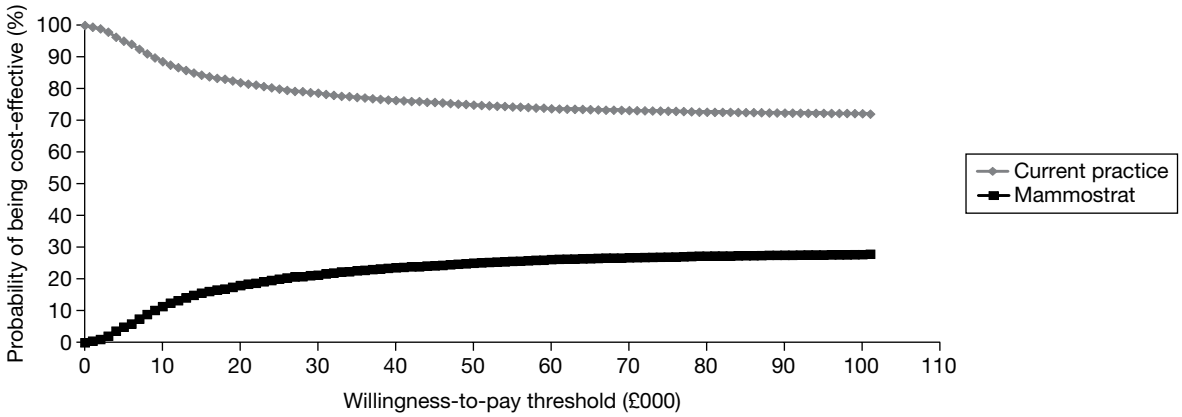


FIGURE 16 Cost-effectiveness acceptability curve for the exploratory analysis comparing Mammostrat with current clinical practice assuming that the test is given only to women with a NPI score > 3.4.

TABLE 74 Results of the univariate sensitivity analysis for the exploratory analysis assuming that the Mammostrat test is offered to all women with ER+, LN–, HER2– early breast cancer in England and Wales

	Current clinical practice		Mammostrat		
	QALYs	Cost (£)	QALYs	Cost (£)	ICER (£)
Base case	12.86	7699	12.91	9040	26,598
No adjustment for risk of recurrence	13.18	6995	13.23	8333	25,729
Interpretation based on assumption	12.86	7699	12.99	9115	10,407
Chemotherapy benefit = LCI	12.89	7626	12.96	8927	18,879
Chemotherapy benefit = UCI	12.78	7863	12.77	9329	**Dominated**
Time in DM = 50 months	12.91	8279	12.96	9593	27,324
Time in DM = 60 months	12.95	8820	13.00	10,110	28,071
Time in DM = 70 months	13.00	9319	13.04	10,586	28,830
Utility value in recurrence = 0.7	12.86	7699	12.91	9040	26,696
Utility value in recurrence = 0.75	12.87	7699	12.92	9040	27,028
Reduction in cost of DM of 20%	12.86	7216	12.91	8579	27,037
Reduction in cost of DM of 40%	12.86	7457	12.91	8810	26,817
Proportion of patients classified as intermediate with the new test undergoing chemotherapy = low-risk value	12.86	7699	12.92	8616	15,500
Proportion of patients classified as intermediate with the new test undergoing chemotherapy = high-risk value	12.86	7699	12.90	9287	34,959

DM, distant metastasis; LCI, lower confidence interval; UCI, upper confidence interval.

TABLE 75 Results of the univariate sensitivity analysis for the exploratory analysis assuming that the Mammostrat test is offered only to women with a NPI score >3.4

	Current clinical practice		Mammostrat		
	QALYs	Cost (£)	QALYs	Cost (£)	ICER (£)
Base case	12.34	9717	12.29	10,985	**Dominated**
No adjustment for risk of recurrence	13.15	7962	13.12	9193	**Dominated**
Interpretation based on assumption	12.34	9717	12.28	10,955	**Dominated**
Chemotherapy benefit = LCI	12.43	9523	12.37	10,815	**Dominated**
Chemotherapy benefit = UCI	12.15	10,146	12.10	11,401	**Dominated**
Time in DM = 50 months	12.41	10,528	12.37	11,819	**Dominated**
Time in DM = 60 months	12.48	11,286	12.43	12,597	**Dominated**
Time in DM = 70 months	12.54	11,984	12.50	13,314	**Dominated**
Utility value in recurrence = 0.7	12.35	9717	12.30	10,985	**Dominated**
Utility value in recurrence = 0.75	12.37	9717	12.32	10,985	**Dominated**
Reduction in cost of DM of 20%	12.34	9040	12.29	10,291	**Dominated**
Reduction in cost of DM of 40%	12.34	9379	12.29	10,638	**Dominated**
Proportion of patients classified as intermediate with the new test undergoing chemotherapy = low-risk value	12.34	9717	12.30	10,471	**Dominated**
Proportion of patients classified as intermediate with the new test undergoing chemotherapy = high-risk value	12.34	9717	12.28	11,484	**Dominated**

DM, distant metastasis; LCI, lower confidence interval; UCI, upper confidence interval.

from £15,500 to £34,959 if the test was given to all women assuming that the probability of receiving chemotherapy was the same as for patients classified as low and high risk respectively.

The ICER ranged from £18,879 to being dominated, using the lower and upper CIs from the Ross *et al.* study,[126] for the impact of chemotherapy in terms of reduction in risk of recurrences if the test was offered to all women with ER+, LN–, HER2– early breast cancer. The ICER was not sensitive to the assumptions about utility values, management costs and the time spent in the recurrence health state (see *Table 74*).

The ICERs for the use of Mammostrat remained dominated under the assumptions examined in sensitivity analysis if the test was offered only to women with a NPI score >3.4 (see *Table 75*).

Exploratory analysis: cost-effectiveness of treatment guided using MammaPrint

Finally, a second exploratory analysis was carried out assessing the cost-effectiveness of MammaPrint. Although there was a greater volume of evidence for MammaPrint than for Mammostrat, there were significant gaps in the evidence available and data that were used to populate the economic model were not considered to be robust. Therefore, any conclusions that can be drawn from this analysis are subject to considerable uncertainty.

Of note, particular concerns exist about the existing evidence on the benefit of chemotherapy as this is likely to have a significant influence on the ICER. Other issues include the lack of UK data and the fact that the data available were derived mainly from premenopausal women, limiting generalisability to the UK population. Because of these issues, the EAG was not sufficiently confident to provide a single ICER but presented a range of ICERs within the CIs for the benefit of chemotherapy, as this was considered to be the main uncertainty in the model. Note that, although a range of ICERs is presented, there were also significant concerns relating to the design of studies in the evidence base, and this range does not capture this uncertainty in study design.

More assumptions have been made within this analysis than within the other analyses and the results are highly uncertain.

Only a limited number of univariate sensitivity analyses were carried out because of the nature of the analysis. A sensitivity analysis was conducted assuming no additional costs for the NHS for the use of fresh tissue. A second sensitivity analysis was conducted assuming that 5% of patients classified as having good prognosis and 95% of patients classified as having poor prognosis received chemotherapy.

In addition to the univariate sensitivity analyses, we performed a multivariate sensitivity analysis examining different values for the benefit of chemotherapy.

No PSA was conducted as there was considered to be significant uncertainties in the studies used that could not be adequately captured in the economic model (for instance limitations in study design, differences in population included in the studies being younger and at higher risk than the population in the economic model, uncertainties that could not be adequately captured by the parameter uncertainty within the PSA).

MammaPrint offers the option of three complementary tests at no additional cost. ER, PR and HER2 status can be provided in the TargetPrint report. The impact of this has not been captured in the economic model.

Deterministic results

Compared with current clinical practice, the incremental cost for treatment guided using MammaPrint was estimated to range between £12,240 and £53,058 per QALY gained, when the benefit of chemotherapy was varied by the upper and lower CI limits from the Knauer *et al.* study,[110] assuming that the test was given to all women with ER+, LN–, HER2– early breast cancer. If MammaPrint was given only to women with a NPI score > 3.4, the ICER ranged between £6053 and £29,569 per QALY gained (*Table 76*).

Of note, the proportion of patients receiving chemotherapy increased significantly with the use of MammaPrint compared with current practice under our base-case assumptions: 44.18% vs. 14.42% in all women and 90.31% vs. 33.60% in women with a NPI score > 3.4.

Univariate sensitivity analysis

As expected, assuming no additional cost for the NHS for the use of fresh tissue samples improved the ICER (more favourable to MammaPrint) (*Table 77*).

Assuming that 5% and 95% of patients classified as having good and poor prognosis with MammaPrint, respectively, received chemotherapy improved the ICER (more favourable to MammaPrint) (*Table 78*).

Multivariate sensitivity analysis examining the benefit of chemotherapy for risk of distant recurrence

An exploratory multivariate sensitivity analysis was conducted examining the effect of different values for the benefit of chemotherapy by risk group in terms of reduction in the risk of distant recurrence, assuming that MammaPrint is given to all women with ER+, LN–, HER2– early breast cancer (*Table 79*).

TABLE 76 Deterministic ICER for the exploratory analysis comparing MammaPrint with current clinical practice in women with ER+, LN–, HER2– early breast cancer in England and Wales

	Mean QALYs	Mean cost (£)	ICER (£)
All patients			
Current practice	13.49–13.39	6408–6629	12,240–53,058
MammaPrint	13.78–13.47	10,017–10,748	
Patients with a NPI score > 3.4			
Current practice	13.07–12.81	8281–8872	6053–29,569
MammaPrint	13.73–12.99	12,278–14,014	

TABLE 77 Sensitivity analysis assuming no additional cost for the NHS for the use of fresh tissue samples

	Mean QALYs	Mean cost (£)	ICER (£)
All patients			
Current practice	13.49–13.39	6408–6629	11,392–49,838
MammaPrint	13.78–13.47	9767–10,498	
Patients with a NPI score > 3.4			
Current practice	13.07–12.81	8281–8,872	5675–28,131
MammaPrint	13.73–12.99	12,028–13,764	

TABLE 78 Sensitivity analysis assuming that chemotherapy is given to 5% and 95% of patients classified as having good and poor prognosis with MammaPrint respectively

	Mean QALYs	Mean cost (£)	ICER (£)
All patients			
Current practice	13.49–13.39	6408 –6629	12,369–48,322
MammaPrint	13.78–13.48	10,045 –10,756	
Patients with a NPI score > 3.4			
Current practice	13.07–12.81	8281 –8872	6115–23,939
MammaPrint	13.63–12.99	11,705 –13,189	

This analysis suggested that, under base-case assumptions about the risk classification, risk of recurrence and interpretation of the test (i.e. which patients would receive chemotherapy), the ICER was < £20,000 per QALY gained if the relative risk reduction in the risk of recurrence was at least 60% for patients with a poor prognosis. However, the conclusions are likely to change if different assumptions are used for the risk classification or the proportion of patients who would receive chemotherapy according to MammaPrint.

Comparison of assumptions and results with the economic models submitted by Genomic Health and Clarient

Comparison with the economic model submitted by Genomic Health

The base-case ICER estimated by Genomic Health for treatment guided using OncotypeDX compared with current clinical practice was £6232 per QALY gained assuming that the test was given to all women with ER+, LN– or single node-positive and HER2+/– early breast cancer (*Table 80*).

The ICER estimated by the EAG was £26,940 (deterministic) assuming that the test was offered to all women with ER+, LN–, HER2– early breast cancer (see *Table 80*).

The main differences between the EAG's economic assessment and the economic model submitted by Genomic Health for OncotypeDX are:

- A shorter time horizon was used in the economic evaluation submitted by Genomic Health (30 years compared with lifetime in the EAG economic assessment). The starting age of the cohort was also different (58.3 years in the EAG economic assessment compared with 60.6 years in the Genomic Health economic model). This partly explains the differences in the mean life-years and QALYs.
- There were differences in the populations under assessment (LN–, HER– only in the EAG economic assessment compared with LN– or single positive node and HER2+/– in the economic model submitted by Genomic Health).
- The risk of distant recurrence was taken from the TransATAC trial in the EAG economic model of a UK population treated with tamoxifen and anastrozole. Data from Paik *et al.*[48] from a US cohort treated with tamoxifen only was used in the Genomic Health economic model. Of note, the manufacturer examined a scenario using the risk of distant recurrence from the TransATAC trial and showed that the ICER increased from about £6232 to about £9160 per QALY gained.
- The EAG assumed that the risk of distant recurrence was halved after 10 years and that no distant recurrences occurred after 15 years. The Genomic Health economic model assumed

TABLE 79 Multivariate sensitivity analysis varying the benefit of chemotherapy for patients classified as having a good or a poor prognosis with MammaPrint

		ICER (£)								
		Reduction in the risk of distant recurrence in patients classified with a poor prognosis using MammaPrint								
		0%	5%	10%	15%	20%	25%	30%	35%	40%
Reduction in the risk of distant recurrence in patients classified with a good prognosis using MammaPrint	0%	**Dominated**	**Dominated**	679,644	169,516	95,893	66,391	50,493	40,552	33,748
	5%	**Dominated**	**Dominated**	620,556	165,504	94,569	65,740	50,107	40,298	33,568
	10%	**Dominated**	**Dominated**	570,873	161,673	93,280	65,102	49,728	40,046	33,389
	15%	**Dominated**	**Dominated**	528,515	158,013	92,024	64,474	49,353	39,797	33,213
	20%	**Dominated**	**Dominated**	491,972	154,510	90,800	63,858	48,983	39,552	33,037
	25%	**Dominated**	**Dominated**	460,125	151,156	89,607	63,254	48,619	39,309	32,864
	30%	**Dominated**	**Dominated**	432,123	147,941	88,444	62,659	48,260	39,068	32,692
	35%	**Dominated**	**Dominated**	407,310	144,857	87,309	62,076	47,905	38,831	32,522
	40%	**Dominated**	**Dominated**	385,170	141,896	86,202	61,502	47,556	38,595	32,353
	45%	**Dominated**	**Dominated**	365,293	139,051	85,121	60,938	47,211	38,363	32,186
	50%	**Dominated**	**Dominated**	347,349	136,314	84,067	60,384	46,870	38,133	32,020
	55%	**Dominated**	**Dominated**	331,070	133,681	83,036	59,840	46,534	37,906	31,856
	60%	**Dominated**	**Dominated**	316,233	131,145	82,030	59,304	46,203	37,680	31,694
	65%	**Dominated**	**Dominated**	302,656	128,701	81,047	58,778	45,876	37,458	31,532
	70	**Dominated**	**Dominated**	290,185	126,343	80,086	58,260	45,553	37,238	31,373
	75%	**Dominated**	**Dominated**	278,689	124,069	79,147	57,751	45,234	37,020	31,215
	80%	**Dominated**	**Dominated**	268,059	121,872	78,229	57,250	44,920	36,804	31,058
	85%	**Dominated**	**Dominated**	258,200	119,750	77,330	56,757	44,609	36,591	30,902
	90%	**Dominated**	**Dominated**	249,032	117,698	76,451	56,272	44,303	36,380	30,748
	95%	**Dominated**	**Dominated**	240,484	115,714	75,591	55,795	44,000	36,171	30,596
	100%	**Dominated**	**Dominated**	232,496	113,793	74,750	55,325	43,701	35,964	30,444

45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%
28,799	25,037	22,081	19,696	17,732	16,087	14,688	13,485	12,438	11,519	10,707	9983
28,665	24,933	21,998	19,629	17,677	16,040	14,648	13,450	12,408	11,493	10,683	9962
28,531	24,830	21,916	19,562	17,621	15,993	14,608	13,415	12,377	11,466	10,659	9940
28,399	24,728	21,834	19,495	17,566	15,946	14,568	13,381	12,347	11,440	10,636	9920
28,268	24,626	21,753	19,429	17,511	15,900	14,528	13,346	12,317	11,413	10,613	9899
28,138	24,525	21,672	19,363	17,456	15,854	14,489	13,312	12,287	11,387	10,589	9878
28,009	24,425	21,592	19,298	17,401	15,808	14,449	13,278	12,258	11,361	10,566	9857
27,881	24,325	21,513	19,233	17,347	15,762	14,410	13,244	12,228	11,334	10,543	9836
27,755	24,226	21,434	19,168	17,293	15,716	14,371	13,210	12,198	11,308	10,520	9816
27,629	24,128	21,355	19,104	17,240	15,671	14,332	13,177	12,169	11,282	10,496	9795
27,504	24,031	21,277	19,040	17,186	15,626	14,293	13,143	12,140	11,256	10,473	9774
27,380	23,934	21,199	18,976	17,133	15,581	14,255	13,110	12,110	11,231	10,451	9754
27,257	23,838	21,122	18,913	17,080	15,536	14,216	13,076	12,081	11,205	10,428	9733
27,135	23,742	21,045	18,850	17,028	15,491	14,178	13,043	12,052	11,179	10,405	9713
27,014	23,648	20,969	18,787	16,975	15,447	14,140	13,010	12,023	11,154	10,382	9693
26,894	23,553	20,893	18,725	16,923	15,403	14,102	12,977	11,994	11,128	10,360	9672
26,775	23,460	20,818	18,663	16,871	15,359	14,064	12,944	11,966	11,103	10,337	9652
26,657	23,367	20,743	18,601	16,820	15,315	14,027	12,912	11,937	11,078	10,315	9632
26,540	23,275	20,669	18,540	16,768	15,271	13,989	12,879	11,908	11,052	10,292	9612
26,423	23,183	20,595	18,479	16,717	15,228	13,952	12,847	11,880	11,027	10,270	9592
26,308	23,092	20,521	18,418	16,667	15,185	13,915	12,814	11,852	11,002	10,247	9572

TABLE 80 Comparison of the EAG and Genomic Health estimates of the ICER for OncotypeDX compared with current clinical practice

	Genomic Health economic model	EAG economic model
Cost (£)	12,735	9094
QALYs	11.54	13.54
Life expectancy (years)	14.89	16.47
Current clinical practice		
Cost (£)	11,847	6519
QALYs	11.39	13.44
Life expectancy (years)	14.73	16.35
ICER (£)	6232	26,940

that the risk of distant recurrence was constant and ongoing over time. Therefore, there is the potential to avoid more recurrences in the Genomic Health economic model, resulting in a more favourable ICER.

- The distribution of patients reclassified using OncotypeDX was derived from the TransATAC trial and cancer registry data in the UK in the EAG economic model, compared with the reclassification from the Holt *et al.* study[78] in the Genomic Health economic model. However, the EAG had concerns regarding the representativeness of patients included in the Holt *et al.* study (discussed in *Chapter 2*, *Results: OncotypeDX test*).
- The proportion of patients receiving chemotherapy under clinical practice was extracted from registry data in the EAG economic model, whereas in the Genomic Health economic model the proportion observed in the Holt *et al.* study[78] was used. About 44% of women received chemotherapy in the manufacturer's model under current clinical practice. Registry data (used in the EAG economic model) suggested that about 14.4% of women with ER+, LN–, HER2– breast cancer received chemotherapy (4.6% among women with a NPI score ≤ 3.4 and 33.6% among women with a NPI score > 3.4). In the Genomic Health model, 43.86% of patients subsequently classified as low risk by OncotypeDX received chemotherapy under current clinical practice. Because those patients have a low risk of distant recurrence and derive limited benefit from chemotherapy, this high estimate of the proportion of patients receiving chemotherapy in the comparator arm in this subgroup is favourable to OncotypeDX.
- There were also structural differences between the models. The EAG modelled patients with a NPI score ≤ 3.4 and patients with a NPI score > 3.4 separately in order to conduct a subgroup analysis but also to account for the prognostic value of current decision-making based on clinicopathological parameters. Indeed, as shown in cancer registry data, patients with a low NPI score are less likely to receive chemotherapy than patients with a NPI score > 3.4. But at the same time, patients with a low NPI score have a lower risk of recurrence than patients with a NPI score > 3.4. The economic model submitted by Genomic Health assumed that the risk of recurrence was constant within each OncotypeDX RS group and used the Holt *et al.* study[78] to estimate the proportion of patients who would receive chemotherapy. The Genomic Health approach ignores the prognostic value of current treatment decision-making using clinicopathological parameters and is therefore more favourable to OncotypeDX.
- The EAG economic model further reclassified patients according to the IHC4 test results to also evaluate the cost-effectiveness of IHC4. A scenario analysis was conducted and showed that the impact of this structural assumption was minimal.

- The costs of chemotherapy and associated short-term adverse events were lower in the EAG economic assessment (£4866) than in the Genomic Health economic assessment (£7728).
- There were also differences between the EAG economic assessment and the Genomic Health economic assessment in the utility estimates for the recurrence-free (0.824 vs. 0.78) and distant recurrence (0.685 vs. 0.60) health states. There were therefore more gains associated with the prevention of a distant recurrence in the Genomic Health economic model than in the EAG model (0.18 vs. 0.14).
- Finally, the EAG economic model also included local recurrences and long-term adverse events due to chemotherapy.

To understand the differences in the results produced by the two models, the EAG economic model was repopulated using the same data inputs and assumptions as in the Genomic Health economic model. The results of this analysis are presented in *Table 81*. Each row shows the impact of introducing a new assumption in addition to the assumptions considered in the rows above.

Using a similar model structure as the Genomic Health economic model reduced the ICER in the EAG model from £26,960 to £18,859 per QALY gained (i.e. modelling three groups of patients according to the OncotypeDX RS classification, with no split by NPI score). When we further assumed a constant risk of recurrence over time using data from the Paik *et al.* study,[49] the ICER decreased to £8311 per QALY gained. Finally, using similar assumptions/data inputs for the proportion of patients receiving chemotherapy, starting age, time horizon, utility values and costs as in the Genomic Health economic model reduced the ICER further to £6276 (compared with £6232 in Genomic Health economic model).

This analysis suggests that the differences in the results are mainly explained by the choice of model structure, assumptions about the risk of recurrence over time and data on risk reclassification and the proportion of patients receiving chemotherapy in clinical practice and after using the new tests.

TABLE 81 Changes in the ICER using Genomic Health assumptions in the EAG model

Assumption	ICER (£)
Base case	26,940
Using data for OncotypeDX only (excluding IHC4)	25,574
Using data for OncotypeDX only (excluding IHC4) and modelling the entire cohort as a single group (no split by NPI score)	18,859
Assuming the risk of distant recurrence to be constant and ongoing	13,874
10-year risk of recurrence extracted from the Paik *et al.* study[49]	8311
Using the classification of patients from the Holt *et al.* study[78]	3953
Using the proportion of patients receiving chemotherapy from the Holt *et al.* study[78]	7032
Starting age = 60.55 years as per the Genomic Health model	7887
Time horizon = 30 years as per the Genomic Health model	8431
No long-term adverse events	8883
No local recurrences	9067
Cost of chemotherapy as per the Genomic Health model	6534
Utility values as per the Genomic Health model	6607
No terminal care cost or decrement in utility	7091
Cost of distant recurrence as per the Genomic Health model	6276

Note: Each row shows the impact of introducing a new assumption in addition to the assumptions considered in the rows above.

Comparison with the economic model submitted by Clarient

(CIC information has been removed.)

Both analyses had to use a large number of assumptions given the gap in the evidence available and therefore need to be interpreted with caution. The EAG economic assessment also showed that the use of Mammostrat in women with a NPI score >3.4 is dominated (i.e. provided less benefit at a higher cost). This may reflect limitations in the reclassification data used.

(CIC information has been removed.) Because of time and resource constraints, the late submission and the nature of this analysis (exploratory), only a brief comparison of the differences is presented for completeness:

- Time horizon: 10 years in the Clarient economic assessment compared with lifetime in the EAG economic assessment.
- Model structure: the EAG economic assessment is simple and assumes that patients enter the recurrence state and remain in that health state until death (using data from Thomas *et al.*[155]). The Clarient economic model is more complex and models recurrence-free survival and OS separately; however, a large number of assumptions have been made and inconsistencies were reported by the manufacturer.
- (CIC information has been removed.)
- (CIC information has been removed.)

Discussion of the independent economic model results

The four tests with the most well-developed clinical evidence base were considered within the economic evaluation. The EAG presented a primary analysis that compared OncotypeDX and IHC4 with current clinical practice in England and Wales. Based on the EAG model the incremental cost for adjuvant chemotherapy guided using OncotypeDX was estimated to be £29,503 per QALY gained compared with current clinical practice, assuming that the test was offered to all woman with ER+, LN–, HER2– early breast cancer under our base-case assumptions. This assumes that the test has predictive ability, that is, patients in the high-risk group benefit relatively more proportionally from chemotherapy than patients in the lower-risk groups. The IHC4 test was dominant compared with current clinical practice, providing more QALYs at a lower cost. The ICER for OncotypeDX increased substantially if the test was assumed to be prognostic only, that is, assuming the same relative reduction in the risk of recurrence from chemotherapy for all patients irrespective of the OncotypeDX RS classification. IHC4 remained dominant under this assumption. In incremental analysis, when the treatment decision using OncotypeDX was compared with the treatment decision using IHC4, the ICER increased to £64,111 per QALY gained. In a second scenario, assuming that the test was offered only to women with a NPI score >3.4, treatment guided using IHC4 remained dominant (i.e. provided more QALYs at a lower cost) compared with current clinical practice. The incremental cost for treatment guided using OncotypeDX was £9774 per QALY gained compared with current clinical practice and £31,125 per QALY gained compared with IHC4 (assuming that the test has predictive ability). However, it should be noted that the evidence base for IHC4 is less well developed and therefore the results should be interpreted with consideration of the additional assumptions used in the evaluation. One-way sensitivity analyses indicated that the ICER was most sensitive to the assumptions about the benefit reduction associated with chemotherapy, the time horizon of the model, the starting age of the cohort, the risk of recurrence and who would receive chemotherapy depending on the result of the test. A key area of uncertainty is whether the new tests are prognostic only or offer predictive ability.

The economic analyses suggested that treatment guided using IHC4 has the greatest potential to be cost-effective at a willingness-to-pay threshold of £20,000. However, the evidence base for IHC4 is less well developed than the evidence base for OncotypeDX and a number of additional assumptions were needed to model the IHC4 test. The IHC4 test provides only a continuous risk score and so it was necessary to derive risk categories solely for the purposes of the analysis. No evidence exists on the predictive ability of the IHC4 test. The benefits of chemotherapy by IHC4 risk group were based on indirect evidence, using the OncotypeDX classification; no additional benefit was assumed for IHC4. In the absence of evidence it was assumed that the likelihood of receiving chemotherapy based on the IHC4 risk classification would be the same as for OncotypeDX RS group, that is, that physicians would interpret the results from OncotypeDX and IHC4 in the same way.

For MammaPrint and Mammostrat there were significant gaps/limitations in the evidence available and data that have been used were not considered to be robust by the EAG. For this reason the analyses that were carried out evaluating the cost-effectiveness of MammaPrint and Mammostrat compared with current clinical practice in England and Wales were considered to be exploratory. Any conclusions that are drawn from these analyses are limited and further clinical evidence will be needed to make the findings more robust. The exploratory analyses suggested that the ICER for Mammostrat was around £28,000 per QALY gained compared with current clinical practice under our base-case assumptions, assuming that the test was offered to all women with ER+, LN–, HER2– early breast cancer, but Mammostrat was dominated if the test was given only to women with a NPI score > 3.4 (i.e. provided less QALYs at a higher cost). The second exploratory analysis indicated that MammaPrint has the potential to be cost-effective, but there were too many uncertainties in the data used and the design of the clinical studies to draw any definitive conclusions. Additionally, MammaPrint offers the option of three complementary tests at no additional cost. Notably, ER, PR and HER2 status can be provided in the TargetPrint report. This has not been captured in the economic model.

We did not perform an incremental analysis including the four tests evaluated because of the heterogeneity in the data used to populate the models and the differences in the quality of evidence between the tests. These differences are not adequately reflected in the PSA. Although this may be considered a limitation, we considered that including MammaPrint and Mammastrat within an incremental analysis could potentially be misleading given the gaps in the evidence base and significant issues relating to the quality of the data used to populate the economic models for these two tests.

No prospective studies that follow patients from initial diagnosis through to final health outcomes have been identified for any of the tests. Two prospective studies, MINDACT[186] (MammaPrint) and TAILORx[187] (OncotypeDX), are ongoing but not due to report for several years. The economic model therefore needed to combine data from different sources to model how the results from the new tests translated into final outcomes in the form of QALYs. This resulted in significant limitations – data used in the model were not always based on UK populations and were not always specifically based on the ER+, LN–, HER2– population of interest. Differences in the ages of study populations and the endocrine and chemotherapy regimens used in the studies compared with those in the model introduced further uncertainty. In addition to the uncertainty in the data derived from each study, there are uncertainties introduced by using separate studies to represent different elements in the model.

These tests will have an impact on the health of patients only if the management of the patients changes. Evidence on how the results of tests change treatment decisions in practice in the UK is limited. We conducted two analyses, one assuming that the test was given to all women and one assuming that the test was given only to women with a NPI score > 3.4, as a proxy for those

at intermediate risk. This group reflects patients for whom the decision whether or not to given chemotherapy is most uncertain. The definition of the subgroup is relatively simplistic, because of data limitations, and may include women at the top end of the NPI distribution who are likely to receive chemotherapy despite the result of the test. It does, however, suggest that generally the cost-effectiveness may be improved by focusing the test in these women (although this was not the case in the exploratory analysis for Mammostrat).

Our analysis focused on women with ER+, LN–, HER2– early breast cancer as this population is supported by the most robust clinical evidence. Other populations, such as women with a small number of positive nodes, might also benefit from the tests, and results are likely to change if the population appraised is extended to women with ER– cancer or with positive nodes.

Evidence used in the model was generally identified from the systematic review of the literature on the clinical validity and utility of each of the tests. However, for the purpose of the economic analysis of patients with ER+, LN–, HER2– early breast cancer in a UK population the published data were sometimes not available in the right format for use within the economic model or the necessary data were not presented. Therefore, on occasion, once we had identified the most relevant data source from the review, we sought additional data to populate the economic model. For instance, data on the risk reclassification and risk of recurrence for patients treated with endocrine therapy in the UK for the main analysis were taken from a reanalysis of a study (TransATAC trial[79]) identified through the systematic review of the literature, as the published data were not specific to the population of interest (ER+, LN–, HER2– early breast cancer). Similarly, data on the impact of OncotypeDX on decision-making were taken from a reanalysis of the Holt *et al.* study[78] (identified through the systematic review), to provide data specific to the population of interest, that is, ER+, LN–, HER2– early breast cancer. On occasion, data outside the systematic review were used, such as UK registry data to inform the current level of chemotherapy in the UK.

Despite the strength of the analysis, there were a number of significant limitations, mostly because of the gaps in the evidence base, the quality of the evidence base in some instances (e.g. issues with trial design) and the necessity to use data taken from non-UK populations when UK data were not available. There were particular concerns with the data used to reflect the benefit associated with chemotherapy by risk group for the new tests. Methodological flaws have been highlighted for the study on the benefits of chemotherapy by MammaPrint risk group. Limitations were identified with the data for OncotypeDX and Mammostrat in terms of how this evidence should be generalised to the UK population and potential biases in the evidence base. In addition, the evidence base on the proportion of patients who would receive chemotherapy after classification with the new tests had limitations or was lacking (in the case of IHC4 and Mammostrat). OncotypeDX was the only test for which there was evidence from a UK setting; however, there were concerns relating to this study, notably the small sample size and the possibility that patients were not representative of typical patients seen in clinical practice in England and Wales. Univariate sensitivity analyses indicated that the ICER was sensitive to the assumptions about the benefit reduction associated with chemotherapy and the proportion of patients who would be offered chemotherapy after categorisation with the new test. There are particular uncertainties relating to whether or not physicians would recommend chemotherapy to patients classified as intermediate risk with the new tests, as the evidence for the benefit of chemotherapy (reduction in the risk of recurrence) in these patients is less clear. Data from the TransATAC trial show that about 26% and 10% of ER+, LN–, HER– women were classified as intermediate risk with OncotypeDX and IHC4 respectively (predicted risk of recurrence between 10% and 20%). The ICER for Mammostrat was very sensitive to the assumption about the proportion of patients who would receive chemotherapy in the intermediate-risk group.

The ICER for OncotypeDX improved if more chemotherapy was given to this intermediate-risk group. In addition, in the evaluation of OncotypeDX and IHC4, the data on risk classification using OncotypeDX followed by further reclassification using IHC4 relied on a very small number of patients and therefore biases could have been introduced.

The exploratory analyses were subject to further uncertainties in the data. The exploratory analysis for Mammostrat used data from a subset of patients included in the Ring *et al.* study;[125] however, the tests (CIC information has been removed). The exploratory analysis for MammaPrint used a wide range of assumptions and it was not possible for the EAG to present an ICER with confidence given the lack of robustness of the data that have been used to populate the economic model.

No direct comparison between tests was possible because of the differences in quality of the evidence. This therefore limits the conclusions that can be drawn from the analysis.

Further uncertainties were introduced into the analysis because of the wide range of assumptions needed in the EAG model. These include:

- The use of UK cancer registry data to inform the proportion of patients receiving chemotherapy in the current practice arm. The registry data allowed us to capture decision-making based on real clinical practice, using current methods (a mix of the NPI, Adjuvant! Online and/or other prognostic tools). It should be noted that NPI was not used as a comparator in the economic model. NPI was used only to separate patients into two groups – those with a NPI score ≤ 3.4 and those with a NPI score > 3.4. This was to allow a subgroup analysis to be conducted and to allow the model to take into account, at least in part, the prognostic value of the treatment decision using clinicopathological parameters. This is a limitation given that it may be less discriminatory than current practice. The decision was taken to model current practice in this way as the evidence available for each test did not reflect the current level of chemotherapy given in the UK. Furthermore, data from only two registries were used (WMCIU and ECRIC), the results are generalisable only if the centres included in these two regions are considered to be representative of the centres across England and Wales.
- The original RS groups were used for OncotypeDX to define patients who are at low, intermediate or high risk of distant recurrence. However, cut-offs have been modified in the ongoing TAILORx trial. The impact of these revised cut-offs cannot be assessed.
- It was assumed that the IHC4 test was reproducible; however, there are issues relating to the reproducibility of the Ki-67 element of the test, which would need to be addressed in the UK before this test could be used by local laboratories in clinical practice.
- The risk of distant recurrence was assumed to be constant over the first 10 years. It is likely that the risk increases over the first few years and then decreases with time. Likewise, we assumed that the risk reduced after 10 years and that no recurrence would occur after 15 years. This is a simplifying assumption.
- The impact of locoregional recurrences was included in the model by applying a one-off cost and a decrement in utility to a proportion of patients developing distant recurrence. This is simplistic but this approach was used because of data limitations.
- Long-term adverse events were modelled using simplifying assumptions. Only AML was included as a long-term adverse event after chemotherapy. The prevalence of CHF following chemotherapy with FEC may be higher but clinical opinion suggested that modelling CHF is complex as some patients remain asymptomatic or have a reversible disease and this would have added further uncertainty into the model.

- A significant proportion of the total cost of chemotherapy (including treatment of adverse events and prevention of febrile neutropenia) is made up of the cost of treatment to prevent febrile neutropenia, which is more uncertain than the cost of the drug or the administration costs. We assumed that 25% of women receive G-CSF for the secondary prevention of neutropenia after chemotherapy in the UK based on clinical opinion.
- In the comparator arm (current clinical practice), the probability of receiving chemotherapy was based on registry data. It was assumed that the probability of receiving chemotherapy was the same irrespective of the reclassification of patients with the new test into the low-, intermediate- or high-risk group. This is likely to be conservative.

In addition, there are potentially some limitations relating to the structure of the model. The model is static in that individuals are separated into risk groups and are assumed to be homogeneous with similar characteristics on average within these groups. For the main analysis, we separated individuals according to NPI, OncotypeDX and IHC4 to allow us to model IHC4 using direct evidence against OncotypeDX. Revised structural assumptions were examined in sensitivity analysis: removing IHC4 from the analysis and therefore separating patients according to NPI and OncotypeDX only or assuming no further reclassification using IHC4 (exclusion of IHC4) and modelling the entire cohort as a single group (no split by NPI). Results of the scenario analysis suggested that our base-case ICER for OncotypeDX was minimally affected by our choice of model structure to accommodate the evaluation of IHC4. Results were more affected when modelling the entire cohort as a single group (no split by NPI and IHC4) as the prognostic value of current decision-making using clinicopathological parameters is ignored (i.e. that patients with a low NPI score have a lower risk of recurrence but are also less likely to receive chemotherapy than patients with a NPI score > 3.4 under clinical practice). It is unclear how the ICER would be affected if a different NPI cut-off was selected or if patients were separated using Adjuvant! Online or other prognostic tools. Additional limitations are imposed as we assumed that tests categorised patients into risk groups and that the groups are homogeneous. However, we did not have access to individual patient-level data to explore the heterogeneity within the risk groups or to explore using different thresholds to define the risk groups.

Finally, the model structure for the exploratory analyses for MammaPrint and Mammostrat was driven by the OncotypeDX analysis, which imposed some constraints in the data that have been used. No economic assessment was provided for PAM50, NPI+, Randox BCA, BluePrint and BCI because of significant gaps in the data.

Chapter 4

Assessment of factors relevant to the NHS and other parties

Central processing

Most GEP tests require samples to be sent to central processing laboratories and therefore time delays will be imposed on patient management pathways. This may also be an issue for Mammostrat, which is likely to require central processing. IHC tests and GEP tests that can be processed locally will provide faster results than assays that need to be processed centrally. There are also legal issues relating to possible litigation costs if errors occur when tests are performed in other legal jurisdictions. The impact of sending large numbers of blocks for central processing in terms of pathology services, tissue tracking, pathologist and technical staff time, data input on receipt, etc. would need to be explored.

Impact on NHS services

Tests that require the use of fresh tissue raise particular service configuration issues. Fresh tissue collection is not routine in the NHS and so there will be additional costs that would be considerable at hospitals where the dissection facilities are already filled to capacity (which is likely to be a significant proportion of hospitals) and where explicit staffing for collection of fresh tissue is not in place. Discussion with local clinicians indicated that capital costs could be at least £75,000 per hospital if new dissection tables are required, which is likely to be the case in many hospitals. If routine fresh tissue sampling is not in place (only a few research centres currently have this working arrangement) then additional staff costs for biomedical scientists and histopathologists will be incurred. If a full fresh tissue service was required and needed to cover all theatre time then additional staff costs could be £20,000–50,000 per year (Simon Cross, July 2011, personal communication).

The impact on the chemotherapy service has not been considered. For instance, if additional women were prescribed chemotherapy as a result of these tests, NHS capacity (compared with current practice) may need to expand. Services are typically already running at full capacity and therefore this might mean delays in chemotherapy or the need for additional staff and beds.

Quality issues relating to immunohistochemistry tests

Lack of reproducibility of IHC assays will need to be taken into account when considering the use of IHC4 in local UK laboratories. Differences in IHC values can occur as a result of variability in several factors, including fixation, antigen retrieval, reagents and interpretation. A quality assurance programme will need to be considered, such as the UK National External Quality Assessment Service (NEQAS), given that these have in the past been shown to lead to marked improvements in between-laboratory agreement. Validation of the IHC4 score when carried out in a range of local laboratories is required. A guideline is currently in preparation (Professor Mitch Dowsett, July 2011, personal communication) to help standardise

the measurement of Ki-67. Guidelines will need to be developed through NEQAS to ensure consistency among all participating UK laboratories.

Patient anxiety

There is evidence to suggest that OncotypeDX improves patient anxiety levels and decisional conflict. This was based on a small study of 89 assessable patients with ER+, LN– breast cancer.[76] Before and after OncotypeDX testing, medical oncologists stated their adjuvant treatment recommendation and confidence in it, and patients indicated their treatment choice. Changes in oncologist treatment recommendations were evaluated and patients completed measures for decisional conflict, anxiety and quality of life. Such improvements in patient anxiety levels are not taken into account within the economic analysis.

Classification of patients in the intermediate group

Some GEP and expanded IHC tests classify a proportion of patients into an intermediate-risk category. Evidence for the benefit of chemotherapy (in terms of reduction in the risk of recurrence) in these patients is less clear. It is also less clear whether or not physicians would recommend chemotherapy in addition to endocrine therapy for patients classified as intermediate risk with GEP or expanded IHC tests. This question is being addressed by the ongoing TAILORx[187] study for OncotypeDX.

Categorical risk compared with continuous risk score

Some of the GEP and expanded IHC tests (MammaPrint, Mammostrat) classify patients into risk group only (categorical) and do not calculate a continuous risk score. This is likely to be less informative than a continuous risk score as it does not differentiate between patients who are at the lower end of the distribution and those who are at the upper end or those who are borderline.

Failure of the test/wrong results

Immunohistochemistry-based tests offer the advantage that biomarker expression is interpreted in situ, which allows the pathologist to ensure that the test is not confounded by expression of biomarkers in non-tumour tissue. Gene expression assays that homogenise the tissue and measure biomarkers that may be expressed in stroma run a greater risk of confounding the interpretation of biomarker expression levels.

Chapter 5

Discussion

Statement of principal findings

Clinical effectiveness

Thirty studies reporting data on analytical validity, clinical validity or clinical utility of the nine included GEP and expanded IHC tests for breast cancer were identified. Thirty-four studies (on OncotypeDX and MammaPrint) that had been included in previous systematic reviews were also retrieved and summarised.

OncotypeDX

The OncotypeDX evidence is the furthest along the validation pathway compared with other similar tests and the evidence base, in particular in relation to the clinical validity (prognostic ability) of the test, was consider to be reasonably sound. This review has identified recent studies supporting the clinical validity of the test. These are generally of moderate to high quality. Our findings indicate that there are no prospective studies reporting the impact of OncotypeDX on long-term outcomes such as OS. Four studies on the impact of OncotypeDX on decision-making indicate that the use of OncotypeDX leads to changes in decision-making for 31.5–38% of patients, but only one of these studies relates to the UK setting. Two studies on the predictive benefit of the test were identified: one was based on the same data used in the Paik *et al.* study[48] and one included LN+ patients. The first evidence relating to improvements in quality of life and reductions in patient anxiety as a result of using the test has been reported, but this is based on small patient numbers and further evidence is required. Key gaps in the evidence remain and few of the studies were considered to be of high quality ($n = 3$). A number of studies in the current review were judged to provide moderate-quality (although retrospective) evidence for OncotypeDX ($n = 9$). Further direct evidence of clinical utility of OncotypeDX is still required. This will be addressed by the ongoing TAILORx trial.

MammaPrint

The evidence base for MammaPrint, in particular in relation to the prognostic ability of the test, is developing but is based on small sample sizes ($n \leq 272$). The evidence for MammaPrint is less robust than that for OncotypeDX. No MammaPrint studies used RCT data, the sample sizes were small and heterogeneous patient populations were studied, making generalisation of the findings difficult. None of the studies used UK-based patients and the data were all based on cohort studies. The test appears to be prognostic at 5 years although the validity of the test to predict longer-term outcomes does not seem to have been established. Robust evidence of clinical utility is needed as it is not yet clear to what extent the use of the MammaPrint test will change the management of patients. It is also unclear to what extent MammaPrint risk groups are predictive of chemotherapy benefit or how the use of MammaPrint will improve patient outcomes through increases in disease-free and overall survival. The evidence for MammaPrint to date is mainly derived from premenopausal women and this evidence may not be generalisable to an older population given that younger women are likely to be at higher risk of recurrence and are more likely to be classified as having poor prognosis using MammaPrint.

PAM50

The PAM50 evidence base, in particular in relation to the prognostic ability of the test, is developing. The limitations of this evidence are based primarily on the fact that currently most of the evidence is unpublished (ARUP Laboratories, Salt Lake City, UT, USA).

Mammostrat

The evidence base for Mammostrat is developing and the evidence relating to the prognostic ability of the test is of reasonably high quality. These initial studies include a large sample size and one study provided external validation of the test in a UK population. A further study provides evidence relating to the benefit of chemotherapy by risk group, indicating that both low- and high-risk groups benefit whereas those in the moderate-risk group do not. Further evidence is needed to clarify these findings. Further evidence of analytical validity and clinical utility is also required. In particular, there was no published evidence on reclassification of risk groups compared with conventional risk classifiers and no evidence on the impact of the test on decision-making.

IHC4

The evidence base for IHC4 is currently limited to one study of clinical validity (prognostic ability). However, this evidence for clinical validity is relatively strong given that the test has been developed using a large cohort of patients and has been validated in an external cohort. This study allowed direct comparison with OncotypeDX. Further evidence is required on the analytical validity and clinical utility of IHC4.

BluePrint, Breast Cancer Index, Nottingham Prognostic Index plus and Randox Breast Cancer Array

The evidence base for the MammaPrint and BluePrint test combined (use of the BluePrint test for subtyping following the MammaPrint test), BCI, NPI+ and Randox BCA is limited to date and no firm conclusions can be drawn about these tests.

Cost-effectiveness

The four tests with the most well-developed clinical evidence base were considered within the economic evaluation. The EAG presented a primary analysis that compared OncotypeDX and IHC4 with current clinical practice in England and Wales. Based on the EAG model the incremental cost for adjuvant chemotherapy guided using OncotypeDX was estimated to be £29,503 per QALY gained compared with current clinical practice, assuming that the test was offered to all woman with ER+, LN–, HER2– early breast cancer under our base-case assumptions (assuming the test to be predictive of the benefit of chemotherapy). The IHC4 test was dominant compared with current clinical practice, providing more QALYs at a lower cost. The ICER for OncotypeDX increased substantially if the test was assumed to be prognostic only, that is, assuming the same relative reduction in the risk of recurrence from chemotherapy for all patients irrespective of the OncotypeDX RS classification. IHC4 remained dominant under this assumption. In incremental analysis, when the treatment decision using OncotypeDX was compared with that using IHC4, the ICER increased to £64,111 per QALY gained. In a second scenario, assuming that the test was offered only to women with a NPI score > 3.4, treatment guided using IHC4 remained dominant (i.e. provided more QALYs at a lower cost) compared with current clinical practice. The incremental cost for treatment guided using OncotypeDX was £9774 per QALY gained compared with current clinical practice and £31,125 per QALY gained compared with IHC4 (assuming that the test had predictive ability). A key area of uncertainty is whether tests are prognostic or also offer predictive ability.

It should be noted that the evidence base for IHC4 is less well developed and therefore the results of this analysis should be interpreted with consideration of the assumptions used in

the evaluation. One-way sensitivity analysis indicated that the ICER was most sensitive to the assumptions about the benefit reduction associated with chemotherapy, the time horizon of the model, the starting age of the cohort, the risk of recurrence and who would receive chemotherapy depending on the result of the test. IHC4 remained dominant compared with current clinical practice (i.e. provided more QALYs at a lower cost) except when the cost of IHC4 was raised to £400 (producing an ICER of £1557 per QALY gained compared with current clinical practice).

The IHC4 test provides a continuous risk score and so it was necessary to assume risk categories for the purposes of analysis. The benefits of chemotherapy by risk group were based on indirect evidence using the OncotypeDX classification; no additional benefit was assumed for IHC4. In the absence of evidence it was assumed that the likelihood of receiving chemotherapy based on the IHC4 risk classification would be the same as for OncotypeDX RS group, that is, that physicians would interpret the results from OncotypeDX and IHC4 in the same way.

There are other issues that need to be considered for these new tests, such as technical and logistical issues. The implementation of the OncotypeDX test will have an impact on pathology services, and issues of tissues tracking and additional pathologist and technical staff time should be considered. There are also issues in terms of turnaround time and legal issues relating to possible litigation costs in case of errors when tests are performed in another legal jurisdiction. There is no morphological correlation and tissues included in the analysis can be heterogeneous and the results will be affected by tumour cellularity. The accuracy of HER2 measurements with OncotypeDX also needs clarification. The IHC4 test, despite the lack of evidence, is promising and can be incorporated more easily into clinical practice and should provide results more quickly. However, there are also issues of variability for this test (time to fixation, different fixatives), and the need for standardisation of the method of Ki-67 assessment and the cut-off to be used. Quality assurance issues would need to be addressed before the implementation of IHC4 in clinical practice.

For MammaPrint and Mammostrat there were significant gaps in the evidence available and data that have been used were not considered to be robust by the EAG. For this reason the analyses that were carried out evaluating the cost-effectiveness of MammaPrint and Mammostrat compared with current clinical practice in England and Wales were considered to be exploratory. Further clinical evidence will be needed to make the findings more robust.

Strengths and limitations of the assessment

Clinical effectiveness

The systematic review was conducted by an independent research team to a prespecified protocol using the latest evidence for nine GEP and expanded IHC tests. Extensive searches were undertaken to identify all literature relating to the clinical effectiveness of GEP and expanded IHC tests to guide the use of chemotherapy in breast cancer management.

The main limitation was the varied nature of the evidence base, relating to the study design for the evidence on clinical validity and clinical utility, making comparisons between tests difficult. None of the clinical studies had a prospective RCT design, although there are currently ongoing RCTs of both OncotypeDX and MammaPrint. Few studies, across all of the tests, used RCT data, with the majority of the evidence based on cohort designs. One of the most characteristic features of the studies across all tests was their heterogeneity. The studies varied considerably in their size, study design, patient populations and objectives. A large proportion of the studies were small and retrospective. Many studies used old archived tumour samples and some included the use of retrospective chart review to elicit treatment recommendations before and after testing. There

was a lack of standardised decision-making tools both within and between studies and non-standardised methods of patient selection were used.

Studies relating to analytical validity, where available, appeared adequate, although for the majority of the tests the data are lacking and further studies are required. For MammaPrint and BluePrint, BCI, Randox BCA, Mammostrat, IHC4 and NPI+, no specific evidence for analytical validity has been reported, and for PAM50 the evidence for analytical validity is only in abstract form.

Economic evaluation

The economic assessment was conducted by an independent research team using the latest evidence for four GEP and expanded IHC tests. The EAG economic assessment has several strengths compared with previous evaluations. The evaluation used UK-specific data whenever possible, including for the baseline use of chemotherapy, risk of distant recurrence/recurrence and reclassification with the new test, so that its conclusions should be relevant to the UK setting. Notably, the EAG model used cancer registry data from ECRIC and WMCIU and data from the TransATAC trial, which are considered to provide the best reflection of current practice in the UK. The risk reclassification with OncotypeDX and IHC4 was taken from the TransATAC trial, and the risk of distant recurrence was taken from the same data source. The EAG economic assessment also considered an analysis of IHC4, which has not previously been undertaken, using direct evidence of the test compared with OncotypeDX. We also modelled women with a NPI score ≤ 3.4 and women with a NPI score > 3.4 separately to account for the prognostic value of the current treatment decision based on clinicopathological parameters and to allow a scenario to be conducted assuming that the test was offered to a subgroup of the population with a NPI score > 3.4. Extensive sensitivity analyses were undertaken to determine the impact of key parameter uncertainties on the cost-effectiveness ratio and a PSA was carried out to account for the joint uncertainty between parameters when appropriate.

Our analysis focuses on women with ER+, LN–, HER2– early breast cancer as this population is supported by the most robust clinical evidence. Other populations, such as women with a small number of positive nodes, might also benefit from the test, and results are likely to change if the population appraised is extended to women with ER– cancer or with positive nodes. We conducted two analyses, one assuming that the test was given to all women and one assuming that the test was given only to women with a NPI score > 3.4. This subgroup analysis was undertaken to explore the impact of targeting the tests at patients at intermediate risk. It is considered likely that the majority of women with a NPI score ≤ 3.4 would be considered low risk and would not receive chemotherapy under current practice or using the new tests and therefore the test would have a limited impact on the management of these women. Although this is relatively simplistic, and includes women at the top end of the NPI distribution who are likely to receive chemotherapy despite the result of the test, it does indicate that generally the cost-effectiveness may be improved by focusing the test in these women (although this was not the case in the exploratory analysis for Mammostrat).

Despite the strength of the analysis there were some significant limitations, mostly because of gaps in the evidence base, the quality of the studies within the evidence base in some instances and the necessity of using data from non-UK populations when UK data were not available. There were particular concerns over the data used to reflect the benefit associated with chemotherapy for the categorisation of patients with the new tests. In addition, the evidence base on the proportion of patients who would receive chemotherapy after classification with the new tests had limitations or was lacking (in the case of IHC4 and Mammostrat). There are particular

uncertainties relating to whether or not physicians would recommend chemotherapy to patients classified as intermediate risk with the new tests, as the evidence for the benefit of chemotherapy (reduction in the risk of recurrence) in these patients is less clear.

The exploratory analyses for Mammostrat and MammaPrint were subject to further uncertainties in the data. The exploratory analysis for Mammostrat used data from a subset of patients included in the Ring *et al.* study;[125] however, the tests (CIC information has been removed). The exploratory analysis for MammaPrint used a wide range of assumptions and it was not possible for the EAG to present an ICER with confidence given the perceived lack of robustness of the data that have been used to populate the economic model.We did not perform an incremental analysis because of these differences in the quality of evidence between tests. These differences are not adequately reflected in the PSA. Although this may be considered a limitation, we considered that including MammaPrint and Mammastrat within an incremental analysis could potentially be misleading given the gaps in the evidence base and significant issues relating to the quality of the data used to populate the economic models for these two tests.

Uncertainty was increased by the model structure used and the significant number of assumptions that had to be made in the EAG model. These are discussed in *Chapter 3*, *Discussion of the independent economic model results*. Extensive sensitivity analyses were carried out to determine the factors that impacted most on the ICER and to determine why the results of our model differed from those of other UK evaluations. The EAG model used UK data whenever possible and modelled patients with low and intermediate or high NPI separately. The results of the EAG analysis for OncotypeDX suggest that the ICER may be higher than that reported by the manufacturer's model. The difference in the ICER between the two models is attributable to the differences in model structure, the assumptions that have been made about the risk of recurrence and the different data sources used. The model developed by the manufacturer was built on data on changes in treatment decisions taken from the Holt *et al.* study.[77] However, there are issues with this study, particularly that patients might not be representative of patients seen in clinical practice in the UK. This study indicated that 36.4% of patients with ER+, LN–, HER2– breast cancer (based on EAG analysis) were offered chemotherapy under current clinical practice, which appears high. Cancer registry data (used in the EAG economic model) suggested that about 14.4% of women with ER+, LN–, HER2– breast cancer currently receive chemotherapy (5% among women with a NPI score ≤ 3.4 and 34% among women with a NPI score > 3.4). The EAG model also used data from the TransATAC trial, which is considered to provide a more robust source of evidence for the risk of distant recurrence for patients treated with endocrine therapy in the UK. This also provided risk reclassification data in a large sample of patients and a direct comparison against IHC4. A key area of uncertainty is whether tests are prognostic only or offer predictive ability, that is, whether or not they identify high-risk patients who will benefit more in relative terms from reductions in the risk of recurrence following chemotherapy than low-risk patients.

A structural assumption was also examined in sensitivity analysis, modelling the population as a single group instead of separating patients by NPI. This was shown to influence the ICER. The explanation is that modelling patients as one group ignores the prognostic value of current treatment decision-making using clinicopathological parameters and therefore will be more favourable to the new test. The base-case analysis separated patients into two subgroups by NPI; it is unclear how the ICER would be affected if patients were separated using Adjuvant! Online.

No modelling work was undertaken on tests providing outputs in terms of intrinsic breast cancer subtype rather than risk of recurrence. This will be an important area for future modelling work.

No economic assessment was provided for PAM50, NPI+, Randox BCA, BluePrint and BCI. This was because of significant gaps in the data and the uncertainty over how the tests would be used to inform clinical decision-making.

No direct comparison between tests was possible because of the differences in quality of the evidence. This therefore limits the conclusions that can be drawn from the analysis.

Uncertainties

The main uncertainties included:

- The varied nature of the clinical evidence base, making comparisons between tests difficult.
- The lack of prospective trials for the tests directly linking the use of the tests with final outcomes in terms of recurrence or survival. The economic model therefore needed to combine data from different sources to model how the results from the new tests translated into final outcomes in the form of QALYs, resulting in significant limitations – data used in the model were not always based on UK populations, were not always specifically taken from the ER+, LN–, HER2– population of interest and tended to be based on younger populations and populations treated with older, less effective, endocrine and chemotherapy regimens than are currently used.
- The lack of data on the ability of the tests to classify patients in the relevant UK population.
- The benefit of chemotherapy in terms of reduction in the risk of distant recurrence/recurrence in patients classified as low, intermediate or high risk according to the new tests. Although evidence was available for three of the tests (OncotypeDX, MammaPrint and Mammostrat), there were limitations with these studies and it is also unclear how this evidence translates specifically to the ER+, LN–, HER2– population in the UK. A key area of uncertainty is therefore whether tests are prognostic only or are predictive of the benefit of chemotherapy.
- The lack of UK data about how the tests will impact on decision-making, that is, the proportion of patients who would receive chemotherapy according to the risk classification with the new test. One small UK study was identified for OncotypeDX but this had some limitations. Also, there is a lack of evidence on how this impact is likely to differ between tests providing a continuous risk score and tests providing only a categorical risk label.
- Some GEP and expanded IHC tests classify a proportion of patients into an intermediate-risk category. Evidence for the benefit of chemotherapy (reduction in the risk of recurrence) in these patients is not clear. It is more uncertain whether or not physicians would recommend chemotherapy in addition to endocrine therapy for patients classified as being at intermediate risk with GEP or expanded IHC tests.
- How the test will be used in UK clinical practice, notably the group of women who are most likely to be offered the new tests.
- The potential acceptance and adoption of the tests by UK physicians.

Other relevant factors

Our analyses do not capture the cost implications of any service reconfiguration issues. The use of fresh tissue by some tests would require a change in practice with regard to the handling of tissues by pathology laboratories. This would have major service reconfiguration and cost issues. The impact on the chemotherapy service has also not been considered. For instance, if additional women were prescribed chemotherapy as a result of these tests, NHS capacity (compared with

current practice) may need to expand. Services are typically already running at full capacity and therefore this might mean delays in chemotherapy or the need for additional staff and beds.

Most GEP tests require samples to be sent to central processing laboratories and therefore time delays will be imposed on patient management pathways. This may also be an issue for Mammostrat, which is likely to require central processing.

Currently, ER and HER2 testing is performed in most hospitals whereas PR testing is performed in a more limited number of hospitals. The potential introduction of the IHC4 test would require quality assurance issues to be addressed for the Ki-67 test. Because the IHC4 test is expected to be carried out locally, full validation would require evaluation of the IHC4 score when carried out in a range of local laboratories. A guideline is currently in preparation (Professor Mike Dowsett, July 2011, personal communication) to help standardise the measurement of Ki-67; however, reproducibility of the test would need to be confirmed and quality assurance programmes put in place.

Some of the tests (MammaPrint, Mammostrat) classify patients into risk group only (categorical) and do not calculate a continuous risk score. This is less informative than a continuous risk score as patients are classified into broad groups, for example low, intermediate or high. This does not allow differentiation between patients who are at the lower end or upper end of the distributions and those who are borderline; the impact that this additional knowledge would have on clinical decision-making is unclear.

Immunohistochemistry-based tests such as Mammostrat offer the advantage that biomarker expression is interpreted in situ, which allows the pathologist to ensure that the test is not confounded by expression of biomarkers in non-tumour tissue. Gene expression assays that require homogenisation of the tissue and measure biomarkers that may be expressed in stroma run a greater risk of confounding the interpretation of biomarker expression levels.

Chapter 6

Conclusions

Clinical effectiveness

Two of the tests (OncotypeDX and MammaPrint) have a reasonably large evidence base although there are some methodological weaknesses relating to this evidence in terms of the heterogeneity of patient cohorts and the retrospective study design. In addition, the MammaPrint evidence is typically based on observational data (small cohort studies) rather than randomised data, increasing the risk of selection bias. Further evidence is required on the clinical utility of all of the tests and specifically in UK-based populations.

The IHC4 and Mammostrat tests also demonstrate promise, presenting early evidence of the prognostic ability of the tests based on large UK-based validation cohorts. There is no predictive evidence for IHC4. PAM50 has an emerging evidence base; however, most of the evidence to date is in abstract form or unpublished. NPI+, Randox BCA, BluePrint and BCI have little evidence to date.

Cost-effectiveness

The economic analysis suggests that the use of the new tests may result in small increases in QALYs compared with currently used prognostic tools, but current limitations in the evidence base produce significant uncertainty in the results. A key area of uncertainty is whether tests are prognostic only or identify high-risk patients who will benefit from larger relative reductions in the risk of recurrence following chemotherapy than lower-risk patients. The economic analysis suggested that, of the four tests considered, treatment guided using IHC4 has the greatest potential to be cost-effective at a willingness-to-pay threshold of £20,000, given the low cost of the test. However, the evidence base to support IHC4 needs further research and the exact cost of using the test in the NHS needs to be investigated further. OncotypeDX has a more robust evidence base but further evidence on its impact on decision-making in the UK and the predictive ability of the test, specifically in an ER+, LN–, HER– population receiving current endocrine and chemotherapy regimens, is needed. For MammaPrint and Mammostrat there were significant gaps in the evidence available and the estimates of cost-effectiveness produced were not considered to be robust by the EAG.

Implications for service provision

The implications for service provision will vary by test. The impact of sending large numbers of blocks to central testing facilities in terms of pathology services, tissue tracking, pathologist and technical staff time, data input on receipt, etc. would need to be explored. The potential introduction of the IHC4 test would require quality assurance issues to be addressed for the Ki-67 element of the test. Currently, ER and HER2 testing is performed in most hospitals whereas PR testing is performed in a more limited number of hospitals. The use of tests requiring fresh tissue would be expected to have more major implications for service reconfiguration within

pathology departments in England and Wales as currently only a minority of centres in the UK have the structure and staff to handle fresh tissue.

Gene expression profiling tests requiring samples to be sent to central processing laboratories will impose time delays of up to 2–3 weeks on patient management pathways. This may also be an issue for Mammostrat, which is likely to require central processing.

Suggested research priorities

Future research priorities common to all tests include:

- Studies investigating the predictive ability of GEP and expanded IHC tests. Do tests identify patients classified at high risk who benefit more in terms of larger relative reductions in risk of recurrence following chemotherapy than those classified at lower risk?
- Prospective studies investigating how the tests will be used in clinical practice within the current decision-making process in England and Wales. Further evidence is needed for all of the tests demonstrating how they will be used in the current decision-making process and, especially, how this will impact on patient management decisions.
- There is a need for pilot studies demonstrating how tests could be introduced in the UK and used within the current decision-making process and highlighting issues that this would raise for the NHS.
- Studies investigating the use of continuous compared with categorical risk scores in terms of clinicians' preferences and the potential differential impact on decision-making.
- Studies providing evidence on how the subtyping information provided by some tests would impact on clinical decision-making in the UK.
- The psychological impact of these tests needs more formal evaluation, in particular the impact of the test results on decision conflict, decision quality and regret for women considering chemotherapy. Quality of life data in women who have access to the tests or not would also be of value.
- Further extension of the clinical evidence base to other populations who may benefit from the use of these tests, including patients with a small number of positive nodes.

Research specific to IHC4 includes:

- Studies on the analytical validity of IHC4. There is a lack of data on the reproducibility of the IHC tests used to compose the IHC4 score, in particular the Ki-67 element. Studies need to investigate whether or not the incorporation of Ki-67 in clinical practice is feasible and whether or not results are reproducible.
- Further studies to confirm the prognostic value of IHC4. There is also a need for studies directly comparing the use of IHC4 against current practice (NPI and Adjuvant! Online) in England and Wales.

Research specific to Mammostrat includes:

- further evidence on analytical validity and risk reclassification.

Research specific to MammaPrint includes:

- studies based on trial data – although the test is promising, most data are based on cohort studies or pooled analyses
- research evidence to confirm the analytical validity and clinical validity of MammaPrint results based on FFPE rather than fresh samples.

Acknowledgements

The following team of clinical advisors gave substantial advice throughout the project and/or provided peer review of the report: Dr Matthew C Winter, Consultant in Medical Oncology, Weston Park Hospital, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield; Dr Simon S Cross, Reader and Honorary Consultant, Academic Unit of Pathology, Faculty of Medicine, Dentistry and Health, University of Sheffield, Sheffield; and Professor Jorge S Reis-Filho, Professor of Molecular Pathology, Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, London.

The authors also wish to thank the Eastern Cancer Registration and Information Centre (ECRIC) and West Midland Cancer Intelligence Unit (WMCIU) for providing data and advice.

The authors wish to thank Gill Rooney and Andrea Shippam for their help in preparing and formatting the report, John Stevens for providing statistical advice, Alice Bessey and Hasan Basarir assisted in deriving input parameters for the economic model, and Mike Paulden, Dr Peter Hall, Dr Matt Stevenson and Dr Eva Kaltenthaler for peer-reviewing the report.

Contribution of authors

Sue Ward was the Assessment Group lead and provided advice for the conduct of the clinical effectiveness review and was involved in the development of the economic model.

Alison Scope undertook the clinical effectiveness review.

Rachid Rafia undertook the cost-effectiveness review and developed the cost-effectiveness model.

Abdullah Pandor and **Sue Harnan** helped undertake the clinical effectiveness review.

Pippa Evans performed the literature searches.

Lynda Wyld provided clinical advice, reviewed the report and was involved in the writing of the report.

About ScHARR

The School of Health and Related Research (ScHARR) is one of the twelve departments that comprise the Faculty of Medicine, Dentistry and Health at the University of Sheffield. ScHARR specialises in health services and public health research, and the application of health economics and decision science to the development of health services and the improvement of public health.

The ScHARR Technology Assessment Group (ScHARR-TAG) synthesises research on the clinical effectiveness and cost-effectiveness of health-care interventions for the National Institute of Health Research (NIHR) Health Technology Assessment programme on behalf of a range of policy-makers, including NICE. ScHARR-TAG is part of a wider collaboration of a number of units from other regions including Southampton Health Technology Assessment Centre (SHTAC), University of Southampton; Aberdeen Health Technology Assessment Group

(Aberdeen HTA Group), University of Aberdeen; Liverpool Reviews & Implementation Group (LRiG), University of Liverpool; Peninsular Technology Assessment Group (PenTAG), University of Exeter; Centre for Reviews and Dissemination (CRD) and Centre for Health Economics (CHE), University of York; Warwick Evidence, University of Warwick; the BMJ Group; and Kleijnen Systematic Reviews.

References

1. Dumitrescu RG, Cotarla I. Understanding breast cancer risk – where do we stand in 2005? *J Cell Mol Med* 2005;**9**:208–21.
2. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58 209 women with breast cancer and 101 986 women without the disease. *Lancet* 2001;**358**:1389–99.
3. Ziegler RG, Hoover RN, Pike MC, Hildesheim A, Nomura AM, West DW, *et al.* Migration patterns and breast cancer risk in Asian-American women. *J Natl Cancer Inst* 1993;**85**:1819–27.
4. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* 2008;**371**:569–78.
5. Lagerros YT, Hsieh SF, Hsieh CC. Physical activity in adolescence and young adulthood and breast cancer risk: a quantitative review. *Eur J Cancer Prev* 2004;**13**:5–12.
6. Fisher B, Bauer M, Wickerham DL, Redmond CK, Fisher ER, Cruz AB, *et al.* Relation of number of positive axillary nodes to the prognosis of patients with primary breast cancer. An NSABP update. *Cancer* 1983;**52**:1551–7.
7. National Institute for Health and Care Excellence, National Collaborating Centre for Cancer. *Early and locally advanced breast cancer: diagnosis and treatment*. Clinical guideline CG80. London: NICE; 2009.
8. Office for National Statistics. *Survival rates in England, patients diagnosed 2001–2006 followed up to 2007*; 2007. URL: www.statistics.gov.uk/StatBase/Product.asp?vlnk=14007 (last accessed September 2011).
9. West Midlands Cancer Intelligence Unit. *0–10 year relative survival for cases of breast cancer by stage diagnosed in the West Midlands 1985–1989 followed up to the end of 1999, as at January 2002*. 2009.
10. Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham Prognostic Index in primary breast cancer. *Breast Cancer Res Treat* 1992;**22**:207–19.
11. Adjuvant! Online. *Adjuvant! Online. Decision making tools for health care professionals*; 2009. URL: www.adjuvantonline.com/index.jsp (last accessed September 2011).
12. Cancer Research UK. *Cancer statistics: breast cancer UK*; 2009. URL: http://publications.cancerresearchuk.org/WebRoot/crukstoredb/CRUK_PDFs/CSBREA09.pdf (last accessed September 2011).
13. Cancer Research UK, National Cancer Intelligence Network. *Cancer incidence and survival by major ethnic group, England, 2002–2006*; 2009. URL: http://publications.cancerresearchuk.org/WebRoot/crukstoredb/CRUK_PDFs/CR-UKNCIN_ETHNIC.pdf (last accessed September 2011).
14. National Cancer Intelligence Network. *Cancer incidence by deprivation England, 1995–2004*. URL: www.ncin.org.uk/docs/081202-NCIN-Incidence_by_Deprivation_95_04.pdf (last accessed September 2011).

15. Welsh Cancer Intelligence and Surveillance Unit. *Cancer incidence, mortality and survival by deprivation in Wales*; 2009. URL: www.wales.nhs.uk/sites3/Documents/242/Deprivation%20in%20Wales%201993–2007.pdf (last accessed September 2011).

16. Cancer Research UK. *Cancer survival and deprivation statistics*; 2005. URL: http://info.cancerresearchuk.org/cancerstats/survival/survivaldeprivation (last accessed September 2011).

17. Office for National Statistics. *Death registrations in England and Wales, 2008, causes*; 2009. URL: www.statistics.gov.uk/pdfdir/dthreg0809.pdf (last accessed September 2011).

18. Blanks RG, Moss SM, McGahan CE, Quinn MJ, Babb PJ. Effect of NHS breast screening programme on mortality from breast cancer in England and Wales, 1990–8: comparison of observed with predicted mortality. *BMJ* 2000;**321**:665–9.

19. International Union against Cancer. *TNM classification of malignant tumours*, 5th edn. New York: Wiley-Liss; 1997.

20. Singletary SE, Connolly JL. Breast cancer staging: working with the sixth edition of the AJCC Cancer Staging Manual. *CA Cancer J Clin* 2006;**56**:37–47.

21. Cancer Research UK. *TNM breast cancer staging*; 2009. URL: www.cancerhelp.org.uk/help/default.asp?page=3316 (last accessed September 2011).

22. Agency for Healthcare Research and Quality. *Surgery choices for women with early-stage breast cancer*; 2009. URL: www.ahrq.gov/consumer/brcanchoice.htm (last accessed September 2011).

23. US National Cancer Institute. *Surgery choices for women with early-stage breast cancer*; 2009. URL: www.cancer.gov/cancertopics/breast-cancer-surgery-choices (last accessed September 2011).

24. American Cancer Society. *How Is Breast Cancer Staged?* URL: www.cancer.org/docroot/CRI/content/CRI_2_4_3X_How_is_breast_cancer_staged_5.asp. 2009 (last accessed August 2011).

25. National Institute for Health and Care Excellence. *Gene expression profiling and expanded immunohistochemistry tests to guide selection of chemotherapy regimes in breast cancer management.* Final scope. Diagnostics Assessment Programme. London: NICE; 2011.

26. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, *et al.* Molecular portraits of human breast tumours. *Nature* 2000;**406**:747–52.

27. Genomic Health. *Oncotype DX*. Information provided by the manufacturers to NICE. 2011.

28. National Institute for Health and Care Excellence. *Guidance on cancer services. Improving outcomes in breast cancer*. London: NICE; 2011.

29. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2011;**365**:1687–717.

30. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. Centre for Reviews and Dissemination; 2009. URL: www.york.ac.uk/inst/crd/systematic_reviews_book.htm (last accessed September 2011).

31. Liberati A, Altman D, Tetzlaff J, Mulrow C, Gotzsche P, Ioannidis J, *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;**339**:b2700.

32. National Institute for Health and Care Excellence. *NICE Diagnostic Assessment Programme interim methods statement (pilot)*. London: NICE; 2010.

33. Marchionni L, Wilson R, Marinopoulos S, Wolff A, Parmigiani G, Bass E, *et al.* Impact of gene expression profiling tests on breast cancer outcomes. Evidence Report/Technology Assessment No. 160. Rockville, MD: Agency for Healthcare Research and Quality; 2008.

34. Smartt P. *A comparison of gene expression profiling tests for breast cancer*. HSAC Report 3(16). Christchurch, New Zealand: Health Services Assessment Collaboration; 2010.

35. Levine MN, Browman GP, Gent M, Roberts R, Goodyear M. When is a prognostic factor useful? A guide for the perplexed. *J Clin Oncol* 1991;**9**:348–56.

36. Simon R. A checklist for evaluating reports of expression profiling for treatment selection. *Clin Adv Hematol Oncol* 2006;**4**:219–24.

37. Altman D. Systematic reviews in health care: systematic reviews of evaluations of prognostic variables. *BMJ* 2001;**323**:224–8.

38. Shea B, Hamel C, Wells G, Bouter L, Kristjansson E, Grimshaw J, *et al.* AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol* 2009;**62**:1013–20.

39. Cronin M, Pho M, Dutta D, Stephans J, Shak S, Kiefer M, *et al.* Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. *Am J Pathol* 2004;**164**:35–42.

40. Cronin M, Sangli C, Liu M, Pho M, Dutta D, Nguyen A, *et al.* Analytical validation of the Oncotype DX genomic diagnostic test for recurrence prognosis and therapeutic response prediction in node-negative, estrogen receptor-positive breast cancer. *Clin Chem* 2007;**53**:1084–91.

41. Habel L, Shak S, Jacobs M, Capra A, Alexander C, Pho M, *et al.* A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Res* 2006;**8**:R25.

42. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, *et al.* A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;**351**:2817–26.

43. Cobleigh M, Tabesh B, Bitterman P, Baker J, Cronin M, Liu M, *et al.* Tumor gene expression and prognosis in breast cancer patients with 10 or more positive lymph nodes. *Clin Cancer Res* 2005;**11**:8623–31.

44. Esteva F, Sahin A, Cristofanilli M, Coombes K, Lee S, Baker J, *et al.* Prognostic role of a multigene reverse transcriptase-PCR assay in patients with node-negative breast cancer not receiving adjuvant systemic therapy. *Clin Cancer Res* 2005;**11**:3315–19.

45. Bryant J. *Toward a more rational selection of tailored adjuvant therapy*. Data from the National Surgical Adjuvant Breast and Bowel Project; Proceedings of the Ninth International St Gallen Breast Cancer Conference, 26–29 January 2005, St Gallen, Switzerland.

46. Hornberger J, Cosler L, Lyman G. Economic analysis of targeting chemotherapy using a 21-gene RT-PCR assay in lymph-node-negative, estrogen-receptor-positive, early-stage breast cancer. *Am J Manag Care* 2005;**11**:313–24.

47. Paik S, Shak S, Tang G. *Risk classification of breast cancer patients by the recurrence score assay: comparison to guidelines based on patient age, tumor size, and tumor grade*. Abstract 104. Paper presented at the Annual San Antonio Breast Cancer Symposium; 2004.

48. Oratz R, Paul D, Cohn A, Sedlacek S. Impact of a commercial reference laboratory test recurrence score on decision making in early-stage breast cancer. *J Oncol Pract* 2007;**3**:182–6.

49. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, *et al.* Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 2006;**24**:3726–34.

50. Goldstein L, Gray R, Badve S, Childs B, Yoshizawa C, Rowley S, *et al.* Prognostic utility of the 21-gene assay in hormone receptor-positive operable breast cancer compared with classical clinicopathologic features. *J Clin Oncol* 2008;**26**:4063–71.

51. Wolf I, Ben Baruch N, Shapira-Frommer R, Rizel S, Goldberg H, Yaal-Hahoshen N, *et al.* Association between standard clinical and pathologic characteristics and the 21-gene recurrence score in breast cancer patients: a population-based study. *Cancer* 2008;**112**:731–6.

52. Asad J, Jacobson A, Estabrook A, Smith S, Boolbol S, Feldman S, *et al.* Does Oncotype DX recurrence score affect the management of patients with early-stage breast cancer? *Am J Surg* 2008;**196**:527–9.

53. Henry L, Stojadinovic A, Swain S, Prindiville S, Cordes R, Soballe PW. The influence of a gene expression profile on breast cancer decisions. *J Surg Oncol* 2009;**99**:319–23.

54. Li L, Xu X, Zhao Y, Liu Z, Shen Z, Jin W, *et al.* Integrated gene expression profile predicts prognosis of breast cancer patients. *Breast Cancer Res Treat* 2009;**113**:231–7.

55. Rayhanabad J, Difronzo L, Haigh P, Romero L. Changing paradigms in breast cancer management: introducing molecular genetics into the treatment algorithm. *Am Surg* 2008;**74**:887–90.

56. Erb C, Fox KR, Patel M, Hook K, DeMichele A, Kaplan C, Domchek S. Evaluation of practice patterns in the treatment of node-negative, hormone-receptor positive breast cancer patients with the use of the oncotype DX assay at the University of Pennsylvania. Paper presented at the 30th Annual San Antonio Breast Cancer Symposium, 2007.

57. Gold JM, Najita JS, Lester S, Richardson AL, Morganstern DE, Chen WY, *et al.* Personalizing treatment in early-stage breast cancer: The role of standard clinical factors and genomic information in adjuvant chemotherapy decision making. *J Clin Oncol* 2009;**27**(Suppl. 15):572.

58. Lo SS, Norton J, Mumby PB, Smerage J, Kash J, Chew HK, *et al.* Prospective multicenter study of the impact of the 21-gene recurrence score (RS) assay on medical oncologist (MO) and patient (pt) adjuvant breast cancer (BC) treatment selection. *J Clin Oncol* 2007;**25**(Suppl. 15):18.

59. Shak S, Palmer G, Baehner FL, Millward C, Watson D, Sledge GW. Molecular characterization of male breast cancer by standardized quantitative RT-PCR analysis: first large genomic study of 347 male breast cancers compared to 82,434 female breast cancers. *J Clin Oncol* 2009:**27**(Suppl. 15):549.

60. Ach R, Floore A, Curry B, Lazar V, Glas A, Pover R, *et al.* Robust interlaboratory reproducibility of a gene expression signature measurement consistent with the needs of a new generation of diagnostic tools. *BMC Genomics* 2007;**8**:148.

61. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, *et al.* Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* 2006;**98**:1183–92.

62. Glas A, Floore A, Delahaye L, Witteveen A, Pover R, Bakx N, *et al.* Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics* 2006;**7**:278.

63. van't Veer LJ, Dai H, van de Vijver M, He Y, Hart A, Mao M, *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;**415**:530–6.

64. van de Vijver M, He Y, van't Veer L, Dai H, Hart A, Voskuil D, *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N Eng J Med* 2002;**347**:1999–2009.

65. Mook S, Schmidt M, Viale G, Pruneri G, Eekhout I, Floore A, *et al.* The 70-gene prognosis-signature predicts disease outcome in breast cancer patients with 1–3 positive lymph nodes in an independent validation study. *Breast Cancer Res Treat* 2009;**116**:295–302.

66. Wittner B, Sgroi D, Ryan P, Bruinsma T, Glas A, Male A, *et al.* Analysis of the MammaPrint breast cancer assay in a predominantly postmenopausal cohort. *Clin Cancer Res* 2008;**14**:2988–93.

67. Bueno-de-Mesquita JM, van Harten W, Retel V, Van't Veer LJ, van Dam F, Karsenberg K, *et al.* Use of 70-gene signature to predict prognosis of patients with node-negative breast cancer: a prospective community-based feasibility study (RASTER). *Lancet Oncol* 2007;**8**:1079–87.

68. Bender RA, Knauer M, Rutgers EJ, Glas AM, de Snoo FA, Linn, SC, *et al.* The 70-gene profile and chemotherapy benefit in 1,600 breast cancer patients. *J Clin Oncol* 2009;**27**(Suppl. 15):512.

69. de Snoo FA, Knauer M, Bender RA, Stork-Sloots L, Rutgers EJ, Glas AM, *et al.* Outcome prediction by the 70-gene profile in the context of the National Comprehensive Cancer Network (NCCN) guidelines. *J Clin Oncol* 2009;**27**(Suppl. 15):535.

70. Glas A, de Snoo F, Dreezen C, Roepman P, Bender R, Van 't Veer, L. *Use of MammaPrint to identify good prognosis group within clinically indeterminate risk patients*. 2008. URL: www.asco.org/ASCOv2/Meetings/Abstracts?&vmview=abst_detail_view&confID=58&abstractID=40205.

71. Knauer M, Mook S, Retèl V, Kok M, Wesseling J, Glas AM, *et al.* Early determination of metastatic potential in breast cancer: The 70-gene signature in small tumors. *J Clin Oncol* 2009;**27**(Suppl. 15):518.

72. Saghatchian M, Mook S, Pruneri G, Viale G, Glas A, Eekhout I, *et al.* 43LBA Combining genomic profiling (70-gene MammaPrint) with nodal status allows to classify patients with primary breast cancer and positive lymph nodes (1-9) into very distinct prognostic subgroups that could help tailor treatment strategies. *Eur J Cancer Suppl* 2009;**7**:19.

73. Henry LR, Stojadinovic A, Swain SM, Prindiville S, Cordes R, Soballe PW. The influence of a gene expression profile on breast cancer decisions. *J Surg Oncol* 2009;**99**:319–23.

74. Li LF, Xu XJ, Zhao Y, Liu ZB, Shen ZZ, Jin WR, *et al.* Integrated gene expression profile predicts prognosis of breast cancer patients. *Breast Cancer Res Treat* 2009;**113**:231–7.

75. Mook S, Schmidt MK, Weigelt B, Kreike B, Eekhout I, van de Vijver MJ, *et al.* The 70-gene prognosis signature predicts early metastasis in breast cancer patients between 55 and 70 years of age. *Ann Oncol* 2010;**21**:717–22.

76. Lo SS, Mumby PB, Norton J, Rychlik K, Smerage J, Kash J, *et al.* Prospective multicenter study of the impact of the 21-gene recurrence score assay on medical oncologist and patient adjuvant breast cancer treatment selection. *J Clin Oncol* 2010;**28**:1671–6.

77. Geffen DB, Amir N, Sion-Vardy N, Ariad S, Kachko L, Bayme M, *et al. Stage I breast cancer in a regional oncology practice in Israel* 2002–2006: clinicopathologic features, risk estimation and planned therapy of 328 consecutive patients. *Breast* 2009;**18**:316–21.

78. Holt S, Bertelli G, Pudney D, Rolles M, Moe M, Brinkworth S, *et al.* Results from a prospective clinical study on the impact of Oncotype DX on adjuvant treatment decision and risk classification by Nottingham prognostic index (NPI) and recurrence score (RS). *Breast* 2011;(Suppl. 1):s43.

79. Dowsett M, Cuzick J, Wale C, Forbes J, Mallon EA, Salter J, *et al.* Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a TransATAC study. *J Clin Oncol* 2010;**28**:1829–34.

80. Mamounas EP, Tang G, Fisher B, Paik S, Shak S, Costantino JP, *et al.* Association between the 21-gene recurrence score assay and risk of locoregional recurrence in node-negative, estrogen receptor-positive breast cancer: results from NSABP B-14 and NSABP B-20. *J Clin Oncol* 2010;**28**:1677–83.

81. Tang G, Shak S, Paik S, Anderson SJ, Costantino JP, Geyer CE, *et al.* Comparison of the prognostic and predictive utilities of the 21-gene recurrence score assay and Adjuvant! for women with node-negative, ER-positive breast cancer: results from NSABP B-14 and NSABP B-20. *Breast Cancer Res Treat* 2011;**127**:133–42.

82. Ademuyiwa FO, Miller A, O'Connor T, Edge SB, Thorat MA, Sledge GW, *et al.* The effects of Oncotype DX recurrence scores on chemotherapy utilization in a multi-institutional breast cancer cohort. *Breast Cancer Res Treat* 2011;**126**:797–802.

83. Albain KS, Barlow WE, Shak S, Hortobagyi GN, Livingston RB, Yeh IT, *et al.* Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol* 2010;**11**:55–65.

84. Cuzick J, Dowsett M, Wale C, Salter J, Quinn E, Zabaglo L, *et al.* Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and Human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. *J Clin Oncol* 2011;**29**:4273–8.

85. Kelly CM, Krishnamurthy S, Bianchini G, Litton JK, Gonzalez-Angulo AM, Hortobagyi GN, *et al.* Utility of Oncotype DX risk estimates in clinically intermediate risk hormone receptor-positive, HER2-normal, grade II, lymph node-negative breast cancers. *Cancer* 2010;**116**:5161–7.

86. Tang G, Constantino J, Crager M, Shak S, Wolmark N. [S4–9] Comparing the prediction of chemotherapy benefit in patients with node-negative, ER-positive breast cancer using the recurrence score and a new measure that integrates clinical and pathologic factors with the recurrence score. Proceedings of the 33rd Annual San Antonio Breast Cancer Symposium, 8–12 December, 2010.

87. Toi M, Iwata H, Yamanaka T, Masuda N, Ohno S, Nakamura S, *et al.* Clinical significance of the 21-gene signature (Oncotype DX) in hormone receptor-positive early stage primary breast cancer in the Japanese population. *Cancer* 2010;**116**:3112–18.

88. Yorozuya K, Takeuchi T, Yoshida M, Mouri Y, Kousaka J, Fujii K, *et al.* Evaluation of Oncotype DX recurrence score as a prognostic factor in Japanese women with estrogen receptor-positive, node-negative primary stage I or IIA breast cancer. *J Cancer Res Clin Oncol* 2010;**136**:939–44.

89. Dundar Y, Dodd S, Dixkson R, Walley T, Haycox A, Williamson PR. Comparison of conference abstracts and presentations with full-text articles in the health technology assessment of rapidly evolving technologies. *Health Technol Assess* 2006;**10**(5).

90. Dowsett M, Cuzick J, Wales C, Forbes J, Mallon L, Salter J, *et al.* Risk of distant recurrence using Oncotype DX in postmenopausal primary breast cancer patients treated with anastrozole or tamoxifen: a TransATAC study. *Cancer Res* 2009;**69**:75S–76S.

91. ClinicalTrials.gov. *Hormone therapy with or without combination chemotherapy in treating women who have undergone surgery for node-negative breast cancer (the TAILORx Trial)*; 2011. URL: http://clinicaltrials.gov/ct2/show/study/NCT00310180?term=oncotype+dx&rank=3 (last accessed September 2011).

92. Kunz G. Use of a genomic test (MammaPrint™) in daily clinical practice to assist in risk stratification of young breast cancer patients. *Arch Gynecol Obstet* 2011;**283**:597–602.

93. Bueno-de-Mesquita JM, Linn SC, Keijzer R, Wesseling J, Nuyten DSA, van Krimpen C, *et al.* Validation of 70-gene prognosis signature in node-negative breast cancer. *Breast Cancer Res Treat* 2009;**117**:483–95.

94. Gevensleben H, Gohring UJ, Buttner R, Heukamp LC, Kunz G, Dimpfl T, *et al.* Comparison of MammaPrint and TargetPrint results with clinical parameters in German patients with early stage breast cancer. *Int J Mol Med* 2010;**26**:837–43.

95. Goldhirsch A, Wood WC, Gelber RD, Coates AS, Thurlimann B, Senn HJ. Meeting highlights: updated international expert consensus on the primary therapy of early breast cancer. *J Clin Oncol* 2003;**21**:3357–65.

96. Ishitobi M, Goranova TE, Komoike Y, Motomura K, Koyama H, Glas AM, *et al.* Clinical utility of the 70-gene MammaPrint profile in a Japanese population. *Jpn J Clin Oncol* 2010;**40**:508–12.

97. Goldhirsch A, Ingle JN, Gelber R, Coates AS, Thuerlimann B. Thresholds for therapies: highlights of the St Gallen international expert consensus on the primary therapy of early breast cancer. *Ann Oncol* 2009;**20**:1319–29.

98. Goldhirsch A, Glick JH, Gelber RD, Senn HJ. Meeting Highlights: international consensus panel on the treatment of primary breast cancer. *J Natl Cancer Inst* 1998;**90**:1601–9.

99. Kok M, Koornstra R, Mook M, Hauptmann M, Fles R, Glas A, *et al.* Additional value of 70-gene signature and levels of ER and PR in the prediction of outcome from tamoxifen-treated ER-positive breast cancer. In *The Breast*. 2012. doi:10.1016/j.breast.2012.04.010.

100. Na KY, Kim KS, Lee JE, Kim HJ, Yang JH, Ahn SH, *et al.* The 70-gene prognostic signature for Korean breast cancer patients. *J Breast Cancer* 2011;**14**:33–8.

101. National Institutes of Health Consensus Development Panel. National Institutes of Health Consensus Development Conference statement: adjuvant therapy for breast cancer. *J Natl Cancer Inst Monogr* 2001;**30**:5–15.

102. Goldhirsch A, Wood WC, Gelber RD, Coates AS, Thurlimann B, Senn HJ. Progress and promise: highlights of the international expert consensus on the primary therapy of early breast cancer 2007. *Ann Oncol* 2007;**18**:1133e44.

103. Goldhirsch A, Glick JH, Gelber R, Senn HJ. Meeting highlights: international consensus panel on the treatment of primary breast cancer. *J Natl Cancer Inst* 1998;**90**:1601–9.

104. Goldhirsch A, Glick JH, Gelber RD, Coates AS, Senn HJ. Meeting highlights: International Consensus Panel on the treatment of primary breast cancer. Seventh International Conference on adjuvant therapy of primary breast cancer. *J Clin Oncol* 2001;**19**:381–727.

105. Nationaal Borstkanker Overleg Nederland. *Richtlijn Behandeling van hetMammacarcinoom 2005*. 2005. URL:www.oncoline.nl.

106. Kwaliteitsinstituut voor de Gezondheidszorg CBO Vereniging voor Integrale Kankercentra. In *Adjuvante Systemische Therapie voor het Operabel Mammacarcinoom*. Richtlijn Behandeling van het Mammacarcinoom; 2005. pp. 46–70.

107. Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, *et al.* Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol* 2001;**19**:980–91.

108. Mook S, Knauer M, Bueno-de-Mesquita JM, Retel VP, Wesseling J, Linn SC, *et al.* Metastatic potential of T1 breast cancer can be predicted by the 70-gene MammaPrint signature. *Ann Surg Oncol* 2010;**17**:1406–13.

109. Bueno-de-Mesquita JM, Sonke G, van der Vijer M, Linn S. Additional value and potential use of the 70-gene prognosis signature in node-negative breast cancer in daily clinical practice. *Ann Oncol* 2011;**22**:2012–30.

110. Knauer M, Cardoso F, Wesseling J, Bedard PL, Linn SC, Rutgers EJT, *et al.* Identification of a low-risk subgroup of HER-2-positive breast cancer by the 70-gene prognosis signature. *Br J Cancer* 2010;**103**:1788–93.

111. Knauer M, Mook S, Rutgers EJT, Bender RA, Hauptmann M, van de Vijver MJ, *et al.* The predictive value of the 70-gene signature for adjuvant chemotherapy in early breast cancer. *Breast Cancer Res Treat* 2010;**120**:655–61.

112. Knauer M, Mook S, Linn S, van't Veer L, Rutgers E. The 70-gene expression profile as MammaPrint predictive markers for a response to (neo) adjuvant chemotherapy in breast cancer. *Onkologie* 2010;**33**:27.

113. Tice JA. *The 70-gene signature (MammaPrint) as a guide for the management of early stage breast cancer*. San Francisco: California Technology Assessment Forum; 2010. URL: www.ctaf.org/content/assessment/detail/1178 (last accessed September 2011).

114. Stork-Sloots L, Krijgsman O, Roepman P, de Snoo FA, Bender RA, Glas AM. Combining multigene profiling of molecular subtypes with the 70-gene profile for classification of breast cancer. *J Clin Oncol* 2009;**27**:15s.

115. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 2009;**27**:1160–7.

116. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, *et al.* A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin Cancer Res* 2010;**16**:5222–32.

117. Cheang M, Voduc D, Tu D, Jiang S, Leung S, Chia SKL, *et al.* The responsiveness of intrinsic subtypes to adjuvant anthracyclines versus nonanthracyclines in NCIC.CTG MA.5 randomized trial. *J Clin Oncol* 2011;**29**(Suppl.):A1032.

118. Ebbert M, Bastien R, Rowe L, Miller P, Anderson D, Boucher KM, *et al.* PAM50 breast cancer intrinsic classifier: clinical validation of a multianalyte laboratory developed test. *J Clin Oncol* 2011;**29**(Suppl.):A10597.

119. Bernard PS, Davis C, Munarriz B. Determining agreement between immunohistochemistry and RT-qPCR for standard biomarkers in breast cancer: validation on GEICAM 9906 clinical trial. *J Clin Oncol* 2011;**29**(Suppl.):A611.

120. Martin M, Rodriguez-Lescure A, Stijleman I, Munarriz B, Ruiz-Borrego M, Davis C, *et al.* PAM50 proliferation index predicts response to weekly adjuvant paclitaxel in node-positive operable breast cancer; 2012. Unpublished.

121. Chia SK, Bramwell VH, Tu D, Shepherd LE, Jiang S, Vickery T, *et al.* A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. *Clin Cancer Res* 2012;**18**:4465–72.

122. Jerevall PL, Ma X-J, Li H, Salunga R, Kesty NC, Erlander MG, *et al.* Prognostic utility of HOXB13:IL17BR and molecular grade index in early-stage breast cancer patients from the Stockholm trial. *Br J Cancer* 2011;**104**:1762–9.

123. Randox Laboratories. Information provided by manufacturer to the National Institute for Health and Care Excellence Diagnostic Assessment programme. 2011.

124. Bartlett JM, Thomas J, Ross DT, Seitz RS, Ring BZ, Beck RA, *et al.* Mammostrat as a tool to stratify breast cancer patients at risk of recurrence during endocrine therapy. *Breast Cancer Res* 2010;**12**:R47.

125. Ring B, Seitz R, Beck R, Shasteen W, Tarr S, Cheang M, *et al.* Novel prognostic immunohistochemical biomarker panel for estrogen receptor-positive breast cancer. *J Clin Oncol* 2006;**24**:3039–47.

126. Ross D, Kim C-Y, Tang G, Bohn O, Beck R, Ring B, *et al.* Chemosensitivity and stratification by a five monoclonal antibody immunohistochemistry test in the NASBP B14 and B20 trials. *Clin Cancer Res* 2008;**14**:6602–9.

127. Bartlett J. Biomarkers and patient selection for PI3K/Akt/mTOR targeted therapies: current status and future directions. *Clin Breast Cancer* 2010;**10**(Suppl. 3):S86–95.

128. Nottingham Prognostics. *NPI+*. Information provided by the manufacturers to NICE. 2011.

129. National Cancer Comprehensive Network (NCCN). *NCCN Clinical Practice Guidelines in Oncology. Breast Cancer. V.2.2010*. URL: www.nccn.org/professionals/physician_gls/PDF/breast.pdf.

130. Drummond M, Jefferson T. Guidelines for authors and peer reviewers of economic submissions to the *BMJ*. The *BMJ* Economic Evaluation Working Party. *BMJ* 1996;**313**:275–283.

131. Eddy DM. Technology assessment: The role of mathematical modelling. In *Assessing Medical Technology*. Washington, DC: National Academy Press; 1985. pp. 144–54.

132. Chen E, Tong KB, Malin JL. Cost-effectiveness of 70-gene MammaPrint signature in node-negative breast cancer. *Am J Manag Care* 2010;**16**:E333–42.

133. Retel VP, Joore MA, Knauer M, Linn SC, Hauptmann M, van Harten WH. Cost-effectiveness of the 70-gene signature versus St. Gallen guidelines and Adjuvant Online for early breast cancer. *Eur J Cancer Suppl* 2010;**8**:1382–91.

134. Medical Advisory Secretariat. *Gene expression profiling for guiding adjuvant chemotherapy decisions in women with early breast cancer: an evidence-based and economic analysis*. Ont Health Technol Assess Ser (Internet); December 2010. URL: www.health.gov.on.ca/english/providers/program/mas/tech/reviews/pdf/gep_20101213.pdf (last accessed September 2011).

135. Tsoi DT, Inoue M, Kelly CM, Verma S, Pritchard KI, Tsoi DT, *et al.* Cost-effectiveness analysis of recurrence score-guided treatment using a 21-gene assay in early breast cancer. *Oncologist* 2010;**15**:457–65.

136. Paulden M, Franek J, Pham B, Krahn M. *Gene expression profiling for guiding adjuvant chemotherapy decisions in women with early breast cancer: a cost-effectiveness analysis of 1000 strategies for the provision of Adjuvant! Online, Oncotype DX and adjuvant chemotherapy*; 2011. URL: http://theta.utoronto.ca/papers/theta_cadth2011_paulden-2.pdf (last accessed September 2011).

137. Hornberger J, Lyman GH, Chien R, Hornberger J, Lyman GH, Chien R. Economic implications of 21-gene recurrence score assay: US multicenter experience. *J Clin Oncol* 2010;**28**:e382.

138. Liaropoulos L, Galanis P, Kaitelidou D, Rovithis D, Siskou O, Theodorou P. Gross financial evaluation of the application of the 21-gene assay Oncotype DX for adjuvant chemotherapy decision-making in estrogen receptor-positive node negative breast cancer in Greece. *Breast* 2011;**20**:S44.

139. Lopes LF, Bacchi CE, Lopes LF, Bacchi CE. HER-2 status in gastrointestinal stromal tumor. *Ann Diagn Pathol* 2008;**12**:283–5.

140. Tsoi DT, Inoue M, Kelly CM, Verma S, Pritchard K. Cost-effectiveness analysis of Oncotype DX-guided treatment in early breast cancer. *J Clin Oncol* 2009;**27**:e11536.

141. Klang SH, Hammerman A, Liebermann N, Efrat N, Doberne J, Hornberger J. Economic implications of 21-gene breast cancer risk assay from the perspective of an Israeli-managed health-care organization. *Value Health* 2010;**13**:381–7.

142. HAYES, Inc. *Oncotype DX for prognosis of breast cancer recurrence* (structured abstract). Lansdale, PA: HAYES, Inc.; 2010.

143. Bacchi CE, Prisco F, Carvalho FM, Ojopi EB, Saad ED, Bacchi CE, *et al.* Potential economic impact of the 21-gene expression assay on the treatment of breast cancer in Brazil. *Rev Assoc Med Bras* 2010;**56**:186–91. URL: www.oncologypractice.com/fileadmin/content_images/tor/issue_archives/2010/ONC_JAN_FEB_2010_lores.pdf.

144. Jancin B. Oncotype DX can predict risk in node-positive patients. *Oncol Rep* 2010;(January–February):13.

145. Annabi BL. A MT1-MMP/NF-B signaling axis as a checkpoint controller of COX-2 expression in CD133(+) U87 glioblastoma cells. *J Neuroinflammation* 2008;**6**:8.

146. Retel VP, Joore MA, Knauer M, Linn SC, Hauptmann M, Harten WH, *et al.* Cost-effectiveness of the 70-gene signature versus St Gallen guidelines and Adjuvant Online for early breast cancer. *Eur J Cancer* 2010;**46**:1382–91.

147. Lidgren M, Jonsson B, Rehnberg C, Willking N, Bergh J. Cost-effectiveness of HER2 testing and 1-year adjuvant trastuzumab therapy for early breast cancer. *Ann Oncol* 2008;**19**:487–95.

148. Lidgren M, Willking N, Jonsson B, Rehnberg C. Health related quality of life in different states of breast cancer. *Qual Life Res* 2007;**16**:1073–81.

149. National Cancer Institute. *Surveillance, Epidemiology and End Results cancer statistics*. May 2009. URL: http://seer.cancer.gov/statistics/ (last accessed 22 December 2009).

150. Early Breast Cancer Trialists' Collaborative Group. Polychemotherapy for early breast cancer: an overview of the randomised trials. *Lancet* 1998;**352**:930–42.

151. Earle CC, Chapman RH, Baker C, Bell CM, Stone PW, Sandberg EA, *et al.* Systematic overview of cost–utility assessments in oncology. *J Clin Oncol* 2000;**18**:3302–17.

152. Hassett MJ, O'Malley AJ, Pakes JR, Newhouse JP, Earle CC. Frequency and cost of chemotherapy-related serious adverse effects in a population sample of women with breast cancer. *J Natl Cancer Inst* 2006;**98**:1108–17.

153. Ara RM, Brazier JE. Using health state utility values from the general population to approximate baselines in decision analytic models when condition-specific data are not available. *Value Health* 2011;**14**:539–45.

154. Hillner BE, Smith TJ. Efficacy and cost effectiveness of adjuvant chemotherapy in women with node-negative breast cancer. *N Engl J Med* 1991;**324**:160–8.

155. Partridge AH, Burstein HJ, Winer EP. Side effects of chemotherapy and combined chemohormonal therapy in women with early-stage breast cancer. *J Natl Cancer Inst Monogr* 2001;**30**:135.

156. Office for National Statistics. 2007-2009 *Interim Life Tables*. 2010. URL: www.statistics.gov.uk/STATBASE/Product.asp?vlnk=14459 (last accessed on 6 April 2011).

157. Thomas R, Williams M, Marshall C, *et al.* The total hospital and community UK costs of managing patients with relapsed breast cancer. *Br J Cancer* 2009;**100**:598–600.

158. Milne R, Heaton-Brown KH, Hansen P, Thomas D, Harvey V, Cubitt A. Quality-of-life valuations of advanced breast cancer by New Zealand women. *Pharmacoeconomics* 2006;**24**:281–92.

159. Peasgood T, Ward SE, Brazier J. Health-state utility values in breast cancer. *Exp Rev Pharmacoecon Outcomes Res* 2011;**10**:553–66.

160. Conner-Spady BL, Cumming C, Nabholtz JM, Jacobs P, Stewart D. A longitudinal prospective study of health-related quality of life in breast cancer patients following high-dose chemotherapy with autologous blood stem cell transplantation. *Bone Marrow Transplant* 2005;**36**:251–9.

161. British Medical Association and Royal Pharmaceutical Society of Great Britain. *British national formulary.* No. 61, March 2011. London: BMA and RPS; 2011.

162. Tappenden P, Jones R, Paisley S, Carroll C. Systematic review and economic evaluation of bevacizumab and cetuximab for the treatment of metastatic colorectal cancer. *Health Technol Assess* 2007;**11**(12).

163. Hind D, Ward S, De Nigris E, Simpson E, Carroll C, Wyld L. Hormonal therapies for early breast cancer: systematic review and economic evaluation. *Health Technol Assess* 2007;**11**(26).

164. Department of Health. *NHS reference costs* 2009–2010. London: Department of Health; 2010.

165. Wolowacz SE, Cameron DA, Tate HC, Bagust A. Docetaxel in combination with doxorubicin and cyclophosphamide as adjuvant treatment for early node-positive breast cancer: a cost-effectiveness and cost–utility analysis. *J Clin Oncol* 2008;**26**:925–33.

166. Blohmer JU, Kuhn T, Rezai M, Kummel S, Warm M, Eggemann H, *et al.* German multicentre decision impact study of Oncotype DX recurrence score (RS) on adjuvant treatment in estrogen receptor positive (ER plus) node negative (N0) and node positive (N plus) early breast cancer. *Breast* 2011;**20**:S46.

167. Demichell R. Time distribution of the recurrence risk for breast cancer patients undergoing mastectomy: further support about the concept of tumor dormancy. *Breast Cancer Res Treat* 1996;**41**:177–85.

168. Sacco JJ, Botten J, Macbeth F, Bagust A, Clark P. The average body surface area of adult cancer patients in the UK: a multicentre retrospective study. *Plos One* 2010;**5**:e8933.

169. Clarient. *Mammostrat*. Information provided by the manufacturers to NICE. 2011.

170. Rosman M, Mylander WC, Tafra L, Rosman M, Mylander WC, Tafra L. What is the value of the 21 gene recurrence score in HER2-negative patients? *J Clin Oncol* 2010;**28**:e647.

171. ARUP Laboratories Inc. *PAM50*. Information provided by the manufacturers to NICE. 2011.

172. Breast Cancer Clinical Outcome Measures (BCCOM) Project. *Analysis of the management of symptomatic breast cancers diagnosed in 2002*; 2006. URL: www.wmpho.org.uk/wmciu/documents/BCCOM_annual_report.pdf (last accessed September 2011).

173. Albanell J, Colomer R, Ruiz-Borrego M, Alba E, Saenz JAG, Martin M, *et al.* Prospective trans-GEICAM study of the impact of the 21-gene recurrence score assay and traditional clinico-pathological factors on clinical decision making in women with estrogen receptor-positive, HER2-negative, node-negative breast cancer. *Breast* 2011;**20**:S43.

174. Gregg X, Belnap T, Rowley B, Rees W. Experience with use of the Oncotype DX gene assay test in a multicenter community-based healthcare system. *Cancer Res* 2009;**69**(Suppl. 3):842S.

175. Department of Health. *NHS reference costs 2010–2011*. London: Department of Health; 2011.

176. Campbell HE, Epstein D, Bloomfield D, Griffin S, Manca A, Yarnold J. *et al.* The cost-effectiveness of adjuvant chemotherapy for early breast cancer: a comparison of no chemotherapy and first, second and third generation regimens for patients with differing prognoses. *Eur J Cancer* 2011;**47**:2517–30.

177. Avon, Somerset and Wiltshire NHS Cancer Services. *Fluorouracil, epirubicin and cyclophosphamide (FEC75) for early breast cancer*. Avon, Somerset and Wiltshire NHS Cancer Services; 2011. URL: www.avon.nhs.uk/aswcs-chemo/STCP/Breast%202009and2010/FEC75%20ASWCS09%20BR001-Final.pdf (last accessed August 2011).

178. Curtis L. *Unit costs of health and social care 2010*. Canterbury: PPSRU, University of Kent; 2010. URL: www.pssru.ac.uk/pdf/uc/uc2010/uc2010.pdf (last accessed September 2011).

179. Roche H, Fumoleau P, Spielmann M, Canon J, Delozier T, Serin D, *et al.* Sequential adjuvant epirubicin-based and docetaxel chemotherapy for node-positive breast cancer patients: the FNCLCC PACS 01 trial. *J Clin Oncol* 2006;**24**:5664–71.

180. Praga C, Bergh J, Bliss J, Bonneterre J, Cesana B, Coombes RC, *et al.* Risk of acute myeloid leukemia and myelodysplastic syndrome in trials of adjuvant epirubicin for early breast cancer: correlation with doses of epirubicin and cyclophosphamide. *J Clin Oncol* 2005;**23**:4179–91.

181. Edlin R, Connock M, Round J, Tubeuf S, Fry-Smith A, Hyde C, *et al. Azacitidine for the treatment of myelodyplastic syndrome, chronic myelomonocytic leukaemia, and acute myeloid leukaemia. A single technology appraisal*. West Midlands Health Technology Assessment Collaboration. 2009. URL: www.nice.org.uk/nicemedia/live/12036/45019/45019.pdf (last accessed September 2011).

182. Karnon J, Kerr GR, Jack W, Papo NL, Cameron DA. Health care costs for the treatment of breast cancer recurrent events: estimates from a UK-based patient-level analysis. *Br J Cancer* 2007;**97**:479–85.

183. Younis T, Rayson D, Skedgel C. Adjuvant chemotherapy in breast cancer: is TC a cost-effective regimen compared to AC? 2008. URL: www.abstracts2view.com/sabcs/viewp.php?nu=p6110 (last accessed September 2011).

184. Office for National Statistics. *Cancer registration statistics England 2007*. Series MB1 no. 38; 2010. URL: www.statistics.gov.uk/StatBase/Product.asp?vlnk=7720&Pos=1&ColRank=1&Rank=272 (last accessed September 2011).

185. de Bock G, Putter H, Bonnema J, van der Hage J, Bartelink H, van de Velde C. The impact of loco-regional recurrences on metastatic progression in early-stage breast cancer: a multistate model. *Breast Cancer Res Treat* 2009;**117**:401–8.

186. Bueno-de-Mesquita JM. Use of 70-gene signature to predict prognosis of patients with node-negative breast cancer: a prospective community-based feasibility study (RASTER). *Lancet Oncol* 2007;**8**:1079–87.

187. MINDACT trial. *Microarray In Node negative Disease may Avoid ChemoTherapy. A prospective, randomised study comparing the 70-gene expression signature with common clinical-pathological criteria in selecting patients for adjuvant chemotherapy in node-negative breast cancer. 2011*. URL: http://www.eortc.be/services/unit/mindact/documents/MINDACT_trial_outline.pdf (last accessed 11 September 2012).

188. TAILORx trial. *Hormone therapy with or without combination chemotherapy in treating women who have undergone surgery for node-negative breast cancer (the TAILORx trial)*. 2011. URL: http://clinicaltrials.gov/ct2/show/study/NCT00310180?term=oncotype+dx&rank=3 (last accessed 11 September 2012).

189. Early Breast Cancer Trialists' Collaborative Group. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005;**365**:1687–717.

190. Schor A, Carvalho F, Ojopi E, Mangabeira A, Saad E, Bacchi C. 21-gene recurrence score (RS) in 100 consecutive Brazilian Patients. *Cancer Res* 2009;**69**:842S–843S.

191. Bighin C, Del Mastro L, Canavese G, Carli F, Taveggia P, Levaggi A, *et al.* Use in current clinical practice of 70-gene signature in early breast cancer. *Int J Cancer* 2010;**127**:2736–7.

192. Espinosa E, Sanchez-Navarro I, Gamez-Pozo A, Marin AP, Hardisson D, Madero R, *et al.* Comparison of prognostic gene profiles using qRT-PCR in paraffin samples: a retrospective study in patients with early breast cancer. *Plos One* 2009;**4**:5911.

193. Ma XJ, Salunga R, Dahiya S, Wang W, Carney E, Durbecq V, *et al.* A five-gene molecular grade index and HOXB13:IL17BR are complementary prognostic factors in early stage breast cancer. *Clin Cancer Res* 2008;**14**:2601–8.

194. European Organisation for Research and Treatment of Cancer. MINDACT (Microarray In Node negative Disease may Avoid ChemoTherapy). A prospective, randomised study comparing the 70-gene expression signature with common clinical-pathological criteria in selecting patients for adjuvant chemotherapy in node-negative breast cancer. 2011. URL: www.eortc.be/services/unit/mindact/documents/MINDACT_trial_outline.pdf (last accessed September 2011).

Appendix 1

Search strategy

Update search for OncotypeDX and MammaPrint

Date limits: January 2009–May 2011
Filter: human studies only

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 medullary)).mp.
9. 1 or 2 or 3 or 6 or 7 or 8
10. MammaPrint.mp.
11. 70-gene.mp.
12. gene70.mp.
13. gene?seventy.mp.
14. seventy?gene.mp.
15. amsterdam profile.mp.
16. Oncotype.mp.
17. Oncotype DX.mp.
18. 21-gene.mp.
19. gene21.mp.
20. gene?twentyone.mp.
21. twentyone?gene.mp.
22. GHI Recurrence score.mp.
23. GHI-RS.mp.
24. 92-gene.mp.
25. gene92.mp.
26. gene?ninetytwo.mp.
27. ninetytwo?gene.mp.
28. RT-PCR (adj 5) 21.mp.
29. or/10–28
30. 9 and 29

Search for Randox Breast Cancer Array, BluePrint, PAM50, Breast Cancer Index, IHC4, Mammostrat, and NPI+

Date limits: 2002–May 2011
Filter: human studies only

31. exp Breast Neoplasms/
32. exp mammary neoplasms/
33. exp "Neoplasms, Ductal, Lobular, and Medullary"/
34. exp breast/
35. exp neoplasms/
36. 4 and 5
37. (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.
38. (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 medullary)).mp.
39. 1 or 2 or 3 or 6 or 7 or 8
40. Randox.mp.
41. Blueprint.mp.
42. 80-gene.mp.
43. gene80.mp.
44. gene?eighty.mp.
45. eighty?gene.mp.
46. PAM50.mp.
47. 50-gene.mp.
48. gene50.mp.
49. gene?fifty.mp.
50. fifty?gene.mp.
51. breast bioclassifier.mp.
52. Breast Cancer Index.mp.
53. Breast cancer gene expression ratio.mp.
54. 2-gene.mp.
55. Two-gene-index.mp.
56. 2-gene-index.mp.
57. Two?gene.mp.
58. gene?two.mp.
59. H?I.mp.
60. H:I.mp.
61. 5-gene.mp.
62. gene5.mp.
63. gene?five.mp.
64. five?gene.mp.
65. 7-gene.mp.
66. seven-gene.mp.
67. gene7.mp.
68. gene?seven.mp.
69. Theros.mp.
70. Biotheranostics.mp.
71. Theros breast cancer index.mp.
72. HOXB13$.mp.
73. homeobox?13$.mp.
74. interleukin?17B$.mp.
75. IL17BR.mp.
76. mammostrat.mp.

77. five-biomarker-assay.mp.
78. IHC4.mp.
79. NPI+.mp.
80. Nottingham prognostic index plus.mp.
81. Nottingham prognostic index +.mp.
82. or/10–51
83. 9 and 52

Appendix 2

Example of the quality assessment checklist applied to included studies

A framework for assessing the internal validity of articles describing prognostic factor studies

Study feature	Qualities sought
Sample of patients	Inclusion criteria defined
	Sample selection explained
	Adequate description of diagnostic criteria
	Clinical and demographic characteristics fully described
	Representative
	Assembled at a common (usually early) point in the course of their disease
	Complete
Follow-up of patients	Sufficiently long
Outcome	Objective
	Unbiased (e.g. assessment blinded to prognostic information)
	Fully defined
	Appropriate
	Known for all or a high proportion of patients
Prognostic variable	Fully defined, including details of method of measurement if relevant
	Precisely measured
	Available for all or a high proportion of patients
	If relevant, cut-off point(s) defined and justified
Analysis	Continuous predictor variable analysed appropriately
	Statistical adjustment for all important prognostic factors
Intervention subsequent to inclusion in cohort	Fully described
	Intervention standardised or randomised

Source: Altman *et al.*[37]

Appendix 3

Assessment of multiple systematic reviews (AMSTAR): a measurement tool to assess the methodological quality of systematic reviews

	Marchionni *et al.*[33]	Smartt[34]
1. Was an 'a priori' design provided? The research question and inclusion criteria should be established before the conduct of the review	Yes	Yes
2. Was there duplicate study selection and data extraction? There should be at least two independent data extractors and a consensus procedure for disagreements should be in place	Yes	Unclear
3. Was a comprehensive literature search performed? At least two electronic sources should be searched. The report must include years and databases used (e.g. CENTRAL, EMBASE and MEDLINE). Keywords and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialised registers or experts in the particular field of study, and by reviewing the references in the studies found	Yes	Yes
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion? The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review) based on their publication status, language, etc.	Yes	Unclear
5. Was a list of studies (included and excluded) provided? A list of included and excluded studies should be provided	Yes	Yes (only for included studies)
6. Were the characteristics of the included studies provided? In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all of the studies analysed, for example age, race, sex, relevant socioeconomic data, disease status, duration, severity or other diseases, should be reported	Yes	Yes
7. Was the scientific quality of the included studies assessed and documented? 'A priori' methods of assessment should be provided [e.g. for effectiveness studies if the author(s) chose to include only randomised, double-blind or placebo-controlled studies, or allocation concealment, as inclusion criteria; for other types of studies alternative items will be relevant]	Yes	Yes
8. Was the scientific quality of the included studies used appropriately in formulating conclusions? The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations	Yes	Yes
9. Were the methods used to combine the findings of studies appropriate? For the pooled results, a test should be carried out to ensure that the studies were combinable, to assess their homogeneity (i.e. chi-squared test for homogeneity, I^2). If heterogeneity exists a random-effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e. is it sensible to combine?)	Yes	Yes
10. Was the likelihood of publication bias assessed? An assessment of publication bias should include a combination of graphical aids (e.g. funnel plots, other available tests) and/or statistical tests (e.g. Egger regression test)	No	No
11. Was the conflict of interest stated? Potential sources of support should be clearly acknowledged in both the systematic review and the included studies	Yes	Yes

Source: Shea *et al.*[38]

Appendix 4

Summary of evidence relating to OncotypeDX

Summary of evidence relating to OncotypeDX reported in the Marchionni *et al.* systematic review[33]

Analytical validity	Clinical validity	Clinical utility
Reported in four studies.[39–42] Technical and operational aspects were reported in two studies[39,40] and test and assay variability and reproducibility were reported in three studies.[40–42] Conclusion: Preanalytic issues relating to sample storage and preparation appeared to play a larger role than within-laboratory variation Six studies reported overall success rate,[41–44,48,49] which ranged from 78.9% to 98.9% Not all of the studies provided detailed descriptions of the reasons for assay failure. When failures were reported they were mainly ascribed to an insufficient number of cancer cells in the specimens, poor RNA quality and, in a few cases, failure of the RT-PCR technique Systematic review conclusion: Evidence existed for some of the operational characteristics of this test but there was limited evidence for the reproducibility of the test. Reasonable reproducibility of the test across different samples of the same block, and samples from different blocks. No direct evidence was available about the effect of sample preparation. There was indirect evidence that the overall success rate of extracting analysable mRNA was fairly high. Centralisation was considered to be a current strength of OncotypeDX with regard to reproducibility	Reported in four studies in relation to the determination of recurrence risk (prognosis) Paik *et al.*[42] studied the prognostic validity of OncotypeDX in an independent tamoxifen-treated population. The RS was shown to be significantly correlated with DFS ($p<0.001$) and OS ($p<0.001$). RS alone was a better predictor of distant recurrence at 10 years than traditional clinicopathological predictors Esteva *et al.*[44] failed to find a correlation between RS and distant breast cancer recurrence in untreated node-negative (LN0) patients. In the reverse of what was expected, well-differentiated tumours were correlated with poorer survival than higher-grade tumours Cobleigh *et al.*[43] reported that the RS score was significantly correlated with DRFS in a training set of LN0 patients. As this data set related to training and not validation, it was considered to present minimal evidential value Habel *et al.*[41] assessed the risk of breast cancer-specific mortality among women in a large case–control study of ER+, LN0 breast cancer patients treated with tamoxifen. The 10-year risk of death from breast cancer was 3% for patients with a low RS, 12% for patients with an intermediate RS and 27% for patients with a high RS. Multivariate analysis showed that RS and tumour size were independent risk predictors of breast cancer death in ER+, tamoxifen-treated patients (RR (relative risk) for RS (risk score) per 50 units = 7.6, $p<0.001$) and untreated patients (RR (relative risk) for RS (risk score) per 50 units = 4.1, $p<0.001$). The RS score also showed some prognostic value in ER– patients Three posters describing studies that compared risk predictions provided by OncotypeDX assays and standard risk classification methods were reported.[45,46,47] The data presented in these posters suggested that optimal predictions may come from a combination of gene expression tests and standard risk assessment methods Systematic review conclusion: Fairly strong support for the clinical validity of the OncotypeDX test over and above standard clinical predictors in ER+, LN0 and tamoxifen-treated patients with a clear treatment indication for adjuvant chemotherapy. The authors noted, however, that it was not clear (1) how much the test added to the management of patients, (2) what proportion of patients would benefit from the use of the OncotypeDX test and (3) the stability of the observed risk categories in other populations, particularly those treated with current therapies	No published studies reported demonstrating clinical utility (direct evidence) Two studies reported that provided preliminary evidence of the potential predictive power of OncotypeDX (indirect evidence) Paik *et al.*,[49] using specimens and data from an existing trial (NSABP B20), addressed the potential value of the RS in predicting chemotherapy benefit in a population of ER+, LN0 patients. This study compared a group of patients treated with tamoxifen and chemotherapy with a group of patients who were randomised to tamoxifen only. The RS was found to be correlated with chemotherapy benefit, defined in terms of 10-year DRFS, with a significant benefit from the use of chemotherapy in the high RS group ($p=0.001$). However, in a multivariate analysis the benefit from chemotherapy was unclear because of large CIs in the low and intermediate RS risk groups Oratz *et al.*[48] reported that knowledge of the RS changed the clinicians' treatment recommendations for 21% of patients and the actual administered treatment for 25% of patients. They did not report what the patients (or doctors) were told or understood about the risk of recurrence Systematic review conclusion: the Paik *et al.* study[49] represented the strongest evidence derived from already existing data regarding the clinical utility of the OncotypeDX test. This study also noted that, although prospective confirmation of these findings was required, the evidence provided reasonable justification in the interim for the use of the test by women in this specific population

Systematic review summary

The studies assessed in this review were heterogeneous in focus and quality. Few of the publications addressed technical aspects of the tests. A number of the reports focused on prognostic prediction. Only one study examined the prediction of treatment benefit. Most of the published evidence available for OncotypeDX was obtained using the marketed assay. Overall, the evidence presented for the clinical validity of OncotypeDX/21-gene signature in the systematic review was considered to have provided fairly strong support for the clinical performance of the test compared with standard predictors in a well-defined population (ER+, LN0, tamoxifen-treated women). It was considered that there was strong enough evidence of the clinical utility of the test in retrospectively collected data from one large clinical trial to provide reasonable justification for the interim use of the test in women in the same population group as the trial patients. There was little information about the impact of the test on clinical decision-making

Summary of evidence relating to OncotypeDX reported in the Smartt systematic review[34]

Clinical validity	Clinical utility
Two studies reported on the clinical validity of the test[50,51] The purpose of the Goldstein *et al.* study[49] was to evaluate the prognostic value of OncotypeDX in hormone receptor-positive, LN0 or LN+ patients and to determine whether or not it could better predict outcome at 5 years than a modified Adjuvant! Online algorithm. The 21-gene assay was a more accurate predictor of relapse than standard clinical features for individual patients with hormone receptor-positive operable breast cancer treated with chemotherapy/hormonal therapy and provides information that is complementary to features typically used in anatomic staging, such as tumour size and LN involvement. The 21-gene assay may be used to select low-risk patients for abbreviated chemotherapy regimens similar to those used in our study or high-risk patients for more aggressive regimens or for clinical trials In the Wolf *et al.* study[51] the authors sought to assess the correlation between standard clinical and pathological breast cancer characteristics and the RS in a cohort of Israeli breast cancer patients and to compare the stratification of patients using RS with that of commonly used clinical guidelines. High tumour grade, low PR expression, infiltrating ductal histology and HER2 overexpression were found to be associated with a high RS. Patient age, tumour size, ER expression, and LN micrometastasis were found to correlate poorly with the RS. The ability of any of these variables, either alone or in combination, to predict the RS was limited. Similarly, none of the guidelines nor the Adjuvant! Online software could predict the RS. This study reported on a selected population of patients who were referred to undergo the OncotypeDX test. No association was noted between the RS and patient age or ER intensity and only a modest association was noted between the RS and tumour size. The clinical utility of these comparisons was not made clear Summary of reported conference abstracts: Shak *et al.*[59] demonstrated that the distribution of RS was similar for men and women with breast cancer	Four studies reported on the clinical utility (indirect evidence) of the test[52–55] The purpose of the Asad *et al.* study[52] was to determine whether or not the results of OncotypeDX influence the decision to administer chemotherapy. The OncotypeDX results influenced the decision for chemotherapy in 37 (44%) patients; four patients classified as low risk by the NCCN guidelines[129] (tumours < 1 cm) were advised to have chemotherapy and 33 patients classified as high risk by the NCCN guidelines (tumours ≥ 1 cm) were advised to undergo hormone treatment only. The authors concluded that the OncotypeDX RS is significantly related to tumour grade and HER2/neu status. **Comment:** There was no evidence that OncotypeDX changed clinical outcomes The Henry *et al.* study[53] reported on the functional and clinically relevant impact of the RS on the adjuvant therapy administered to 29 patients with ER+, LN0 breast cancer, as well as its influence on a panel of five expert breast oncologists. They concluded that the RS contributed to chemotherapy changes in 31% of patients, with more changes made against than for adjuvant chemotherapy. The RS increased consensus recommendations by 10% but did not appear to increase the reported strength of panellists' recommendations. **Limitations:** The small sample size increased the likelihood of a type 2 error (false-negative result) and the study lacked statistical power to draw definitive conclusions. Determination of therapy received was retrospective and may have been subject to the well-established biases (e.g. selection bias, information bias) associated with this methodology. Panellists were the same medical oncologists who administered chemotherapy and panellists may have remembered their recommendations from when they were actually managing these patients. The 2-month washout period may have been insufficient to erase all recollections of previous recommendations (recall bias). Although the RS predicts only distant relapse, Adjuvant! Online includes distant and local relapse, thus the estimate of recurrence for Adjuvant! Online was much higher (90%) than that for the RS and the chemotherapy decision for 54% of patients was changed with RS information. One patient was male Li *et al.*[54] hypothesised that an integrated gene expression profile could a predict patient's response to chemotherapy. The main purpose of this study was the validation of a new gene signature, which overlapped in part with OncotypeDX and the 70-gene signature. The authors reported that their integrated signature was a stronger prediction of chemotherapy outcome than the single signatures (OncotypeDX and the 70-gene signature). **Comment:** Neither OncotypeDX nor the 70-gene signature formed the main focus of this study. Both signatures were used in populations that were very different from those that the tests were validated for. The follow-up was short The purpose of the Rayhanabad *et al.* study[55] was to examine the utility of OncotypeDX in the prediction of recurrence and the degree of benefit from chemotherapy. Treatment received after OncotypeDX testing was compared with treatment based on NCCN guidelines.[129] A total of 13 out of 18 high-risk NCCN, low-risk RS patients did not receive chemotherapy ($p < 0.001$); 11 patients with an intermediate RS received chemotherapy. OncotypeDX results changed management in 15 (26%) patients ($p = 0.05$). The authors concluded that the use of gene assays altered recurrence risk stratification and the decision for chemotherapy in a significant number of patients. This allowed better individualised treatment for patients, reserving chemotherapy for those at high risk of recurrence, whereas low-risk patients were spared the morbidity associated with chemotherapy. However, Smartt reported that there were a number of serious limitations in this study, which threaten the validity of the reported results Summary of reported conference abstracts: Most studies reported examined or modelled the impact of the RS on clinical decision-making in relation to adjuvant chemotherapy. Erb *et al.*[56] reported a significant decline in the use of adjuvant chemotherapy after the introduction of the test in the authors' institution. Gold *et al.*,[57] reporting on how clinicians integrated RS into their decision-making, found that RS, tumour grade and size were all independent predictors of chemotherapy administration. Lo *et al.*[58] examined the effect of knowledge of the RS on both patients and medical oncologists in relation to their adjuvant therapy choice. In total, 22% of oncologists and 10% of patients changed from chemotherapy to hormone therapy. The change in the other direction (i.e. from hormone therapy to chemotherapy) occurred in 3% and 8% respectively

Systematic review summary

There were no additional studies reporting on the analytical validity of the test and this remains an area of weakness in the evidence story to date. In contrast to the studies reported in the original systematic review,[33] the majority of these studies primarily addressed questions relating to the clinical utility of OncotypeDX, some reported further on the clinical validity or validity and utility of the test and one study reported, for the first time, on the use of the test in male breast cancer. The additional studies reporting on the clinical validity of OncotypeDX further endorsed the advantages of the test compared with standard clinicopathological assessment of risk and extended the examination of its prognostic value beyond clinical trial populations to a general population, as well as the cohort of male breast cancer patients. The studies reporting on the clinical utility of the test examined its ability to predict response to treatment or its impact on clinical decision-making. The latter studies all reported a positive impact of the test on clinical decision-making and generally claimed that there was a reduction in the number of patients who were or would have been considered for chemotherapy. However, the studies generally had methodological weaknesses that were likely to have overestimated the effect/influence of the test and were not designed to assess the effect of the test on clinical outcomes. Studies examining the ability of OncotypeDX to predict response to adjuvant and neoadjuvant endocrine therapy and chemotherapy generally reported that OncotypeDX was predictive, to a greater or lesser extent, of response to therapy; however, as the design of the studies precluded any firm conclusions about the ability of the test to predict response to therapy, these studies did not materially add to the body of evidence in this area

Appendix 5

Summary of evidence relating to MammaPrint

Summary of evidence relating to MammaPrint reported in the Marchionni *et al.* systematic review[33]

Analytical validity	Clinical validity	Clinical utility
Two technical studies[60,61] provided evidence relating to the analytical validity of MammaPrint. Repeated gene expression measurements over time, within and across individual microarrays and across different laboratories, protocols, instruments and operators, provided data on the variability and reproducibility of the test. Buyse *et al.*[61] reported an overall success rate of the assay of 80.9% The systematic review concluded that the studies that used the 70-gene signature provided useful information about the validity of the biological correlations underlying the profile. However, although these studies suggested that MammaPrint could be used in a clinical setting, they could not be considered to be direct validations of the assay. The review also noted that evidence underpinning the analytical validity of the test was obtained from a limited number of patients and a moderate number of replications. The only validation study using the MammaPrint assay (rather than the underlying 70-gene signature) showed that only about 80% of fresh-frozen specimens were analysable	van't Veer *et al.*[63] reported on the development data for the 70-gene panel that formed the basis for the MammaPrint test. Using multivariate analysis, the 70-gene signature was found to be an independent predictor of metastases within 5 years, with an OR = 18 (95% CI 3 to 94) van de Vijver *et al.*[64] reported the first major validation of the 70-gene signature in a young (< 52 years) population with small (< 5 cm) tumours that were heterogeneous with respect to LN positivity, ER status, chemotherapy and tamoxifen treatment. Multivariate analysis showed that the MammaPrint prognosis group, tumour size and adjuvant chemotherapy were the strongest predictors of distant metastases. The 'poor prognosis' MammaPrint group had the largest HR (4.6, 95% CI 2.3 to 9.2). The authors demonstrated the prognostic value of the gene signature using survival curves stratified by conventional clinical indexes. The analyses showed substantial separation between 70-gene prognostic groups that were either low or high risk by clinical indices. **Optimal prediction was achieved when the gene index and conventional clinical predictors were combined** Buyse *et al.*[61] compared the MammaPrint assay with conventional clinical combination risk predictors in an independent, multicentre validation study. The specificity and sensitivity of the MammaPrint assay and the Adjuvant! Online algorithm were compared for prediction of distant metastases within 5 years and for death within 10 years. Similar sensitivities were found in both methods, but a higher specificity was demonstrated for MammaPrint. The areas under the receiver operating characteristic (ROC) curves were comparable for MammaPrint and Adjuvant! Online (0.68 vs. 0.66 for distant metastases at 5 years). However, with ROC values much closer to 0.50 than 1.00 neither prediction was particularly accurate Glas *et al.*[62] compared the commercial MammaPrint assay results with those obtained with a generic 70-gene signature test using the same patients as van't Veer and van de Vijver. The results of the 70-gene signature used in the original cohorts applied equally to the commercial MammaPrint assay based on the signature **Summary:** The authors concluded that, overall, the available published evidence supported MammaPrint as a better predictor of the 5-year risk of distant recurrence than traditional clinical predictors. However, the cohorts used for the development and validation of MammaPrint were considerably more clinically heterogeneous than those used for the OncotypeDX test. Despite this, MammaPrint had an 80% concordance with the OncotypeDX array-based RS classification when applied to the same patients. There was some evidence to suggest that the commercial MammaPrint test and a generic 70-gene signature assay produced comparable results	No studies on clinical utility were reported The systematic review did not identify any published studies evaluating the ability of the 70-gene signature or the commercial MammaPrint test to predict chemotherapy benefit

Systematic review summary

The review found studies that tested the MammaPrint assay, as well as studies about the 70-gene signature that the assay is based on. The studies that use the gene signature cannot be considered as validation of the assay itself. In terms of analytical validity, two recent papers looked at reproducibility between laboratories and found a good degree of agreement. RNA labelling emerged as a possible source of variation, and the question of reproducibility remains open. The only validation study using the MammaPrint assay itself showed that only 80% of fresh-frozen samples were useable, although it is hoped that the success rate would increase with the use of the assay. Studies of clinical validity overall show MammaPrint to be a better predictor of 5-year risk of distant recurrence than traditional algorithms and characteristics, although the validation and derivation cohorts were clinically more heterogeneous than those used for the OncotypeDX test. It remains to be seen how well it predicts in cohorts with greater homogeneity as used in the development of OncotypeDX. No studies that evaluated clinical utility were found

To conclude, the literature on the 70-gene signature includes numerous studies that focused more on its biological underpinning and less on the clinical implications of the gene expression profile. It is not yet clear which are the optimal patient populations for the use of this test, exactly what its performance is in those populations and how many of its predictions would result in different therapeutic decisions. Larger independent validation studies in therapeutically homogeneous groups are needed. Studies that test MammaPrint alongside standard predictors, develop the use of risk categories rather than a continuous scale and assess the assay's stability in different populations are also needed

Summary of evidence relating to MammaPrint reported in the Smartt systematic review[34]

Clinical validity	Clinical utility
Two studies on clinical validity were reported	One study on clinical utility (indirect) was reported
Mook *et al.*[65]	**Bueno-de-Mesquita *et al.*[67]**
Rational and objective: Patients with axillary LN metastases are generally considered to have a poor prognosis and most will be treated with adjuvant chemotherapy; however, up to 30% of these patients would remain free of distant metastases without adjuvant chemotherapy (Early Breast Cancer Trialists' Collaborative Group (EBCTCG),[189] 2005). In this study the authors sought to validate the prognostic value and accuracy of MammaPrint in an independent cohort of 241 patients with axillary LN metastases	*Rationale and objective:* In most hospitals tumour samples are routinely fixed in formalin and embedded in paraffin blocks. MammaPrint requires fresh tumour samples and one of the potential difficulties in the implementation of the test in daily clinical practice is the ease with which sample requirements can be met. In this prospective multicentre study the authors set out to evaluate (1) whether or not MammaPrint was suitable for use in routine clinical practice in the Netherlands, (2) the effect of the test on the use of adjuvant systemic treatment, (3) the proportion of patients with 'poor' compared with 'good' prognosis and (4) the concordance between risk predicted by MammaPrint and risk predicted by commonly used clinicopathological tools
Results: 41% of patients in the independent cohort ($n = 241$) had a good prognosis gene signature and 59% had a poor prognosis gene signature. There was a significant difference in DMFS (as the first event) and BCSS between the good and poor prognosis gene signature groups at both 5 and 10 years ($p < 0.001$). The poor prognosis signature group was associated with a shorter BCSS (HR 5.70; 95% CI 2.01 to 16.23; $p < 0.001$). The probability of distant metastases as the first event was significantly greater in the poor gene signature group (HR 4.13; 95% CI 1.71 to 9.96; $p = 0.002$)	*The patient population and eligibility criteria:* Patients were enrolled in this prospective multicentre study if they had unilateral primary operable invasive adenocarcinoma of the breast (TNM classification = T1–4, N0, M0) and were < 61 years of age. Sixteen participating Dutch hospitals contributed 812 women to the trial between 2004 and 2006. In total, 81 patients had breast-conserving surgery, 70% had small (< 2 cm) tumours, 81% had ductal histology, 80% had grade II–III tumours, 80% were ER+, 84% *ERBB2* negative and 85% LN–. Adjuvant systemic treatment varied: 39% of patients received no adjuvant treatment, 18% received chemotherapy, 13% received endocrine treatment and 29% received both chemotherapy and endocrine therapy. The median age of patients was 49 years and the median follow-up was 14 months (range 0.3–36.4 months). Hospitals were eligible to participate only if they had structured multidisciplinary breast cancer care, used standard operating procedures, treated at least 100 patients a year and had a dedicated physician as the local co-ordinator
In univariate analysis significant predictors of BCSS were the number of positive nodes, tumour grade, ER status, HER2 status endocrine treatment and MammaPrint risk group. Only the number of positive nodes, endocrine therapy and MammaPrint risk group remained significant predictors in multivariate analysis. MammaPrint was the most powerful independent predictor in this analysis (HR 7.17; 95% CI 1.81 to 28.43; $p = 0.005$)	*Endpoints and analyses:* Differences between MammaPrint and commonly used histopathological guidelines were assessed using Pearson's chi-squared test and the Cochrane–Armitage test for trends. The level of agreement between different risk assessment techniques was assessed using Cohen's kappa. In addition to MammaPrint, the CBO guidelines,[105] Adjuvant! Online, the NPI and the St Gallen guidelines were used to assess clinical risk. MammaPrint analyses were carried out blinded to clinical data and an initial recommendation for treatment using clinical criteria carried out before disclosure of the MammaPrint results
Predictors of DMFS in univariate analysis were the number of positive nodes, tumour size, histological grade, ER and HER2 status, endocrine therapy and MammaPrint risk group. Only endocrine therapy was a significant independent predictor of DMFS in multivariate analysis (HR 0.31, 95% CI 0.12 to 0.80, $p = 0.02$). MammaPrint risk group and number of positive nodes tended to be prognostic with HR = 2.99 (95% CI 0.996 to 8.99; $p = 0.051$) and HR = 2.29 (95% CI 0.99 to 5.29; $p = 0.053$) respectively	*Results:* Of the original 812 enrolled patients, 585 (72%) were eligible for the study. MammaPrint profiles were obtained in 427 (73%) of eligible patients. During follow-up five patients had distant metastases as the first event. According to MammaPrint, 51% of patients had a good prognosis signature compared with 57%, 31%, 58% and 17%, respectively, for the CBO,[105] Adjuvant! Online, NPI and St Gallen risk assessments
Adjuvant! Online classified 13% of patients as low risk and 87% as high risk; Adjuvant! Online and MammaPrint risk assessments were discordant for 77 patients (32%); 72 of these discordant patients were assessed as having a high risk of relapse by Adjuvant! Online and a good prognosis gene signature	
When 209 Adjuvant! Online high-risk patients were stratified by MammaPrint the 10-year BCSS probability was 94% for the good prognosis gene signature group and 76% for the poor prognosis gene signature group (HR 4.12; 95% CI 1.45 to 11.76; $p = 0.008$). Subgroup analysis suggested that MammaPrint was predictive for BCSS in patients in different treatment groups and patients with ER+ tumours	

Clinical validity	Clinical utility
The second cohort of 106 previously studied patients[64] (with one to three positive nodes) differed significantly from the independent cohort in terms of age (younger), axillary procedures, adjuvant systemic therapy and overall and median survival (10.3 years, range 1.6–21.2 years). The 10-year BCSS probability was 98% for the good prognosis gene profile and 64% for the poor prognosis gene profile. The poor prognosis signature was associated with shorter BCSS (HR 6.60; 95% CI 1.97 to 22.10; $p=0.002$) and a multivariate HR of 3.63 (95% CI 0.88 to 14.76; $p=0.07$)	Clinical and molecular risk assessments were discordant in 27%–39% of patients depending on the clinical assessment tool used. The amount of discordance between the clinical guidelines themselves was between 7% and 40%. Adjuvant treatment was recommended for 48% of patients based on the Dutch guideline alone; this increased to 62% when the guideline was used with the prognostic gene signature. Overall, and once patient preferences had been taken into account, adjuvant systemic treatment was administered to 61% of patients. An increase in systematic therapy occurred in patients whose risk according to the Dutch guidelines and MammaPrint were discordant. In the final analysis, 50 (12%) more patients received endocrine treatment, 54 (13%) patients had endocrine treatment added and 4 (1%) patients had endocrine treatment withheld. Sixteen (4%) more patients had chemotherapy, in 35 (8%) patients chemotherapy was added and it was withheld in 19 (4%) patients
Conclusion: MammaPrint predicted disease outcome better than traditional clinical prognostic factors in patients with one to three positive nodes and was able to accurately identify LN+ patients with an excellent prognosis. The potential clinical utility of MammaPrint was demonstrated in 72 (34%) clinically high-risk patients with a good prognosis signature who had a 10-year BCSS of 94% and therefore might be spared chemotherapy	*Limitations:* There was an early protocol change reducing the age of eligibility to < 55 years. It was not clear how representative the hospital sample was and the short follow-up time and low number of events precluded survival analyses
Wittner *et al.*[66]	*Quality:* This was a well-conducted prospective clinical trial that demonstrated the feasibility of conducting the MammaPrint test routinely in Dutch hospitals. As reported, the study fulfilled 35 of 44 (80%) REMARK criteria for the reporting of tumour marker prognostic studies indicating a high level of adherence to the reporting guidelines
Rationale and objectives: Most patients with breast cancer are older and present with smaller early-stage ER+ tumours than the cohorts of patients used to define and evaluate the MammaPrint gene signature. Decisions relating to the use of adjuvant chemotherapy in these older patients may be complicated by comorbidity. To explore these issues the authors carried out a retrospective evaluation of the prognostic value of MammaPrint in 100 older patients diagnosed and treated at the Massachusetts General Hospital (MGH) between 1985 and 1997. The study cohort of 100 patients was compared with the original Dutch cohort (NKI) of 151 LN0 patients used to validate the MammaPrint signature[64]	*Conclusion:* The study demonstrated a lack of congruence between well-known clinical guidelines for risk assessment in breast cancer. In approximately one-third of patients there was discordance between MammaPrint and clinical guidelines in the assessment of risk. The addition of MammaPrint to the standard Dutch clinical assessment of risk (modified by patient preference) increased by 20 the number of patients receiving adjuvant systemic therapy. However, although the study was able to demonstrate that MammaPrint had an impact on clinical decision-making the follow-up was not long enough to provide evidence of its effect on clinical end points such as DMFS or its utility in predicting treatment benefit
The patient population and eligibility criteria: Eligible MGH patients were consecutively diagnosed and treated patients with LN0 breast cancer and frozen primary tumour samples for whom histopathological and clinical information could be retrieved. The median age of the cohort was 62.5 years and the median duration of follow-up was 11.3 years (range 1.2–18.5 years). In total, 72% of patients had small tumours (≤ 2 cm), 94% were of histological grade II–III. A total of 21% of patients received chemotherapy and 24% hormonal therapy. Surgery included mastectomy (56%) and breast conservation (44%)	One study published as a conference abstract reported on clinical utility
Results: The MGH cohort was significantly older ($p<0.001$) than the original MammaPrint cohort.[64] There were also significant differences ($p<0.005$) in tumour size, histological grade and the proportion of patients undergoing systemic treatment	**Bender *et al.*[68]**
MammaPrint classified 27% of the MGH patients as low risk and 73% as high risk of distant metastases as the first event. The cohort had a significantly lower event rate than the original NKI cohort ($p<0.001$); there was no difference in OS in the older MGH cohort because of death from other causes. Survival analysis discriminated between the high- and low-risk gene signature with non-overlapping CIs; however, because of the low event rate the difference was not significant. This contrasted with the significant difference between the low- and high-risk groups reported for the original Dutch NKI cohort	In this study the authors present the results of a meta-analysis of 1637 patients with MammaPrint outcomes (T1–2, LN–/+ invasive breast cancer and median follow-up 7.1 years) to determine the chemotherapy benefit of patients treated with adjuvant chemotherapy in addition to endocrine therapy. Patient samples were recruited from seven large data sets from multiple institutions across Europe
	MammaPrint assigned 772 patients (47%) to a low-risk category and 865 (53%) to a high-risk category. In total, 349 patients (21%) were treated with endocrine therapy and 226 (14%) were treated with chemotherapy and endocrine therapy. In patients with a poor prognosis MammaPrint profile the 5-year DMFS improved from 69% to 88% (HR 0.28, 95% CI 0.14 to 0.56, $p<0.001$) when chemotherapy was added to hormone therapy. In multivariate analysis patients classified by MammaPrint as having good prognosis had no significant benefit from chemotherapy ($p=0.962$)

Clinical validity	Clinical utility
The NPV of MammaPrint in the MGH cohort was 100% (overall and at 5 and 10 years) compared with 88% in the original NKI cohort. The PPV was only 12% in the MGH cohort (because of the large number of patients classified as high risk who did not have distant metastases as the first event) compared with 52% in the NKI cohort. Sensitivity analysis varying the cut-off/classification threshold of MammaPrint did not improve the PPV. In a comparison between the Adjuvant! Online 10-year relapse risk for each MGH patient and MammaPrint, the latter identified an additional 21 patients who did not develop distant metastases as the first event, and an additional five patients when considering DMFS per se *Conclusion:* MammaPrint had a high NPV and provided some information that was additional to that provided by Adjuvant! Online. However, with an extremely low PPV and insignificant differences in OS between MammaPrint high- and low-risk patients the prognostic utility of MammaPrint in this population remained unproven. Moreover, although MammaPrint classified a significant proportion of study patients as high risk, few of these developed metastatic disease Four studies published as conference abstracts reported on clinical validity **Glas *et al.*[70]** Patients with ER+, LN0 from the original validation series[63] were analysed for MammaPrint outcome according to grade. Kaplan–Meier analysis of 106 patients for DMFS at 10 years showed a significant difference between low risk (56 patients, 53%) and high risk (50 patients, 47%) with a HR of 4.7 (95% CI 2.1 to 10.4). Good prognosis (low-risk) patients had a 10-year survival of 86%. In patients with grade II, ER+, LN0 breast cancer a significant separation of patients with good or poor prognosis according to MammaPrint was observed ($p = 0.001$). The probability of developing distant metastasis in the good prognosis group was $< 10\%$; in the poor prognosis group it was 44%. MammaPrint provided a significant separation in recurrence risk in these patients, which improved guidance for the requirement of adjuvant therapy **de Snoo *et al.*[69]** A total of 566 tumour samples from women with ER+, LN0, HER2– breast cancer from five previously reported studies were classified using MammaPrint and the NCCN guidelines,[129] and the 10-year BCSS determined according to each MammaPrint classified 380 (57%) samples as having a good prognosis and 186 (33%) as having a poor prognosis. The NCCN guidelines[129] classified 7% as low risk and 93% as high risk. MammaPrint also identified approximately 66% of NCCN high-risk patients as having a good prognosis. There was an overall discordance between the two tools in 62% of cases. In total, 349 (62%) patients received no adjuvant treatment, 17% received hormone treatment only, 2% chemotherapy only and 20% both	It was concluded that MammaPrint poor-prognosis/high-risk patients demonstrated a benefit when adjuvant chemotherapy was added to hormone therapy. Patients classified by MammaPrint as good prognosis/low risk for recurrence do not appear to benefit from the addition of chemotherapy to hormone treatment

Clinical validity	Clinical utility
MammaPrint predicted a 10-year BCSS of 91% vs. 67% for the good and poor prognosis groups respectively (HR 4.0, 95% CI 2.0 to 7.9, $p<0.001$). NCCN guidelines[129] predicted a BCSS of 86% vs. 83% for the low- and high-risk groups respectively (HR 1.11, 95% CI 0.3 to 4.6, $p=0.888$). Median follow-up was 3.5 years (range 0.1–21.1 years). In multivariate analysis (adjusted for known prognostic factors and adjuvant therapy), only MammaPrint and histological grade were independent predictors for 10-year BCSS with HRs of 2.8 (95% CI 1.3 to 6.1, $p=0.008$) and 1.9 (95% CI 1.1 to 3.1, $p=0.015$) respectively. It was concluded that MammaPrint was a strong and independent prognostic indicator in ER+, LN0, HER2– breast cancer	
Knauer *et al.*[71]	
In this study the authors used MammaPrint to assess prognosis, BCSS and DMFS in 965 pT1 breast cancer tumour samples from seven previous studies. MammaPrint classified 526 patients (55%) as having a good prognosis and 439 (45%) as having a poor prognosis. In total, 562 patients (59%) received no adjuvant treatment, 19% received hormone treatment only, 10% received chemotherapy only and 12% both hormone therapy and chemotherapy. MammaPrint accurately predicted differences in 10-year DDFS (HR 2.7, 95% CI 1.9 to 3.9, $p<0.01$) and BCSS (HR 4.0, 95% CI 2.6 to 6.3, $p<0.01$) for all T1 tumours. Similar results were obtained in multivariate analysis for all patients, adjusted for known prognostic factors and adjuvant therapy, as well as for adjuvant therapy-untreated patients. For the pT1a/b subgroup ($n=140$), 10-year DDFS was 93% vs. 78% for the good and poor prognosis groups (HR 3.9, 95% CI 1.0 to 15.2, $p=0.048$), whereas in the T1c subgroup ($n=825$) DDFS was 86% vs. 72% respectively (HR 2.6, 95% CI 1.8 to 4.0, $p<0.01$). BCSS was 87% vs. 73% for the good and poor prognosis groups in the T1a/b subgroup (HR 2.4, 95% CI 0.8 to 7.7, $p=0.128$) and 92% vs. 72% for the good and poor prognosis groups in the T1c subgroup (HR 4.4, 95% CI 2.7 to 7.1, $p<0.01$) It was concluded that MammaPrint was a strong and independent prognostic indicator in small breast tumours	
Saghatchian *et al.*[72]	
It has been shown that MammaPrint predicts disease outcome in patients with one to three positive nodes and four to nine positive nodes. In this study the authors report a further analysis of 519 LN+ patients from a consecutive series of patients from two hospitals based on adjuvant treatment received. Female patients diagnosed between 1984 and 1995 with LN+, unilateral T1, T2 or operable T3 primary invasive breast carcinoma who received mastectomy or breast-conserving therapy and for whom fresh-frozen tumour material was available were eligible for the study In total, 346 patients had one to three positive lymph nodes and 173 had four to nine positive lymph nodes. Tumours were classified by MammaPrint as good prognosis/low risk in 212 patients (41%) and poor prognosis/high risk in 307 patients (59%) with strictly equal proportions among the two LN groups. With a median follow-up of 10.3 years, distant metastases occurred in 141 (27%) patients (116 as first event), and 103 (20%) died of their disease. It was concluded that combining nodal status and MammaPrint profiling allowed patients to be stratified for tailored treatment strategies. Patients with an elevated number of LNs and high genomic risk had a very poor prognosis and might need to be considered for stronger treatment combinations	

Systematic review summary

This review updates the review by Marchionni *et al.*[33] and found an additional 11 studies, some journal publications and some conference abstracts. Analytical validity remains a weakness of the evidence base for MammaPrint, with no new studies identified. The majority of studies found across the two reviews provide evidence relating to the clinical validity of the test in heterogeneous populations. The additional studies reporting on the clinical validity of the test sought to validate the prognostic value and accuracy of MammaPrint in an independent cohort and to extend previous experience of the test in older patients with small tumours. Four studies reported subset analyses of data reported in previous studies examining the use of the test in very heterogeneous populations. The evidence relating to the clinical validity of MammaPrint was not always conclusive or supportive of the prognostic value of the test. Four studies suggested that the ratio could predict prognosis, one study failed to verify the prognostic utility of the test and in another the methods and results were at variance with those of other studies. Three studies focusing on clinically utility were identified: one journal article and two conference abstracts. The fully reported study provided important evidence of the potential impact of MammaPrint on decision-making in Dutch hospitals and the concordance between the gene profile and commonly used clinicopathological tools for risk prediction. A second study, published as an abstract only, presented initial results of a meta-analysis of 1637 patients from seven large multinational data sets to determine the benefit of adding adjuvant chemotherapy to endocrine therapy. The encouraging results of this study may eventually provide strong enough evidence to provide reasonable justification for the interim use of the test in women in the same population group as the trial patients. One study examined the budgetary impact of MammaPrint. As in the original review, the evidence for the clinical implications of using MammaPrint remains unclear

Appendix 6

Studies identified by the electronic searches and other searches and excluded at the full paper stage for reasons not immediately apparent from the full text

Study	Reason for exclusion
Schor *et al.* (2009)[190]	Not available within study timescale
Jancin (2010)[144]	Not available within study timescale
Bighin *et al.* (2010)[191]	Letter without sufficiently detailed data
Espinosa *et al.* (2009)[192]	Study of the research version rather than the commercial version of the 70-gene signature
Mook *et al.* (2010)[108]	Pooled analysis, therefore did not meet inclusion criteria
Knauer *et al.* (2010)[110]	Pooled analysis, therefore did not meet inclusion criteria
Knauer *et al.* (2010)[111]	Pooled analysis, therefore did not meet inclusion criteria
Bueno-de-Mesquita *et al.* (2011) (online version 2009 used)[109]	Pooled analysis, therefore did not meet inclusion criteria
Ma *et al.* (2008)[193]	Study of a previous version of the BCI

Appendix 7

OncotypeDX test: quality assessment and summary of results

Methodological quality assessment of studies investigating the OncotypeDX test

Study feature	Qualities sought	Ademuyiwa *et al.* (2011)[82]	Albain *et al.* (2010)[83]	Cuzick *et al.* (2011)[84]	Dowsett *et al.* (2010)[79]	Geffen *et al.* (2009)[77]	Holt *et al* (2011)[78] (abstract only)	Kelly *et al.* (2010)[85]	Lo *et al.* (2010)[76]	Tang *et al.* (2011);[81] Mamounas *et al.* (2010)[80]	Tang *et al.* (2010)[86] (abstract only)	Toi *et al.* (2010)[87]	Yorozuya *et al.* (2009)[88]
Sample of patients	Inclusion criteria defined	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Sample selection explained	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Adequate description of diagnostic criteria	N	Y	Y	Y	Y	N	N	N	Y	Y	Y	Y
	Clinical and demographic characteristics fully described	Y	Y	N	Y	Y	N	Y	Y	N	N	Y	Y
	Representative (selected by random selection or as consecutive cases)	Y	U	U	U	Y	U	Y	Y	U	U	U	N
	Assembled at a common (usually early) point in the course of their disease	Y (ER+, HER2–, LN–)	U (LN+, ER+, postmenopausal)	U (all ER+ or PR+)	U (postmenopausal HR+ women only)	Y (T1N0M0)	Y	U (HR+ cancers only)	Y (LN–, ER+ only)	Y (LN–, ER+)	Y (LN–, ER+)	Y (early stage, ER+, LN–)	Y (LN–, ER+, stage I or IIA)
	Complete (all eligible patients were included)	Y	N	Y	N	N	N	Y	N	N	U	N	Y
Follow-up of patients	Sufficiently long	Y	Y	Y	Y	N	U	Y	N	Y	U	U	N

Study feature	Qualities sought	Ademuyiwa *et al.* (2011)[82]	Albain *et al.* (2010)[83]	Cuzick *et al.* (2011)[84]	Dowsett *et al.* (2010)[79]	Geffen *et al.* (2009)[77]	Holt *et al* (2011)[78] (abstract only)	Kelly *et al.* (2010)[85]	Lo *et al.* (2010)[76]	Tang *et al.* (2011);[81] Mamounas *et al.* (2010)[80]	Tang *et al.* (2010)[86] (abstract only)	Toi *et al.* (2010)[87]	Yorozuya *et al.* (2009)[88]
Outcome	Objective	Y	U	Y	Y	U	Y	Y	Y	Y	Y	U	Y
	Unbiased (e.g. assessment blinded to prognostic information)	Y	U	U	U	U	Y	N	Y	U	U	U	Y
	Fully defined	Y	Y	Y	N	U	Y	Y	Y	Y	U	Y	N
	Appropriate	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y
	Known for all or a high proportion of patients	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Prognostic variable	Fully defined, including details of method of measurement if relevant	Y	Y	Y	Y	N	Y	N	N	Y	U	N	Y
	Precisely measured	Y	Y	Y	Y	Y	Y	U	Y	Y	U	Y	Y
	Available for all or a high proportion of patients	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y
	If relevant, cut-point(s) defined and justified	Y (reference provided)	Y (reference provided)	Y (reference provided)	Y (reference provided)	Y (reference provided)	Y (reference provided)	Y (reference provided)	Y (reference provided)	Y (reference provided)	U	Y (reference provided)	Y (reference provided)

Study feature	Qualities sought	Ademuyiwa *et al.* (2011)[82]	Albain *et al.* (2010)[83]	Cuzick *et al.* (2011)[84]	Dowsett *et al.* (2010)[79]	Geffen *et al.* (2009)[77]	Holt *et al* (2011)[78] (abstract only)	Kelly *et al.* (2010)[85]	Lo *et al.* (2010)[76]	Tang *et al.* (2011);[81] Mamounas *et al.* (2010)[80]	Tang *et al.* (2010)[86] (abstract only)	Toi *et al.* (2010)[87]	Yorozuya *et al.* (2009)[88]
Analysis	Continuous predictor variable analysed appropriately	U	Y	Y	Y	U	Y	Y	U	Y	U	U	Y
	Statistical adjustment for all important prognostic factors	Y	U	Y	Y	N	U	U	U	Y	U	Y	Y
Intervention subsequent to inclusion in cohort	Fully described	Y	Y	Y	Y	U	U	N	U	Y	Y	U	N
	Intervention standardised or randomised	N	Y	Y	Y	U	U	N	N	Y	Y	U	N

HR, hormone receptor; N, no; U, unclear/not reported; Y, yes.

Summary of results: OncotypeDX test (new data)

Study	Outcomes/end points	Results	Author conclusions	Comments
Ademuyiwa *et al.* (2011)[82]	1. Impact on clinical decision-making in terms of recommending chemotherapy (CT)	**1a. OncotypeDX (ODX)-blinded recommendation vs. ODX risk group and actual treatment received** Low (0–17): $n=142$; recommended CT: 52 (37%); actually received CT: 13 (9%) Intermediate (18–30): $n=110$; recommended CT: 52 (47%); actually received CT: 52 (47%) High (>30): $n=24$; recommended CT: 21 (87%); actually received CT: 23 (96%) **1b. ODX blinded recommendation vs. ODX score-based actual treatment** ODX-blinded 'no' and ODX-based 'no': 117/276 (42.3%) ODX-blinded 'yes' and ODX-based 'no': 71/276 (25.7%) ODX-blinded 'no' and ODX-based 'yes': 34/276 (12.3%) ODX-blinded 'yes' and ODX-based 'yes': 54/276 (19.7%) 37 fewer patients (71 − 34) received CTX using ODX score to help decide CTX use 38% of patients (25.7% + 12.3%) had a change in management as a result of ODX score **1c. ODX-blinded recommendation vs. NPI category** Low (0–17): $n=142$; excellent/good NPI: 123; moderate NPI: 19 Intermediate (18–30): $n=110$; excellent/good NPI: 86; moderate NPI: 24 High (>30): $n=24$; excellent/good NPI: 11; moderate NPI: 13 $p<0.001$	The ODX score had a significant impact on the receipt of adjuvant CT and altered management for 38% of women	

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Author conclusions</th><th>Comments</th></tr>
<tr><td>Albain et al. (2010)[83]</td><td>1. The degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes</td><td>
1a. RS for DFS
In tamoxifen (TAM)-alone group stratified by number of positive nodes, log-rank, $p=0.017$
DFS estimate at 10 years: low RS: 60%; intermediate RS: 49%; high RS: 43%
Cox regression model, continuous RS highly significant, $p=0.006$, HR = 2.64 (95% CI 1.33 to 5.27) for 50-point difference
Proportional hazards showed test not consistent over time ($p=0.0016$)
HR for those surviving beyond 5 years = 0.86 (95% CI 0.27 to 2.74, $p=0.8$)
DFS HRs adjusted for number of positive nodes, for chemotherapy benefit, by RS over time
All years interaction p-value = 0.053
5 years interaction p-value = 0.029
10 years interaction p-value = 0.58 (i.e. RS not good predictor for chemotherapy benefit over 5 years)
Treatment effect overall: DFS HRs (95% CIs) adjusted for number of positive nodes, for chemotherapy benefit, by RS over time:
<table>
<tr><th></th><th>All years HR</th><th>5 years HR</th><th>After 5 years HR</th></tr>
<tr><td>Entire RS sample</td><td>0.72 (0.51 to 1.00)</td><td>0.79 (0.51 to 1.23)</td><td>0.63 (0.39 to 1.04)</td></tr>
<tr><td>At selected RS values</td><td></td><td></td><td></td></tr>
<tr><td>10</td><td>0.95 (0.59 to 1.52)</td><td>1.24 (0.62 to 2.48)</td><td>0.72 (0.38 to 1.36)</td></tr>
<tr><td>18</td><td>0.83 (0.56 to 1.22)</td><td>1.03 (0.58 to 1.81)</td><td>0.67 (0.40 to 1.14)</td></tr>
<tr><td>25</td><td>0.74 (0.53 to 1.04)</td><td>0.87 (0.53 to 1.42)</td><td>0.64 (0.39 to 1.05)</td></tr>
<tr><td>31</td><td>0.67 (0.48 to 0.93)</td><td>0.75 (0.48 to 1.18)</td><td>0.61 (0.35 to 1.04)</td></tr>
<tr><td>40</td><td>0.57 (0.39 to 0.83)</td><td>0.61 (0.38 to 0.96)</td><td>0.56 (0.28 to 1.11)</td></tr>
</table>
There are data looking at the validity of RS in the cyclophosphamide, doxorubicin and fluorouracil followed by tamoxifen (CAF-T) group alongside the data for the TAM group, but these data seem to show that chemotherapy has a benefit over TAM alone; do not give HR for CAF-T group alone
RS was a strong predictor of benefit from CAF-T for DFS; only those in high-risk groups gain benefit (CAF-T vs. TAM DFS at 10 years, stratified by number of nodes, log-rank test):
Low RS: not significantly different ($p=0.97$): 64% survival in CAF-T group, 60% in TAM group
Intermediate RS: not significantly different ($p=0.48$)
High RS: significantly different ($p=0.033$): 55% survival in CAF-T group, 43% in TAM group
</td><td></td><td></td></tr>
</table>

Study	Outcomes/end points	Results	Author conclusions	Comments
		1b. RS for OS		
		In TAM-alone group, stratified by number of positive nodes, log-rank, $p = 0.003$		
		DFS estimate at 10 years: low RS: 77%; intermediate RS: 68%; high RS: 51%		
		HR after adjustment for number of positive nodes = 4.42 (95% CI 1.96 to 9.97, $p = 0.0006$) for 50-point difference		
		Proportional hazards showed not consistent over time ($p = 0.0005$)		
		RS was a strong predictor of benefit from CAF-T for OS; only those in high-risk groups gain benefit (CAF-T vs. TAM OS at 10 years, stratified by number of nodes, log-rank test):		
		Low RS: not significantly different ($p = 0.63$)		
		Intermediate RS: not significantly different ($p = 0.85$)		
		High RS: significantly different ($p = 0.027$)		
		1c. RS for BCSS		
		RS was a predictor of benefit from CAF-T for BCSS; only those in high-risk groups gain benefit (CAF-T vs. TAM BCSS at 10 years, stratified by number of nodes, log-rank test):		
		Low RS: not significantly different ($p = 0.56$)		
		Intermediate RS: not significantly different ($p = 0.89$)		
		High RS: significantly different ($p = 0.033$)		

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Author conclusions</th><th>Comments</th></tr>
<tr><td>Cuzick et al.[84] (2011)</td><td>Distant recurrence (within 10 years)
TTDR</td><td>G1 cohort: 195 recurrences of which 145 distant recurrences; in LN– women 101 recurrences of which 67 distant recurrences
The mean change in likelihood ratio chi-squared (95% CI) for addition of GHI-RS to the classical score in the validation halves of 100 random splits of the data (higher values indicate more added prognostic information):
<table>
<tr><th colspan="2">TTDR (months)</th><th colspan="2">Time to recurrence (all recurrences)</th></tr>
<tr><th>All patients</th><th>LN–</th><th>All patients</th><th>LN–</th></tr>
<tr><td>25.3 (25.2–25.9)</td><td>20.9 (20.7–21.6)</td><td>25.6 (25.2–25.9)</td><td>25.7 (25.4–26.4)</td></tr>
</table>
9-year distant recurrence probabilities for 25th and 75th percentiles of GHI-RS scores for different grades and nodal status for women aged >65 years with a 1–2 cm tumour treated with anastrozole:
<table>
<tr><th></th><th></th><th colspan="3">Grade (%)</th></tr>
<tr><th>Nodal status</th><th>Percentile</th><th>Poor or undifferentiated</th><th>Moderate</th><th>Well differentiated</th></tr>
<tr><td>Negative</td><td>25</td><td>8.3</td><td>5.8</td><td>2.5</td></tr>
<tr><td></td><td>75</td><td>12.1</td><td>8.4</td><td>3.6</td></tr>
<tr><td>Positive</td><td>25</td><td>12.1</td><td>8.4</td><td>3.6</td></tr>
<tr><td></td><td>75</td><td>17.3</td><td>12.2</td><td>5.3</td></tr>
</table></td><td>NR for GHI-RS alone</td><td></td></tr>
</table>

GHI-RS, Genomic Health Recurrence Score.

Study	Outcomes/end points	Results	Author conclusions	Comments
Dowsett *et al* 2010[79]	1. Degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes	**1a. RS and risk of distant recurrence (DR)** Risk score for a 50-point change (e.g. RS = 55 vs. RS = 5) was significantly associated with risk of DR (HR 3.92; 95% CI 2.08 to 7.39; $\Delta\chi^2 = 15.5$; $p < 0.001$) when adjusted for the effects of tumour size, local grade, age and treatment When local grade replaced with central grade in multivariate analysis, adjusted RS also significantly associated with risk of DR (HR 5.25; 95% CI 2.84 to 9.73; $\Delta\chi^2 = 22.7$; $p < 0.001$) **1b. RS and TTDR** In N0 patients: HR = 5.25 (95% CI 2.84 to 9.73); $\Delta\chi^2 = 22.7$; $p < 0.001$ In N+ patients: HR = 3.47 (95% CI 1.64 to 7.38); $\Delta\chi^2 = 9.4$; $p < 0.002$ **1c. Differences in absolute DR rates for N0 and N+ patients** *DR at 9 years N0 patients* RS < 18: 4% (95% CI 3% to 7%) RS 18–30: 12% (95% CI 8% to 18%) RS ≥ 31: 25% (95% CI 17% to 34%) HR adjusted for clinical variables (tumour size, grade, age, treatment and number of positive nodes): between high and low RS groups = 5.2 (95% CI 2.7 to 10.1); between intermediate and low RS groups = 2.5 (95% CI 1.3 to 4.5) *DR at 9 years N+ patients* RS < 18: 17% (95% CI 12% to 24%) RS 18–30: 28% (95% CI 20% to 39%) RS ≥ 31: 49% (95% CI 35% to 64%) HR adjusted for clinical variables (tumour size, grade, age, treatment and number of positive nodes): between high and low RS groups = 2.7 (95% CI 1.5 to 5.1); between intermediate and low RS groups = 1.8 (95% CI 1.0 to 3.2)	This study confirmed the performance of RS in postmenopausal hormone receptor-positive patients treated with tamoxifen in a large contemporary population and demonstrated that RS is an independent predictor of DR in N0 and LN+ hormone receptor-positive patients treated with anastrozole, adding value to estimates with standard clinicopathological features	

Study	Outcomes/end points	Results	Author conclusions	Comments
		1d. OS at 9 years for N0 and N+ patients *N0 patients* RS < 18: 88% (95% CI NR) RS 18–30: 84% (95% CI NR) RS ≥ 31: 73% (95% CI NR) HR adjusted for clinical variables (tumour size, grade, age, treatment and number of positive nodes): between high and low RS groups = 2.5 (95% CI 1.5 to 4.0); between intermediate and low RS groups = 1.2 (95% CI 0.8 to 1.9) *N+ patients* RS < 18: 74% (95% CI NR) RS 18–30: 69% (95% CI NR) RS ≥ 31: 54% (95% CI NR) HR adjusted for clinical variables (tumour size, grade, age, treatment and number of positive nodes): between high and low RS groups = 2.1 (95% CI 1.2 to 3.8); between intermediate and low RS groups = 1.4 (95% CI 0.9 to 2.4) Data to show that treatment group (tamoxifen vs. anastrozole) did not interact with RS prediction of DR **1e. RS, Adjuvant! Online and DR** Correlation between RS-predicted DR and Adjuvant! Online-predicted recurrence was low but statistically significant by central grade (Spearman's rank correlation = 0.23, $p < 0.001$) or local grade (Spearman's rank correlation = 0.22, $p < 0.001$). Only approx. 5% of variability explained by each other, therefore have independent prognostic value		
Geffen *et al* 2009[77]	1. Impact on clinical decision-making	**1. Impact on clinical decision-making** 25 patients had RS assay; nine patients' (36%) treatment recommendations were changed based on the scores, six from chemotherapy to no chemotherapy	NR for this outcome	

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Author conclusions</th><th>Comments</th></tr>
<tr><td>Holt et al 2011[78]</td><td>1. Impact on clinical decision-making</td><td>1a. Change in initial recommendations pre RS assay to post RS assay
All patients have hormone therapy as standard
No change no chemotherapy (CT): 49 (46.23%)
Change CT to no CT: 25 (23.6%)
Change no CT to CT: 10 (9.43%)
No change CT: 22 (20.75%)
1b. Change in patient choices pre RS assay to post RS assay by NPI score
<table>
<tr><th>NPI</th><th>No CT (unchanged) (n)</th><th>CT to no CT (changed) (n)</th><th>No CT to CT (changed) (n)</th><th>CT (unchanged) (n)</th></tr>
<tr><td><2.4</td><td>9</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>2.4–3.4</td><td>31</td><td>8</td><td>4</td><td>5</td></tr>
<tr><td>3.4–4.4</td><td>8</td><td>15</td><td>5</td><td>10</td></tr>
<tr><td>4.4–5.4</td><td>1</td><td>2</td><td>0</td><td>6</td></tr>
<tr><td>>5.4</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>Total</td><td>49</td><td>25</td><td>10</td><td>22</td></tr>
</table>
1c. Spearman's rank correlation comparing RS with individual components of NPI
Of size, LN status and grade, only grade was significantly correlated</td><td>Early results of study suggest that OncotypeDX is applicable and feasible to perform in the UK setting with a reduction in the use of adjuvant CT consistent with the findings of other reported studies. RS added prognostic information beyond that from NPI alone</td><td></td></tr>
<tr><td>Kelly et al 2010[85]</td><td>1. Correlation with Adjuvant! Online
2. Risk prediction</td><td>1. Correlation between predicted risk of recurrence and death after 5 years of tamoxifen therapy vs. RS = 0.13 and 0.18 respectively
2. Assumes cohort of patients sent for OncotypeDX testing are clinically intermediate patients. Of these, OncotypeDX was able to dichotomise 52% (n = 160) to low-risk group and 9% (n = 27) to high-risk group; 39% (n = 122) were judged at intermediate risk</td><td>Authors concluded that OncotypeDX yielded potentially informative risk assignments in patients who may be considered at indeterminate risk by routine clinical variables. However, 40% of the time they remain intermediate risk using RS thresholds; this increases to 66% when using revised TAILORx thresholds</td><td></td></tr>
</table>

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Author conclusions</th><th>Comments</th></tr>
<tr><td>Lo et al 2010[76]</td><td>Impact of the 21-gene RS assay on clinical decision-making and patient preferences. End points include (1) changes in physician treatment recommendations, (2) physician self-assessed changes in long-term adjuvant treatment, (3) patient anxiety, (4) quality of life, (5) relapse data</td><td>1a. Whole cohort – changes in physician treatment recommendations
From hormone therapy (HT) to chemotherapy and hormone therapy (CHT): 3/89 (3.4%)
From CHT to HT: 20/89 (22.5%)
From HT to equipoise:[a] 3 (3.4%)
From CHT to equipoise:[a] 2 (2.2%)
No change HT: 40 (44.9%)
No change CHT: 20 (22.5%)
No change equipoise:[a] 1 (1.1%)
1b. By RS category – changes in physician treatment recommendations
<table><tr><th>Physician pre- to post-RS assay treatment recommendation</th><th>Low RS</th><th></th><th>Intermediate RS</th><th></th><th>High RS</th><th></th><th>Total</th><th></th></tr><tr><th></th><th>n</th><th>%</th><th>n</th><th>%</th><th>n</th><th>%</th><th>n</th><th>%</th></tr><tr><td>HT to HT</td><td>21</td><td>52.5</td><td>19</td><td>47.5</td><td>0</td><td>0</td><td>40</td><td>100</td></tr><tr><td>HT to CHT</td><td>0</td><td>0</td><td>0</td><td>0</td><td>3</td><td>100</td><td>3</td><td>100</td></tr><tr><td>CHT to HT</td><td>12</td><td>60</td><td>8</td><td>40</td><td>0</td><td>0</td><td>20</td><td>100</td></tr><tr><td>CHT to CHT</td><td>3</td><td>15</td><td>11</td><td>55</td><td>6</td><td>30</td><td>20</td><td>100</td></tr><tr><td>HT to equipoise</td><td>1</td><td>33.3</td><td>2</td><td>66.7</td><td>0</td><td>0</td><td>3</td><td>100</td></tr><tr><td>HT to equipoise</td><td>1</td><td>50</td><td>1</td><td>50</td><td>0</td><td>0</td><td>2</td><td>100</td></tr><tr><td>Equipoise to equipoise</td><td>0</td><td>0</td><td>1</td><td>100</td><td>0</td><td>0</td><td>1</td><td>100</td></tr><tr><td>Total</td><td>38</td><td>42.7</td><td>42</td><td>47.2</td><td>9</td><td>10.1</td><td>89</td><td>100</td></tr></table>Difference between mean RS for recommendation of CHT vs. HT alone: 29 vs. 16 ($p=0.0001$)
Difference between mean RS for recommendation of CHT vs. equipoise: 29 vs. 19 ($p=0.001$)
Difference between mean RS for HT alone vs. equipoise: 16 vs. 19 ($p=0.288$)
1c. Correlation between treatment and RS category
High-risk RS: 9/9 (100%) CHT
Intermediate RS: 11 (26.2%) CHT
Low-risk RS: 3 (7.9%) CHT</td><td>The RS assay impacts significantly on physician and patient adjuvant treatment decision-making. Most of the treatment changes were from a pretreatment recommendation of CHT to HT alone for both physicians and patients. In addition, RS results have an enduring impact on physician confidence in their treatment recommendations, patient satisfaction and patient anxiety</td><td></td></tr>
</table>

Study	Outcomes/end points	Results	Author conclusions	Comments
		2. Physician self-assessed changes in long-term adjuvant treatment 16 (94%) physicians completed a follow-up questionnaire; 15/16 (94%) of these stated that the assay provided additional information for adjuvant decision-making; 14/16 believed that it had influenced their recommendations; 16/16 (100%) would use it again **3. DCS and anxiety** Mean DCS pre RS: 1.99 (SD 0.62); mean DCS post RS: 1.69 (SD 0.5) ($p<0.001$) STAI pre RS, post RS and at 12-month follow-up: state: 39.6 (SD 14.5), 36 (SD 12.6), 34 (SD 11.5) ($p=0.007$); trait: 32.2 (SD 14.5), 31.7 (SD 13.3), 33.2 (SD 11.0) ($p=0.27$) **4. Quality of life** FACT-B pre RS: mean 112.2 (SD 17.4), FACT-B 12 months post RS: mean 114.3 (SD 18.6) ($p=0.55$) FACT-G pre RS: mean 88.7 (SD 12.3), FACT-G 12 months post RS: mean 87.6 (SD 14.9) ($p=0.49$) **5. Relapse data** Of the 67 patients who completed the 12-month questionnaire, none had experienced a relapse. The status of the remaining 22 is unknown		

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Author conclusions</th><th>Comments</th></tr>
<tr>
<td>Mamounas et al 2010[80]</td>
<td>1. The degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes</td>
<td>
1a. Association between RS and locoregional recurrence by treatment group
All groups showed significant associations
Kaplan–Meier estimates and 95% CIs of the proportion of patients with locoregional recurrence at 10 years for 355 placebo-treated patients (NSABP B14), 895 tamoxifen-treated patients (NSABP B14 and B20) and 424 tamoxifen plus chemotherapy-treated patients (NSABP B20)
<table>
<tr><th>Treatment group and RS group</th><th>10-year Kaplan–Meier estimate (%)</th><th>95% CI</th><th>Log-rank p-value</th><th>No. of events/no. at risk</th></tr>
<tr><td>Placebo</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Low (<18)</td><td>10.8</td><td>5.8 to 15.8</td><td>0.022</td><td>19/171</td></tr>
<tr><td>Intermediate (18–30)</td><td>20.0</td><td>9.9 to 30.0</td><td></td><td>15/85</td></tr>
<tr><td>High (≥31)</td><td>18.4</td><td>9.5 to 27.4</td><td></td><td>19/99</td></tr>
<tr><td>Tamoxifen</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Low (<18)</td><td>4.3</td><td>2.3 to 6.3</td><td>0.001</td><td>24/473</td></tr>
<tr><td>Intermediate (18–30)</td><td>7.2</td><td>3.4 to 11.0</td><td></td><td>6/194</td></tr>
<tr><td>High (≥31)</td><td>15.8</td><td>10.4 to 21.2</td><td></td><td>33/228</td></tr>
<tr><td>Chemotherapy + tamoxifen</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Low (<18)</td><td>1.6</td><td>0.0 to 3.5</td><td>0.028</td><td>4/218</td></tr>
<tr><td>Intermediate (18–30)</td><td>2.7</td><td>0.0 to 6.4</td><td></td><td>2/89</td></tr>
<tr><td>High (≥31)</td><td>7.8</td><td>2.6 to 13.0</td><td></td><td>8/117</td></tr>
</table>
Note: Results are given for all patients and for the prespecified RS risk categories.
</td>
<td>Similar to the association between RS and risk for distant recurrence, a significant association exists between RS and risk for locoregional recurrence. This information has biologic consequences and potential clinical implications relative to locoregional therapy decisions for patients with LN– and ER+ breast cancer</td>
<td>Locoregional relapse, not distant relapse</td>
</tr>
</table>

Study	Outcomes/end points	Results	Author conclusions	Comments

1b. Multivariate Cox regression analysis of predictors of locoregional recurrence

Cohort of 895 tamoxifen-treated patients from NSABP trials B14 and B20

Variable	Hazard	95% CI	Wald test *p*-value
Age (≥ 50 vs. < 50)	0.40	0.25 to 0.65	0.0002
Mastectomy vs. L + XRT	0.62	0.39 to 0.99	0.047
Clinical tumour size (> 2 vs. ≤ 2 cm)	0.98	0.61 to 1.59	0.933
Tumour grade (moderate vs. well)	1.10	0.54 to 1.92	0.113
Tumour grade (poor vs. well)	1.76	0.89 to 3.48	
Recurrence score[a]	2.16	1.26 to 3.68	0.005

L, lumpectomy; LRR, locoregional recurrence; XRT, radiation therapy.

a RS was a continuous variable, with the HR for LRR calculated relative to an increment of 50 units (chosen to dichotomise the RS and thus improve comparability of the HR with the HRs based on the clinical covariates). The *p*-value for the likelihood ratio test on RS is 0.007.

Study	Outcomes/end points	Results	Author conclusions	Comments
Tang *et al* 2011[81]	1. The degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes	**1a. Comparison between point estimates for RS risk group and recurrence interval (RI) (Adjuvant! Online) risk group for DRFI in NSABP B14 tamoxifen-treated patients ($n=668$)** RS low overall: $n=338$; RS low, RI low: $n=216$; RS low, RI intermediate: $n=57$; RS low, RI high: $n=65$ RS intermediate overall: $n=149$; RS intermediate, RI low: $n=84$; RS intermediate, RI intermediate: $n=24$; RS intermediate, RI high: $n=41$ RS high overall: $n=181$; RS high, RI low: $n=52$; RS high, RI intermediate: $n=43$; RS high, RI high: $n=86$ Concordance between RS and RI was 0.49, correlation was modest (Spearman's correlation coefficient of 0.38) RI low overall: $n=332$; RI low, RS low ($n=216$) point estimate distant recurrence (DR) 10 years: 5.6%; RI low, RS intermediate ($n=84$) point estimate DR 10 years: 10%; RI low, RS high ($n=52$) point estimate DR 10 years: 18.2% RI intermediate overall: $n=146$; RI intermediate, RS low ($n=57$) point estimate DR 10 years: 13.4%; RI intermediate, RS intermediate ($n=24$) point estimate DR 10 years: 13.9%; RI intermediate, RS high ($n=43$) point estimate DR 10 years: 43.2% RI high overall: $n=190$; RI high, RS low ($n=65$) point estimate DR 10 years: 5%; RI high, RS intermediate ($n=41$) point estimate DR 10 years: 23.4%; RI high, RS high ($n=86$) point estimate DR 10 years: 31.5% **1b. Cox models of HRs in B14 tamoxifen-treated patients ($n=668$)** RI percentile as sole predictor, using 50-point increment in score, HR = 2.87 (95% CI 1.95 to 4.23) RS percentile as sole predictor, using 50-point increment in score, HR = 3.61 (95% CI 2.49 to 5.24)		

Study	Outcomes/end points	Results	Author conclusions	Comments

1c. Multivariate Cox models assessing relative associations of RI and RS using 50-point increment in score in B14 tamoxifen-treated patients ($n=668$)

Model 1 – not relevant

Model 2 – not relevant

Model 3 – RS percentile using 50-point increment in score, HR = 3.51 (95% CI 2.49 to 5.24), $p<0.001$

Model 4 – RI and RS percentiles using 50-point increment in score, HR for RS = 2.83 (95% CI 1.91 to 4.18), $p<0.001$

Model 5 – RI and RS percentiles using 50-point increment in score, age, tumour size, grade (moderate vs. well), grade (poor vs. well), HR for RS = 2.37 (95% CI 1.58 to 3.55)

Model 6 – as model 5 but without RI percentile using 50-point increment in score, HR for RS = 2.34 (95% CI 1.56 to 3.5)

Model 7 – not relevant

1d. Multivariate Cox models assessing relative associations of RI and RS using risk groups in B14 tamoxifen-treated patients ($n=668$)

Model	Variables	HR (95% CI)	*p*-value
1	RI intermediate vs. low	2.51 (1.55 to 4.21)	0.001
	RI high vs. low	2.01 (1.25 to 3.23)	
	RS intermediate vs. low	2.21 (1.28 to 3.81)	<0.001
	RS high vs. low	3.8 (2.36 to 6.1)	
2	Age (>50 vs. ≤50 years)	0.76 (0.52 to 1.13)	0.173
	Tumour size	1.2 (1.07 to 1.36)	0.003
	Grade moderate vs. well	1.51 (0.75 to 3.05)	0.003
	Grade poor vs. well	3.18 (1.42 to 7.15)	
	RI intermediate vs. low	1.51 (0.82 to 2.78)	0.176
	RI high vs. low	0.95 (0.52 to 1.76)	
	RS intermediate vs. low	2.07 (1.18 to 3.61)	<0.001
	RS high vs. low	2.88 (1.74 to 4.76)	

Study	Outcomes/end points	Results	Author conclusions	Comments
		1e. Multivariate Cox models assessing relative associations of RS using 50-point increment in score (RS/50) in B14 tamoxifen-treated patients with breast cancer-specific mortality as the end point		
		RS/50 alone, HR = 3.32 (95% CI 2.29 to 4.81), $p<0.001$		
		RS/50 (RI/50 in model), HR = 2.45 (95% CI 1.66 to 3.61), $p<0.001$		
		RS/50 (age, tumour size, grade and RI/50 in model), HR = 2.02 (95% CI 1.35 to 3.0), $p<0.001$		
		RS/50 (age, tumour size, grade in model), HR = 2.01 (95% CI 1.35 to 2.98), $p<0.001$		
		1f. Multivariate Cox models assessing relative associations of RS/50 in B14 tamoxifen-treated patients with OS as the end point		
		RS/50 alone, HR = 1.95 (95% CI 1.51 to 2.52), $p<0.001$		
		RS/50 (RI/50 in model), HR = 1.77 (95% CI 1.35 to 2.33), $p<0.001$		
		RS/50 (age, tumour size, grade and RI/50 in model), HR = 1.67 (95% CI 1.26 to 2.22), $p<0.001$		
		RS/50 (age, tumour size, grade in model), HR = 1.65 (95% CI 1.24 to 2.19), $p<0.001$		
		1g. Multivariate Cox models assessing relative associations of RS/50 in B14 tamoxifen-treated patients with DFS as the end point		
		RS/50 alone, HR = 1.77 (95% CI 1.44 to 2.18), $p<0.001$		
		RS/50 (RI/50 in model), HR = 1.75 (95% CI 1.4 to 2.18); $p<0.001$		
		RS/50 (age, tumour size, grade and RI/50 in model), HR = 1.69 (95% CI 1.34 to 2.14); $p<0.001$		
		RS/50 (age, tumour size, grade in model), HR = 1.67 (95% CI 1.32 to 2.11); $p<0.001$		

Study	Outcomes/end points	Results	Author conclusions	Comments

1h. Cox models assessing relative associations of RI and RS using risk groups in B20 chemotherapy patients (n=651) and outcomes of DRFI, OS and DFS

End point and cohort		B20 patients with RS assessment (n=651)		All B20 patients with tumour grade (n=1952)	
		HR for benefit from MF/CMF (95% CI)	Pa (interaction)	HR for benefit from MF/CMF (95% CI)	Pa (interaction)
DRFI	Overall	0.56 (0.34 to 0.91)		0.62 (0.47 to 0.81)	
	RS low	1.31 (0.46 to 3.78)	0.031	NA	
	RS intermediate	0.61 (0.24 to 1.59)			
	RS high	0.26 (0.13 to 0.53)			
	Adjuvant! Online low	0.58 (0.23 to 1.42)	0.99	0.92 (0.53 to 1.62)	0.219
	Adjuvant! Online intermediate	0.54 (0.2 to 1.46)		0.52 (0.29 to 0.93)	
	Adjuvant! Online high	0.53 (0.25 to 1.1)		0.53 (0.36 to 0.77)	
OS	Overall	0.76 (0.49 to 1.17)		0.74 (0.58 to 0.95)	
	RS low	1.37 (0.63 to 3.01)	0.011	NA	
	RS intermediate	0.94 (0.4 to 2.25)			
	RS high	0.31 (0.16 to 0.6)			
	Adjuvant! Online low	1.16 (0.55 to 2.45)	0.311	1.26 (0.81 to 1.95)	0.009
	Adjuvant! Online intermediate	0.7 (0.3 to 1.61)		0.53 (0.31 to 0.9)	
	Adjuvant! Online high	0.53 (0.26 to 1.07)		0.57 (0.4 to 0.82)	

Study	Outcomes/end points	Results					Author conclusions	Comments
				B20 patients with RS assessment ($n=651$)		All B20 patients with tumour grade ($n=1952$)		
		End point and cohort		HR for benefit from MF/CMF (95% CI)	p-value[a] (interaction)	HR for benefit from MF/CMF (95% CI)	p-value[a] (interaction)	
		DFS	Overall	0.73 (0.54 to 0.99)		0.75 (0.63 to 0.89)		
			RS low	0.91 (0.57 to 1.45)	0.082	NA		
			RS intermediate	0.79 (0.43 to 1.47)				
			RS high	0.41 (0.23 to 0.71)				
			Adjuvant! Online low	0.97 (0.59 to 1.61)	0.357	0.91 (0.69 to 1.21)	0.099	
			Adjuvant! Online intermediate	0.6 (0.33 to 1.09)		0.75 (0.5 to 1.13)		
			Adjuvant! Online high	0.62 (0.36 to 1.05)		0.59 (0.45 to 0.78)		

NA, not applicable.
a From likelihood ratio test.

Study	Outcomes/end points	Results	Author conclusions	Comments
		All RS tests are significant; Adjuvant! Online tests are significant when larger cohort is used **1i. Cox models assessing relative associations of RI and RS using risk groups in B20 chemotherapy patients ($n=651$) and outcome of breast cancer-specific mortality**		

Cohort	HR for benefit from MF/CMF (95% CI)	*p*-value[a]	*p*-value[b] (interaction)
Overall	0.62 (0.36 to 1.06)	0.081	
RS low	1.86 (0.38 to 9.19)	0.449	0.025
RS intermediate	0.94 (0.32 to 2.82)	0.918	
RS high	0.27 (0.13 to 0.55)	< 0.001	
Adjuvant! Online low	1.03 (0.35 to 3.01)	0.96	0.463
Adjuvant! Online intermediate	0.62 (0.23 to 1.71)	0.358	
Adjuvant! Online high	0.44 (0.19 to 1.02)	0.054	

a From Wald tests.
b From likelihood ratio tests.

Tang 2010[86] (abstract only)	Distant recurrence Value of RSPC in the prediction of chemotherapy benefit in reducing risk of recurrence	60/625 distant recurrences occurred RS showed a significant interaction with chemotherapy treatment ($p=0.037$) with a standardised HR of 0.836. Interaction of RSPC with treatment not significant ($p=0.10$) although trend was in the same direction as RS (HR 0.833)	RS used alone remains the best predictor of chemotherapy benefit in ER+, LN- breast cancer	

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Author conclusions</th><th>Comments</th></tr>
<tr><td>Toi et al 2010[87]</td><td>1. The degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes</td><td>
1a. Kaplan–Meier plot of DRFI by RS
<table>
<tr><th>RS category</th><th>No. in category at year 0</th><th>No. of distant recurrences 0–5 years</th><th>No. in category at year 5</th><th>No. of distant recurrences 5–10 years</th><th>No. in category at year 10</th></tr>
<tr><td>Low</td><td>95</td><td>2</td><td>90</td><td>1</td><td>70</td></tr>
<tr><td>Intermediate</td><td>40</td><td>0</td><td>40</td><td>0</td><td>31</td></tr>
<tr><td>High</td><td>65</td><td>9</td><td>52</td><td>6</td><td>36</td></tr>
</table>
Low-risk category patients had a significantly lower risk of distant recurrence than patients in the high-risk category ($p<0.001$, log-rank test)
No recurrences in the intermediate RS group
1b. Univariate Cox proportional hazards model of DRFI – continuous risk score
50-point increase in RS, HR = 6.20 (95% CI 2.27 to 17.0)
1c. Multivariate cox model adjusting for age (< 50 vs. ≥ 50 years) and clinical tumour size (≤ 2 cm vs. > 2 cm)
50-point increase in RS, HR = 6.03 (95% CI 2.17 to 16.7)
1d. Kaplan–Meier estimates of other event rates by RS group
<table>
<tr><th>End point</th><th>Event</th><th>Low (RS < 18) (n = 95), % (95% CI)</th><th>Intermediate (RS 18–30) (n = 40), % (95% CI)</th><th>High (RS ≥ 31) (n = 65), % (95% CI)</th></tr>
<tr><td>DRFI</td><td>Distant recurrence</td><td>3.3 (1.1 to 10.0)</td><td>0 (NA)</td><td>24.8 (15.7 to 37.8)</td></tr>
<tr><td>RFI</td><td>Recurrence</td><td>5.5 (2.3 to 12.8)</td><td>2.5 (0.4 to 16.5)</td><td>24.6 (15.6 to 37.6)</td></tr>
<tr><td>RFS</td><td>Recurrence or death</td><td>9.6 (5.1 to 17.6)</td><td>5.1 (1.3 to 18.8)</td><td>23.4 (14.8 to 35.9)</td></tr>
<tr><td>OS</td><td>Death</td><td>6.4 (2.9 to 13.6)</td><td>2.6 (0.4 to 16.8)</td><td>19.1 (11.3 to 31.3)</td></tr>
</table>
NA, not available.
1e. Cox proportional hazards models, adjusting for age (< 50 vs. ≥ 50 years) and clinical tumour size (≤ 2 cm vs. > 2 cm)
Risk of recurrence: HR = 3.38 (95% CI 1.32 to 8.69)
Risk of recurrence or death: HR = 2.09 (95% CI 0.84 to 5.20)
Risk of death: HR = 2.67 (95% CI 0.93 to 7.62)
</td><td></td><td></td></tr>
</table>

Study	Outcomes/end points	Results	Author conclusions	Comments
Yorozuya *et al.* 2009[88]	1. Degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes	**1a. Difference in mean RS value between cases and control subjects** Cases: mean RS = 40.0 (95% CI 21.1 to 58.9); control subjects: mean RS = 17.8 (95% CI 13.8 to 21.9) ($p < 0.001$) **1b. Proportion of patients in risk category groups** Cases: low risk (0 to <18): 3 (30%), intermediate risk (18–30): 1 (10%), high risk (RS ≥ 31): 6 (60%); control subjects: low risk (0 to <18): 19 (63%), intermediate risk (18–30): 8 (27%), high risk (RS ≥ 31): 3 (10%) ($p = 0.005$) **1c. Multivariate logistic regression analysis of age, ER score, PR score, RS, histological grade, Ly vs. distant metastases** RS is not significant but study concludes that odds ratio indicates that it has value	Both histological grade and risk category classification were effective in identifying women at risk of developing distant metastases after initial therapy for ER+, LN– stage I or IIA breast cancer. These patients may benefit from the addition of adjuvant therapy at diagnosis	

Variable	*p*-value	Odds ratio (95% CI)
Age at diagnosis	0.195	0.90 (0.764 to 1.057)
ER score	0.651	1.33 (0.389 to 4.53)
PR score	0.378	0.65 (0.246 to 1.702)
RS ≥ 50 vs. RS < 50	0.579	2.85 (0.07 to 115.552)
Histological grade II vs. histological grade I	0.369	7.48 (0.093 to 602.504)
Histological grade II vs. histological grade I	0.041	222.0 (1.243 to 39,647.336)
Ly(+) vs. Ly(–)	0.557	0.37 (0.013 to 10.312)

Ly, lymphatic invasion.

NR, not reported; SD, standard deviation.

a Equipoise defined as equal options of CHT, HT or enrolment onto the TAILORx clinical trial (random assignment to HT alone or CHT then HT).

Appendix 8

Ongoing trials

The TAILORx[188] and MINDACT[187] trials aim to address the gaps in the literature (clinical utility data).

Anticipated data from TAILORx and MINDACT

TAILORx and MINDACT were recently initiated to prospectively evaluate the clinical utility (i.e. provide direct evidence that these tests in breast cancer patients lead to improvement in outcomes) of OncotypeDX and MammaPrint respectively. Definitive high-quality evidence of the effect of these tests on patient outcomes and their ability to predict treatment response is expected. TAILORx will provide information on the appropriate RS threshold for recommending adjuvant chemotherapy, and will not directly assess the effect of clinical decision-making with and without the test (as all patients will receive the test). The data generated may allow indirect inferences to be made. MINDACT will allow more direct inferences about the clinical utility as it will be compared directly with the use of a conventional risk index. For both trials, patient health outcomes will be end points.

TAILORx

This multicentre, partially randomised trial aims to assess whether hormone therapy alone or hormone therapy with combination chemotherapy is better for women who have a RS of 11–25 (an intermediate risk score) when tested using OncotypeDX.[91] The trial will also allow for the generation of new data on patients with very low RSs. Patients at the low end of the RS spectrum will be compared with a prespecified target of 95% recurrence-free survival.[91] It should be noted that the cut-off values used in the TAILORx trial are different from those delineated in other studies of OncotypeDX.[32]

Population

Patients with ER+ and/or PR+, HER2/neu-negative tumours who are LN– (and who will be treated with tamoxifen) are eligible for inclusion.

Key aspects of the study design

- Patients showing low RSs (≤ 10) by OncotypeDX testing will receive endocrine therapy alone.
- Patients with high RSs (≥ 26) by OncotypeDX testing will receive endocrine therapy and adjuvant chemotherapy.
- Patients with mid-range RSs (11–25) by OncotypeDX will receive endocrine therapy and be randomly assigned to chemotherapy or no chemotherapy.

After completion of the study treatment, patients will be followed up for up to 20 years.

Objectives

The primary objective is to assess whether or not women with an intermediate OncotypeDX score have better outcomes (DFS, DMFS, RFI and OS) when treated with either hormone therapy alone or hormone therapy with combination chemotherapy.

The secondary objectives include assessing whether or not low-risk patients (score ≤ 10) can safely be treated with hormone therapy alone (expect 95% to have DFS); to determine the DFS, DRFI, RFI and OS of patients with OncotypeDX RSs of ≤ 10; to compare the outcomes projected

at 10 years using classical pathological information with those made by the OncotypeDX test; to estimate failure rates as a function of OncotypeDX RS separately in patients treated with combination chemotherapy and in patients treated with no chemotherapy; to determine the prognostic significance of the OncotypeDX RS and of the individual RS gene groups (proliferation gene group, HER2 gene group, ER gene group, invasion gene group and other genes) in patients treated with these regimens.[91]

This study will not provide direct evidence for the value of OncotypeDX but will indicate whether or not adjuvant chemotherapy is of value within the trial's intermediate RS range. This will provide better estimates of the degree of benefit gained by using the test, but cannot ascertain what therapeutic choices would have been made and what clinical outcomes would have occurred if only standard risk prediction methods were used. Information about what choices would have been made could be inferred by applying other prognostic methods retrospectively.[33]

Completion

TAILORx commenced in April 2006.The trial is currently still recruiting and has a primary completion date of April 2014. The target for recruitment is 11,248 participants and the study currently has 280 centres recruiting in the USA, Canada, Australia and Peru.

MINDACT

A partially randomised trial, MINDACT[194] has recently been activated. The multicentre, prospective, phase III randomised trial will compare two different ways of assessing the risk of cancer recurrence and making therapeutic decisions: a 'traditional method' using Adjuvant! Online and the MammaPrint assay. The rationale for this study is that many women who actually have low-risk tumours are currently classified as average or high risk and therefore are recommended to receive adjuvant chemotherapy that ultimately may be of no benefit.

Population

Patients with histologically confirmed unilateral invasive breast cancer with T1–T3 operable disease, up to three positive lymph nodes and no distant metastases are eligible for inclusion. It situ tumours were allowed. Patients must have undergone breast-conserving surgery or a mastectomy with a sentinel node procedure or full axillary clearance, and appropriate radiotherapy.

Key aspects of the study design

- Patients at low risk by both MammaPrint and standard clinicopathological criteria will not receive chemotherapy.
- Patients at high risk by both criteria receive chemotherapy.
- Patients with discordant criteria, in which the clinicopathological prognosis using Adjuvant! Online is different to the gene expression prognosis using the 70-gene signature (which is estimated to be the case for 1920 patients), will be randomised to use either MammaPrint only or standard criteria to decide treatment. This is achieved by randomising patients to either receive or not receive chemotherapy. This will directly test whether or not the choice of chemotherapy guided by MammaPrint provides benefit over that guided by the Adjuvant! Online criteria.
- All those who go on to have chemotherapy (i.e. those at high risk by both prognostic tests, as well as those with discordant criteria who went on to receive chemotherapy) are then eligible for further randomisation to treatment with anthracycline-based chemotherapy or docetaxel/capecitabine-based chemotherapy.

- All hormone receptor-positive patients, regardless of previous randomisations and risk categorisations, are eligible for randomisation to two different endocrine treatment regimens, namely letrozole only or tamoxifen followed by letrozole.
- Patients will be followed up for DMFS at 5 years and DFS. Follow-up will be for a minimum of 15 years after completion of the study treatment.

Objectives

The main objective of the trial is to confirm that patients with low-risk molecular prognosis and high-risk clinical prognosis can be safely spared chemotherapy without affecting DMFS and to demonstrate the superiority of the molecular profiling approach over the usual clinical assessment in assigning risk categories.

The trial has two further main objectives: (1) a comparison of docetaxel and capecitabine regimens (which are possibly associated with increased efficacy and reduced long-term toxicities) with existing commonly used anthracycline-based chemotherapy regimens and (2) to determine the best endocrine treatment strategy between a single-agent upfront aromatase inhibitor (letrozole) for 7 years and the sequential endocrine strategy of 2 years of tamoxifen followed by 5 years of letrozole.

Completion

The trial is currently still recruiting and has a primary completion date of March 2019. The target for recruitment was recently increased from 6000 to 6600 participants and the projected proportion of patients who will fall into the discordant group is 32%.

Comparative summary of the design and characteristics of the TAILORx and MINDACT trials

Variable	TAILORx	MINDACT
Trial	Hormone therapy with or without combination chemotherapy in treating women who have undergone surgery for LN– breast cancer	A prospective, randomised study comparing the 70-gene expression signature with common clinicopathological criteria in selecting patients for adjuvant chemotherapy in LN– breast cancer (EORTC Protocol 10041 – BIG 3–04)
Trial type	Prospective, controlled, partially randomised	Prospective, controlled, partially randomised, open label
	Clinical utility	Clinical utility
	Non-inferiority design	
Test	OncotypeDX	MammaPrint
Gene signature	21-gene	70-gene
Tissue sample type	FFPE	Fresh tissue
Non-molecular clinical profiling technique/prognostic tool (comparator)	Adjuvant! Online	Adjuvant! Online
Sponsor	NCI (co-ordinated by ECOG)	EORTC/TRANSBIG
Countries participating	USA and Canada	Europe
Target for recruitment	11,248 [7887 recruited to date, 4500 randomised (45%?)]	6600 [2100 (32%) randomised]
Date of trial start/activation	April 2006	September 2006
		Estimated accrual time of 3 years and a total duration of 6 years

Appendix 9

MammaPrint test: quality assessment and summary of results

Methodological quality assessment of studies investigating the MammaPrint test

Study feature	Qualities sought	Bueno-de-Mesquita *et al.* (2009)[93]	Gevensleben *et al.* (2010)[94]	Ishitobi *et al.* (2010)[96]	Kok *et al.* (2010)[99]	Kunz *et al.* (2011)[92]	Mook *et al.* (2010)[75]	Na *et al.* (2011)[100]
Sample of patients	Inclusion criteria defined	Y	Y	Y	Y	Y	Y	Y
	Sample selection explained	Y	Y	Y	Y	Y	Y	Y
	Adequate description of diagnostic criteria	Y	Y	?	Y	Y	Y	Y
	Clinical and demographic characteristics fully described	Y	Y	Y	Y	Y	Y	Y
	Representative (selected by random selection or as consecutive cases)	Y (consecutive)	Y (consecutive)	U	U	U	Y (consecutive)	U
	Assembled at a common (usually early) point in the course of their disease	Y (pT1–2, N0)	?	Y (all N0)	?	Y (T1-T3, N0–3)	Y (T1–2, N0)	Y (T1–2, N0, M0)
	Complete (all eligible patients were included)	U	N	N	U	N	N	N
Follow-up of patients	Sufficiently long	Y (median 5.8 years)	U	Y (median 7.1 years)	Y (9.6–11.1 years)	U	Y (median 11.6 years)	U
Outcome	Objective	Y	Y	Y	Y	Y	Y	Y
	Unbiased (e.g. assessment blinded to prognostic information)	Y	?	?	?	?	Y	?
	Fully defined	Y	? (risk classification)	N	Y	? (risk classification)	Y	? (risk classification)
	Appropriate	Y (OS, DMFS)	? (risk classification)	Y (DMFS, risk classification)	Y (BCSS)	? (risk classification)	Y (DMFS, BCSS, risk classification)	? (risk classification)
	Known for all or a high proportion of patients	Y	Y	Y	Y	Y	Y	Y

Study feature	Qualities sought	Bueno-de-Mesquita *et al.* (2009)[93]	Gevensleben *et al.* (2010)[94]	Ishitobi *et al.* (2010)[96]	Kok *et al.* (2010)[99]	Kunz *et al.* (2011)[92]	Mook *et al.* (2010)[75]	Na *et al.* (2011)[100]
Prognostic variable	Fully defined, including details of method of measurement if relevant	Y (MammaPrint, St Gallen, NPI, Adjuvant! Online)	Y (MammaPrint, St Gallen, Adjuvant! Online)	Y (MammaPrint, St Gallen)	Y (MammaPrint, endocrine response category)	Y (MammaPrint, St Gallen, Adjuvant! Online)	Y (MammaPrint, Adjuvant! Online)	Y (MammaPrint, St Gallen, NIH, Adjuvant! Online)
	Precisely measured	Y	Y	Y	Y	?	?	Y
	Available for all or a high proportion of patients	Y	Y	Y	Y	Y	Y	Y
	If relevant, cut-point(s) defined and justified	Y (reference provided)	Y (reference provided)	Y	Y (reference provided)	N	Y (reference provided)	Y
Analysis	Continuous predictor variable analysed appropriately	Y	N	N	Y	?	Y	N
	Statistical adjustment for all important prognostic factors	Y	N	N	Y	?	Y	N
Intervention subsequent to inclusion in cohort	Fully described	N	N	N	N	?	N	N
	Intervention standardised or randomised	N	N (96% had chemotherapy/ endocrine therapy; no further details provided)	N (retrospective study: 73% had adjuvant hormonal therapy, 28% adjuvant chemotherapy; no further details provided)	N [retrospective study: 100% tamoxifen treated (no further details provided) and no neoadjuvant therapy; consecutive series: 100% tamoxifen naive]	? (prospective study)	N [retrospective study: 18% had adjuvant endocrine (tamoxifen) therapy; no further details provided]	N (retrospective study: 73% had chemotherapy; no further details provided)

N, no; U, unclear/not reported; Y, yes.

Summary of results: MammaPrint test (new data)

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Authors' conclusions</th><th>Comments</th></tr>
<tr><td>Bueno-de-Mesquita et al. (2009)[93]</td><td>Time from surgery to distant metastasis as first event (counted as failures)
OS (defined as time from surgery to death)</td><td>Classification and disease outcome
Univariate analysis
OS at 5 years
<table>
<tr><th>Method</th><th>HR</th><th>95% CI</th><th>p-value</th></tr>
<tr><td>MammaPrint</td><td>3.4</td><td>1.2 to 9.6</td><td>0.021[a]</td></tr>
<tr><td>Adjuvant! Online</td><td>2.5</td><td>0.59 to 11</td><td>0.22</td></tr>
<tr><td>NPI</td><td>2.8</td><td>0.99 to 7.8</td><td>0.053</td></tr>
<tr><td>CBO[102]</td><td>2.3</td><td>0.84 to 6.6</td><td>0.11</td></tr>
<tr><td>St Gallen</td><td>3.0</td><td>0.4 to 22</td><td>0.29</td></tr>
</table>
a The probability of OS (as first event) was 97% (±SE 2%) for good and 82% (±SE 5%) for poor prognosis signature patients (p-value not reported).

Distant metastasis (as first event) at 5 years
<table>
<tr><th>Method</th><th>HR</th><th>95% CI</th><th>p-value</th></tr>
<tr><td>MammaPrint</td><td>5.7</td><td>1.6 to 20</td><td><0.007[a]</td></tr>
<tr><td>Adjuvant! Online</td><td>4.6</td><td>0.61 to 35</td><td>0.14</td></tr>
<tr><td>NPI</td><td>2.2</td><td>0.78 to 6.5</td><td>0.14</td></tr>
<tr><td>CBO[102]</td><td>1.8</td><td>0.64 to 5.3</td><td>0.26</td></tr>
<tr><td>St Gallen</td><td>2.5</td><td>0.34 to 19</td><td>0.37</td></tr>
</table>
a The probability of remaining free of distant metastasis (as first event) was 98% (±SE 2%) for good and 78% (±SE 6%) for poor prognosis signature patients (p-value not reported).</td><td>The MammaPrint test is also an independent prognostic factor in node-negative breast cancer patients for women diagnosed in recent years</td><td>Data on NPI, OS, distant metastasis as first event. Additional data on updated follow-up results (median 10.2 years) reported by van de Vijver et al.[64]</td></tr>
</table>

Study	Outcomes/end points	Results	Authors' conclusions	Comments

Multivariate analysis with clinical risk indicators

OS, adjusted for performance of MammaPrint test (poor/high vs. good/low)

Method	HR	95% CI	*p*-value
Adjuvant! Online	3.0	1.0 to 8.9	0.04
NPI	2.7	0.87 to 8.1	0.09
CBO[102]	2.9	0.98 to 8.6	0.06
St Gallen	3.1	1.0 to 9.2	0.04

Distant metastasis (as first event), adjusted for performance of MammaPrint test (poor/high vs. good/low)

Method	HR	95% CI	*p*-value
Adjuvant! Online	4.8	1.3 to 17	0.018
NPI	5.4	1.4 to 21	0.015
CBO[102]	5.6	0.98 to 8.6	0.011
St Gallen	5.8	1.5 to 22	0.010

Gene expression profiling comparison between MammaPrint test and risk assessment based on (A) Adjuvant Online!, (B) St Gallen guidelines, (C) NPI and (D) CBO guidelines 2004[105]

A.

Clinical risk (Adjuvant! Online)	MammaPrint, *n* (%)			Discordant finding
	Good	Poor	Total	
Low	23 (19)	6 (5)	29 (24)	38% (47/123), 95% CI 30% to 47%, kappa: 0.252
High	41 (33)	53 (43)	94 (76)	
Total	64 (52)	59 (48)	123 (100)	

B.

Clinical risk (St Gallen guidelines)	MammaPrint, *n* (%)			Discordant finding
	Good	Poor	Total	
Low	15 (12)	1 (1)	16 (13)	41% (50/123), 95% CI 32% to 49%, kappa: 0.211
Intermediate/high	49 (40)	58 (47)	107 (87)	
Total	64 (52)	59 (48)	123 (100)	

Study	Outcomes/end points	Results	Authors' conclusions	Comments

C.

Clinical risk (NPI)	MammaPrint, *n* (%)			Discordant finding
	Good	Poor	Total	
Low	46 (37)	14 (11)	60 (49)	26% (32/123), 95% CI 18% to 34%, kappa: 0.480
Moderate/high	18 (15)	45 (37)	63 (51)	
Total	64 (52)	59 (48)	123 (100)	

D.

Clinical risk (CBO guidelines[102])	MammaPrint, *n* (%)			Discordant finding
	Good	Poor	Total	
Low	41 (33)	14 (11)	55 (45)	30% (37/123), 95% CI 22% to 38%, kappa: 0.401
High	23 (19)	45 (37)	68 (55)	
Total	64 (52)	59 (48)	123 (100)	

Study	Outcomes/end points	Results	Authors' conclusions	Comments
Gevensleben *et al.* (2010)[94]	Comparison of risk prediction by the MammaPrint test with that using St Gallen criteria and Adjuvant! Online	**Gene expression profiling compared with current risk classifications**	The MammaPrint test provides improved prediction of recurrence risk compared with currently used guidelines	

	MammaPrint ($n=140$)		
	Low risk ($n=78$)	High risk ($n=62$)	Discordant finding
St Gallen criteria[a]			
Low ($n=7$)	6[b]	1	–
Intermediate ($n=123$)	70	53	
High ($n=10$)	2	8[c]	
Adjuvant! Online[c]			
Low ($n=45$)	33[b]	12	57 (41%)
High ($n=95$)	45	50[b]	

a St Gallen risk classification according to Goldhirsch *et al.*[95]
b These values were summed to obtain concordant findings.
c Adjuvant! Online risk classification according to Ravdin *et al.*[107]

Treatment advice

For 59/62 patients with a poor prognosis signature identified by the MammaPrint test, the clinical treatment was recorded. In total, 19 (32%) of these patients did not receive adjuvant systemic treatment other than endocrine therapy and were potentially undertreated. In contrast, 35/77 patients who were classified as having a good prognosis by the MammaPrint test, and for whom treatment was known, received chemotherapy and were potentially overtreated. As a result, the MammaPrint test would have resulted in altered treatment advice for 40% of patients

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Authors' conclusions</th><th>Comments</th></tr>
<tr><td>Ishitobi et al. (2010)[96]</td><td>DMFS (not defined)
Correlation between the MammaPrint test risk category and clinicopathological parameters (St Gallen criteria)</td><td>Classification and disease outcome
DMFS
At 5 years (probability): G1 (low risk): 100% vs. G2 (high risk): 94%, HR not reported
Risk classification and distant metastasis
Among the 102 patients, 20 were classified as low risk and 82 as high risk. Based on the 1998 St Gallen criteria[103] all patients were classified as intermediate or high risk. The 2009 St Gallen criteria[97] use more refined criteria to define the low-risk group and classify seven patients as low risk. This is still lower (7/100) than the number identified by the MammaPrint test (20/102) ($p = 0.009$). See table for further details
<table>
<tr><th rowspan="2">Group</th><th rowspan="2">Patients ($n = 102$)</th><th rowspan="2">DM</th><th rowspan="2">DF</th><th rowspan="2">p-value</th><th colspan="2">Predictive value (%)</th></tr>
<tr><th>PPV</th><th>NPV</th></tr>
<tr><td>MammaPrint</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>High risk</td><td>82</td><td>8</td><td>74</td><td>0.143</td><td>9.8</td><td>100</td></tr>
<tr><td>Low risk</td><td>20</td><td>0</td><td>20</td><td></td><td></td><td></td></tr>
<tr><td>St Gallen 1998[103] version[a]</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Intermediate or high risk</td><td>100</td><td>8</td><td>92</td><td>NA</td><td>8.0</td><td>100</td></tr>
<tr><td>Low risk</td><td>0</td><td>0</td><td>0</td><td></td><td></td><td></td></tr>
<tr><td>St Gallen 2009[97] version[a]</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Intermediate or high risk</td><td>93</td><td>8</td><td>85</td><td>0.419</td><td>8.6</td><td>100</td></tr>
<tr><td>Low risk</td><td>7</td><td>0</td><td>7</td><td></td><td></td><td></td></tr>
</table>
DF, disease free; DM, distant metastasis; NA, not available.
a Criteria with missing data.</td><td>The MammaPrint test accurately identified Japanese breast cancer patients at low risk of developing recurrences. In fact, 100% of the individuals in the low-risk group remained metastasis free for the duration of the observation period</td><td></td></tr>
</table>

Study	Outcomes/end points	Results	Authors' conclusions	Comments
Kok *et al.* (2010)[99]	BCSS (defined as time from surgery to breast cancer-related death)	**Classification and disease outcome** BCSS in patients treated with adjuvant tamoxifen according to the MammaPrint test At 5 years: G1 (low risk): 96.2% (±SE 2.2%) vs. G2 (high risk): 72.5% (±SE 7.4%); p=NR; univariate HR: NR At 10 years: G1 (low risk): 80.6% (±SE 4.7%) vs. G2 (high risk): 63.4% (±SE 8.2%); p=NR; univariate HR: 2.78 (95% CI 1.30 to 5.94; p=0.008) BCSS in patients treated with adjuvant tamoxifen according to endocrine response categories (St Gallen consensus: highly endocrine responsive: ER and PR ≥ 50%; incompletely endocrine responsive: ER and/or PR low or with either one absent) At 5 years: G1 (high response): 98.0% (±SE 2.0%) vs. G2 (incomplete response): 82.1% (±SE 4.7%); p=NR; univariate HR: NR At 10 years: G1 (high response): 93.1% (±SE 3.9%) vs. G2 (incomplete response): 61.3% (±SE 6.3%); p= NR; univariate HR: 7.22 (95% CI 2.17 to 24.00; p=0.001) BCSS in patients treated with no adjuvant systemic treatment according to MammaPrint test At 5 years: G1 (low risk): 97.6% (±SE 1.6%) vs. G2 (high risk): 80.9% (±SE 5.0%); p=NR; univariate HR: NR At 10 years: G1 (low risk): 90.2% (±SE 3.3%) vs. G2 (high risk): 63.3% (±SE 6.3%); p=NR; univariate HR: 4.52 (95% CI 2.01 to 10.2; p<0.001) BCSS in patients treated with no adjuvant systemic treatment according to endocrine response categories At 5 years: G1 (high response): 92.9% (±SE 2.6%) vs. G2 (incomplete response): 85.7% (±SE 5.0%); p=NR; univariate HR: NR At 10 years: G1 (high response): 82.0% (±SE 4.0%) vs. G2 (incomplete response): 72.6% (±SE 6.5%); p=NR; univariate HR: 1.78 (95% CI 0.86 to 3.66; p=0.118)		Data also available for combined use of the MammaPrint test and endocrine response categories but not extracted

Study	Outcomes/end points	Results	Authors' conclusions	Comments
Kunz *et al.* (2011)[92]	Comparison of risk prediction by the MammaPrint test with that by the St Gallen guidelines 2007/9 and Adjuvant! Online	**GEP compared with current risk classifications** Method / Low risk (*n*) / Intermediate risk (*n*) / High risk (*n*) MammaPrint: 29 / – / 15 St Gallen criteria[a]: 4 / 34[b] / 6 Adjuvant! Online[c]: 19 / – / 25 a St Gallen risk classification according to Goldhirsch *et al.*[97] b In the group of women with intermediate risk, the MammaPrint test assigned 23 patients to low risk and 11 to high risk. c Patients were classified as having low clinical risk when the 10-year OS rate as predicted by Adjuvant! Online was > 88% for ER+ tumours and > 92% for ER– tumours. **Risk assessment by the MammaPrint test according to nodal status in women with early breast cancer** Low risk (*n*) / High risk (*n*) LN+ disease: 13 / 5 LN– disease: 19 / 9 **Comparison of the clinicopathological features with those of previous validation studies** Data reported but not extracted	Using gene expression analysis as an additional tool, patients with an intermediate clinical risk can be accurately separated into low- and high-risk groups. The gene expression analysis provides more accurate information on recurrence risk than conventional clinicopathological criteria	

DOI: 10.3310/hta17440

Study	Outcomes/end points	Results	Authors' conclusions	Comments
Mook *et al.* (2010)[75]	DMFS (defined as time from surgery to distant metastasis as first event; counted as failures) BCSS (defined as time from surgery to breast cancer-related death) Comparison of risk prediction by the MammaPrint test with that by Adjuvant! Online	**Classification and disease outcome** *DMFS* At 5 years: G1 (low risk): 93% (±SE 3%) vs. G2 (high risk): 72% (±SE 6%); $p=0.07$; univariate HR: 4.6 (95% CI 1.8 to 12.0; $p=0.001$) At 10 years: G1: 80% (±SE 5%) vs. G2: 67% (±SE 7%); $p=$NR; univariate HR: NR Over entire follow-up period: univariate HR: 1.8 (95% CI 0.9 to 3.5; $p=0.07$) *BCSS* At 5 years: G1: 99% (±SE 1%) vs. G2: 80% (±SE 5%); $p=0.036$; univariate HR: 19.1 (95% CI 2.5 to 148; $p=0.005$) At 10 years: G1: 90% (±SE 4%) vs. G2: 69% (±SE 6%); $p=$NR; univariate HR: NR Over entire follow-up period: univariate HR: 2.0 (95% CI 1.0 to 4.0; $p=0.04$) **Prediction of early BCSD** *BCSS* At 5 years: adjusted HR: 14.4 (95% CI 1.7 to 122.2; $p=0.01$) At 10 years: adjusted HR: 4.4 (95% CI 1.4 to 13.6; $p=0.01$) *Subgroup analyses: BCSS in hormonal therapy-naive patients (untreated)* At 5 years: adjusted HR: 10.8 (95% CI 1.2 to 94.7; $p=0.03$) **GEP compared with current risk classifications**	The MammaPrint test can accurately select postmenopausal patients at low risk of breast cancer-related death within 5 years of diagnosis and can be of clinical use in selecting postmenopausal women for adjuvant chemotherapy	Data on distant metastasis as first event

	MammaPrint test ($n=148$) (n)		
Adjuvant! Online[a]	**Low risk ($n=91$)**	**High risk ($n=57$)**	**Discordant finding**
Low risk ($n=74$)	62[b]	12	41 (28%)
High risk ($n=74$)	29	45[b]	

a Patients were classified as having low clinical risk when the 10-year OS rate as predicted by Adjuvant! Online was > 88% for ER+ tumours and > 92% for ER– tumours.
b These values were summed to obtain concordant findings.

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Authors' conclusions</th><th>Comments</th></tr>
<tr><td>Na et al. (2011)[100]</td><td>Comparison of risk prediction by the MammaPrint test with that by the St Gallen criteria, NIH guidelines[101] and Adjuvant! Online</td><td>Gene expression profiling compared with current risk classifications
The MammaPrint test identified five patients with a low-risk prognosis signature and 31 patients with a high-risk prognosis signature. Clinical risk was concordant with the prognosis signature for 29 patients according to the St Gallen guidelines; 30 patients according to the NIH guidelines[101] and 23 patients according to Adjuvant! Online
<table>
<tr><th></th><th colspan="2">MammaPrint test (n=36) (n)</th><th></th></tr>
<tr><th></th><th>Low risk (n=5)</th><th>High risk (n=31)</th><th>Discordant finding</th></tr>
<tr><td>St Gallen criteria[a]</td><td></td><td></td><td></td></tr>
<tr><td>Low risk (n=6)</td><td>2[b]</td><td>4</td><td>7 (19%)</td></tr>
<tr><td>High risk (n=30)</td><td>3</td><td>27[b]</td><td></td></tr>
<tr><td>[c]NIH guidelines[101]</td><td></td><td></td><td></td></tr>
<tr><td>Low risk (n=5)</td><td>2[b]</td><td>3</td><td>6 (17%)</td></tr>
<tr><td>High risk (n=31)</td><td>3</td><td>28[b]</td><td></td></tr>
<tr><td>Adjuvant! Online[d]</td><td></td><td></td><td></td></tr>
<tr><td>Low risk (n=14)</td><td>3[b]</td><td>11</td><td>13 (36%)</td></tr>
<tr><td>High risk (n=22)</td><td>2</td><td>20[b]</td><td></td></tr>
</table>
a St Gallen risk classification guideline according to Goldhirsch et al.:[95] a low clinical risk was defined as possessing all of the following criteria: ER+ and/or PR+ status, tumour size ≤2 cm, histological grade I and age ≥35.
b These values were summed to obtain concordant findings.
c Low risk for the LN– group was defined as a tumour size <1 cm and a favourable histological subtype such as tubular and mucinous cancer.
d Patients were classified as having low clinical risk when the 10-year OS rate as predicted by Adjuvant! Online was >88% for ER+ tumours and >92% for ER– tumours.

Comparison of the clinicopathological features with those of previous validation studies
Data reported but not extracted</td><td>The results of the MammaPrint test for Korean patients with breast cancer were somewhat different from those identified in Europe. This difference should be studied to determine whether or not there is a gene disparity between Asians and Europeans. Further large-scale studies with a follow-up evaluation are required to assess whether or not the use of the MammaPrint test can predict the prognosis of Korean patients with breast cancer</td><td></td></tr>
</table>

NR, not reported.

Appendix 10

MammaPrint and BluePrint tests: quality assessment and summary of results

Methodological quality assessment of studies investigating the MammaPrint and BluePrint tests

Study feature	Qualities sought	Stork-Sloots *et al.* (2009) (abstract)[114]
Sample of patients	Inclusion criteria defined	Y
	Sample selection explained	U
	Adequate description of diagnostic criteria	N
	Clinical and demographic characteristics fully described	N
	Representative (random or consecutive sample)	U
	Assembled at a common (usually early) point in the course of their disease	U
	Complete (all eligible patients were included)	U
Follow-up of patients	Sufficiently long	Y (5 years)
Outcome	Objective	U
	Unbiased (e.g. assessment blinded to prognostic information)	U
	Fully defined	U
	Appropriate	U
	Known for all or a high proportion of patients	U
Prognostic variable	Fully defined, including details of method of measurement if relevant	U
	Precisely measured	U
	Available for all or a high proportion of patients	U
	If relevant, cut-point(s) defined and justified	U
Analysis	Continuous predictor variable analysed appropriately	U
	Statistical adjustment for all important prognostic factors	U
Intervention subsequent to inclusion in cohort	Fully described	U
	Intervention standardised or randomised	U

N, no; U, unclear/not reported; Y, yes.

Summary of results: MammaPrint and BluePrint tests

Study	Outcomes/end points	Results	Authors' conclusions	Comments
Stork-Sloots *et al.* (2009)[114] (abstract)	Five-year survival	Profile classified: 66% (712) luminal-like; 18% (194) *ERBB2*-like; 16% (173) basal-like 13% of the samples positive for ER/PR did not express a luminal-like gene profile *ERBB2*-like or basal-like profiles showed equally poor 5-year survival rates of ~65% *ERBB2*-like subset of MammaPrint low-risk patients (15%) showed an 89% (95% CI 71% to 100%) survival rate without trastuzumab treatment Luminal-like subtypes separated into high and low risk by MammaPrint showed survival rates of 56% (95% CI 46% to 68%) for high risk and 94% (95% CI 90% to 99%) for low risk	The developed multigene profile can classify breast tumours into luminal-, *ERBB2*- and basal-like subgroups. By combining this molecular subtyping with the MammaPrint risk classification, specific groups of patients can be recognised who are at high risk of recurrence. The low-risk patients within the luminal- and *ERBB2*-like subclasses have a very low risk of recurrence. Implementation of this knowledge can improve the clinical management of breast cancer patients	

Appendix 11

PAM50 test: quality assessment and summary of results

Methodological quality assessment of studies investigating the PAM50 test

Study feature	Qualities sought	Bernard *et al.* (2011)[119] (abstract: analytical validity) [Additional data from Martin *et al.*[120] (abstract: clinical utility)]	Cheang *et al.* (2011)[117]	Chia *et al.* (2011)[121]	Ebbert *et al.* (2011)[118] (abstract)	Nielsen *et al.* (2010)[116]	Parker *et al.* (2009)[115]
Sample of patients	Inclusion criteria defined	U	Y	Y	U	Y	Y
	Sample selection explained	U	Y	Y	U (breast samples – no further details provided)	Y	Y
	Adequate description of diagnostic criteria	U	Y	Y	U	Y	Y
	Clinical and demographic characteristics fully described	U	Y	Y	U	Y	Y
	Representative (random or consecutive sample)	U	Y (random – data and samples from a RCT)	Y (random – data and samples from a RCT)	U	U (cohort – cases with complete outcomes and representative sample)	U (cohort – tissues collected under approved protocols)
	Assembled at a common (usually early) point in the course of their disease	U (all LN+)	Y (all LN+)	U (75% LN+)	U	U (65% LN+)	U (>85% LN–)
	Complete (all eligible patients were included)	U	Y	N	U	N	U

Study feature	Qualities sought	Bernard *et al.* (2011)[119] (abstract: analytical validity) [Additional data from Martin *et al.*[120] (abstract: clinical utility)]	Cheang *et al.* (2011)[117]	Chia *et al.* (2011)[121]	Ebbert *et al.* (2011)[118] (abstract)	Nielsen *et al.* (2010)[116]	Parker *et al.* (2009)[115]
Follow-up of patients	Sufficiently long	Y (8.7 years)	Y	Y (median 9.7 years)	NA	Y	U
Outcome	Objective	Y	Y	Y	NA	Y	Y
	Unbiased (e.g. assessment blinded to prognostic information)	U	Y	U	NA	U	U
	Fully defined	U	Y	N	NA	Y	N
	Appropriate	Y (DFS, OS)	Y (RFS, OS)	Y (RFS, OS)	NA (analytical)	Y (RFS, DSS)	Y (RFS)
	Known for all or a high proportion of patients	U	Y	Y	NA	Y	Y
Prognostic variable	Fully defined, including details of method of measurement if relevant	U	Y	Y	U	Y	Y
	Precisely measured	U	Y	Y	U	Y	Y
	Available for all or a high proportion of patients	U	Y	Y	U	Y	Y
	If relevant, cut-point(s) defined and justified	U	Y (detailed)	Y (detailed)	U	Y (detailed)	Y (provided reference)
Analysis	Continuous predictor variable analysed appropriately	U	Y	Y	U	Y	Y
	Statistical adjustment for all important prognostic factors	U	Y (includes multivariate analysis)	Y (includes multivariate analysis)	U	Y (includes multivariate analysis)	Y (includes multivariate analysis)
Intervention subsequent to inclusion in cohort	Fully described	U	Y	Y	U	N	N
	Intervention standardised or randomised	U	Y	Y	U	N	N

DSS, disease-specific survival; N, no; NA, not applicable; U, unclear/not reported; Y, yes.

Summary of results: PAM50 test

Study	Outcomes/end points	Results	Authors Conclusions	Comments
Bernard *et al.*, 2011[119] (abstract) Additional data from Martin *et al.*[120] (abstract)	Analytical outcomes including accuracy and reproducibility	**Bernard *et al.*[119]** There was good agreement between RT-qPCR gene expression and IHC scoring for the clinical markers (gene/protein) *ESR1*/ER, *PGR*/PR and *ERBB2*/HER2. The accuracy was significantly lower for *MKI67*/Ki-67, *EGFR*/EGFR and *KRT5*/CK5/6. Discrepancies between the Herceptest and CISH for a test score of 2+ and 3+ samples showed that RT-qPCR agreed better with the Herceptest (AUC 0.95 vs. 0.93) **Martin *et al.*[120]** Concerning predictive factors, exploratory analyses showed that fluorouracil, epirubicin, cyclophosphamide and paclitaxel (FEC-P) was better than FEC in the low PR group (HR: 0.68, $p=0.033$) and not in the high PR group (HR: 0.83, $p=0.245$); interaction test $p=0.358$. Similarly, FEC-P was better in the low *ERBB2* group (HR: 0.67, $p=0.005$) and not in the high *ERBB2* group (HR: 0.92, $p=0.707$); interaction test $p=0.256$. In addition, superiority of FEC-P was observed for the low proliferation signature group (HR: 0.58, $p=0.014$) in contrast to the high proliferation signature group (HR: 0.93, $p=0.633$); interaction test $p=0.069$. The FEC-P group showed improved outcomes in all genomic intrinsic subtypes, although no subtype alone reached statistical significance	Calling cut-points based on RT-qPCR expression across subtypes is reproducible across data sets and has good agreement with expression by IHC for clinically used biomarkers. In addition, the PAM50 proliferation signature could be predictive of benefit for adding weekly paclitaxel to the adjuvant chemotherapy FEC regimen. These results need further validation in an independent study	

Study	Outcomes/end points	Results	Authors Conclusions	Comments
Cheang *et al.* (2011)[117] (additional data from unpublished manuscript)	Responsiveness of intrinsic subtypes to adjuvant anthracyclines vs. non-anthracyclines	**Intrinsic subtyping using the PAM50 assay** Subtype: Tumour samples analysed ($n=476$), n (%) Luminal A: 146 (30.7) Luminal B: 110 (23.1) HER2 enriched: 105 (22.1) Basal-like: 94 (19.7) Normal-like: 21 (4.4) **Association of intrinsic subtypes with survival** Intrinsic subtypes were significantly associated with RFS ($p=0.0005$) and OS ($p<0.0001$) on the combined cohort. The HER2-enriched subtype demonstrated the greatest benefit from FEC vs. CMF, with an absolute difference of more than 20% in both 5-year RFS and OS, whereas there was a < 2% difference for the non-HER2-enriched tumours (interaction $p=0.03$ for RFS and $p=0.02$ for OS). Within tumours defined clinically as HER2+ by IHC or fluorescence in situ hybridisation, 79% (72/91) were classified as the HER2-enriched subtype by genomics and these tumours were also significantly associated with better response to CEF vs. CMF (62% vs. 22%, $p=0.0006$). In contrast, basal-like tumours ($n=94$) did not benefit from the substitution of methotrexate for epirubicin with a HR of 1.1 for RFS and 1.3 for OS in favour of methotrexate, but the test for interaction was not significant	The HER2-enriched assignment strongly predicted anthracycline sensitivity. The chemotherapy-sensitive basal-like tumours showed no benefit for CEF, suggesting that non-anthracycline regimens should be further investigated in this subtype	Additional data available but not extracted

Study	Outcomes/end points	Results	Authors Conclusions	Comments
Chia *et al.* (2011)[121]	Intrinsic subtyping, disease-free survival, OS, risk of relapse modelling	**Intrinsic subtyping using the PAM50 assay** Subtype / Tumour samples analysed ($n=398$), n (%)[a] Luminal A: 135 (34) Luminal B: 76 (19) HER2 enriched: NR Basal-like: NR Normal-like: NR NR, not reported. a All numbers calculated based on reported percentages. **Intrinsic subtyping comparing the PAM50 assay with IHC** The concordance for intrinsic subtypes among 348 patients who could be classified by both the IHC and PAM50 classifiers was 70.8%, 86.8%, 80.2% and 93.4% for luminal A, luminal B, HER2 enriched (and basal-like breast cancers, respectively, with an overall kappa of 0.57 (95% CI 0.51 to 0.64) **Association of intrinsic subtypes with survival** Intrinsic subtypes as classified by the PAM50 assay were prognostic for both DFS ($p=0.0003$) and OS ($p=0.0002$), with the HER2-enriched subtype having the lowest and the luminal A subtype the highest 5-year survival values (DFS: 52.8% vs. 84.2%; OS: 68.1% vs. 95.7%). The prognostic value remained significant for both DFS ($p=0.02$) and OS ($p=0.02$) in multivariate analysis. Classification by the IHC panel was not statistically significant **Prediction of tamoxifen benefit** Luminal subtype by PAM50 was predictive of tamoxifen benefit (DFS: HR 0.52, 95% CI 0.32 to 0.86 vs. HR 0.80, 95% CI 0.50 to 1.29 for non-luminal subtype), although the interaction was not significant ($p=0.24$). Neither subtyping by central IHC nor by local ER status was predictive	Intrinsic subtype classification with the PAM50 assay was superior to IHC profiling for both prognosis and prediction of benefit from adjuvant tamoxifen	Additional data available but not extracted

Study	Outcomes/end points	Results	Authors Conclusions	Comments
Ebbert *et al.*, (2011)[118] (abstract)	Analytical outcomes including accuracy and reproducibility	Within-platform cross-validation of the clinical subtype predictor showed 91.6% concordance. There was 100% reproducibility in subtype predictions across 46 runs testing different subtypes. Subtype predictions across platforms showed 88.1% concordance. Dilution experiments, introducing 'normal' breast tissue RNA into breast cancer RNA, showed a systematic switch towards the 'normal' signature, with luminal A and luminal B subtypes being most susceptible	The PAM50 Breast Cancer Intrinsic Classifier is highly reproducible within and across platforms. The clinical test has utility in the management of ER+ and ER– invasive breast cancers of all stages. It provides a necessary tool for identifying differences in tumour biology that are important for guiding patient care	
Nielsen *et al.*, (2010)[116]	Numbers assigned to each intrinsic subtype, risk of relapse score Comparators: clinical, IHC (ER, PR, HER2, Ki-67) Adjuvant! Online used to generate breast cancer RFS and disease-specific survival (DSS) estimates for each patient	**Prognosis – intrinsic subtypes and risk of relapse by PAM50 and comparators** *Intrinsic subtyping of ER+, tamoxifen-treated breast cancer using the PAM50 assay* **Subtype** / **ER+, tamoxifen-treated systemic therapy ($n=786$), n (%)** Luminal A: 372 (47.3) Luminal B: 329 (41.9) HER2 enriched: 64 (8.1) Basal-like: 5 (0.6) Normal-like: 16 (2.0) *Kaplan–Meier survival analysis of intrinsic subtypes and ROR-S, as determined by PAM50* The included patients were considered to be at high risk with overall 10-year RFS of 62% and DSS of 72%. Those assigned to luminal A had significantly a better outcome (10-year RFS 74%; DSS 83%) than those assigned to luminal B, HER2-enriched and basal-like tumours In Cox models incorporating standard prognostic variables, HRs for BCSS over the first 5 years of follow-up, relative to the most common luminal subtype, were 1.99 (95% CI 1.09 to 3.64) for the luminal B subtype, 3.65 (95% CI 1.64 to 8.16) for the HER2-enriched subtype and 17.71 (95% CI 1.71 to 183.33) for the basal-like subtype ($p=0.0018$)	C-index, Kaplan–Meier analysis and Cox model analyses show that IHC approaches do work and provide significant prognostic information; however, PAM50 is superior in terms of adding significant additional information and in its capacity to identify a particularly low-risk group. PAM50 can be applied to a large sample of FFPE breast cancers and gives more prognostic information than clinical factors and IHC	Authors highlight that the studied population was biased towards higher-risk breast cancers and thus underestimates the broader N0 population for whom adjuvant tamoxifen would represent adequate treatment Additional data available but not extracted

ROR-S, risk of recurrence score based on subtype.

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Authors Conclusions</th><th>Comments</th></tr>
<tr><td>Parker et al. (2009)[115]</td><td>Distribution of instrinsic subtypes in comparison with clinical marker status
Risk of relapse models for prognosis in LN– breast cancer</td><td>Intrinsic subtyping using the PAM50 assay
<table><tr><th>Subtype</th><th>No adjuvant systemic therapy (n=761), n (%)[a]</th></tr><tr><td>Luminal A</td><td>269 (35.3)</td></tr><tr><td>Luminal B</td><td>168 (22.1)</td></tr><tr><td>HER2 enriched</td><td>120 (15.8)</td></tr><tr><td>Basal-like</td><td>128 (16.8)</td></tr><tr><td>Normal-like</td><td>76 (10.0)</td></tr></table>a All numbers calculated based on reported percentages.

Comparison of relapse prediction models
The intrinsic subtypes showed prognostic significance (for RFS) in untreated (no systemic therapy) patients and remained significant in multivariable analyses that incorporated clinical covariates (ER status, histological grade, tumour size and LN status)</td><td>The intrinsic subtype and risk predictors based on the PAM50 gene set added significant prognostic and predictive value to pathological staging, histological grade and standard clinical molecular markers</td><td>Neoadjuvant data available in paper but not extracted</td></tr>
</table>

Appendix 12

Breast Cancer Index: quality assessment and summary of results

Methodological quality assessment of study investigating the Breast Cancer Index

Study feature	Qualities sought	Jerevall *et al.* (2011)[122]
Sample of patients	Inclusion criteria defined	Y
	Sample selection explained	Y
	Adequate description of diagnostic criteria	Y
	Clinical and demographic characteristics fully described	Y
	Representative (random or consecutive)	Y (random)
	Assembled at a common (usually early) point in the course of their disease	Y
	Complete (all eligible patients included)	N
Follow-up of patients	Sufficiently long	Y
Outcome	Objective	Y
	Unbiased (e.g. assessment blinded to prognostic information)	U
	Fully defined	Y
	Appropriate	Y
	Known for all or a high proportion of patients	Y
Prognostic variable	Fully defined, including details of method of measurement if relevant	Y
	Precisely measured	Y (detail provided)
	Available for all or a high proportion of patients	Y
	If relevant, cut-point(s) defined and justified	Y (detail provided)
Analysis	Continuous predictor variable analysed appropriately	Y
	Statistical adjustment for all important prognostic factors	Y
Intervention subsequent to inclusion in cohort	Fully described	Y
	Intervention standardised or randomised	Y

N, no; U, unclear/not reported; Y, yes.

Summary of results: Breast Cancer Index

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Authors' conclusions</th><th>Comments</th></tr>
<tr><td>Jerevall et al. (2011)[122]</td><td>Time to distant metastasis
DMFS
BCSD</td><td>
Development and testing of HOXB13:IL17BR (H:I) + MGI as a continuous index (BCI)
Developed a continuous algorithm using G1 (training set) and tested on G2 (test set)
G1: BCI classified 59% as low risk (rate of distant recurrence: 1.7%, 95% CI 0% to 3.5%; rate of death: 1.1%, 95% CI 0% to 2.6%); 22% as intermediate risk (rate of distant recurrence: 17.8%, 95% CI 7.6% to 26.8%; rate of death: 14.5%, 95% CI 5.2% to 22.9%); and 18.4% as high risk (rate of distant recurrence: 20.0%, 95% CI 8.7% to 30.0%; rate of death: 14.7%, 95% CI 4.7% to 23.6%)
G2: BCI classified 53% as low risk (rate of distant recurrence: 8.3%, 95% CI 4.7% to 14.4%; rate of death: 5.1%, 95% CI 1.3% to 8.7%); 27% as intermediate risk (rate of distant recurrence: 22.9%, 95% CI 14.5% to 35.2%; rate of death: 19.8%, 95% CI 10.0% to 28.6%); and 20% as high risk (rate of distant recurrence: 28.5%, 95% CI 17.9% to 43.6%; rate of death: 28.8%, 95% CI 15.3% to 40.2%)
BCI is a strong prognostic factor for distant recurrence and BCSD independent of tumour size, grade, HER2 status and PR status (although tumour size did contribute prognostic value to distant recurrence)
<table>
<tr><th></th><th>St Gallen</th><th>BCI</th></tr>
<tr><td>G1</td><td></td><td></td></tr>
<tr><td>Low risk (rate of recurrence, 95% CI) (%)</td><td>22 (5.2, –0.5 to 10.9)</td><td>59 (7.1, 0 to 3.5)</td></tr>
<tr><td>Intermediate risk (rate of recurrence, 95% CI) (%)</td><td>78 (8.5, 4.9 to 12.1)</td><td>22 (17.8, 7.6 to 26.8)</td></tr>
<tr><td>High risk (rate of recurrence, 95% CI) (%)</td><td>NA</td><td>18.4 (20.0, 8.7 to 30.0)</td></tr>
<tr><td>G2</td><td></td><td></td></tr>
<tr><td>Low risk (rate of recurrence, 95% CI) (%)</td><td>19 (8.8, 1.4 to 16.2)</td><td>5.3 (8.3, 4.7 to 14.4)</td></tr>
<tr><td>Intermediate risk (rate of recurrence, 95% CI) (%)</td><td>81 (17.0, 11.7 to 22.3)</td><td>27 (22.9, 14.5 to 35.2)</td></tr>
<tr><td>High risk (rate of recurrence, 95% CI) (%)</td><td>NA</td><td>20 (28.5, 17.9 to 43.6)</td></tr>
</table>
NA, not available.

Prognostic utility of BCI also assessed in comparison to Adjuvant! Online for G2. Both BCI and Adjuvant! Online were significant predictors of BCSD (BCI: HR 2.3, 95% CI 1.5 to 3.7, $p<0.001$; Adjuvant Online: HR 1.4, 95% CI 1.0 to 1.8, $p<0.04$) and distant recurrence (BCI: HR 2.0, 95% CI 1.3 to 3.1, $p<0.001$; Adjuvant! Online: HR 1.4, 95% CI 1.0 to 1.8, $p<0.003$) (calculated HRs are relative to an increment of their interquartile ranges: 2.484 for BCI, 6 for Adjuvant Online!)
</td><td>Retrospective analysis of this randomised, prospective trial cohort validated the prognostic utility of H:I + MGI and was used to develop and test a continuous risk model that enables prediction of distant recurrence risk at the patient level</td><td>Only data related to the BCI have been extracted (rather than data on H:I + MGI)</td></tr>
</table>

Appendix 13

Mammostrat test: quality assessment and summary of results

Methodological quality assessment of studies investigating the Mammostrat test

Study feature	Qualities sought	Bartlett *et al.* (2010)[124]	Ring *et al.* (2009)[125]	Ross *et al.* (2008)[126]
Sample of patients	Inclusion criteria defined	Y	Y	Y
	Sample selection explained	Y	Y	Y
	Adequate description of diagnostic criteria	Y	Y	Y
	Clinical and demographic characteristics fully described	Y	N (NA for one of the cohorts used)	Y
	Representative (random or consecutive sample)	Y (consecutive)	U (unclear if either)	Y (from a RCT)
	Assembled at a common (usually early) point in the course of their disease	Y	U	U
	Complete (all eligible patients were included)	Y	U	U
Follow-up of patients	Sufficiently long	Y	U	Y
Outcome	Objective	Y	Y	Y
	Unbiased (e.g. assessment blinded to prognostic information)	Y	U	Y
	Fully defined	Y	Y	Y
	Appropriate	Y	Y	Y
	Known for all or a high proportion of patients	Y	U	U
Prognostic variable	Fully defined, including details of method of measurement if relevant	Y	Y	Y
	Precisely measured	Y (detail provided)	Y (detail provided)	Y (detail provided)
	Available for all or a high proportion of patients	Y	Y	U
	If relevant, cut-point(s) defined and justified	Y (reference provided)	Y (detail provided)	Y (detail provided)
Analysis	Continuous predictor variable analysed appropriately	Y	Y	Y
	Statistical adjustment for all important prognostic factors	U	Y	Y
Intervention subsequent to inclusion in cohort	Fully described	Y	N	Y (from prespecified treatment arms)
	Intervention standardised or randomised	N	N	Y

N, no; NA, not available; U, unclear/not reported; Y, yes.

Summary of results: Mammostrat test

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Authors' conclusions</th><th>Comments</th></tr>
<tr><td>Bartlett et al. (2010)[124]</td><td>DRFS
RFS
OS</td><td>
Assignment to risk groups
<table>
<tr><th></th><th>All cases (n=1540), n (%)</th><th>G1: all ER+ (n=1189), n (%)</th><th>G2: ER+, tamoxifen only (n=831), n (%)</th><th>G3: ER+, N0, tamoxifen only (n=657), n (%)</th></tr>
<tr><td>Low risk</td><td>717 (46.6)</td><td>643 (54.1)</td><td>444 (53.4)</td><td>341 (51.9)</td></tr>
<tr><td>Moderate risk</td><td>305 (19.8)</td><td>244 (20.5)</td><td>175 (21.1)</td><td>139 (21.2)</td></tr>
<tr><td>High risk</td><td>278 (18.1)</td><td>168 (14.1)</td><td>112 (13.5)</td><td>88 (13.4)</td></tr>
<tr><td>Missing</td><td>240 (15.6)</td><td>134 (11.3)</td><td>100 (12.0)</td><td>89 (13.5)</td></tr>
</table>
Significantly more cases assigned to high-risk group in ER– vs. ER+ (45% vs. 16%, $p<0.001$). No differences between other groups

All cases ($n=1300$): Significant association between risk score and RFS, DRFS and OS (all $p<0.001$). Multivariate analysis: risk score independent predictor of RFS ($p<0.001$), DRFS ($p<0.001$) and OS ($p<0.01$) (along with clinicopathological predictors)

G1 ($n=1055$): Significant association between risk score and DRFS ($p<0.001$), RFS ($p<0.001$) and OS ($p<0.001$). Multivariate analysis: risk score independent predictor of RFS ($p<0.05$), DRFS ($p<0.01$) and OS ($p<0.01$) (along with clinicopathological predictors)

G2 ($n=731$): Significant association between risk score and DRFS ($p<0.001$), RFS ($p<0.01$) and OS ($p<0.01$). Multivariate analysis: risk score independent predictor of DRFS ($p<0.05$), OS ($p<0.05$), trend for RFS ($p=0.064$) (along with clinicopathological predictors)

G3 ($n=568$): Significant association between risk score and RFS ($p<0.05$), DRFS ($p<0.01$) and trend for OS. Multivariate analysis: trend towards Mammostrat score to predict RFS ($p=0.076$) and DRFS ($p=0.092$) (along with clinicopathological predictors)
</td><td>Mammostrat can act as an independent prognostic tool for ER+, tamoxifen-treated breast cancer. This study revealed a possible association with outcome regardless of LN status and ER– tumours. These data provide further support for the use of this antibody panel as an aid to patient management in early breast cancer</td><td>Further data within the three groups were not extracted

Data on small number of ER– and untreated cases not extracted (reported to show a similar pattern to other groups)</td></tr>
</table>

Study	Outcomes/end points	Results	Authors' conclusions	Comments
Ring *et al.* (2006)[125]	DFS at 5 years	G1 (training/validation cohort, ER+): Cox model identified a group of patients as having either poor or moderate outcomes with a 5-year DFS of approximately 75% as opposed to patients classified as good with a 5-year DFS of approximately 95% ($p<0.001$) G2 (ER+ patients): Cox model identified poor patients with a 5-year DFS of 50% compared with approximately 70% for patients classified as moderate and 87% for patients classified as good ($p=0.008$) G3 (ER+ patients): Cox model identified ER+ patients classified as poor with OS of 55% compared with 75% for patients classified as moderate and 90% for patients classified as good ($p=0.0039$) In both cohorts the Cox model was independent of stage, grade and LN status Combined G2 and G3: for patients with poor or good prognosis (82%), sensitivity for poor prognosis in predicting disease progression was 38% whereas specificity was 88%. The PPV of poor prognosis was 38% (95% CI 32% to 44%) whereas the NPV was 88% (95% CI 84% to 92%)	The test can significantly improve on traditional prognosticators in predicting outcome for ER+ breast cancer patients	Only data relating to validation were extracted (information about training results in paper) Only data relating to the Cox models (not tree models) were extracted
MS – reclassification data All CIC	[CIC information has been removed]	[CIC information has been removed]	[CIC information has been removed]	

Study	Outcomes/end points	Results	Authors' conclusions	Comments
Ross *et al.* (2008)[126]	RFI DRFI BCSD	**Association between clinical outcomes and stratification by test** *Tamoxifen treated (n = 711)* ~58% low risk, 21% moderate risk, 21% high risk Significant association between patients stratified by test and RFI (HR 1.3, 95% CI 1.1 to 1.6, $p = 0.006$). Low risk vs. moderate risk not significant (log-rank, $p = 0.05$); low risk vs. high risk significant (HR 1.8, 95% CI 1.2 to 2.6) Significant association between patients stratified by test and DRFI (HR 1.4, 95% CI 1.1 to 1.7, $p = 0.001$). Low risk vs. moderate risk not significant; high risk vs. low risk significant (HR 2.1, 95% CI 1.4 to 3.1, $p = 0.0004$) Significant association between patients stratified by test and BCSD (HR 1.5, 95% CI 1.2 to 1.9, $p = 0.0003$). Low risk vs. moderate risk not significant; high risk vs. low risk significant (HR 2.3, 95% CI 1.5 to 3.5, $p < 0.0001$) Kaplan–Meier estimate of proportion of patients recurrence free after 10 years Overall: 82% (95% CI 79% to 85%) Low risk: 85% (95% CI 81% to 88%) Moderate risk: 85% (95% CI 80% to 91%) High risk: 73% (95% CI, 65% to 80%) Multivariate Cox model – significant prognostic power independent of age and tumour size (HR 1.3, 95% CI 1.1 to 1.6, $p = 0.007$) **Chemotherapy responsiveness** *Tamoxifen and cytotoxic chemotherapy treated (n = 269) vs. NSABP B20 tamoxifen only (n = 161)* Kaplan–Meier estimate of RFI Low-risk: improved by 5% from 86% to 91% (HR 0.4, 95% CI 0.2 to 0.8, $p = 0.01$) High risk: improved by 21% from 64% to 85% (HR 0.4, 95% CI 0.2 to 0.9, $p = 0.02$) Moderate risk: not significant Interaction: not significant *Placebo treated (n = 287)* Non-significant association between RFI at 10 years and stratification by test	The risk index was significantly associated with clinical outcome among the ER-expressing, LN–, tamoxifen-treated patients. It seems that the test may be able to identify patients who have greater absolute benefit from adjuvant chemotherapy compared with unstratified patient populations	Data on subsets of patients (e.g. by age) within the treatment arms have not been extracted

Appendix 14

IHC4 test: quality assessment and summary of results

Methodological quality assessment of the study investigating the IHC4 test

Study feature	Qualities sought	Cuzick *et al.* (2011)[84]
Sample of patients	Inclusion criteria defined	Y
	Sample selection explained	Y
	Adequate description of diagnostic criteria	Y
	Clinical and demographic characteristics fully described	Y
	Representative (random or consecutive sample)	Y
	Assembled at a common (usually early) point in the course of their disease	Y
	Complete (all eligible patients were included)	Y
Follow-up of patients	Sufficiently long	Y
Outcome	Objective	Y
	Unbiased (e.g. assessment blinded to prognostic information)	U
	Fully defined	U
	Appropriate	Y
	Known for all or a high proportion of patients	Y
Prognostic variable	Fully defined, including details of method of measurement if relevant	Y (reference provided)
	Precisely measured	Y
	Available for all or a high proportion of patients	Y
	If relevant, cut-point(s) defined and justified	Y
Analysis	Continuous predictor variable analysed appropriately	Y
	Statistical adjustment for all important prognostic factors	Y
Intervention subsequent to inclusion in cohort	Fully described	Y
	Intervention standardised or randomised	Y

U, unclear/not reported; Y, yes.

Summary of results: IHC4

Study	Outcomes/ end points	Results	Authors' conclusions	Comments
Cuzick *et al.* (2011)[84]	Distant recurrence (within 10 years) TTDR	G1: 195 recurrences of which 145 were distant recurrences; in LN– women there were 101 recurrences of which 67 were distant recurrences The median IHC4 score for all patients was –4.2 (IQR –29.9 to 29.9). The HR for a change from the 25th to the 75th percentile of the IHC4 score for all patient was 5.7 (95% CI 3.4 to 9.7) in univariate analysis and 3.9 (95% CI 2.4 to 6.7) when added to clinical score G2: IHC4 score was highly significantly predictive of outcome for a change from the 25th to 75th percentile in a univariate analysis (HR 4.8, 95% CI 2.2 to 10.2), and gave similar results when added to clinical score (HR 4.4, 95% CI 2.0 to 9.3, $\Delta\chi^2 = 26.61$, $p < 0.0001$)	Additional studies are needed to determine the general applicability of the IHC4 score	

Appendix 15

NPI+ test: quality assessment and summary of results

Methodological quality assessment of the studies investigating the NPI+ test

Study feature	Qualities sought	Green *et al.* and Nottingham Prognostics[128]	Nottingham Prognostics (abstract)[128]
Sample of patients	Inclusion criteria defined	Y	U
	Sample selection explained	U	U
	Adequate description of diagnostic criteria	N	U
	Clinical and demographic characteristics fully described	N	U
	Representative (random or consecutive)	U	U
	Assembled at a common (usually early) point in the course of their disease	U	U
	Complete (all eligible patients were included)	N	U
Follow-up of patients	Sufficiently long	Y	U
Outcome	Objective	Y	U
	Unbiased (e.g. assessment blinded to prognostic information)	U	U
	Fully defined	N	U
	Appropriate	Y	U
	Known for all or a high proportion of patients	Y	U
Prognostic variable	Fully defined, including details of method of measurement if relevant	Y	U
	Precisely measured	Y (detail provided)	U
	Available for all or a high proportion of patients	Y	U
	If relevant, cut-point(s) defined and justified	Y (detail provided)	U
Analysis	Continuous predictor variable analysed appropriately	Y	U
	Statistical adjustment for all important prognostic factors	Y	U
Intervention subsequent to inclusion in cohort	Fully described	U	U
	Intervention standardised or randomised	U	U

N, no; U, unclear/not reported; Y, yes.

Summary of results: NPI+ test

Study	Outcomes/end points	Results	Authors' conclusions	Comments
Green *et al.* and Nottingham Prognostics[128] (All AIC)	[CIC information has been removed]	[CIC information has been removed]	[CIC information has been removed]	[CIC information has been removed]
Nottingham Prognostics (abstract)[128] (All AIC)	[CIC information has been removed]	[CIC information has been removed]	[CIC information has been removed]	

NR, not reported.

Appendix 16

Tabulated summary of cost-effectiveness studies addressing the use of MammaPrint to guide the selection of adjuvant chemotherapy regimes in breast cancer management

Parameter	Retel *et al* (2010)[133]	Chen *et al* (2010)[132]
Country	Netherlands	USA
Perspective (costs)	Health care	Payer perspective
Comparators (NPI, Adjuvant! Online)	Adjuvant! Online vs. MammaPrint St Gallen vs. MammaPrint	Adjuvant! Online vs. MammaPrint
Starting age in the model	50 years	Unclear
Population	Early operable breast cancer, LN–, ER+, HER2+/–	≤60 years, ER+/–, LN–
Model structure (type, health states)	Markov model with four mutually exclusive health states (disease free, relapse, distant metastasis and death)	Markov model with three mutually exclusive health states (no recurrence, death from cancer, death from other causes)
Definition of relapse	Includes local, regional recurrence, secondary primary and contralateral breast cancer	Relapse included in terms of cost only (local, regional, contralateral, distant)
Time horizon	20 years	Lifetime (99% of patients dead)
Endocrine therapy regime	All patients are assumed to receive 2.5 years of tamoxifen followed by 2.5 years of aromatase inhibitor	Endocrine therapy only given to ER+ patients; tamoxifen for patients receiving endocrine therapy
Chemotherapy regime	80% receive 6 cycles of FEC; 10% receive 6 cycles of docetaxel, doxorubicin and cyclophosphamide (TAC); 10% receive AC + paclitaxel (4 + 12 cycles) in combination with trastuzumab	Cost based on the following chemotherapy regimens: alkylating agents (58%), anthracyclines (51%), taxanes (25%) and antimetabolites (18%)
Benefit of chemotherapy	HR for trastuzumab: 0.64 (95% CI 0.54 to 0.76)	Relative risk: 26% in ER+; 32% in ER–
Adverse events	CHF was included	Implicitly included in the cost of chemotherapy and QALYs
Quality of life	EQ-5D; utilities extracted from Lidgren *et al.*[148] No adjuvant systemic treatment (first year): 0.935 DFS (years 2–20): 0.935 Chemotherapy (year 1): 0.620 Endocrine therapy (years 1–5): 0.744 Trastuzumab (year 1): 0.620 CHF: 0.700 Relapse: 0.779 Distant recurrence: 0.685	Chemotherapy: 0.70 (6 months) Recurrence free: 0.98

Parameter	Retel *et al* (2010)[133]	Chen *et al* (2010)[132]
Costs and resources used	Cost expressed in 2005 euros; costs of health states extracted from Lidgren *et al.*[144] Chemotherapy: €8596 Endocrine therapy: €822 Trastuzumab: €36,298 CHF: €3453 Relapse (year 1): €12,181 Relapse (after): €2359 Distant metastasis (year 1): €14,303 Distant metastasis (after): €6813 MammaPrint: €2675	Cost expressed in 2007 US dollars Chemotherapy: $35,964 Cost no recurrence (per year): $5928 Recurrence: $57,424 Terminal (cancer): $76,557 Terminal (other): $65,016 MammaPrint: $4200 Endocrine therapy (per year): $1383
Discounting	Costs: 4%; benefits: 1.5%	Costs: 3%; benefits: unclear
% of HER2+	10%	HER2+ excluded as assumed to receive trastuzumab anyway

Results for St Gallen or NCCN[129] are not presented in this table.

Appendix 17

Critical appraisal checklist of the economic model comparing MammaPrint with Adjuvant! Online

	Modelling assessments should include:	Retel *et al.* (2010)[133]	Chen *et al.* (2010)[132]
1	A statement of the problem	Yes	Yes
2	A discussion of the need for modelling vs. alternative methodologies	Yes	Yes
3	A description of the relevant factors and outcomes	Yes	Yes
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 submodel	Yes	Yes
5	A description of the 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	Yes	Yes
6	A list of assumptions pertaining to the structure of the model (e.g. factors included, relationships and distributions) and the data	Yes	Yes
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	Yes	Yes
8	The results derived from applying the model for the base case	Yes	Yes
9	The results of the sensitivity analyses: unidimensional, best/worst case, multidimensional (Monte Carlo/parametric), threshold	Yes	Yes
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	Yes	Yes
11	A description of the validation undertaken, including the concurrence of experts, internal consistency, external consistency and predictive validity	Unclear	Unclear
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	Unclear	Unclear
13	A description of research in progress that could yield new data that could alter the results of the analysis	Yes	Unclear

Appendix 18

Tabulated summary of cost-effectiveness studies addressing the use of OncotypeDX to guide the selection of adjuvant chemotherapy regimes in breast cancer management

Parameters	Tsoi *et al.* (2010)[135]	OHTA (2011)[134,136]
Country	Canada	
Perspective (costs)	Health care	
Comparators (NPI, Adjuvant! Online)	Adjuvant! Online vs. Adjuvant! Online + RS	
Starting age in the model	50 years	
Population	ER+, LN–, HER2– early breast cancer	
Model structure (type, health states)	Markov model with five health states (risk reclassification, chemotherapy, recurrence free, distant recurrence and death)	
Definition of relapse	Distant metastases only	Distant metastases only
Time horizon	Lifetime	
Endocrine therapy regime	5 years of tamoxifen	
Chemotherapy regime	Four cycles of AC – in the base case	TC or FEC-D
Benefit of chemotherapy	30% relative risk reduction	Low: no benefit; intermediate: 39% reduction; high: 74% reduction
Adverse events	Minor (60%), major (5%)	
Definition of high risk	Intermediate was grouped with high risk for RS For Adjuvant! Online, arbitrary decision so that same proportion of low cases between Adjuvant! Online and RS	Adjuvant! Online: low risk: mortality < 9%; intermediate risk: 9% ≤ mortality < 17%; high risk: mortality ≥ 17%
Quality of life	Different valuation methods (VAS – SG) Adjuvant! Online/RS: 0.94 Major toxicity: 0.8 Minor toxicity: 0.9 No toxicity: 0.94 Recurrence free after chemotherapy: 0.98 Distant recurrence: 0.75	

VAS–SG, visual analogue scale – standard gamble.

Parameters	Tsoi *et al.* (2010)[135]	OHTA (2011)[134,136]
Costs and resources used	2008 Canadian dollars; inflation rate assumed to be 5% per year Oncotype: C$4,404 Chemotherapy per cycle: C$768.3 Major non-fatal toxicity: C$2459 Major fatal toxicity: C$28,385 Recurrence free (yearly): C$444 Tamoxifen (5 years): C$678 Distant metastases (21 months): C$35,023 Terminal care (last 3 months): C$21,367	The majority of costs have been adapted from Tsoi *et al.*[135] to reflect 2010 prices. The costs of OncotypeDX (C$4191) and chemotherapy have been updated
Discounting	5% for both costs and benefits	

Appendix 19

Critical appraisal checklist of the economic model comparing OncotypeDX with Adjuvant! Online

Note that Tsoi *et al.*[135] (2010) and the OHTA analysis[134,136] were assessed together as they were based on the same economic model.

	Modelling assessments should include	Tsoi *et al.* (2010)[135]/ OHTA[134,136]
1	A statement of the problem	Yes
2	A discussion of the need for modelling vs. alternative methodologies	Yes
3	A description of the relevant factors and outcomes	Yes
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 submodel	Yes
5	A description of the 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	Yes
6	A list of assumptions pertaining to the structure of the model (e.g. factors included, relationships and distributions) and the data	Yes
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	Yes
8	The results derived from applying the model for the base case	Yes
9	The results of the sensitivity analyses: unidimensional, best/worst case, multidimensional (Monte Carlo/parametric), threshold	Yes
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	Yes
11	A description of the validation undertaken, including the concurrence of experts, internal consistency, external consistency and predictive validity	Unclear
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	Unclear
13	A description of research in progress that could yield new data that could alter the results of the analysis	Yes

Appendix 20

Final scope

NATIONAL INSTITUTE FOR HEALTH AND CLINCIAL EXCELLENCE

Diagnostics Assessment Programme

Gene expression profiling and expanded immunohistochemistry tests to guide selection of chemotherapy regimes in breast cancer management

Final scope

April 2011

1 Introduction

The Medical Technologies Advisory Committee identified the Randox Breast Cancer Array (Randox BCA), a gene expression profiling test, as potentially suitable for evaluation by the Diagnostics Assessment Programme (DAP) on the basis of a briefing note. The Randox BCA is manufactured by Randox Laboratories Limited. This document has been updated following feedback from attendees at the scoping workshop held on 2 March 2011 and the assessment subgroup meeting held on 11 April 2011.The scope has been extended to include gene expression profiling and expanded immunohistochemistry tests for guiding selection of chemotherapy regimes in breast cancer management. The final scope outlines the approach for assessing the clinical and cost effectiveness components of this evaluation.

2 Target condition/indication

2.1 Breast cancer background

Breast cancer is the most common cancer in women in England. In 2008 there were 39,681 new cases diagnosed, an increase of 1,633 cases compared with 2007 (4%). Just over 10,000 women died from breast cancer in England in 2008, a rate of 26 deaths per 100,000 women. It is the second most common cause of cancer death in women, after lung cancer.

One in eight women will develop breast cancer at some point in their lives. Age is a known risk factor for developing breast cancer. Four out of every five new cases are diagnosed in women aged 50 and over, with cases peaking in the 60 to 64 age group (14% of all new cases).

Earlier detection and improved treatment for breast cancer have meant that survival rates have risen. Although incidence rates for breast cancer increased by more than 85 per cent between 1971 and 2008, mortality rates have fallen by 33% since 1971. Survival from breast cancer is higher than that for cervical cancer and much higher than that of other major cancers in women – lung, colorectal and ovarian.

2.2 Diagnosis

Sections 2.2 through 2.5 have been adapted from NICE clinical guideline – CG80 – Breast cancer (early & locally advanced).

In most cases, whether suspected at breast screening or through presentation to the GP, diagnosis in the breast clinic is made by triple assessment (clinical assessment, mammography and/or ultrasound imaging with core biopsy and/or fine needle aspiration cytology).

2.3 Primary systemic therapy (neoadjuvant therapy)

Neoadjuvant treatment in oncology is defined as additional treatment preceding the main therapy option; surgery is the main therapy option. Optimal management of breast cancer includes local control in the breast and the prevention of metastatic spread. Some patients will have developed occult metastatic spread before clinical or radiological detection of the primary tumour. There are also patients whose tumours at presentation are too large to be considered appropriate for breast conservation. Primary systemic therapy of invasive breast cancer may be offered in an attempt to enable breast conserving treatment and subsequent surgery (mastectomy or wide local excision). Histological examination is usually conducted to inform the treatment decision. Radiotherapy may then be offered according to similar criteria to those patients presenting de novo. Primary systemic treatment involves the use of systemic therapy, either chemotherapy or endocrine therapy, after diagnosis but before definitive surgery. Primary systemic therapy (also referred to as neoadjuvant therapy) can be successfully used to shrink the size of the primary tumour such that breast conservation may be achieved with a good cosmetic result but with a slightly higher risk of local recurrence comapred with mastectomy. Primary systemic therapy can also identify the efficacy of the systemic treatment regimen since the primary tumour is available to monitor response to the therapy. This option is of course not available if the primary tumour has been removed surgically. The use of primary systemic treatment allows targeting of occult metastatic tumour deposits at an earlier stage than the conventional approach of postoperative chemotherapy. Randomised trials of primary systemic therapy have failed to show a significant survival benefit, but more recent studies using current chemotherapy regimens have been able to identify subgroups of patients, such as those achieving complete pathological response at surgery, that have a survival advantage.

2.4 Surgery

Surgery is the mainstay of treatment for invasive breast cancer and is usually used as the first treatment option.

2.5 Postoperative assessment and adjuvant treatment planning

Following surgery, further information is obtained by histological examination, which provides prognostic information including histological grade, nodal status and tumour size. Factors predicting response to specific targeted therapies including hormone receptor and the human epidermal growth factor receptor 2 (HER2) statuses are also evaluated. These prognostic and predictive factors, together with patient characteristics, enable subsequent treatment planning to be undertaken by the breast cancer multidisciplinary team (MDT).

2.5.1 Predictive factors

Hormone receptors

Approximately 70% of invasive breast cancers are oestrogen receptor alpha (ER) positive and the level of ER assessed immunohistochemically provides useful predictive information regarding efficacy of endocrine therapy. ER status therefore forms part of the UK minimum dataset for histopathology reporting of invasive breast cancer. ER status is routinely determined on all invasive breast cancers and reported using a standardised technique (such as the Allred scoring

system). The prediction of likelihood of response of a breast cancer to endocrine therapies using ER assessment is not, however, precise; some patients with ER-positive disease will not respond to endocrine therapies. Additional discriminatory markers to predict response to endocrine agents with greater accuracy may prove useful. Progesterone receptor (PR) status has been considered as such an additional marker, but it does not appear to add useful information in ER-positive tumours. Divergent ER and PR status is uncommon (for example < 5% of cases are ER-negative but PR-positive) and the value of the addition of PR status in this situation in predicting likelihood of response to endocrine therapy is also unclear. Nevertheless, PR examination is routinely performed on all invasive tumours by some laboratories.

HER2 status

The clinical importance of amplification of the human epidermal growth factor receptor gene HER2 in breast cancer was recognised in 1987 and an association with poorer patient outcome was subsequently reported. HER2 positivity (protein over-expression or gene amplification) is seen in approximately 15% of early invasive breast cancer. Women whose breast cancers are HER2-positive may benefit from Trastuzumab therapy. Therefore, the HER2 status of an invasive breast cancer has become an essential part of selection of this therapy. Diagnostic tests for HER2 over-expression and gene amplification include immunohistochemistry (IHC) and fluorescence in situ hybridisation (FISH). Breast cancers are reported as HER2-negative or HER2-positive according to standardized guidelines (i.e. those scoring 3+ by IHC, or 2+ and FISH amplified, as positive).

Determining hormone recptor and HER2 status – Immunohistochemistry

IHC is used to identify specific molecules in the breast cancer sample. Specifically, IHC is commonly used to show whether or not the cancer cells have hormone receptors (ER and/or PR) and/or HER2 receptors on their surface. The tissue is treated with antibodies that bind to the specific molecule. These are made visible under a microscope by using a colour reaction, a radioisotope, colloidal gold, or a fluorescent dye.

- IHC for hormone receptor testing: guidelines for pathology reporting of breast disease recommend that results for the ER/PR be reported as negative or positive and accompanied by an Allred score. This score is based on the sum of two measures including 1) a percentage that tells you how many cells out of 100 stain positive for hormone receptors – a number between 0% (none have receptors) and 100% (all have receptors) is given and 2) a number between 0 and 3 is given to indicate the intensity of their staining. '0' means that no receptors are present, '1' a small number present, '2' a medium number, and '3' a large number.
- IHC for HER2 receptor testing: guidelines for pathology reporting of breast disease recommend that results for HER2 be reported as a semi-quantitative system based on the intensity of reaction product and percentage of membrane positive cells, giving a score range of 0–3+. Samples scoring 3+ are regarded as unequivocally positive, and those scoring 0/1+ as negative. Borderline scores of 2+ require confirmation using another analysis system, ideally fluorescence in situ hybridisation.
- Fluorescence in situ hybridisation (FISH): a laboratory technique used to look at genes or chromosomes in cells and tissues. Pieces of DNA that contain a fluorescent dye are made in the laboratory and added to cells or tissues on a glass slide. When these pieces of DNA bind to specific genes or areas of chromosomes on the slide, they light up when viewed under a microscope with a special light. HER2 FISH testing results are conventionally expressed as the ratio of HER2 signal to chromosome 17 signal. Tumours showing a ratio > 2 should be considered as positive.

Expanded IHC tests are defined as those tests that measure biomarkers other than or in addition to ER, PR and HER2. These tests aim to provide similar information to gene expression profiling tests, in particular, the likelihood of cancer recurrence.

2.5.2 Adjuvant treatment planning

Adjuvant treatment in oncology is defined as additional treatment following the main therapy option; surgery is the main therapy option. While defined in this way, adjuvant treatment is viewed as an integral part of breast cancer management. Such adjuvant therapy typically consists of one or more of radiation, chemotherapy, and/or endocrine therapy/biological therapy. Planning adjuvant treatment is complex and incorporates a variety of prognostic and predictive factors. There are a number of tools to help the MDT with decisions on adjuvant treatment planning which assess prognosis and may estimate potential treatment benefit. These are described in the section on comparators (section 4.3).

2.6 Care pathway

The care pathway for this assessment can be ascertained from existing guidelines. NICE clinical guideline – CG80 – 'Breast cancer (early & locally advanced): diagnosis and treatment' should be used in the first instance. Other guidelines that may provide supplementary information include:

- St Gallen consensus recommendations
- National Comprehensive Cancer Network guidelines (NCCN).

3 Gene expression profiling

Greater understanding of the human genome, and subsequently, the genetic determinants of cancer and other diseases, has led to an array of genetic tests for use in health care. Gene expression profiling (GEP) is a relatively new technology for identifying genes whose activity may be helpful in assessing disease prognosis and guiding therapy.

GEP tests assess the identity and number of messenger RNA (mRNA) transcripts in a specific tissue sample. As only a fraction of the genes encoded in the genome of a cell are expressed by being transcribed into mRNA, GEP provides information about the activity of genes that give rise to these mRNA transcripts. Given that mRNA molecules are translated into proteins, changes in mRNA levels are ultimately related to changes in the protein composition of the cells, and consequently to changes in the properties and functions of tissues and cells (both normal and malignant) in the body.

Various assays are used in the management of breast cancer. These assays investigate the expression of specific panels of genes (also known as a gene profile or gene signature). They work by making use of different techniques to measure mRNA levels in breast cancer specimens including real-time reverse transcription-polymerase chain reaction (RT-PCR) and DNA microarrays. Many of these assays have been designed to measure the risk of cancer recurrence. Other uses of the assays include breast cancer sub-typing (using molecular classification systems), predicting the likely benefit from certain types of therapy (e.g. chemotherapy), or diagnosing breast cancer.

There are various ways of preparing the RNA, and different protocols used to prepare the specimens (e.g. formalin-fixed, paraffin-embedded, snap-frozen and fresh samples). Furthermore, there are varying algorithms that can be used to combine the raw data to obtain a summary measure. All of these factors can affect the reproducibility and reliability of GEP tests.

The complexity of gene profiling has led to numerous efforts to develop IHC markers that are able to provide similar information to that given by GEP tests. One such test is IHC4, which looks for the presence of a proliferation marker, Ki67 in addition to testing for ER, PR and HER2.

The detailed use of gene expression profile tests, for improving chemotherapy choices for breast cancer is not currently covered in NICE guidance.

3.1 Improving chemotherapy choices

Systemic therapy options for breast cancer management include endocrine treatments, targeted biological agents and chemotherapy.

The decision about whether or not to use chemotherapy is a major challenge in breast cancer management. Chemotherapy is defined as the use of cytotoxic medications with the intention of preventing cancer recurrence in patients. Chemotherapy regimens containing Anthracycline have been used routinely in the adjuvant setting. It should be noted that, for the purposes of this assessment, chemotherapy does not include other forms of systemic therapy such as endocrine treatments or targeted biological therapy.

Although chemotherapy can reduce the likelihood of cancer recurrence and death for women with breast cancer, it has considerable adverse effects. Many women with early-stage breast cancer are advised to undergo chemotherapy, however, not all will benefit from it and some may remain free of disease recurrence at 10 years without it.

GEP and expanded IHC tests may be capable of better identifying those patients that are likely and unlikely to benefit from chemotherapy than conventional clinical and pathological risk assessment. Two types of information are most likely to be useful in this context. These are the molecular sub-type of the breast tumour and an indication of the likelihood of cancer recurrence. As well as providing information on the likely outcome/course of the cancer (prognostic information), molecular sub-typing and recurrence risk may also provide information on the likelihood of the patient benefitting from chemotherapy (predictive information). Predictive and prognostic information may be used to inform chemotherapy decisions in breast cancer management. Information on molecular sub-typing and recurrence risk can be found below.

3.1.1 Breast tumour sub-typing using molecular classification systems

Micro-array-based gene expression studies have revealed that, in addition to being clinically heterogeneous, breast cancer is also a molecularly heterogeneous disease. As a result, distinct molecular sub-types of breast cancer that exhibit different gene expression patterns and clinical outcomes have been developed. The prognosis and chemotherapy sensitivity of the various molecular sub-types are different. Luminal-like cancers tend to have the most favourable long-term survival compared with the others, whereas basal-like and HER2-positive tumours have significantly worse long-term survival and are more sensitive to chemotherapy. However, it is important to note that these correlations are expected as there is a strong association between the molecular sub-type and conventional histopathologic variables (namely, ER and HER2 status).

Numerous classification systems have been published. The first of these was described by Perou and colleagues in 2000. Since then, this classification system has been refined to distinguish the luminal group into luminal A and luminal B, and the classification of normal-like is less commonly used as it is believed to be a potential artefact from the initial study. This classification system is commonly cited in the literature and includes the following sub-types, the IHC approximation is provided in brackets (ER = oestrogen receptor, PR= progesterone receptor, HER2 = human epidermal growth factor receptor 2):

- Luminal A (ER positive and generally HER2 negative)
- Luminal B (ER positive (but a lower number of receptors than luminal A) and generally HER2 negative)
- HER2 amplified (predominantly HER2 positive and ER negative)
- Basal-like (generally ER, PR and HER2 negative (triple negative))
- Unclassified/5NP (generally ER, PR, HER2, EGFR and CK5 negative).

Initial work to identify the molecular sub-types used hierarchical clustering to design a classification model (single sample predictor (SSP)) that allows a breast cancer to be classified using a nearest centroid classifier. Essentially, this means that new tumours are sub-typed based on how similar their gene profile is to tumours used and sub-typed in the initial data-set for the SSP. Several limitations of SSPs have been posited in the literature. These include the effect of the breast tumour samples and genes selected in defining the molecular sub-types. Consequently, it has been observed that different SSPs may not reliably assign the same tumour to the same molecular sub-type. More recently, a sub-type classification model based on a parametric clustering technique defined by three gene modules has been suggested to overcome the challenges of SSPs.

Although there is a body of literature on molecular classification systems, GEP tests used for molecular sub-typing, in most cases, are at the early stages of the validation pathway. Generally, studies of diagnostic test accuracy in defining the molecular sub-types when comapred with the classification based on ER, PR and HER2 status can be found in the literature.

Clinical experts contacted during scoping felt that molecular classification systems showed great potential, however, their views on the impact of these classification systems on treatment decisions (compared with current clinical practice) were mixed. Some experts felt that molecular classification systems were useful for predicting non-response to neoadjuvant chemotherapy. In addition, the basal-like classification captured other individuals with a poor prognosis who may be missed if only using the triple negative (ER/PR/HER2 negative) diagnosis by IHC. Other experts felt that little was known about the concordance between these molecular classifications with their prognostic and predictive value. Clinical experts also felt that if molecular classification systems were to be used in the clinical setting, they would do so as an adjunct to current clinical practice rather than replacing any part of it.

The impact of molecular classification systems on breast cancer management, when added to current clinical practice, is difficult to determine from the published literature. The literature on the use of molecular signatures in predicting non-response (or response) to neoadjuvant chemotherapy suggests that different molecular sub-types respond differently to neoadjuvant chemotherapy. However, it may also be possible to use IHC as a surrogate marker for the molecular classifications. An area of potential benefit may be that of individuals with basal-like breast cancer who are not identified using the triple negative (ER/PR/HER2 negative) diagnosis by IHC. Although figures in the literature vary, triple negatives may account for approximately 7–20% of all breast cancers and it is thought that approximately 85% of all basal type tumours may be triple negatives. The literature suggests that many breast cancer researchers believe that molecular classification systems will change with further subdivision of these sub-types.

At present, molecular classification systems are not routinely used by physicians in the NHS in England. Guidelines on the use of molecular classification systems in breast cancer management were not identified during scoping.

3.1.2 Recurrence risk

Therapeutic decisions for breast cancer management are based on risk estimates. Tests that improve such estimates have the potential to affect clinical outcomes in breast cancer patients either by avoiding unnecessary chemotherapy with its attendant morbidity or by employing it where it might not otherwise have been used, thereby reducing recurrence risk.

Much of the literature on gene expression profile test validation focuses on the analytical validity and clinical validity of those tests that measure recurrence risk. Some tests are further down the validation pathway and may have evidence on the clinical utility of the technology.

Tests measuring recurrence risk combine the measurements of gene expression levels within the tumour to produce a number associated with the risk of disease recurrence. These tests aim to improve on risk stratification schemes based on clinical and pathological factors currently used in clinical practice (see section 4.3).

Existing breast cancer guidelines have recommended the use of gene expression profile tests to help guide chemotherapy treatment decisions. For example, the 2009 (11th) St Gallen consensus meeting publication states 'the Panel supported the use of a validated multigene-profiling assay, if readily available, as an adjunct to high-quality phenotyping of breast cancer in cases in which the indication for adjuvant chemotherapy remained uncertain.'

At present, GEP tests measuring recurrence risk are not routinely used by physicians in the NHS in England.

3.2 Scoping workshop feedback

Scoping workshop attendees felt that both molecular sub-typing and recurrence risk measurements may be used to stratify patients when considering chemotherapy. Attendees felt that these tests may be used with current clinical practice as opposed to replacing any part of current clinical practice.

The extensive use of chemotherapy in breast cancer management was discussed. Attendees felt that patients were over-treated with chemotherapy as it is difficult to identify those patients who are less likely to benefit from its use. This has been noted both anecdotally and in the scientific literature.

Therefore, the scope has been expanded from an evaluation of Randox BCA to include other gene expression profiling tests that are likely to influence the use of chemotherapy in breast cancer management. In addition, attendees felt it was important to include IHC tests that may fulfil this purpose.

Details of the interventions can be found in section 4.2 – *Table 1*.

4 Scope of the evaluation

The assessment has been expanded to include gene expression profiling tests and expanded immunohistochemistry tests that are likely to influence the use of chemotherapy in breast cancer management.

4.1 Population

People diagnosed with early invasive breast cancer.

Note: Although the population for the assessment is broad, some GEP and expanded IHC tests may only be used in a sub-population. For example, women with early-stage invasive breast cancer (stage I, II or III), lymph node negative or positive (up to 3), oestrogen receptor positive or negative and HER2 positive or negative. Additionally, men with breast cancer should also be included in the assessment if data are available on the use of these technologies in men.

4.2 Interventions

Several GEP and expanded IHC tests that are likely to impact the use of chemotherapy in breast cancer management exist. Technologies identified during scoping are summarised in *Table 1*.

TABLE 1 Interventions identified during scoping

Test	Manufacturer	Purpose	Description	Target population*
Gene expression profiling tests				
Randox BCA	Randox Laboratories	Molecular Sub-typing + Recurrence risk	Low density biochip array 23 gene array	All women with breast cancer
Breast Cancer Index	bioTheranostics	Recurrence risk	RT-PCR Assessment of H/I ratio (*HOXB13*:*IL17BR*) and MGI (Molecular Grade Index)	ER+, LN–
MammaPrint	Agendia	Recurrence Risk	MICROARRAY 70 gene array	Early-stage (stage I or II), LN– or LN+ (up to 3), ER+ or ER–
MammaPrint + BluePrint	Agendia	Recurrence risk + Molecular Sub-typing	MICROARRAY 70 gene array + 80 gene array	Early-stage (stage I or II), LN– or LN+ (up to 3), ER+ or ER–
OncotypeDX	Genomic Health	Recurrence risk and Predictive of chemotherapy benefit	RT-PCR 21 gene assay	Early-stage (stage I or II), LN–, ER+ patients who will be treated with hormone therapy
PAM50	ARUP Laboratories Inc.	Recurrence risk and Predictive of chemotherapy benefit	RT-qPCR 55-gene assay	Early-stage (stage I or II), LN–, ER+ patients who will be treated with hormone therapy
Expanded immunohistochemistry tests				
IHC4	N/A	Recurrence risk	IHC test based on ER, PgR, HER2 and Ki67 Plus clinical factors (age, nodal status, tumour size, grade, randomised treatment)	ER+
Mammostrat	Clarient	Recurrence risk	IHC test based on P53, HTF9C, CEACAM5, NDRG1 and SLC7A5 markers	Early-stage (stage I or II), LN–, ER+ patients who will be treated with hormone therapy
NPI+	Nottingham Prognostics	A clinical decision making tool kit for all operable breast cancer patients providing prognostic and therapeutic predictive outputs	A multistep approach combining biological assessed by immunocytochemistry and traditional pathological and clinical variables	All patients with early (stage I or II) invasive breast cancer

*ER+/– = oestrogen receptor positive or negative, LN+/– = lymph node positive or negative.

4.3 Comparators

Two existing algorithms are in use for predicting survival and the utility of adjuvant therapy in breast cancer and should serve as comparators. These include:

1. Nottingham Prognostic Index
2. Adjuvant! Online

Nottingham Prognostic Index (NPI): the NPI is a well-established, validated and widely used method of predicting survival for operable primary breast cancer. This index was based on a retrospective analysis of 9 factors in 387 patients. Only 3 of the factors (tumour size, stage of disease, and tumour grade) remained significant on multivariate analysis. The NPI is calculated as: lymph node (LN) stage (1–3) + grade (1–3) + maximum tumour diameter, giving an observed range of NPI from 2.08 (LN negative, grade 1, 0.4 cm) to 6.8 (LN stage 3, grade 3, size 4.9 cm).

Adjuvant! Online: the Adjuvant! Online computer program is designed to provide estimates of the benefits of adjuvant endocrine therapy and chemotherapy. A version of Adjuvant! Online that will include HER2 status and the potential benefit of Trastuzumab is in development. It is believed that the current version (version 8) may underestimate the risk of mortality and does not take into account the negative impact of HER2 positivity or how this may be affected by Trastuzumab. Patient and tumour characteristics are entered into the program and provide an estimate of the baseline risk of mortality or relapse for patients without adjuvant therapy. Information about the efficacy of different therapy options is derived from Early Breast Cancer Trialists Collaborative Group meta-analyses in order to provide estimates of reduction in risk at 10 years of breast cancer related death or relapse for selected treatments. Results may be displayed and printed in graphical form to aid shared decision-making. Attendees at the scoping workshop suggested that there were some difficulties in applying the Adjuvant! Online data to the UK population.

4.4 Health outcomes

The outcomes of interest are the morbidity and mortality associated with invasive breast cancer and its treatment. These may include:

- Distant recurrence free survival – 10 years
- Health-related quality of life, such as, adverse events associated with chemotherapy

Note: The health outcomes stated above are preferred for use in the assessment. However, the available data may be limited. In such cases, other data may be used. For example, total disease recurrence at 5 years or pathological complete response.

4.5 Healthcare setting

These tests will be assessed for use in the adjuvant setting and are expected to be used in secondary and tertiary care.

Note: the neoadjuvant setting was considered for inclusion in the scope, however, it was anticipated that evidence on the use of these tests in the neoadjuvant setting would be lacking. Therefore, it was decided that the assessment should focus on the adjuvant setting only.

5 Modelling approach

Tests to be included in the economic modelling will need to have sufficient data to allow modelling to proceed. The level of data required will be set by the external assessment group

(EAG). Both predictive and prognostic 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.

5.1 Modelling possibilities

5.1.1 Molecular sub-typing

Guidelines recommending treatment decisions based on molecular sub-typing have not been uncovered during scoping. To allow the modelling of the role of sub-typing tests it would be necessary to link the accuracy of a diagnostic test to final health outcomes. Distinct molecular sub-types of breast cancer that exhibit different gene expression patterns and clinical outcomes have been developed. The prognosis and chemotherapy sensitivity of the various molecular sub-types are different. However, GEP tests used for molecular sub-typing, in most cases, are at the early stages of the validation pathway. Likely changes in treatment planning resulting from the results of sub-typing tests are as yet unclear.

5.1.2 Recurrence risk

Validation studies exist for the diagnostic technologies dealing with recurrence risk. Data on analytical validity, clinical validity, clinical utility and economic evaluations (described in section 5.2 below) are available in the published literature for certain diagnostic technologies. The availability of these data is expected to make it possible to conduct a thorough assessment.

5.2 Existing Models

Economic models for certain diagnostic technologies exist in the published literature (e.g. for MammaPrint and OncotypeDX). These economic evaluations seek to reclassify the risk category of patients who were initially defined by existing guidelines (e.g. NCCN) using the test in question. Resulting quality adjusted life years (QALYs) and costs have been reported.

5.3 Model structure

Published studies that measure the clinical utility of gene expression profile tests using a prospective study design that follow patients from initial diagnosis through to final health outcomes have not been identified during the scoping phase. Two prospective studies, MINDACT (MammaPrint) and TAILORx (Oncotype), are ongoing. Consequently, it is likely that a linked evidence approach will need to be used in the modelling. That is, outcomes of the diagnostic tests to be assessed will need to be related to changes in final heath outcomes.

5.4 Cost considerations

The Randox BCA is processed locally using the Randox Evidence Investigator Analyser. This analyser can be used to process other biochip arrays available from Randox Laboratories (e.g. ovarian cancer therapy response prediction assay, multiplex pathogen detection arrays for STIs and respiratory infections and drug metabolism SNP assays). At present, this analyser is not widely available in the NHS. Therefore, the Randox BCA will incur non-recurrent set-up costs to purchase the necessary equipment needed to process the test.

Generally, other gene expression profile tests for breast cancer are processed centrally by the manufacturer.

Protocols used to prepare the tumour specimens can vary. These include formalin-fixed, paraffin-embedded, snap-frozen and fresh samples. The costs between these protocols vary significantly and should be considered in the assessment.

5.5 Health outcomes

QALYs will need to be calculated in the economic modelling.

6 Equality issues

None identified during scoping. The population in the scope falls within the provisions of the Equality Act 2010 once a diagnosis of cancer has been made.

7 Implementation

Support tools are developed by the implementation team at NICE. The implementation team does not get involved in developing the guidance recommendations but works alongside the guidance-producing programme, the communications team and field based teams to, among other things, ensure intelligent dissemination of NICE guidance to the appropriate target audiences.

Commissioners will need to know whether there are significant non-recurrent set-up costs associated with the introduction of the interventions listed in *Table 1*, particularly where these are likely to influence the location of services or the size of population they would need to serve.

Appendix A Glossary

Adjuvant therapy

Adjuvant therapy is treatment that is given in addition to (proceeding) the primary (initial) treatment. It is designed to help reach the primary treatment goal (for example, disease eradication). Adjuvant therapy for cancer usually refers to surgery followed by chemotherapy or radiotherapy to help decrease the risk of the cancer recurring (coming back). Adjuvant therapy is considered as an integral part of treatment and is viewed as a non-surgical oncology treatment of (primary) breast cancer by clinicians.

Allred score

The Allred score is a composite of the percentage of cells that stained and the intensity of their staining.

Amplification

In genetics, an increase in the frequency of replication of a DNA segment

Analytic validity

Analytical validity in this context refers to a test's ability to accurately and reliably measure the expression of messenger ribonucleic acid (mRNA) by breast cancer tumour cells. It is usually assessed by determining how much observed measurements provided by the test/technology differ from expected values derived from a standard reference. In the measurement of gene expression, however, there are no standard reference tests and an assessment of the analytical validity of the assays has to be obtained by more indirect methods. This involves an examination of test variability arising from tumour sampling, specimen handling, specimen preparation and biologic variation within and between different samples of the same tumour, and the effect of this on the reproducibility of test when repeated in the same patient, over time.

Biomarkers

A biological molecule used as a marker for a substance or process of interest.

Breast conserving surgery

Surgery in which the cancer is removed together with a margin of normal breast tissue. The whole breast is not removed.

Breast reconstruction

The formation of a breast shape after a total mastectomy, using a synthetic implant or tissue from the woman's body.

Chemotherapy

The use of medication(s) (drugs) that are toxic to cancer cells, given with the aim of killing the cells or preventing or slowing their growth.

Clinical utility

The clinical utility of a gene expression profile relates to its ability to discriminate between those who will have more or less benefit from a therapeutic intervention: the focus in the assessment of clinical utility is outcome. Other utilities which may be considered to be important include the effect of the test on clinical decision making (for example, choice of therapy).

Direct evidence of clinical utility of a gene expression profile can only be provided in context of a randomized clinical trial where benefit can be measured in terms of an improvement of clinical outcomes such as overall survival, disease-free survival, chemotherapy toxicity, or quality of life. Prognostic estimates, though not direct estimates of benefit per se, may provide a crude estimate of benefit which may be relevant for patient decision making. They can also provide an upper limit on the degree of clinical benefit that may be expected.

Clinical validity

Clinical validity is usually defined as the degree to which a test accurately predicts the risk of an outcome (for example, time to distant metastases), as well as its ability to separate/discriminate patients with different outcomes into separate (high and low) risk classes. This is usually reported as the clinical sensitivity and specificity of the test.

Cytotoxic

Toxic to living cells.

DNA microarray

A DNA microarray (also commonly referred to as 'gene chip,' 'DNA chip') is a collection of microscopic DNA spots (defined 'features'), commonly representing single genes or transcripts, arrayed on a solid surface by covalent attachment to chemically suitable matrices, or directly synthesized on them. DNA microarrays use DNA as part of their detection system. Qualitative or quantitative measurements with DNA microarrays use the selective nature of DNA–DNA or DNA–RNA hybridisation under high-stringency conditions and fluorophore-based detection. DNA arrays are commonly used for gene expression profiling, i.e., monitoring expression levels of thousands of genes simultaneously, or for comparative genomic hybridisation.

Endocrine therapy

Treatment of cancer by removing and/or blocking the effects of hormones which stimulate the growth of cancer cells.

External assessment group

An independent group of researchers commissioned by NICE to review the evidence on a group of technologies. The external assessment group includes researchers who assess the quality of studies on the treatments, and health economists who look at whether the treatments are good value for money. The Diagnostics Assessment Committee bases its discussions on the diagnostics assessment report produced by the external assessment group.

Gene expression

Gene expression refers to the translation of the information encoded in a gene into an RNA transcript. Expressed transcripts include messenger RNAs (mRNA) translated into proteins, as well as other types of RNA, such as transfer RNA (tRNA), ribosomal RNA (rRNA), micro RNA (miRNA), and non-coding RNA (ncRNA), that are not translated into protein. Gene expression is a highly specific process by which cells switch genes on and off in a timely manner, according to their state. The study of mRNA expression in a cell is an indirect way to study the proteins counterpart.

Gene expression profiling

This term refers to any genomic techniques that measure the fraction of the genes that is expressed in a specific sample. This definition refers to techniques that allow the assessment of more than one gene at a time, especially microarray and real time RT-PCR.

Gene expression profile/pattern

This is any set of genes for which the expression in a specific sample is known. A gene expression profile may account for a variable number of genes, and the corresponding expression values may be obtained by different techniques. Gene expression profiles can be associated, by various techniques, to phenotypes.

Gene expression signature

This is an equivalent term currently in use to refer to a specific 'gene expression profile,' usually associated with a specific phenotype.

Grading

Assessing the degree of aggressiveness of a malignant tumour based usually on the appearance of its cells under the microscope.

Hierarchical clustering

A method which seeks to build a hierarchy of clusters that involves highly complex computation. In order to decide which clusters should be combined (for agglomerative clustering), or where a cluster should be split (for divisive clustering), a measure of dissimilarity between sets of observations is required.

Histology

An examination of the cellular characteristics of a tissue using a microscope.

Hormone receptor

Proteins with a cell that bind to specific hormones.

Human epidermal growth factor receptor

A molecule on the surface of a cell which interacts with a specific growth factor and helps to control how rapidly the cells grow.

Immunohistochemistry

A technique that uses antibodies to identify specific molecules in tissues which are examined and scored by a pathologist using a microscope.

Invasive breast cancer

Breast cancer where the malignant cells have broken through the lining layer of the normal tissues and extend into the fat and fibrous tissue of the breast.

Lymph nodes

Small structures which act as filters of the lymphatic system. Lymph nodes close to the primary tumour are generally the first site to which cancer spreads.

Malignant

Cancerous cells which can invade into nearby tissue and spread to other parts of the body.

Mammography

The process of taking a mammogram – a soft tissue x-ray of the breast which may be used to evaluate a lump or which may be used as a screening test in women with no signs or symptoms of breast cancer.

Mastectomy

Surgical removal of the breast.

Metastases

Deposits of cancer elsewhere in the body.

Metastasis

Spread of cancer away from the primary site to elsewhere in the body via the bloodstream or the lymphatic system.

Multidisciplinary team

A team with members from different healthcare professions (including for example, oncology, pathology, radiology, nursing).

Nearest centroid classifier

This method computes a standardized centroid for each class. This is the average gene expression for each gene in each class divided by the within-class standard deviation for that gene. Nearest centroid classification takes the gene expression profile of a new sample, and compares it to each of these class centroids. The class whose centroid that it is closest to, in squared distance, is the predicted class for that new sample.

Neoadjuvant therapy

Neoadjuvant therapy is treatment that is given prior to the primary (initial) treatment. Surgery is regarded as the primary treatment in breast cancer.

Occult

Hidden, or difficult to observe directly.

Oestrogen receptor

A protein within breast cancer cells that binds to oestrogens. It indicates that the tumour may respond to endocrine therapies. Tumours rich in oestrogen receptors have a better prognosis than those which are not.

Predictive values/markers

A molecule that is assessed to predict the likely response to a specific treatment, for example oestrogen receptor to predict the likely response to endocrine therapy.

Primary systemic therapy

Systemic therapy given before surgery or radiotherapy.

Progesterone receptor

A protein within cells that binds to progesterone.

Prognosis

A prediction of the likely outcome or course of a disease; the chance of recovery, recurrence or death.

Prognostic factors

Disease characteristics that are correlated with the course of the disease and which are used to predict the likely outcomes.

Real time reverse transcriptase polymerase chain reaction (RT-PCR)

Real-time RT-PCR is a molecular biology technique that allows the amplification and the quantification in real time of defined RNA molecules from specific specimens. This technology has been used for several years in research and clinical settings to measure RNA molecules. In the first step DNA, copies of the investigated RNA molecules present in the template are obtained by a reaction named reverse transcription. Then DNA amplification is obtained using PCR, while the quantification of the accumulating DNA product is accomplished by the use of specific fluorescent reagents. The quantification of the target RNA molecule is based on the analysis of the accumulation curve of the complementary DNA, as measured by the fluorescence detected at each cycle of the reaction.

Reverse transcription

In biochemistry, reverse transcription is the enzymatic reaction induced on by the RNA dependent DNA polymerase. This enzyme, also known as reverse transcriptase, is a DNA polymerase enzyme that copies single-stranded RNA into DNA. This process is the reverse of normal transcription, which involves the synthesis of RNA from DNA.

Single sample predictor

A classification model that enables the sub-type of a single tumour to be identified using a nearest centroid classifier based on the initial hierarchical clustering of a small (typically) data set.

Staging

Clinical description of the size and spread of a patient's tumour, allocated by internationally agreed categories.

Systemic therapy/treatment

Medicine, usually given by mouth or injection, to treat the whole body rather than targeting one specific area.

Transcription

In genetics, the process by which genetic information on a strand of DNA is used to synthesize a strand of complementary RNA.

Translation

In genetics, the process by which a messenger RNA molecule specifies the linear sequence of amino acids on a ribosome for protein synthesis.

Appendix B Abbreviations

BCA	Breast cancer array
CG	Clinical guideline
DAP	Diagnostics Assessment Programme
DNA	Deoxyribonucleic acid
ER	Oestrogen receptor
FISH	Fluorescence in situ hybridisation
GEP	Gene expression profiling
GP	General practitioner
HER2	Human epidermal growth factor receptor 2
IHC	Immunohistochemistry
LN	Lymph node
MDT	Multidisciplinary team
MINDACT	Microarray in node negative and 1-3 positive lymph node disease may avoid chemotherapy
mRNA	Messenger ribonucleic acid
NCCN	National Comprehensive Cancer Network Guidelines
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NPI	Nottingham Prognostic Index
PR	Progesterone receptor
QALY	Quality adjusted life year
RT-PCR	Reverse transcription-polymerase chain reaction
SSP	Single sample predictor
TAILORx	Trial assigning individualised options for treatment (Rx)

Appendix C Related NICE Guidance

Refer to http://guidance.nice.org.uk/Topic/Cancer/Breast.

Appendix D References

1. Andre, F., Benbrahim, Z., and Delaloge, S. Breast cancer classification to select patients for cancer treatment. *Current Breast Cancer Reports* 2 (3) ()(pp 152–158), 2010. Date of Publication: September 2010. (3), 152–158. 2010.
2. Brenton, J. D., Carey, L. A., Ahmed, A. A., and Caldas, C. Molecular classification and molecular forecasting of breast cancer: ready for clinical application?. [Review] [66 refs]. *Journal of Clinical Oncology* 23(29), 7350–7360. 10-10-2005.
3. Cianfrocca, M., Gradishar, W. New Molecular Classifications of Breast Cancer. *CA Cancer J Clin* 59;303–313, 2009.

4. Cosler, L. E., Lyman, G. H. Economic Analysis of gene Expression Profile Data to Guide Adjuvant Treatment in Women with Early-Stage Breast Cancer. *Cancer Investigation* 27:953–959, 2009.

5. De Ronde, J. J., Hannemann, J., Halfwerk, H., Mulder, L., Straver, M. E., Vrancken, Peeters M. J., Wesseling, J., Van, De, V, Wessels, L. F. A., and Rodenhuis, S. Concordance of clinical and molecular breast cancer sub-typing in the context of preoperative chemotherapy response. *Breast Cancer Research and Treatment* 119 (1) ()(pp 119–126), 2010. Date of Publication: January 2010. (1), 119–126. 2010.

6. Goldhirsch, A., Ingle, J. N., Gelber, R. D., Coates, A. S., Thurlimann, B., Senn, H.-J., and panel members. Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2009. *Annals of Oncology* special article.

7. Haibe-Kains, B. Classification models for breast cancer molecular subtyping: what is the best candidate for a translation into clinic. *Women's Health* 6(5), 623–625, 2010.

8. Hergueta-Redondo, M., Palacios, J., Cano, A., and Moreno-Bueno, G. 'New' molecular taxonomy in breast cancer. [Review] [61 refs]. *Clinical & Translational Oncology*: Official Publication of the Federation of Spanish Oncology Societes & of the National Cancer Institute of Mexico 10(12), 777–785. 2008.

9. Huber, K. E., Carey, L. A., and Wazer, D. E. Breast cancer molecular subtypes in patients with locally advanced disease: impact on prognosis, patterns of recurrence, and response to therapy. [Review] [55 refs]. *Seminars in Radiation Oncology* 19(4), 204–210. 2009. MEDLINE.

10. Kaklamani, V. G. and Gradishar, W. J. Gene expression in breast cancer. [Review] [17 refs]. *Current Treatment Options in Oncology* 7(2), 123–128. 2006.

11. Lin, S.-X., Chen, J., Mazumdar, M., Poirier, D., Wang, C., Azzi, A., and Zhou, M. Molecular therapy of breast cancer: Progress and future directions. *Nature Reviews Endocrinology* 6 (9) ()(pp 485–493), 2010. Date of Publication: September 2010. (9), 485–493. 2010.

12. Marchionni L, Wilson RF, Marinopoulos SS, Wolff AC, Parmigiani G, Bass EB, Goodman SN. *Impact of Gene Expression Profiling Tests on Breast Cancer Outcomes*. Evidence Report/ Technology Assessment No. 160. (Prepared by The Johns Hopkins University Evidence based Practice Center under contract No. 290-02-0018). AHRQ Publication No. 08-E002. Rockville, MD: Agency for Healthcare Research and Quality. *January* 2008.

13. Morris, S. R. and Carey, L. A. Molecular profiling in breast cancer. *Reviews in Endocrine and Metabolic Disorders* 8 (3) ()(pp 185–198), 2007. Date of Publication: Sep 2007. (3), 185–198. 2007.

14. Pathology reporting of breast disease. *A Joint Document Incorporating the Third Edition of the NHS Breast Screening Programme's Guidelines for Pathology Reporting in Breast Cancer Screening and the Second Edition of The Royal College of Pathologists' Minimum Dataset for Breast Cancer Histopathology*. NHS BSP Publication 58. *January* 2005.

15. Perou *et al.* Molecular Portraits of Human Breast Tumours. *Nature* 406, 2000.

16. Rakha, E. A., Reis-Filho, J. S., Ellis, I. O. Basal-Like Breast Cancer: A Critical Review. *Journal of Clinical Oncology* 26:2568–2581, 2008.

17. Smartt, P. A comparison of gene expression profiling tests for breast cancer. *HSAC Report* 2010; 3(16).

18. Sorlie, T. Introducing Molecular Subtyping of Breast Cancer Into the Clinic? *Journal of Clinical Oncology* 27(8); 1153–1154, 2009.

19. Sorlie, T., Perou, C. M., Fan, C., Geisler, S., Aas, T., Nobel, A., Anker, G., Akslen, L. A., Botstein, D., Borresen-Dale, A. L., and Lonning, P. E. Gene expression profiles do not consistently predict the clinical treatment response in locally advanced breast cancer. *Molecular Cancer Therapeutics* 5(11), 2914–2918. 2006.

20. Tibshirani, R. *et al.* Diagnosis of multiple cancer types by shrunken centroids of gene expression. *PNAS* 99(10): 6567–6572, 2002.

21. Vallejos, H. Gomez, W. Cruz, J. Pinto, R. Dyer, R. Velarde, J. Suazo, S. Neciosup, M. Leon, M. De La Cruz, and C. Vigil. Breast cancer classification according to immunohistochemistry markers: Subtypes and association with clinicopathologic variables in a peruvian hospital database. *Clinical Breast Cancer* 10 (4) ()(pp 294–300), 2010. Date of Publication: 01 Aug 2010. (4):294–300, 2010.

22. Wirapati *et al.* Met-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Research* 10:R65, 2008.

Appendix E Equality Impact Assessment

The impact on equality has been assessed during this assessment according to the principles of the NICE Equality scheme.

1. Have any potential equality issues been identified during the scoping process (scoping workshop discussion, assessment subgroup discussion), and, if so, what are they?
None identified

2. What is the preliminary view as to what extent these potential equality issues need addressing by the Committee?
N/A

3. Has any change to the draft scope been agreed to highlight potential equality issues?
N/A

4. Have any additional stakeholders related to potential equality issues been identified during the scoping process, and, if so, have changes to the stakeholder list been made?
Additional stakeholders have not been identified

Approved by Associate Director (name): Nick Crabb

Date: 26/04/2011

Appendix F Attendees of the assessment subgroup meeting

The following people were in attendance at the assessment subgroup meeting held on 11th April 2011:

	Name of representative	Job Title	Organisation
Standing Committee Members	Ian Cree	Director, NETSCC-EME	National Institute for Health Research
	Christopher Hyde	Professor of Public Health and Clinical Epidemiology	Peninsula Technology Assessment Group (PenTAG)
Specialist Committee Members	Carole Farrell	Nurse Clinician	The Christie NHS Foundation Trust
	Louise Jones	Consultant Clinical Scientist	Health Service Research Unit, University of Aberdeen
	Simon Pain	Consultant Breast and Endocrine Surgeon	Department of General Surgery, Norfolk & Norwich University Hospital
	Rob Stein	Consultant and Senior Lecturer in Oncology	Department of Oncology UCL Hospitals
	Ursula Van Mann	Principal Clinical Scientist	Health Service Research Unit, University of Aberdeen
External Assessment Group arriving at 13:00	Sue Ward	Project Manager & supervisor for economic modelling	ScHARR, The University of Sheffield
	Rachid Rafia	Economic Modeller	
	Alison Scope	Systematic Reviewer	

NICE staff in attendance:

Name	Title
Prof Adrian Newland	Chair, Diagnostics Advisory Committee
Nick Crabb	Associate Director, Diagnostics Assessment Programme
Hanan Bell	Technical Advisor
Jackson Lynn	Project Manager, Diagnostics Assessment Programme
Gurleen Jhuti	Technical Analyst, Diagnostics Assessment Programme
Farouk Saeed	Technical Analyst, Diagnostics Assessment Programme

Appendix 21

Protocol

Technology Assessment Report commissioned by the NIHR HTA Programme on behalf of the National Institute for Health and Clinical Excellence – Protocol

10 May 2011

1. Title of the project:

Gene expression profiling tests and expanded immunohistochemistry tests to guide selection of chemotherapy regimes in breast cancer management

2. Name of TAR team and 'lead'

ScHARR, University of Sheffield

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

Project Lead: Sue Ward

Tel: (+44) (0)114 222 0816
Fax: (+44) (0)114 272 4095
E-mail: s.e.ward@sheffield.ac.uk

Address for correspondence

Major documentation should be sent to the project lead (Sue Ward, s.e.ward@sheffield.ac.uk), the project administrator (Gill Rooney, g.rooney@sheffield.ac.uk) and the managing director of ScHARR-TAG (Eva Kaltenthaler, e.kaltenthaler@sheffield.ac.uk).

3. Plain English Summary

[This will be used on the HTA Programme website and for any appropriate research registers.]

Breast cancer is the most commonly diagnosed cancer in women in England. In 2008 there were 39,681 new cases diagnosed, an increase of 1,633 cases comapred with 2007 (4%). Just over 10,000 women died from breast cancer in England in 2008, a rate of 26 deaths per 100,000 women. It is the second most common cause of cancer death in women, after lung cancer (ONS,

2010). Treatment usually involves surgery to remove the primary tumour and any involved lymph nodes: this may be followed by radiation therapy, endocrine therapy, Trastuzumab and/or chemotherapy depending on tumour and patient variables.

To help guide treatment decision making, several guidelines have been established. The guidelines used in England include the Nottingham Prognostic Index (NPI) and Adjuvant! Online. These guidelines assist clinicians in the selection of the most appropriate treatment for a particular patient. They provide information about prognosis which is largely based on pathological parameters (e.g., tumour size, grade and lymph node status) for NPI with the addition of ER receptor status, age and co-morbidity for Adjuvant! OnLine. However, it has been suggested that these clinical tools do not predict outcome and response to treatment particularly well (Paik, 2007). Different guidelines can give different results and it has been suggested that a large proportion of women with early stage breast cancer are over-treated. This may result in unnecessary use of toxic and expensive chemotherapy for women who derive no benefit or avoidable deaths in women for whom chemotherapy was withheld.

This presents a great challenge to clinicians in estimating prognosis and making therapeutic decisions particularly relating to the decision about whether or not to use adjuvant chemotherapy (chemotherapy after surgery) in women with early stage breast cancer. While chemotherapy may prevent relapse in some, not all women with early stage breast cancer will benefit and some women remain recurrence free at 10 years without chemotherapy. However, a subset of patients with a 'good' prognosis may still develop recurrence after curative surgery and adjuvant therapy.

Detailed multi-parameter cancer profiling, using either gene expression profiling or protein expression profiling (with immunohistochemistry) has been proposed as an approach to address these issues by identifying genes or proteins whose activity may be helpful in assessing disease prognosis and guiding therapy in this group of patients. Improved information on baseline risk (i.e. prognostic risk) and response to chemotherapy (i.e. predictive benefit) may help target chemotherapy on those patients who will benefit the most. Avoiding chemotherapy in patients at low risk of recurrence and who will therefore obtain limited benefit offers the potential for cost savings (in terms of avoided chemotherapy and avoided treatment of adverse events associated with chemotherapy) and the benefit of avoiding the disutility associated with adverse events. Accurately identifying those patients at highest risk of recurrence will maximise the survival gains from chemotherapy.

Since the systematic reviews by Marchionni *et al.* (2008) (search date from 1990 to January 2007) and Smartt (2009) (search date from 2007 to September 2009) several other studies of gene expression profiling have become available.

The aim of this review is to systematically evaluate and appraise the potential clinical and cost effectiveness of using gene or protein expression profiling tests to guide selection of chemotherapy regimes in breast cancer management.

4. Decision problem

[This will appear on the HTA Programme website and appropriate research registers]

4.1 Purpose of the decision to be made

The aim of the assessment is to answer the following research question:

By guiding the selection of chemotherapy regimes in breast cancer management, will using gene or protein expression profiling tests in patients with early stage breast cancer improve health outcomes and quality of life comapred with currently used decision making protocols?

4.2 Clear definition of the intervention

Nine tests have been identified by NICE and will be included in this assessment: six are based on gene expression profiling and three on immunohistochemistry.

The gene expression profiling tests which are included are as follows;

- The Randox Assay (BCA) (Randox Laboratories) is a cDNA-based expression biochip assay that aims to accurately define the clinical sub-types of breast cancer tumours prior to initiating treatment. The target population is all individuals with diagnosed breast cancer.
- MammaPrint (Agendia) is based on microarray technology which uses a 70-gene expression profile. MammaPrint is intended as a prognostic test for women 61 years or younger with primary invasive ER+, or ER-negative (ER–) LN0 breast cancer.
- Blueprint (Agendia) used is used in addition to MammaPrint for molecular sub-typing, is an 80 gene microarray, the target population is patients with early-stage (stage I or II), LN– or LN+ (up to 3), ER+ or ER– breast cancer.
- PAM50 gene expression assay (ARUP Laboratories Inc.) identifies the major intrinsic biological subtypes of breast cancer and generates risk of recurrence (ROR) score.
- OncotypeDX (Genomic Health) quantifies gene expression for 21 genes in breast cancer tissue by RT-PCR. It is intended to predict the likelihood of recurrence in women of all ages with newly diagnosed Stage I or II, ER-positive (ER+) lymph node negative (LN0) breast cancer treated with tamoxifen. The test assigns the breast cancer a recurrence score. The test also looks at the expression of hormone receptor genes, both the estrogen receptor (ER) and progesterone receptor (PR) and can provide an indication of how responsive the cancer is likely to be to hormonal therapy.
- Breast Cancer Index (Biotheranostics) is a RT-PCR assessment of the ratio of expression of 2 genes, HOXB13 and IL17BR and the Molecular Grade Index (MGI) and gives an indication of recurrence risk. The target population are those with ER+ and LN– breast cancer.

The expanded immunohistochemistry tests for protein expression which are included are the IHC4, Mammostrat and Nottingham Prognostic Indicators plus (NPI+).

- IHC4 assesses levels of four key proteins in a breast cancer sample, ER, PgR, HER2 and Ki-67. This permits broad categorisation into the 5 main tumour subtypes which determine treatment and prognosis.
- The Mammostrat® test uses five immunohistochemical markers (SLC7A5, HTF9C, P53, NDRG1, and CEACAM5) to stratify patients into risk groups to inform treatment decisions. These markers are independent of one another and do not directly measure either proliferation or hormone receptor status.
- NPI+ is a biomarker based prognostic assay which integrates 10 predictive biomarkers of long term survival and therapeutic response with existing clinical and molecular pathology knowledge to support individualised clinical decision making.

5. Report methods for synthesis of evidence of clinical effectiveness

A systematic review of the evidence on the clinical effectiveness of gene and protein expression profiling tests to guide selection of chemotherapy regimes in breast cancer management will be

conducted. For two of the tests MammaPrint and OncotypeDX a recent systematic review exists (Smartt, 2009) therefore a summary of this review will be provided plus an update of this review will be conducted by searching for evidence on each of the two named tests and alternative names for each test for the period January 2009 to present date, and from 2002 on the product names and alternative names for the seven remaining tests. The review will be conducted following the general principles recommended in CRD's guidance (CRD, 2009), the PRISMA statement (Liberati *et al*., 2009), and The NICE Diagnostic Assessment Programme Interim Methods Statement (NICE, 2010).

Unpublished information received from manufacturers will be summarised separately. Unpublished information will only be considered if presented in a structured format, and the method reported in a sufficient detail. Due to the time constraints of the project priority will be given to peer-review articles in press, or submitted to peer-review journals, Other types of unpublished data, including research reports, databases and other non-peer reviewed materials will be considered only if deemed to provide important information by the Assessment Team and if time/resource constraints allow.

5.1 Inclusion and exclusion criteria

The titles and abstracts of records identified by the search strategy will be examined for relevance by one reviewer. Full papers of any potentially relevant records will be obtained where possible and screened by one reviewer. The relevance of each study to the review and the decision to include/exclude studies will be made according to the inclusion criteria detailed below. Any studies which give rise to uncertainty will be reviewed by a second reviewer with involvement of a third reviewer when necessary.

Population

Inclusion criteria: People diagnosed with early invasive breast cancer. Some tests may only be used in a sub-population. For example, women with early-stage invasive breast cancer (stage I, II or III), lymph node negative or positive (up to 3), oestrogen receptor positive or negative and HER2 positive or negative.

Interventions

Inclusion criteria: The assessment will include the gene expression profiling tests and expanded immunohistochemistry tests that have been identified by NICE. Tests to be included are: Randox Breast Cancer Array, MammaPrint + BluePrint, PAM50, MammaPrint, OncotypeDX, Breast Cancer Index, IHC4, Mammostrat and NPI+.

Comparators

The comparator will be current UK clinical practice. This includes the use of Adjuvant! Online or the Nottingham Prognostic Index (NPI), in combination with pathological parameters (eg, tumour size, grade and lymph node status), to predict survival and the utility of adjuvant therapy in breast cancer.

Outcomes

- Analytic validity (ie the ability of the test to accurately and reliably measure the expression of mRNA or proteins by breast cancer tumour cells),
- Clinical validity (ie the degree to which the test could accurately predict the risk of an outcome and discriminate patients with different outcomes),
- Clinical utility in relation to harm, impact on clinical decision making, evidence of improvement in outcomes and health care costs.

- Primary clinical outcomes to include: distant recurrence free survival at 10 years, local recurrence free survival at 10 years, total disease recurrence at 5 years, pathological complete response.
- Secondary outcomes to include: Health-related quality of life, including the impact of adverse events associated with chemotherapy. Reduction in overall chemotherapy use.

Setting

Tests which are used in secondary and tertiary care to make decisions about adjuvant chemotherapy treatment.

Study designs

Inclusion criteria: for the review of clinical effectiveness the best available level of evidence will be included, with priority given to controlled studies if available.

Exclusion criteria: studies will be excluded if they do not meet the inclusion criteria, appear to be methodologically unsound, or do not report methods and/or results in the necessary detail. The following will also be excluded:

- animal models
- preclinical and biological studies
- editorials and opinion pieces
- studies only published in languages other than English unless no other comparable data exist
- reports published as meeting abstracts will be excluded unless comparable data do not exist in full published studies and in such a case will only be included where sufficient methodological details are reported to allow critical appraisal of study quality
- studies applied only to breast cancer biology
- studies relating to these tests only in the neo-adjuvant treatment setting

5.2 Literature searching

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

- Searching of electronic databases;
- Contacting manufacturers;
- Contact with experts in the field;
- Scrutiny of bibliographies of included papers;
- Citation Searching of key papers.

The databases that will be searched include the following:

- MEDLINE and MEDLINE in Process (for latest publications);
- EMBASE;
- The Cochrane Library (including the Cochrane Systematic Reviews Database, Cochrane Controlled Trials Register, CENTRAL, and NHSEED)
- BIOSIS previews;
- Web of Knowledge.

Recent relevant conference proceedings including the St Gallen International Breast Cancer will be screened. In addition, relevant reviews and guidelines will be identified through the following resources: Clinical Evidence, National Institute for Health and Clinical Evidence (NICE) website, NHS Evidence – National Library of Guidelines, SIGN Guidelines, the Guidelines International Network website and the Medicines and Healthcare products Regulatory Agency.

Search terms will take into account product names and any alternative names for each of the tests. Product and alternative product names will be sought from information from manufacturers and their websites, searching full text of potentially included articles, review papers and their reference lists. A draft MEDLINE search strategy is included in Appendix 9.1)

The clinical and cost effectiveness searches will be limited by date from January 2009 to present for the OncotypeDX and MammaPrint (the search strategies from the existing systematic reviews appear to be of good quality and clearly reported and as a result all studies prior to September 2009 should have been identified). A 9 month window of overlap will be used when updating the literature search of these reviews to account for any publications that may not have yet been indexed in major science literature databases when Smartt (2009) conducted her literature search. For the other tests searches will be conducted from 2002 to present date. This date has been identified as a suitable start date by checking previous systematic reviews and submissions of reference lists from manufacturers. The first evidence for the tests included in the previous systematic review (MammaPrint or OncotypeDX) was reported in 2002. As these tests are the most established tests and furthest along the validation pathway, evidence for the subsequent tests will not predate this.

Reference lists of included papers will be assessed for additional relevant studies and where necessary, authors of eligible studies will be contacted for further information. All searches will be limited to human studies. No limits relating to study design will be applied to the searches.

5.3 Study selection and data extraction strategy

Data will be extracted by one reviewer using a standardised data extraction form and checked by another. Discrepancies will be resolved by discussion, with involvement of a third reviewer when necessary. If time constraints allow, attempts will be made to contact authors for any missing data. Data from multiple publications of the same study will be extracted as a single study. Moreover, as part of this systematic review is an update of two existing reviews, all relevant data will be extracted from the reviews in the first instance, but will be cross checked for accuracy with the original papers. If necessary, additional data will be extracted from the original papers. Supplementary information received directly from manufacturers will be summarised and tabulated separately.

5.4 Quality assessment strategy

The nature of the quality assessment which will be undertaken will depend on the types of studies identified, but will be undertaken using appropriate and established tools.

Although there are no validated tools for the assessment of the quality of tumour marker/gene expression profiling studies, Smartt (2009) used the general principles of the reporting recommendations for tumour marker prognostic studies (REMARK) to assess the quality of the studies. The REMARK guidelines were developed to encourage transparent and relevant reporting of study design, pre-planned hypotheses, patient and specimen characteristics, assay methods, and statistical analysis methods, in order to help others judge the usefulness of the data presented (McShane, Altman, Sauerbrei *et al.*, 2005). However, these guidelines are not fully suited to genetic risk prediction studies. Recently Janseens *et al* (2011) developed a checklist for strengthening the reporting of the genetic risk prediction studies (GRIPS) by building on the principles established by prior reporting guidelines (STREGA, REMARK, STARD). For this review, we will assess the study quality using the relevant sections of the GRIPS reporting guidelines (Janseens *et al.*, 2011). The assessment will be performed by one reviewer,

and independently checked by another. Discrepancies will be resolved by discussion, with involvement of a third reviewer when necessary.

5.5 Methods of analysis/synthesis

The results of data extraction will be tabulated and discussed as a narrative summary. If sufficient clinically and statistically homogenous data are available, data will be pooled using appropriate meta-analytic techniques to estimate a summary measure of effect on relevant outcomes. Clinical, methodological and statistical heterogeneity will be investigated.

6. Report methods for synthesising evidence of cost-effectiveness

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

6.1 Identifying and systematically reviewing published cost effectiveness studies

Databases to be searched are shown in section 5.2. Cost-effectiveness studies will be identified using an economic search filter. A draft MEDLINE search strategy is presented in Appendix 1 and will be adapted for use in other databases. In addition, relevant cost papers identified from the clinical effectiveness searches will be included in the economic review.

6.2 Evaluation of costs and cost effectiveness

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 9.2).

6.3 Development of a health economic model

Preliminary discussion with clinical experts indicates that patients diagnosed with breast cancer follow the diagnosis/treatment pathway described in *Figure 1*. GEP and expanded IHC tests aim to improve the use of chemotherapy in breast cancer by stratifying patients and identifying those patients who will gain most benefit from chemotherapy. These tests may report two types of information – breast cancer sub-types and/or risk of recurrence. Tests developed to provide information on sub-types might be used either before surgery for informing decisions on neo-adjuvant therapy or after surgery for informing decisions on adjuvant chemotherapy. Tests predicting the risk of recurrence in a specific population are likely to be used further down in the treatment pathway after surgery, in conjunction with other information available about tumour size, grade etc, 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 in this evaluation. 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 will be required on the impact on adjuvant chemotherapy treatment decisions of the new test, compared with current clinical practice (adjuvant online or NPI). Tests validated for use in predicting chemotherapy benefit will be distinguished from those using prognostic information as a proxy for chemotherapy benefit. Both predictive and prognostic 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.

A preliminary review of the evidence suggests that less robust data are available for the effect of molecular sub-typing tests comapred with the risk of recurrence tests. The potential role of sub-typing tests would be to add additional information into the existing decision making process. For instance information on luminal status may provide an indication of the likelihood of patients responding to chemotherapy. However, it is expected that evidence on the impact of sub-typing on decision-making will be limited or even lacking completely.

We anticipate the appropriate comparators for the risk of recurrence after surgery to guide the use of chemotherapy is expected to be the NPI score, Adjuvant! Online or any adaptation of these tools in clinical practice. It is expected that there might be some variation in clinical practice in the UK.

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 Interim Methods Statement (NICE, 2010). The model will adopt the perspective of the UK NHS and personal social services (PSS) 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 where required. Tests needing fresh samples (such as MammaPrint) may require significant re-organisation of pathology services, with resulting costs. Quality of life data will be reviewed and 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.

The development of the model is likely to be an iterative process. A conceptual model will be developed in conjunction with clinical experts to capture the current pathway of care for the diagnosis and management of breast cancer and how the new tests would change the pathway if routinely available in the NHS. The conceptual model will indicate the data requirements which will be sought both from the published literature and within commercial in confidence data held by the manufacturers. The model is likely to evolve following discussions with project stakeholders and the specialist committee members (SCMs), and according to the availability of data. 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. A range of scenarios will be presented varying main model assumptions to identify parameters that impact the most the ICER and to represent the uncertainty in parameters estimate. Furthermore, Probabilistic sensitivity analysis (PSA) will also be carried out using Monte Carlo simulation. The uncertainty in each parameter will be represented using a probability distribution. 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.

7. Handling the company submission(s)

All relevant data submitted by the manufacturers/sponsors will be considered if received by the TAR team no later than 27 May 2011. Data arriving after this date is unlikely to be considered, except data specifically requested by the Assessment team. 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.

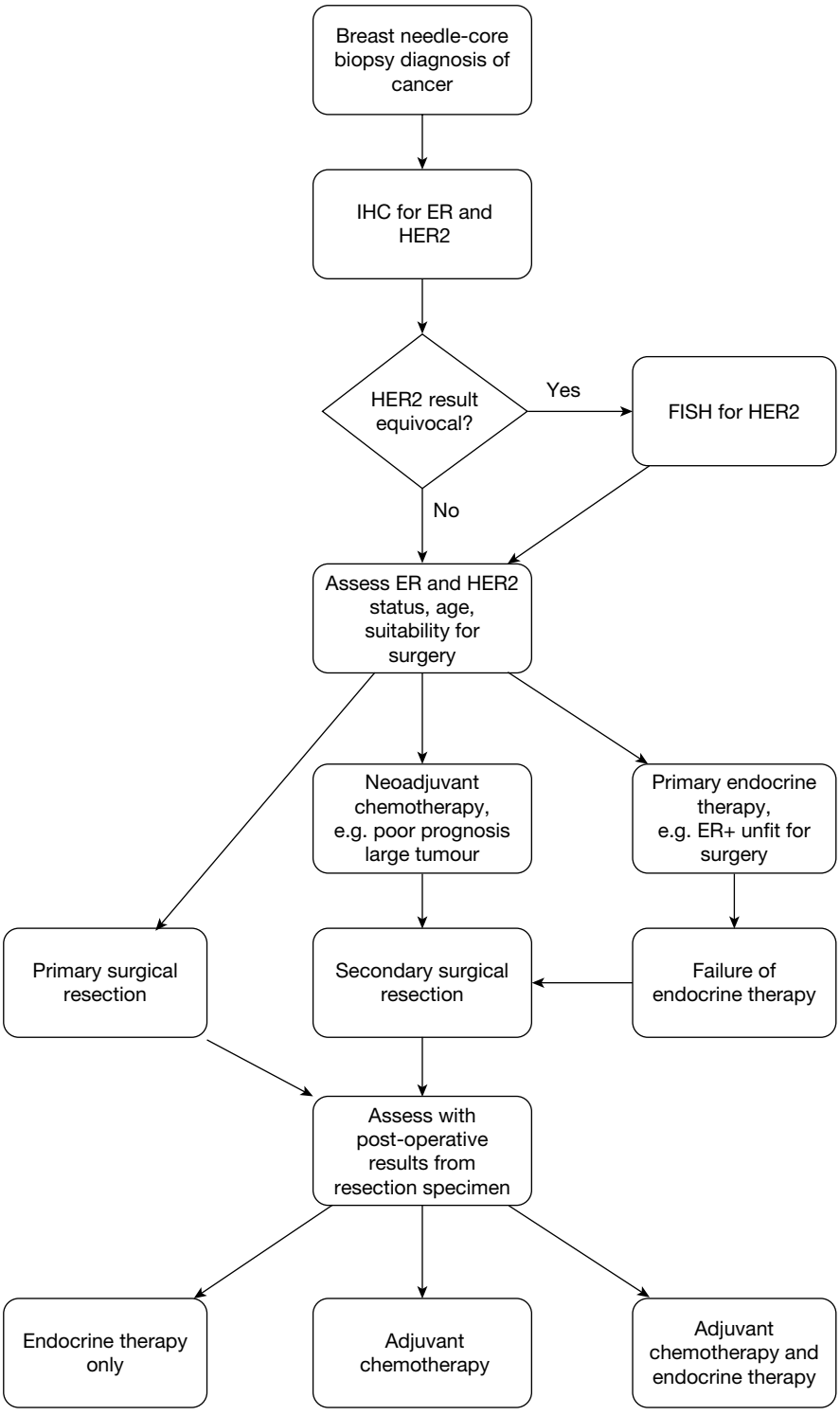


FIGURE 1 Diagnosis and management pathway in breast cancer.

Any economic evaluations included in the company submission, provided it complies with NICE's advice on presentation, will be assessed for clinical relevance, reasonableness of assumptions and appropriateness of the data used in the economic model. If the TAR team judge that the existing economic evidence is not robust, then further work will be undertaken, either by adapting what already exists or developing a de-novo model

Any 'commercial in confidence' data taken from a company submission, and specified as confidential in the check list, 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 manufacturers, and specified as such, will be highlighted in yellow and underlined in the assessment report. Any confidential data used in the cost-effectiveness models will also be highlighted.

8. Competing interests of authors

None

9. Appendices

9.1 Draft search strategy

Update search for OncotypeDX, and MammaPrint

Date limits = January 2009 – date
Filter = human studies only

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 medullary)).mp.
9. 1 or 2 or 3 or 6 or 7 or 8
10. MammaPrint.mp.
11. 70-gene.mp.
12. gene70.mp.
13. gene?seventy.mp.
14. seventy?gene.mp.
15. amsterdam profile.mp.
16. Oncotype.mp.
17. Oncotype DX.mp.
18. 21-gene.mp.
19. gene21.mp.
20. gene?twentyone.mp.
21. twentyone?gene.mp.
22. GHI Recurrence score.mp.
23. GHI-RS.mp.
24. 92-gene.mp.
25. gene92.mp.
26. gene?ninetytwo.mp.
27. ninetytwo?gene.mp.
28. RT-PCR (adj 5) 21.mp.
29. or/10–28
30. 9 and 29

Search for Randox, Blueprint, PAM50, Breast Cancer Index, IHC4, Mammostrat, and NPI+

Date limits = 2002 – date
Filter = human studies only

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 medullary)).mp.
9. 1 or 2 or 3 or 6 or 7 or 8
10. Randox.mp.
11. Blueprint.mp.
12. 80-gene.mp.
13. gene80.mp.
14. gene?eighty.mp.
15. eighty?gene.mp.
16. PAM50.mp.
17. 50-gene.mp.
18. gene50.mp.
19. gene?fifty.mp.
20. fifty?gene.mp.
21. breast bioclassifier.mp.
22. Breast Cancer Index.mp.
23. Breast cancer gene expression ratio.mp.
24. 2-gene.mp.
25. Two-gene-index.mp.
26. 2-gene-index.mp.
27. Two?gene.mp.
28. gene?two.mp.
29. H?I.mp.
30. H:I.mp.
31. 5-gene.mp.
32. gene5.mp.
33. gene?five.mp.
34. five?gene.mp.
35. 7-gene.mp.
36. seven-gene.mp.
37. gene7.mp.
38. gene?seven.mp.
39. Theros.mp.
40. Biotheranostics.mp.
41. Theros breast cancer index.mp.
42. HOXB13$.mp.
43. homeobox?13$.mp.
44. interleukin?17B$.mp.
45. IL17BR.mp.

46. mammostrat.mp.
47. five-biomarker-assay.mp.
48. IHC4.mp.
49. NPI+.mp.
50. Nottingham prognostic index plus.mp.
51. Nottingham prognostic index +.mp.
52. or/10–51
53. 9 and 52

9.2 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	

Details of TAR team

1. Ward Sue Ms (Project Manager & supervisor for economic modelling)

Senior Research Fellow
Health Economic & Decision Science (HEDS)
School of health & related research (ScHARR)
The University of Sheffield
Regent Court
30 Regent Street
Sheffield
S1 4DA

Tel: (+44) (0)114 222 0816
Fax: (+44) (0)114 272 4095
E-mail: s.e.ward@sheffield.ac.uk

2. Rafia, Rachid Mr (Economic modeller)

Research Associate, HEDS
Address – as above

Tel: (+44) (0)114 222 0739
Fax: (+44) (0)114 272 4095
E-mail: r.rafia@sheffield.ac.uk

3. Scope, Alison Dr (Systematic Reviewer)

Systematic reviewer, HEDS.

Tel: (+44) (0)114 222 0670
Fax: (+44) (0)114 272 4095
E-mail: a.scope@sheffield.ac.uk

4. Evans, Pippa (Information Specialist)

Information Specialist, HEDS,

Tel: (+44) (0)114 222 0801
Fax: (+44) (0)114 272 4095
E-mail: p.evans@sheffield.ac.uk

Major documentation should be sent to the project lead (Sue Ward, s.e.ward@sheffield.ac.uk), the project administrator (Gill Rooney, g.rooney@sheffield.ac.uk) and the managing director of ScHARR-TAG (Eva Kaltenthaler, e.kaltenthaler@sheffield.ac.uk).

Timetable/milestones

Progress report (to NETSCC, HTA who forward it to NICE within 24hr): 15 July 2011.

Draft assessment report (simultaneously to NICE and NETSCC, HTA): 22 August 2011.

Assessment Report (simultaneously to NICE and NETSCC, HTA): 19 September 2011.

References

1. Centre for Reviews and Dissemination. *Systematic reviews: CRD's guidance for undertaking reviews in health care.* York: University of York; 2009. URL: www.york.ac.uk/inst/crd/systematic_reviews_book.htm.
2. Drummond, M and Jefferson, TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ* 1996;**313**:275–283.
3. Eddy, D. M. The role of mathematical modeling. 1985;144–154.
4. Janssens ACJW, Ioannidis JPA, van Duijn CM, Little J, Khoury MJ. Strengthening the Reporting of Genetic Risk Prediction Studies: The GRIPS Statement. *Ann Intern Med* 2011;**154**:421-W141.
5. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009:**339**:b2700.
6. Marchionni L, Wilson RF, Marinopoulos SS, Wolff AC, Parmigiani G, Bass EB, Goodman SN. *Impact of Gene Expression Profiling Tests on Breast Cancer Outcomes*. Evidence Report/Technology Assessment No. 160. (Prepared by The Johns Hopkins University Evidence based Practice Center under contract No. 290–02–0018). AHRQ Publication No. 08-E002. Rockville, MD: Agency for Healthcare Research and Quality. *January* 2008.
7. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Instit* 2005;**97**:1180–1184.
8. NICE Diagnostic Assessment Programme Interim Methods Statement (Pilot). *February* 2010. URL: www.nice.org.uk/media/09E/D5/DiagnosticsAssessmentProgrammeInterimMethodsStatement.pdf (accessed 31 March 2011).
9. Office for National Statistics, Cancer Statistics registrations: *Registrations of cancer diagnosed in 2007, England*. Series MB1 no.38. 2010. National Statistics: London.
10. Paik, S. Molecular assays to predict prognosis of breast cancer. *Clin Adv Hematol Oncol* 2007;**5**:681–682.
11. Smartt, P. A comparison of gene expression profiling tests for breast cancer. *HSAC Report* 2010;**3**(16).

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library