Diagnostics Assessment Report commissioned by the NIHR HTA Programme on behalf of the National Institute for Health and Care Excellence

Tumour profiling tests to guide adjuvant chemotherapy decisions in people with breast cancer (update of DG10)

Produced by School of Health and Related Research (ScHARR), The University of Sheffield

Authors Sue Harnan, Senior Research Fellow, ScHARR, University of Sheffield

Paul Tappenden, Reader in Health Economic Modelling, ScHARR, University of Sheffield

Katy Cooper, Senior Research Fellow, ScHARR, University of Sheffield

John Stevens, Reader in Decision Science, ScHARR, University of Sheffield

Alice Bessey, Research Associate, ScHARR, University of Sheffield

Rachid Rafia, Research Fellow, ScHARR, University of Sheffield

Sue Ward, Senior Research Fellow, ScHARR, University of Sheffield

Ruth Wong, Information Specialist, ScHARR, University of Sheffield

Robert C Stein, Consultant Oncologist and Professor of Breast Oncology, NIHR University College Hospitals Biomedical Research Centre, London and Research Department of Oncology, UCL, London

Janet Brown, Professor of Medical Oncology, Department of Oncology and Metabolism, University of Sheffield

Correspondence Author Paul Tappenden, Reader in Health Economic Modelling, ScHARR, University of Sheffield, Sheffield, UK

Date completed Date completed 18th October 2017

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 16/30/03.

Declared competing interests of the authors

Professor Rob Stein is the Chief Investigator of the OPTIMA (Optimal Personalised Treatment of breast cancer using Multi-parameter Analysis) trial and has published results from the OPTIMA prelim study. None of the other authors have any conflicts of interest to declare.

Acknowledgements

We would like to thank Professor Matt Stevenson, Dr John Graham, and Professor John Bartlett for providing comments on the draft report, and Andrea Shippam and Gill Rooney, ScHARR, for providing administrative support and in preparing and formatting the report. We also thank the TransATAC team, the NCRAS and the UK Breast Cancer Group for providing bespoke analyses and data to inform this assessment.

Rider on responsibility for report

This report was commissioned by the NIHR HTA Programme. The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Harnan S, Tappenden P, Cooper K, Stevens J, Bessey A, Rafia R, Ward S, Wong R, Stein R, Brown J. Tumour profiling tests to guide adjuvant chemotherapy decisions in people with breast cancer (update of DG10). *Technology Assessment Report: Final report to the National Institute for Health and Care Excellence, 2017.*

Contributions of authors

Sue Harnan and Katy Cooper undertook the clinical evidence review. Paul Tappenden, Sue Ward, Alice Bessey and Rachid Rafia undertook the health economic analysis. John Stevens provided statistical advice. Ruth Wong undertook the literature search. Rob Stein and Janet Brown provided clinical advice. All authors were involved in drafting and commenting on the final report.

About ScHARR

The School of Health and Related Research (ScHARR) is one of the four Schools that comprise the Faculty of Medicine at the University of Sheffield. ScHARR brings together a wide range of medical- and health-related disciplines including public health, general practice, mental health, epidemiology, health economics, management sciences, medical statistics, operational research and information science. It includes the Sheffield unit of the Trent Institute for Health Services Research, which is funded by NHS R&D to facilitate high-quality health services research and capacity development.

The ScHARR Technology Assessment Group (ScHARR-TAG) synthesises research on the clinical effectiveness and cost-effectiveness of healthcare interventions for the NIHR Health Technology Assessment Programme on behalf of a range of policy makers, including the National Institute for Health and Care Excellence (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; the NHS Centre for Reviews and Dissemination, University of York; Warwick Evidence, The University of Warwick; the BMJ Group and Kleijnen Systematic Reviews.

Table of contents

List of tables

List of figures

List of boxes

1 LIST OF ABBREVIATIONS

$CAN	Canadian dollars
3rdHNC	Third Hospital of Nanchang
ABCSG	Austrian Breast and Colorectal Cancer Study Group
AE	Adverse event
AiC	Academic-in-confidence
AJCC	American Joint Committee on Cancer
AML	Acute myeloid leukaemia
AOL	Adjuvant! Online
ASCO	American Society of Clinical Oncology
AUC	Area under the curve
BCSS	Breast cancer-specific survival
BNF	British National Formulary
BRCA1	Breast Cancer Gene One
BRCA2	Breast Cancer Gene Two
CAF	Cyclophosphamide, doxorubicin and fluorouracil
CALGB	Cancer and Leukemia Group B
CCRCT	Cochrane Central Register of Controlled Trials
CCSYSU	Cancer Centre of Sun Yat-sen University
cDNA	Complementary deoxyribonucleic acid
CDSR	Cochrane Database of Systematic Reviews
CE	Conformité Européene
CEAC	Cost-effectiveness acceptability curve
CG	Clinical guideline
CHF	Congestive heart failure
CI	Confidence interval
CINV	Chemotherapy-induced nausea and vomiting
CLP	Clinical linear predictor
CMF	Cyclophosphamide, methotrexate, and fluorouracil
CMF-T	Tamoxifen plus cyclophosphamide, methotrexate and fluorouracil
CP	Clinico-pathological
CPCI	Conference Proceedings Citation Index
CT	Chemotherapy
CTS	Clinical Treatment Score
DARE	Database of Abstracts of Reviews of Effects
DBCG	Danish Breast Cancer Cooperative Group
DCIS	Ductal carcinoma *in situ*
DDFS	Distant disease-free survival
DES	Discrete event simulation
DFS	Disease-free survival
DG	Diagnostic Guidance
DM	Distant metastases
DMFI	Distant metastasis-free interval
DMFS	Distant metastasis-free survival
DRFI	Distant recurrence-free interval
DRFS	Distant recurrence-free survival
DSA	Deterministic sensitivity analysis
EAG	External Assessment Group
EBC	Early breast cancer
EBCTCG	Early Breast Cancer Trialists' Collaborative Group
ECOG	Eastern Cooperative Oncology Group
EEACT	Economic evaluation alongside a clinical trial
EMBASE	Excerpta Medica dataBASE
EP	EndoPredict

EQ-5D	Euroqol 5-Dimensions
ER	Oestrogen receptor
ESBC	Early stage breast cancer
ESMO	European Society for Medical Oncology
ET	Endocrine therapy
FACT-B	Functional Assessment of Cancer Therapy-Breast cancer
FACT-G	Functional Assessment of Cancer Therapy-General
FEC100-Pw	Fluorouracil, epirubicin, cyclophosphamide and weekly paclitaxel
FEC100-T	Fluorouracil, epirubicin, cyclophosphamide and docetaxel
FEC75	Fluorouracil, epirubicin and cyclophosphamide
FFPE	Formalin-fixed, paraffin-embedded
FFT	Fresh frozen tissue
FN	Febrile neutropenia
G-CSF	Granulocyte colony-stimulating factor
GEICAM	Grupo Espanol de Investigation en Cancer de Mama
GEO	Gene Expression Omnibus
GEP	Gene expression profiling
GPG	Good prognosis group
HAAMMS	Hospital Affiliated Academy of Military Medical Science
HCHS	Hospital and Community Health Services
HER2	Human epidermal growth factor receptor 2
HR	Hazard ratio
HRQoL	Health-related quality of life
HTA	Health Technology Assessment
ICER	Incremental cost-effectiveness ratio
IDFS	Invasive disease-free survival
IES	Intergroup Exemestane Study
IHC	Immunohistochemistry
IHC4+C	IHC4 plus clinical factors
InT	Insufficient tissue
IPD	Individual patient data
IR	Intermediate-risk
ITT	Intention-to-treat
LN+	Lymph node positive
LN0	Lymph node negative
LR	Low-risk
LYG	Life year gained
mAOL	Modified Adjuvant! Online
MDS	Myelodysplastic syndromes
MDT	Multidisciplinary team
MEDLINE	Medical Literature Analysis and Retrieval System Online
MeSH	Medical subject headings
MF-T	Tamoxifen plus methotrexate and fluorouracil
Mg	Milligram
MINDACT	Microarray In Node-negative Disease may Avoid Chemotherapy
mm	Millimetre
MP/MMP	MammaPrint
MPG	Moderate prognosis group
mRNA	Messenger RNA
MV	Multivariate
NCBI	National Centre for Biotechnology Information
NCCN	National Comprehensive Cancer Network
NCRAS	National Cancer Registration and Analysis Service
NHS	National Health Service
NHS EED	NHS Economic Evaluation Database

NICE	National Institute for Health and Care Excellence
NPI	Nottingham Prognostic Index
NR	Not reported
NSABP	National Surgical Adjuvant Breast and Bowel Project
O-DX	Oncotype DX
ONS	Office for National Statistics
OPTIMA prelim	Optimal Personalised Treatment of early breast cancer usIng Multiparameter Analysis preliminary
OR	Odds ratio
OS	Overall survival
PAM50	Prediction Analysis of Microarray 50
PAS	Patient Access Scheme
PCR	Polymerase chain reaction
PGP	Poor prognosis group
PICOS	Population, intervention, comparator, outcome, study design
PR	Progesterone receptor
PROBAST	Prediction model study Risk Of Bias Assessment Tool
PROSPERO	International prospective register of systematic reviews
PSA	Probabilistic sensitivity analysis
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	Quality-adjusted life year
qRT-PCR	quantitative reverse transcription polymerase chain reaction
RASTER	MicroarRAy-prognoSTics-in-breast-cancER
RATHER	RAtional THerapy for breast cancer study
RCT	Randomised controlled trial
RdGG	Reinier de Graaf Hospital
RFI	Recurrence-free interval
RFS	Relapse-free survival
RNA	Ribonucleic acid
ROR	Risk of recurrence
ROR	Risk of recurrence
ROR-PT	PAM50 subtype call, proliferation score, and risk of recurrence score
RR	Relative risk
RRR	Relative risk reduction
RS	Recurrence score
RSPC	Recurrence score pathology-clinical
RT-PCR	Reverse transcription polymerase chain reaction
RxPONDER	Rx for Positive Node, Endocrine Responsive breast cancer
SAE	Serious adverse event
ScHARR	School of Health and Related Research
SCI-E	Science Citation Index Expanded
SD	Standard deviation
SE	Standard error
SEER	Surveillance, Epidemiology, and End Results
SfT	Sent for test
STA	Single Technology Appraisal
STAI	State-Trait Anxiety Inventory
SWOG	Southwest Oncology Group
SYSMH	Sun Yat-sen Memorial Hospital
TAILORx	The Trial Assigning IndividuaLized Options for Treatment (Rx)
TC	Docetaxel and cyclophosphamide
TEAM	Tamoxifen vs Exemestane Adjuvant Multinational
TF	Test failure

TM	Tamoxifen
TM+C	Tamoxifen plus chemotherapy
TNM	Tumour node metastases
TransATAC	Translational substudy of the Arimidex, Tamoxifen, Alone or in Combination
TRANSBIG	Translating molecular knowledge into early breast cancer management: building on the BIG (Breast International Group) network for improved treatment tailoring
TTDM	Time to distant metastasis
UCL	University College London
UICC	Union International Contre le Cancer
UK	United Kingdom
UKBCG	UK Breast Cancer Group
US	United States
VOI	Value of information
WHO	World Health Organization
WSG Plan B	West German Study Group Plan B
WSG-AGO-Doc	West German Study Group epirubicine and cyclophosphamide-Doc
WTP	Willingness-to-pay

2 EXECUTIVE SUMMARY

2.1 Background

Breast cancer is the most commonly diagnosed cancer in women in England and Wales. In 2014, there were 55,222 new cases of breast cancer 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. A proportion of patients also receive neo-adjuvant therapy prior to surgery. Although chemotherapy can reduce the likelihood of cancer recurrence and death for women with breast cancer, it may have considerable adverse effects. Improved information on a patient's risk of recurrence (i.e. prognostic risk) and/or likely response to chemotherapy (i.e. predictive benefit) may help target chemotherapy to those patients who will benefit the most. Avoiding chemotherapy in patients at low-risk of recurrence, who would therefore obtain limited benefit, avoids the unpleasant side effects of chemotherapy and reduces expenditure on both the chemotherapy itself and the treatment of these adverse effects. Tumour profiling tests aim to improve the use of chemotherapy in breast cancer by improving the categorisation of patients according to risk and the identification of those patients who will gain most benefit from chemotherapy.

2.2 Objectives

The overall aim of the assessment is to address the question "*Do tumour profiling tests used for guiding adjuvant chemotherapy decisions in patients with early stage breast cancer represent a clinically effective and cost-effective use of NHS resources?"* This includes an update of the systematic review and cost-effectiveness analysis that informed NICE DG 10.

The objectives of the assessment are as follows:

- To conduct a systematic review of the published evidence on the effectiveness and cost effectiveness of five tumour profiling tests with or without clinicopathological factors (EndoPredict, Oncotype DX, MammaPrint, IHC4 and Prosigna) to guide decisions about adjuvant chemotherapy.
- To develop a health economic model to assess the cost-effectiveness associated with the use of tumour profiling tests compared with current prognostic tools to guide the use of adjuvant chemotherapy in early breast cancer from the perspective of the NHS and Personal Social Services (PSS).

2.3 Methods

2.3.1 Clinical evidence review methods

A systematic review was undertaken, including searching of nine databases in February 2017 plus other sources including a previous review published in 2013. The review included studies assessing clinical effectiveness of the five tumour profiling tests to guide decisions about adjuvant chemotherapy in people with early breast cancer, with a focus on those with ER-positive HER-2 negative stage I-II cancer with 0 to 3 positive lymph nodes. Outcomes included prognostic performance (whether recurrence and survival outcomes differ between test risk groups); prediction of chemotherapy benefit (whether effect of chemotherapy differs between test risk groups); clinical utility (impact of prospective use of the test on recurrence and survival); and decision impact (changes in chemotherapy recommendations pre/post-test).

2.3.2 Cost-effectiveness methods

The EAG undertook a review of existing economic analyses published since NICE DG10. The EAG also reviewed and critically appraised economic analyses of Oncotype DX, MammaPrint and EndoPredict which were provided during the course of the appraisal.

In addition, the EAG developed a *de novo* health economic model to assess the cost-effectiveness of Oncotype DX, MammaPrint, Prosigna, EPClin and IHC4+C, each versus current practice. The health economic analysis was undertaken from the perspective of the NHS and PSS and was largely based on the model developed to inform NICE DG10. The EAG model adopts a hybrid decision tree – Markov structure. The model parameters were informed by a number of sources including a bespoke analysis of the TransATAC trial, the MINDACT trial, a bespoke analysis of the NCRAS dataset, a bespoke survey disseminated by the UKBCG, the NHS England Access Scheme Database, standard costing sources and other literature.

2.4 Results

2.4.1 Clinical evidence results

The review included 153 studies across all five tests and across all outcomes listed in the NICE scope.

Among studies of LN0 patients receiving endocrine monotherapy, percentages categorised as high-risk ranged from 9-33% across all five tests. In LN+ patients, three tests (Prosigna/ROR-PT, EPClin [EndoPredict Clinical] and IHC4+C [IHC4 + clinical score]) categorised far more ([redacted] lymph node positive (LN+) than lymph node negative (LN0) patients as high-risk among studies of endocrine monotherapy, whilst Oncotype-DX categorised a similar

number as high-risk in LN0 and LN+ groups. However, Oncotype DX categorised more patients as low-risk in LN+ than other tests (57% in Oncotype DX versus 4% to [redacted]% in other tests), but with worse 10-year distant-recurrence free survival/interval (DRFS/DRFI) outcomes (82% in Oncotype DX versus 95% to 100% in other tests).

In terms of prognostic performance, all tests had statistically significant prognostic power in unadjusted analyses in LN0 and LN+ populations. However, recurrence score pathology-clinical (RSPC) was only validated in LN0 patients, and unadjusted analyses using clinical cut-offs were not reported in the validation sets for IHC4 or IHC4+C. All tests provided additional prognostic information over most commonly used clinicopathological factors and over clinical treatment score (CTS) and Nottingham Prognostic Index (NPI) in LN0. Results were more varied in LN+ patients.

There was some evidence of differential chemotherapy benefit between risk groups for Oncotype DX as shown by significant interaction tests between risk group and chemotherapy treatment in unadjusted analyses, but interaction tests sometimes became non-significant when clinicopathological factors were adjusted for. Oncotype DX RSPC (Oncotype DX plus age, tumour size and grade) was prognostic but not statistically significantly predictive for chemotherapy benefit, indicating that the incorporation of CP factors to Oncotype DX may reduce prediction of chemotherapy benefit.

Evidence relating to the ability of MammaPrint to predict benefit from chemotherapy was extremely limited. Although the effect of chemotherapy was significant in high-risk groups and not in low-risk groups, interaction tests between risk groups and chemotherapy treatment were not significant, suggesting no statistically significant difference in effect of chemotherapy between risk groups.

For Oncotype DX and MammaPrint, evidence from observational, non-comparative studies assessing the impact of the test used prospectively in clinical practice suggested that recurrence/survival outcomes in low-risk groups were acceptable even with low rates of chemotherapy. There was no similar evidence relating to the other tests.

The MINDACT randomised controlled trial (RCT) for MammaPrint reported that for patients who were high-mAOL, low-MammaPrint risk, chemotherapy gave an absolute benefit of 1.5% in 5 year DRFI. This raises the possibility of avoiding chemotherapy in these patients. In patients who were low-mAOL, high-MammaPrint risk, chemotherapy gave an absolute

benefit of 0.8%. This could be interpreted to mean MammaPrint would not be a useful test in mAOL low-risk patients, as it would not alter treatment decisions.

Decision impact studies from the UK and Europe reported that the percentage of patients with any change in chemotherapy recommendation or decision pre-/post-test ranged from 27% to 49% across UK studies (included Oncotype DX, EndoPredict and IHC4+C) and from 5% to 70% across European studies (included all tests except IHC4). The net change in the percentage of patients with a chemotherapy recommendation or decision pre-/post-test ranged from an increase of 1% to a decrease of 23% among UK studies, and a decrease of 0% to 64% across European studies.

Concordance between tests was not fully reviewed, but one UK study (OPTIMA prelim) which compared Oncotype DX, MammaPrint, Prosigna and IHC4 concluded that whilst tests assigned similar proportions of patients to low/intermediate and high-risk categories, test results for an individual patient could differ markedly depending on which test was used.

Data relating to anxiety and health-related quality of life (HRQoL) was limited as most studies did not include a comparator, instead adopting a pre-test/post-test design. Anxiety generally reduced post-test, but it is unclear if this would occur equally after a treatment decision made according to clinical factors. HRQoL improved in some analyses.

Microarray studies support conclusions from studies using the commercial versions of the assays in suggesting that Oncotype DX, MammaPrint and EndoPredict can discriminate between high- and low-risk patients regardless of LN status (there were no relevant microarray studies for EndoPredict or IHC4).

2.4.2 Cost-effectiveness results

The EAG's base case model suggests the following results.

Oncotype DX: Within the LN0 NPI≤3.4 subgroup, the ICER for Oncotype DX versus current practice is expected to be £122,725 per QALY gained (£34,245 per QALY gained assuming a predictive benefit). Within the LN0 NPI>3.4 and LN+ (1-3 nodes) subgroups, Oncotype DX is expected to be dominated by current practice (conversely, Oncotype DX dominates current practice if a predictive benefit is assumed). The results generated using the EAG model are primarily driven by the modelled reduction in the use of adjuvant chemotherapy using the Oncotype DX test. When based on the same evidence sources, the Genomic Health model produces broadly similar results.

IHC4+C: Within the LN0 NPI≤3.4 subgroup, the ICER for IHC4+C versus current practice is expected to be £2,654 per QALY gained. Within the LN0 NPI>3.4 and LN+ (1-3 nodes) subgroups, IHC4+C is expected to dominate current practice.

Prosigna: Within the LN0 NPI≤3.4 subgroup, the ICER for Prosigna versus current practice is expected to be £91,028 per QALY gained. Within the LN0 NPI>3.4 and LN+ (1-3 nodes) subgroups, the ICERs for Prosigna versus current practice are £26,058 and £28,731 per QALY gained, respectively.

EPClin: Within the LN0 NPI≤3.4 subgroup, the ICER for EPClin versus current practice is expected to be £147,419 per QALY gained. Within the LN0 NPI>3.4 subgroup, the ICER for EPClin versus current practice is expected to be £46,788 per QALY gained. Within the LN+ (1-3 nodes) subgroup, the ICER for EPClin versus current practice is expected to be £21,458 per QALY gained.

MammaPrint: Within the overall MINDACT population, the ICER for MammaPrint versus current practice is expected to be £131,482 per QALY gained. Within the mAOL high-risk subgroup, MammaPrint is expected to be dominated by current practice. Within the mAOL low-risk subgroup, the ICER for MammaPrint versus current practice is expected to be £414,202 per QALY gained.

2.5 Discussion

Strengths and limitations in the clinical evidence base

The evidence base was large, and included multiple reanalyses of RCTs, which are generally considered to be a high quality source of data. However, nearly all studies excluded patients who did not have enough tissue sample.

There were some key gaps in the literature for IHC4+C and RSPC. Notably, the IHC4+C and RSPC has only been validated in one cohort each. There are known problems with conducting the analyses required for IHC4, and it is unclear whether the absolute IHC4 values obtained would be similar across centres.

Much of the evidence base relates to unadjusted analyses, which do not assess the crucial question of whether a test has additional value over clinicopathological factors. Where adjusted analyses were performed, the clinicopathological variables included were not always consistent, and it is unclear if all important factors were included in all analyses.

There were relatively limited data relating to the ability of Oncotype DX and MammaPrint to predict benefit from chemotherapy, and some of the analyses conducted were also subject to criticisms relating to adjustment for all relevant variables.

Many studies were observational in nature, and these are subject to confounding whereby patients who received, or who were selected on the basis of indication for, chemotherapy are likely to be systematically different in terms of known (and potentially unknown) prognostic variables.

The evidence base relating to impact of tests on treatment decisions (decision impact studies) was limited in that use of chemotherapy differs across countries and there were no UK studies for two tests (MammaPrint; Prosigna), and only one UK study for 2 tests (EndoPredict; IHC4+C).

Strengths and limitations relating to the health economic analysis

The EAG model has a number of strengths; in particular: (i) for all tests, risk classification and DMFS probabilities are derived from the same source (TransATAC or MINDACT); (ii) within the LN0 intermediate-risk subgroup (NPI>3.4, analysis of 3-level tests), the probability of receiving chemotherapy with and without the test is based on the NHS England Access Scheme dataset – this is likely to best reflect how the 3-level tumour profiling tests would be used in clinical practice in England; (iii) the model structure is consistent with that of other published models of tumour profiling tests - when similar data inputs are used, the EAG model produces similar results to the previous EAG model and the Genomic Health model, and (iv) extensive deterministic sensitivity analyses have been conducted to explore the impact of uncertainty on the model results.

However, the model is also subject to several limitations, most of which stem from uncertainties in the evidence base. The main limitations and uncertainties relating to the cost-effectiveness analysis are: (i) with the exception of Oncotype DX in the LN0 NPI>3.4 group (clinical intermediate risk), the evidence surrounding the pre- and post-test chemotherapy probabilities is subject to considerable uncertainty – this has the propensity to influence the conclusions regarding the cost-effectiveness of all tests; (ii) there is uncertainty regarding whether Oncotype DX and MammaPrint are predictive of chemotherapy benefit – the inclusion of such effects are likely to strongly influence economic conclusions drawn from the analysis; (iii) the analysis of MammaPrint is based on a different data source than the other four tests; and (iv) the TransATAC study used to estimate test risk classification and

DMFS probabilities was the derivation study for IHC4 – as such, there is potential for the overestimation of prognostic performance for this test.

2.6 Implications for service provision

The per test costs for Prosigna provided by NanoString (used in the EAG economic analyses) are based on an efficient level of throughput. This may not hold if centres do not undertake the anticipated number of tests (for example, in smaller centres, or if multiple tumour profiling tests are available). Furthermore, as NanoString does not offer a centralised testing service, local testing services will need to be established.

IHC4 is not currently commercially available. Standardisation of IHC4 and quality assurance programs are required before this test may be used routinely within the NHS.

2.7 Suggested research priorities

- There is uncertainty regarding whether Oncotype DX and MammaPrint are predictive of chemotherapy benefit. Further studies are required which adjust for all relevant clinico-pathological factors.
- There is limited evidence demonstrating long-term impacts resulting from the use of the five tumour profiling tests. Future studies assessing the comparative long-term impact of the tests compared with risk prediction tools commonly used in clinical practice would be valuable.
- There is uncertainty regarding the cost-effectiveness of all five tests included in the NICE scope. It is noteworthy that under the assumption of no predictive chemotherapy benefit, the inclusion of additional data collected through the NHS England Access Scheme Dataset has a significant impact upon the conclusions previously drawn from the Oncotype DX analysis within NICE DG10 (moving from an ICER of £22,572 per QALY gained to a situation in which Oncotype DX is dominated in the LN0 NPI>3.4 subgroup). Additional UK-based data collection relating to pre- and post-test chemotherapy use for EPClin, IHC4+C, Prosigna and MammaPrint may be important in reducing existing uncertainty surrounding the cost-effectiveness of these tests.

3. BACKGROUND AND DEFINITION OF THE DECISION PROBLEM

3.1 Condition and aetiology

Breast cancer is the most commonly diagnosed cancer in women in England and Wales. In 2014, there were 55,222 new cases of breast cancer diagnosed.[1] 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. A proportion of patients also receive neo-adjuvant therapy prior to surgery.

3.1.1 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.[2]

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.[3] Genetic predisposition is mediated by high-penetrance genes such as Breast Cancer Gene One (BRCA1) and Breast Cancer gene two (BRCA2), which are responsible for around 80-90% of hereditary cancers, and low-penetrance genes which confer increased and decreased risk.[2]

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, whilst Asian-Americans born in the West have incidence rates approximating the US average.[4] 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 which decrease risk include higher folate intake, higher number of pregnancies, breast feeding, and younger age at first birth.[2] Obesity increases risk of breast cancer in post-menopausal women.[5] The picture is less clear for pre-menopausal women whereby risk may be lower, but prognosis is poorer. Physical activity in adolescence and young adulthood confers a decreased risk of breast cancer,[6] which may be mediated hormonally.

Pathology

Breast cancer starts with genetic changes in a single or small group of cells in the epithelia of the ducts or the lobules of the breast. The genetic change allows cells to reproduce uncontrollably, resulting in a tumour. Tumours that have not yet spread to surrounding tissue

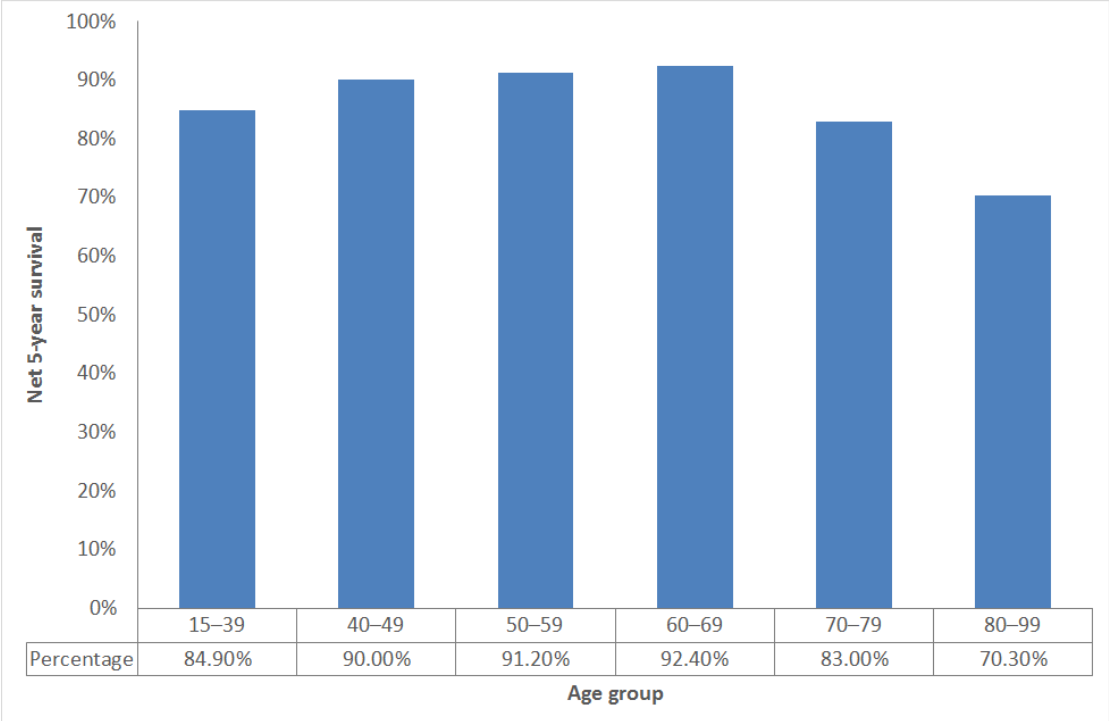
are known as "carcinoma *in situ.*" Once spread to surrounding tissue begins, a tumour is known as "invasive". More rapid growth and spread occurs once a blood supply is secured. Cancer spreads via the lymphatic system or the bloodstream. Lymphatic spread is usually first to the axillary lymph nodes. Spread via the bloodstream can lead to distant metastases in the bone or viscera which are incurable.

The presence or absence of axillary lymph node metastases is a key indicator of disease and prognosis and adjuvant therapy is, in part, planned based on their presence and extent.[7] They are caused when a single or small numbers of cells detach from the main tumour, travel via the lymphatic system and establish themselves in the tissue of the axillary lymph nodes. Axillary metastases occur in approximately 41% of cases;[8] prognosis is better where there is no axillary spread. Where metastases are present, axillary clearance is indicated in order to prevent further spread and ensure local disease control.

Prognosis

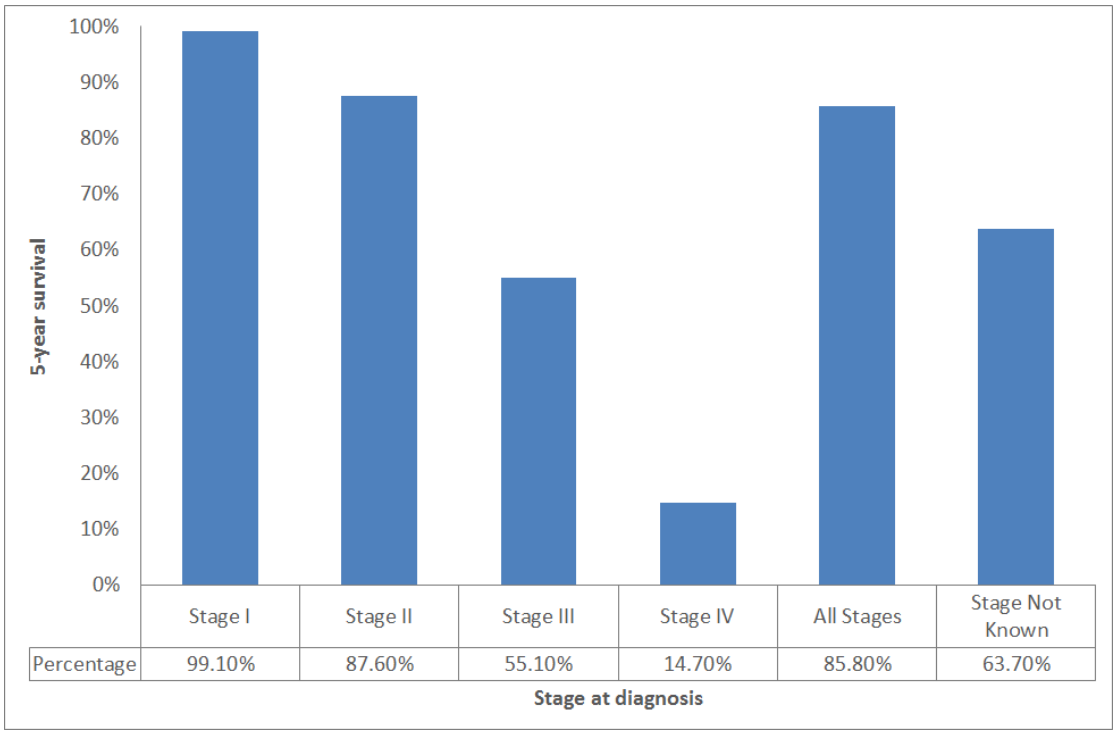
Overall, 5-year, age-standardised survival rates for women with breast cancer are 86.3%.[9] Survival varies with age (Figure 1) and stage of disease (Figure 2).

Figure 1: 5-year net survival by age, women, England, 2009-2013



Source: Cancer Research UK[10]

Figure 2: Survival by stage, women, five-year relative survival by stage, women (aged 15-99 years), former Anglia Cancer Network, 2002-2006



Source: Cancer Research UK[1]

Other factors can also affect prognosis. Clinicians may use tools such as the Nottingham Prognostic Index (NPI),[11] PREDICT or Adjuvant! Online (AOL) to predict disease course and treatment options, although it should be noted that AOL is in the process of being updated and is not currently available. These tools take into account different patient and tumour factors and may give different risk predictions for the same patient.

In general, good prognosis is associated with small tumour size, lymph node-negative (LN0) status, younger age, oestrogen receptor positive (ER+) status and progesterone receptor positive (PR+) status. Overexpression of human epidermal growth factor receptor 2 (HER2) is associated with poorer prognosis.

3.1.2 Epidemiology and incidence

Incidence varies most according to gender. Women are considerably more likely to develop breast cancer than men. For both genders, incidence varies with age (see Table 1). Over 81% of cases occur in women aged 50 years and over. Based on 2014 data, the highest incidence rates for women were reported in the 60-70 year age range.[12]

Table 1: Incidence per 100,000 for England by age group & gender, 2014

Age	Men	Women
All ages	**319**	**45,764**
0-24	0	21
25-29	0	191
30-34	4	593
35-39	3	1,071
40-44	7	2,299
45-49	11	4,369
50-54	17	5,386
55-59	23	4,589
60-64	30	5,072
65-69	57	6,502
70-74	45	4,436
75-79	52	3,889
80-84	42	3,419
85+	28	3,927

Incidence 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. The rate ratios are 0.65, 0.75, 0.49 and 0.58, respectively, when compared to the white group.[13]

Based on data for the period 2006-2010, the incidence of female breast cancer was highest in the least deprived 20% of the population; however, the more deprived had statistically significantly higher mortality.[14] It is unclear why this is, but may be due to lower levels of screening compliance, worse overall general health status and lower levels of treatment due to access and compliance issues.

3.1.3 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 34.6 per 100,000 women. In 2014, this constituted 11,360 deaths for women in the UK.[1]

3.1.4 Measurement of disease

Breast cancer has few obvious symptoms and can easily go undetected for a few years. Amongst 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 via presentation to a general practitioner. Women between the ages of 50 and 70 are routinely invited to attend regular screening; the NHS is currently in the process of extending the programme as a trial, offering

screening to some women aged 47-73 years. A recent case control study within the English Breast Screening Program reported that attendance at breast screening resulted in a breast cancer mortality reduction of 39% (odds ratio [OR], 0.61; 95% confidence interval [CI]: 0.44, 0.85) after self-selection correction.[15] Screening increases the proportion of tumours detected in the early, more curable stages.

The breast mass and axillary areas are investigated clinically through palpation and by mammography or ultrasound, and the status of the tumour is confirmed by histology of a percutaneous tissue biopsy. 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, whilst those with distant metastases are considered incurable. Patients with high-risk early breast cancer also undergo a CT of the chest and abdomen and a bone scan to assess any distant metastases.

Current methods for staging of breast cancer

Three main factors are used to stage breast cancer: (i) tumour size; (ii) metastases to the regional lymph nodes, and (iii) distant metastases. The tumour/node/metastases (TNM) staging system was developed and is maintained by the American Joint Committee on Cancer (AJCC) and the Union International Contre le Cancer (UICC).[16] T stage is classified according to 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 (see Table 2). The overall TNM stage of the cancer is defined as shown in Table 3. Early breast cancer is generally defined as cancer which has not spread beyond the breast or the ipsilateral axillary lymph nodes, and is confined to stages I, II or IIIA.

Table 2: Breast cancer staging, AJCC, version 7[17]

Primary tumour (T)	
Tx	Primary tumour cannot be assessed
T0	No evidence of primary tumour
Tis	Carcinoma *in situ*
Tis (DCIS)	Ductal carcinoma *in situ*
Tis (LCIS)	Lobular carcinoma *in situ*
Tis (Paget's)	Paget's disease of the nipple NOT associated with invasive carcinoma and/or carcinoma *in situ* (DCIS and/or LCIS) in the underlying breast parenchyma. Carcinomas in the breast parenchyma associated with Paget's disease are categorised based on the size and characteristics of the parenchymal disease, although the presence of Paget's disease should still be noted
T1	Tumour ≤ 20 mm in greatest dimension
T1mi	Tumour ≤ 1 mm in greatest dimension
T1a	Tumour > 1 mm but ≤ 5 mm in greatest dimension
T1b	Tumour > 5 mm but ≤ 10 mm in greatest dimension
T1c	Tumour > 10 mm but ≤ 20 mm in greatest dimension
T2	Tumour > 20 mm but ≤ 50 mm in greatest dimension
T3	Tumour > 50 mm in greatest dimension
T4	Tumour of any size with direct extension to the chest wall and/or to the skin (ulceration or skin nodules) Note: Invasion of the dermis alone does not qualify as T4
T4a	Extension to the chest wall, not including only pectoralis muscle adherence/invasion
T4b	Ulceration and/or ipsilateral satellite nodules and/or edema (including peau d'orange) of the skin, which do not meet the criteria for inflammatory carcinoma
T4c	Both T4a and T4b
T4d	Inflammatory carcinoma
Distant Metastases (M)	
M0	No clinical or radiographic evidence of distant metastases
cM0(i+)	No clinical or radiographic evidence of distant metastases, but deposits of molecularly or microscopically detected tumour cells in circulating blood, bone marrow, or other nonregional nodal tissue that are no larger than 0.2 mm in a patient without symptoms or signs of metastases
M1	Distant detectable metastases as determined by classic clinical and radiographic means and/or histologically proven larger than 0.2 mm
Regional Lymph Nodes (N)	
Clinical	
NX	Regional lymph nodes cannot be assessed (for example, previously removed)
N0	No regional lymph node metastases
N1	Metastases to movable ipsilateral level I, II axillary lymph node(s
N2	Metastases in ipsilateral level I, II axillary lymph nodes that are clinically fixed or matted; or in clinically detected ipsilateral internal mammary nodes in the absence of clinically evident axillary lymph node metastases
N2a	Metastases in ipsilateral level I, II axillary lymph nodes fixed to one another (matted) or to other structures
N2b	Metastases only in clinically detected ipsilateral internal mammary nodes and in the absence of clinically evident level I, II axillary lymph node metastases
N3	Metastases in ipsilateral infraclavicular (level III axillary) lymph node(s) with or without level I, II axillary lymph node involvement; or in clinically detected ipsilateral internal mammary lymph node(s) with clinically evident level I, II axillary lymph node metastases; or metastases in ipsilateral supraclavicular lymph node(s) with or without axillary or internal mammary lymph node involvement
N3a	Metastases in ipsilateral infraclavicular lymph node(s)
N3b	Metastases in ipsilateral internal mammary lymph node(s) and axillary lymph node(s)
N3c	Metastases in ipsilateral supraclavicular lymph node(s)

Pathologic (PN)	
pNX	Regional lymph nodes cannot be assessed (for example, previously removed, or not removed for pathologic study)
pN0	No regional lymph node metastasis identified histologically Note: Isolated tumour cell clusters (ITC) are defined as small clusters of cells not greater than 0.2 mm, or single tumour cells, or a cluster of fewer than 200 cells in a single histologic cross-section. ITCs may be detected by routine histology or by immunohistochemical (IHC) methods. Nodes containing only ITCs are excluded from the total positive node count for purposes of N classification but should be included in the total number of nodes evaluated.
pN(i-)	No regional lymph node metastases histologically, negative IHC
pN0(i+)	Malignant cells in regional lymph node(s) no greater than 0.2 mm (detected by H&E or IHC including ITC)
pN0(mol−)	No regional lymph node metastases histologically, negative molecular findings (RT-PCR)
pN0(mol+)	Positive molecular findings (reverse transcriptase/polymerase chain reaction, RT-PCR), but no regional lymph node metastases detected by histology or IHC
pN1	Micrometastases; or metastases in 1–3 axillary lymph nodes; and/or in internal mammary nodes with metastases detected by sentinel lymph node biopsy but not clinically detected
pN1mi	Micrometastases (greater than 0.2 mm and/or more than 200 cells, but none greater than 2.0 mm)
pN1a	Metastases in 1–3 axillary lymph nodes, at least one metastasis greater than 2.0 mm
pN1b	detected by sentinel lymph node biopsy but not clinically detected
pN1c	Metastases in 1–3 axillary lymph nodes and in internal mammary lymph nodes with micrometastases or macrometastases detected by sentinel lymph node biopsy but not clinically detected
pN2	Metastases in 4–9 axillary lymph nodes; or in clinically detected internal mammary lymph nodes in the absence of axillary lymph node metastases
pN2a	Metastases in 4–9 axillary lymph nodes (at least one tumour deposit greater than 2.0 mm)
pN2b	Metastases in clinically detected internal mammary lymph nodes in the absence of axillary lymph node metastases
pN3	Metastases in 10 or more axillary lymph nodes; or in infraclavicular (level III axillary) lymph nodes; or in clinically detected ipsilateral internal mammary lymph nodes in the presence of one or more positive level I, II axillary lymph nodes; or in more than three axillary lymph nodes and in internal mammary lymph nodes with micrometastases or macrometastases detected by sentinel lymph node biopsy but not clinically detected; or in ipsilateral supraclavicular lymph nodes
pN3a	Metastases in 10 or more axillary lymph nodes (at least one tumour deposit greater than 2.0 mm); or metastases to the infraclavicular (level III axillary lymph) nodes
pN3b	Metastases in clinically detected ipsilateral internal mammary lymph nodes in the presence of one or more positive axillary lymph nodes; or in more than three axillary lymph nodes and in internal mammary lymph nodes with micrometastases or macrometastases detected by sentinel lymph node biopsy but not clinically detected
pN3c	Metastases in ipsilateral supraclavicular lymph nodes

Table 3: Summary of TNM stages[17]

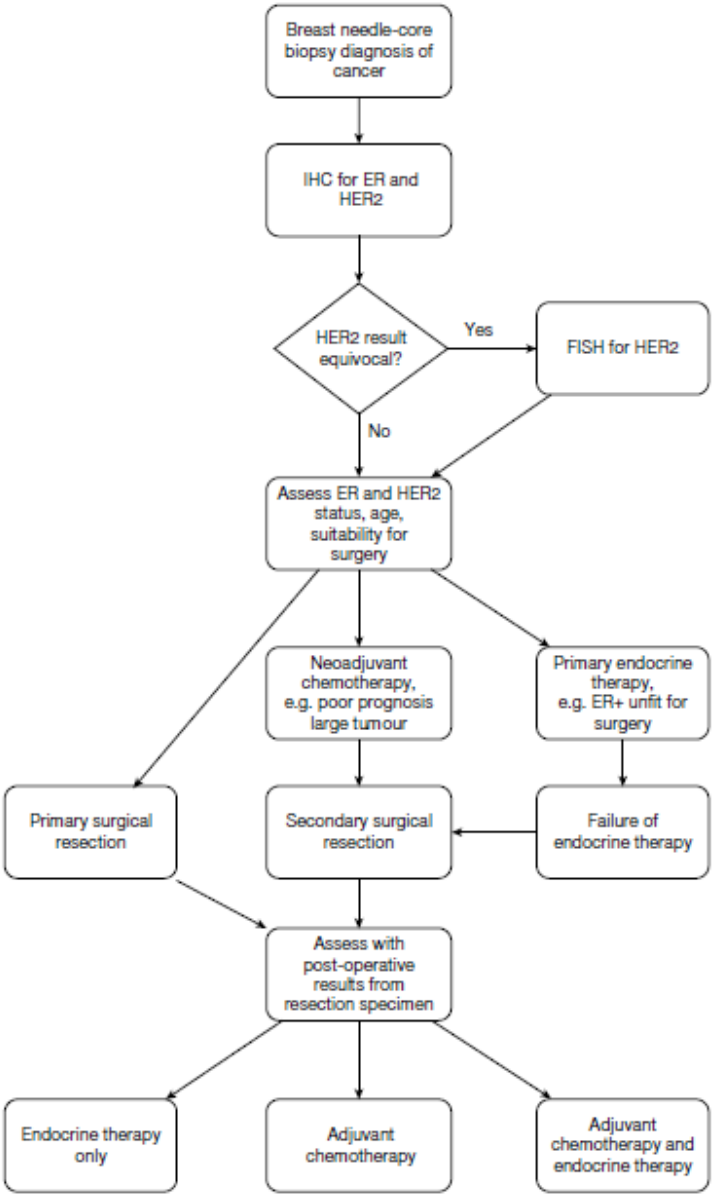
Stage	T	N	M
Stage 0	Tis	N0	M0
Stage IA	T1	N0	M0
Stage IB	T0	N1mi	M0
	T1	N1mi	M0
Stage IIA	T0	N1	M0
	T1	N1	M0
	T2	N0	M0
Stage IIB	T2	N1	M0
	T3	N0	M0
Stage IIIA	T0	N2	M0
	T1	N2	M0
	T2	N2	M0
	T3	N1	M0
	T3	N2	M0
Stage IIIB	T4	N0	M0
	T4	N1	M0
	T4	N2	M0
Stage IIIC	Any T	N3	M0
Stage IV	Any T	Any N	M1

3.2 Current service provision

3.2.1 Management of early breast cancer

Patients diagnosed with early breast cancer currently follow the diagnosis/treatment pathway described in Figure 3.

Figure 3: Diagnosis and management pathway in breast cancer[18]



Notes:
For post-menopausal women whose tumours are greater than grade 1, many centres also use adjuvant bisphosphonates for up to 3 years.
Patients may also be treated with adjuvant radiotherapy depending on whether they have had a wide local excision or mastectomy and depending on the characteristics of the primary tumour; this may not only include radiotherapy to the breast but may also include the chest wall, supraclavicular foca and lymph node and axillar.
Neo-adjuvant treatment may include pertuzumab and trastuzumab
Adjuvant chemotherapy may be given alongside biological therapy

3.2.2 Use of adjuvant chemotherapy

Since 2002, NICE have recommended that women at intermediate- or high-risk of recurrence who have not had neo-adjuvant chemotherapy should normally be offered a multi-agent chemotherapy which includes anythracyclines.[19] 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 (e.g. trastuzumab).

Meta-analyses of randomised clinical trials by the Early Breast Cancer Trialists' Collaborative Group (EBCTCG) have indicated that the use of adjuvant chemotherapy (chemotherapy following surgery) is associated with a reduction in the risk of cancer recurrence and death in women with early stage breast cancer.[20] However, chemotherapy may have considerable adverse effects (AEs). Short- and long-term AEs will affect a proportion of patients receiving chemotherapy, imposing additional costs and reducing health-related quality of life (HRQoL). Short-term AEs that occur during chemotherapy are usually temporary and reversible. The most common AEs include nausea, vomiting, mouth soreness, diarrhoea, tiredness, hair loss and temporary lowering of the blood counts. Long-term AEs such as damage to the heart, and a small increase in the risk of leukaemia are not reversible. Whilst 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 considerable 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.

Improved information on a patient's risk of recurrence (i.e. prognostic risk) and/or likely response to chemotherapy (i.e. predictive benefit) may help target chemotherapy at those patients who will benefit the most. Avoiding chemotherapy in patients at low-risk of recurrence, who would therefore obtain limited benefit, avoids the unpleasant side effects of chemotherapy and reduces expenditure on both the chemotherapy itself and the treatment of these adverse effects.

3.2.3 Current guidelines

NICE Clinical Guideline (CG) 80 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, the potential benefits and side effects of the treatment.[8] 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. The NICE guideline indicates that decisions regarding adjuvant therapy should be made following discussion of these factors with the patient and recommends consideration of the use of AOL to support estimations of individual

prognosis and the absolute benefit of adjuvant treatment for patients with early invasive breast cancer.[8] Whilst there is variation between centres, the NPI and PREDICT are also commonly used as the basis for many local guidelines on decisions regarding whether to use chemotherapy for patients with early breast cancer. These risk prediction tools include different patient and tumour characteristics and may give different predictions for the same patient (see Table 4).

The NICE CG80 guideline does not make specific reference to the use of tumour profile tests to aid decision-making. However, the NICE Diagnostics Guidance 10 (DG10) on tumour profiling[21] recommends Oncotype DX as an option for guiding adjuvant chemotherapy decisions for people with ER+ LN0 HER2- early breast cancer at intermediate (clinical) risk if Oncotype DX is likely to help in the decision of whether to use adjuvant chemotherapy.

Table 4: Breast cancer risk prediction tools

Tool	**NPI**	**AOL**	**PREDICT**
Factors included in prediction algorithm	• Tumour size • Nodal status • Tumour grade	• Age at diagnosis • Comorbidity factors • ER status • Tumour size • Tumour grade • Nodal status	• Age at diagnosis • Mode of detection • Tumour size • Tumour grade • Number of positive nodes • ER status • HER2 status • Ki67 status* • Generation of chemotherapy regimen
Outcome(s) predicted	Mortality	Mortality or relapse	Mortality

* PREDICT can also be used without Ki67 status

Adjuvant! Online

The AOL computer programme is designed to provide estimates of the benefits of adjuvant endocrine therapy and chemotherapy. The current version of AOL 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 are derived from the EBCTCG meta-analyses in order to provide estimates of 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. At the time of writing this report (October 2017), AOL was in the process of being updated and was not accessible.

Nottingham Prognostic Index (NPI)

The NPI is a composite prognostic parameter involving both time-dependent factors and aspects of biological aggressiveness. The NPI score is based on a combination of tumour grade, lymph node involvement and tumour size. To calculate the score: ADD numerical grade (1, 2, or 3), lymph node score (–ve = 1, 1 to 3 nodes = 2, >3 nodes = 3) and 0.2* tumour size in cm. Patients can be divided into three prognostic groups on the basis of the NPI: a good prognostic group (GPG; NPI < 3.4), a moderate prognostic group (MPG; 3.4 < NPI < 5.4), and a poor prognostic group (PPG; NPI > 5.4).

PREDICT (v2.0)

PREDICT (v2.0) is an online computer programme designed to help women with breast cancer and their doctors make informed decisions about treatment with chemotherapy or endocrine therapy following breast cancer surgery. PREDICT v2.0 was developed using data from over 5,000 women with breast cancer from England and has been tested on data from another 23,000 women with breast cancer from around the world. Patient and tumour characteristics are entered into the programme, which provides an estimate of the overall survival for patients with or without adjuvant hormone therapy, adjuvant chemotherapy and trastuzumab.

Clinical opinion suggests that there is variation in clinical practice between Trusts in the UK, with some centres using single risk prediction tools, and others using multiple tools in combination, in addition to other clinical parameters.

3.3 Description of technologies under assessment

Tumour profiling tests aim to improve the use of chemotherapy in breast cancer by improving the categorisation of patients according to risk of recurrence or death, and by identifying those patients who will gain most benefit from 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, nodal status and other factors to guide the use of adjuvant chemotherapy. Tests which require samples to be sent away for central review, following surgery, may introduce a short delay (of up to 3 weeks) before the decision can be taken on whether or not to offer chemotherapy.

Five tests were identified in the final NICE scope[22] and are included in this assessment: four are based on gene expression profiling (EndoPredict, MammaPrint, Oncotype DX and Prosigna) and one on immunohistochemistry (IHC4).

Gene expression profiling tests

Gene expression profiling tests investigate the expression of specific panels of genes (also known as a gene profile or gene signature). They do this by assessing 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. Gene expression profiling tests 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.

There are various ways of preparing the RNA, and different protocols may be used to prepare the specimens (e.g. formalin-fixed, paraffin-embedded [FFPE], snap-frozen and fresh samples). The tests included in this assessment use FFPE tissue and do not require the use of 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 gene expression profiling tests. These tests provide an estimate of the risk of recurrence.

Immunohistochemistry (IHC) tests

IHC markers are being developed to provide similar information to that given by gene expression profiling tests. Some of these tests offer the advantage of using existing immunohistochemistry technology (such as ER and HER2 markers) which is routinely available in all UK pathology departments, though methods for quantifying these markers in the format required for the test may not be routinely available.

Summary of tumour profiling tests included in the assessment

The key features of the five tests are summarised below and in Table 5.

EndoPredict (Myriad Genetics)

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

- ER-positive
- HER2-negative
- lymph node (LN)-negative (no positive nodes) or LN-positive (up to 3 positive nodes).

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

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

The test process involves using a reverse transcription-quantitative polymerase chain reaction (RT-qPCR), in which target messenger RNAs are reverse transcribed, amplified and simultaneously detected. The raw data are then exported to online evaluation software (EndoPredict Report Generator) which performs a quality check and calculates the EP score and the EPClin score. The EP score is a number on a scale between 0 and 15. An EP score of less than 5 indicates low-risk of distant disease recurrence reoccurring in the next 10 years. An EP score of 5 or more indicates a high-risk of distant disease recurrence in the next 10 years. The EPClin score is calculated by adding clinical data about tumour size and nodal status to the EP score. From the EPClin score, the probability of metastasis formation within 10 years is estimated, assuming 5 years of hormonal treatment. If the EPClin 10-year risk is less than 10%, the patient is classed as low-risk for metastases recurring in the next 10 years. If the EPClin 10-year risk is 10% or greater the patient is classed as high-risk for metastases recurring in the next 10 years. It takes approximately 2 days to obtain the test results if the test is done in-house. If samples are sent away for testing, the turnaround time for the central service is 4 to 5 working days.

MammaPrint (Agendia)

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

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

The test can be used irrespective of ER and HER2 status, that is, it can be used for tumours that are ER-negative or ER-positive, and HER2-negative or HER2-positive. MammaPrint measures the expression of 70 genes, including genes associated with 7 different parts of the metastatic pathway: (i) growth and proliferation; (ii) angiogenesis; (iii) local invasion; (iv) entering the circulation; (v) survival in the circulation; (vi) entering organs from the circulation, and (vii) adaption to the microenvironment at a secondary site. The MammaPrint test is offered as an off-site service. In Europe, samples are sent for analysis at the Agendia laboratory in Amsterdam, the Netherlands. The test requires a FFPE breast cancer tissue sample from a surgical specimen or core needle biopsy.

The test process involves isolation of RNA from FFPE sample followed by reverse transcription of the RNA to get complementary deoxyribonucleic acid (cDNA). The cDNA is amplified and labelled before being hybridised (bound) to the diagnostic microarray. The microarray is washed and then scanned using an Agilent DNA microarray scanner. The scan file is analysed using Agilent Feature Extraction Software and an algorithm is used to calculate the correlation of the sample profile to a "Low Risk" template profile on a scale of -1.000 to +1.000 with a cut off at 0. The threshold was set such that women with a low-risk result have a 10% risk of developing distant metastases over the next 10 years without any adjuvant hormone or chemotherapy. Test results are available to healthcare professionals within 10 days of submitting the sample.

Oncotype DX Breast Recurrence Score (Genomic Health)

Oncotype DX is designed to assess the risk of distant recurrence within 10 years and predict the likelihood of chemotherapy benefit. The test also reports the underlying tumour biology: ER, PR and HER2 status. The test is intended for use in pre- and post-menopausal women with Stage I or II breast cancer that has the following clinical features:

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

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

Oncotype DX is offered as a test service to the NHS. Samples are processed centrally at the Genomic Health Inc. laboratory in the US, which is accredited by the American Association for Laboratory Accreditation and the College of American Pathologists. The test requires a FFPE breast cancer tissue sample from a biopsy or surgical resection, which can be sent as a paraffin embedded block or as 15 unstained charged slides. The test process is based on RT-qPCR. The test gives a Recurrence Score of between 0 and 100, which is used to quantify the 10-year risk of distant recurrence, assuming 5 years of hormonal (endocrine) therapy. Based on current cut-offs for the Oncotype DX test, a score below 18 indicates low-risk of distant recurrence; a score between 18 and 30 indicates intermediate-risk; and a score of 31 or more indicates high-risk. It should be noted that a number of studies, including the ongoing TAILORx study, are testing the use of different cut-offs for Oncotype DX. The recurrence score may also predict the benefit of chemotherapy. The Oncotype DX results are typically reported within 7 to 10 days after the sample is received at the laboratory.

The Oncotype DX RS can be combined with clinical and pathological factors (tumour size, tumour grade and patient age) using the recurrence score-pathology-clinical (RSPC) calculator; however, this calculator has not been validated.

Prosigna (NanoString Technologies)

Prosigna is a CE marked assay designed to assess distant recurrence-free survival at 10 years. The test is intended for use in post-menopausal women with early stage breast cancer that is:

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

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

Prosigna is based on the PAM50 gene signature. It measures the expression levels of 50 genes used to classify patients into one of four breast cancer subtypes. It also measures the expression of 8 housekeeping genes used for signal normalisation, 6 positive controls, and 8 negative controls. Prosigna classifies samples into the following breast cancer subtypes based on their PAM50 gene expression signatures: luminal A, luminal B, HER2-enriched or basal-like. A Risk Of Recurrence (ROR) score, representing the risk of distant recurrence within 10 years (assuming 5 years of hormonal treatment), is then derived from an algorithm based on the results of the PAM50 gene signature plus clinicopathological factors. Four versions of the ROR score exist in the research setting: ROR based on PAM50 subtype information (ROR-S); ROR-S weighted for a proliferation score (ROR-P); ROR-S plus tumour size (ROR-T or ROR-C); and ROR-S plus proliferation score and tumour size (ROR-PT). The proliferation score is based on a subset of the PAM50 genes associated with the proliferation pathway.

The Prosigna test uses ROR-PT and therefore includes the PAM50 breast cancer subtype, tumour size and proliferation score. Nodal status is also used in converting the score into a risk category. The risk of recurrence score is a numerical score on a 0 to 100 scale. Based on this score and the nodal status, samples are classified into risk categories:

- Node-negative: low risk (0 to 40), intermediate risk (41 to 60), or high-risk (61 to 100)
- Node-positive (up to 3 positive nodes): low risk (0 to 15), intermediate-risk (16 to 40), or high risk (41 to 100)

This assessment includes all studies assessing ROR-PT, whether they use the commercial Prosigna test (using the nCounter system) or other methods (such as qRT-PCR). However, studies assessing ROR-S (subtype), ROR-T/ROR-C (subtype and tumour size) or ROR-P (subtype and proliferation score) are excluded. Studies are also excluded if they only assess PAM50 breast cancer subtypes (luminal A, etc.) rather than ROR-PT score.

Immunohistochemistry 4 (IHC4) test

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

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

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

The IHC4+C test involves an algorithm that calculates a risk score for distant recurrence based on the results of the four IHC assays and clinical factors including: age, nodal status, tumour size, and grade. The algorithm has been published and validated[23, 24] and is freely available, and a calculator is available for use on request. A distant recurrence score of less than 10% is categorised as low-risk for distant recurrence at 10 years; a score of 10% or more but less than 20% is categorised as intermediate-risk, and a score of 20% or more is categorised as high-risk for distant recurrence at 10 years. At the Royal Marsden NHS Foundation Trust, the test is processed with an average estimated turnaround time of 1 week; however results may be made available in 2 working days if they are required urgently.

Table 5: Summary of technologies included in the assessment

Test	EndoPredict	MammaPrint	Oncotype DX	Prosigna	IHC4
Manufacturer	Myriad	Agendia	Genomic Health	NanoString	-
Purpose	Recurrence risk	Recurrence risk and chemotherapy benefit	Recurrence risk and chemotherapy benefit	Recurrence risk and intrinsic subtype	Recurrence risk
Description	12 gene assay (8 cancer genes; RT-qPCR) + clinical factors	70 gene array (microarray)	21 gene assay (16 cancer genes; RT-qPCR)	50 gene assay (50 cancer genes; direct mRNA counting) + clinical factors	4 IHC tests (ER, PR, HER2, Ki67). IHC4+C includes IHC4 plus clinical factors
Testing location	Local laboratory or test service (Germany)	Test service (the Netherlands)	Test service (US)	Local laboratory or test service (UK)	Local laboratory
Stage	Early stage	Early stage (stage I or II)	Early stage (stage I or II)	Early stage (stage I to IIIA)	Early stage
Lymph node status	LN0 and LN+ (up to 3 positive)	LN0 or LN+ (up to 3 positive)	LN0 or LN+ (up to 3 positive)	LN0 and LN+	LN0 and LN+ (1 to 3 positive nodes)
Hormone receptor status	ER+	ER+ or ER-	ER+	ER+	ER+
HER2 status	HER2-	HER2- or HER2+	HER2-	HER2-	HER2- or HER 2+
Menopausal status	Pre- and post-menopausal	Pre- and post-menopausal	Pre- and post-menopausal	Post-menopausal	Post-menopausal
Test result	Low-risk, high-risk	Low-risk, high-risk	Low-risk, intermediate risk, high-risk	Low-risk, intermediate-risk, high-risk Intrinsic subtype	Low-risk, intermediate-risk, high-risk
Assumptions	Score assumes 5 years of hormonal treatment	Assumes no therapy	Score assumes 5 years of hormonal treatment	Score assumes 5 years of hormonal treatment	Score assumes 5 years of hormonal treatment
Commercially available in England?	Yes	Yes	Yes	Yes	No
Cost	£1,500	[redacted] (based on conversion from Euros to Pounds Sterling)	£2,580 (Excludes PAS)	£1,650 (kit cost only)	[redacted] (uplifted to 2016 values)

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

Current usage of tumour profiling tests in the NHS

A previous systematic review and economic evaluation (Ward *et al*[18]) published in 2013 considered the clinical effectiveness and cost-effectiveness of tumour profiling tests compared with current prognostic tools in guiding the use of adjuvant chemotherapy in people with early breast cancer in England and Wales. This report informed the NICE decision to approve the use of Oncotype DX as an option for guiding adjuvant chemotherapy decisions for people with ER+, LN0 HER2- early breast cancer assessed to be at intermediate-risk of recurrence of breast cancer after surgery. The use of the other tumour profiling tests in the NHS remains limited (mainly to clinical trial use).

3.4 Description of decision problem

This assessment aims to assess whether tumour profiling tests used for guiding adjuvant chemotherapy decisions for people with early stage breast cancer represent a clinically effective and cost-effective use of NHS resources.

3.4.1 Interventions

The following tumour profiling tests are included, in combination with current decision-making:

- EndoPredict and EPClin
- MammaPrint
- Oncotype DX Breast Recurrence Score (RS) and Oncotype DX Breast Recurrence Score-Pathology-Clinical (RSPC)
- Prosigna (or ROR-PT, which is equivalent)
- IHC4 and IHC4+Clinical (IHC4+C).

3.4.2 Comparators

The comparator for the assessment is standard UK practice for chemotherapy decision-making, which may include any tool, or clinical and pathological features, used to assess risk. Clinicopathological tools used in current practice include PREDICT, NPI and AOL. The use of these tools varies between centres.

3.4.3 Population and important sub-groups

The intended population for the assessment relates to people with ER-positive (and/or PR-positive), HER2-negative, early stage breast cancer (stages I or II) with 0 to 3 positive lymph nodes.

In practice, it was anticipated that many potentially relevant studies would include a broader population. Therefore, all relevant studies of early breast cancer were eligible for inclusion, and the

findings are interpreted with reference to how closely the study population matched the intended population (see Section 4.1).

The following subgroups are considered within this assessment:

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

These tests will only have an impact on the health of patients 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+ LN0 HER2- early breast cancer for whom prognostic variables 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 of 3.4 or less 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+ LN0 HER2- early breast cancer will have an NPI score above 5.4 and therefore those with an NPI above 3.4 can be considered as being at intermediate-risk. Some LN+ may also be considered to be at intermediate-risk.

Current treatment protocols indicate that women with HER2+, ER- early breast cancer or with more than 3 positive nodes are likely to receive chemotherapy in most centres in the England. Whilst the use of tumour profiling 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; this population was therefore excluded from the NICE scope. Currently patients with micrometastases, who are managed clinically as node-negative are excluded from NHS funded testing by Oncotype DX, even if they fall within the intermediate-risk group. Patients with micrometastases are included in the NICE scope.

Patients with ER+ HER2- early breast cancer, who are either LN0 or have 1-3 positive nodes, are therefore considered to be an important population in which to assess these tests, given the current evidence base. The scope therefore focusses on the ER+ HER2- LN0-3 population. Within this

population, an important subgroup consists of patients at clinically intermediate-risk for whom the decision about whether or not to offer chemotherapy is not clear-cut.

3.4.4 Outcomes

The clinical effectiveness review considers the clinical effectiveness of the tests in relation to the following broad categories. These are described further in Section 4.1, which also lists the relevant study designs for each outcome.

- Analytic validity (i.e. the ability of the test to accurately and reliably measure the expression of mRNA or proteins by breast cancer tumour cells). Due to time constraints, it was not possible to conduct a full review of analytic validity for all tests. A rapid review of IHC4 will follow as an addendum to this report.
- Prognostic ability (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)
- Prediction of chemotherapy benefit (i.e. whether the effect of chemotherapy versus no chemotherapy on patient outcomes differs between test risk groups)
- Clinical utility (this is defined differently throughout the prognostic literature; here, we define clinical utility studies as those that assess the ability of the test to affect patient outcomes (e.g. recurrence and survival) though the prospective use of the test to guide treatment decisions.
- Decision impact (i.e. how the tests influence decision making in terms of which patients will be offered chemotherapy; this design does not include follow-up of clinical outcomes such as recurrence or survival). The review included only UK and European studies since chemotherapy rates differ widely between European and non-European countries
- Health-related quality of life (HRQoL) and anxiety
- Time to test results.

Assessment of the above outcomes involves making comparisons (between study groups or between risk groups for the test) in terms of clinical patient outcomes such as recurrence and survival. Key clinical outcomes included for this purpose are listed in Section 4.1.

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 the incremental cost per quality-adjusted life year (QALY) gained.

3.5 Aims and objectives of the assessment

The overall aim of the assessment is to address the question "*Do tumour profiling tests used for guiding adjuvant chemotherapy decision in patients with early stage breast cancer represent a clinically effective and cost-effective use of NHS resources?"* This includes an update of the systematic review and cost-effectiveness analysis undertaken to inform NICE DG10.[21]

The objectives of the assessment are as follows:

- To conduct a systematic review of the published evidence on the effectiveness and cost-effectiveness of the five tumour profiling tests.
- To develop a health economic model to assess the cost-effectiveness associated with the use of tumour profiling tests compared with current prognostic tools to guide the use of adjuvant chemotherapy in early breast cancer from the perspective of the NHS and Personal Social Services (PSS).

4. CLINICAL EVIDENCE

A systematic review was undertaken to assess the effectiveness of tumour profiling tests for guiding adjuvant chemotherapy decisions in early-stage breast cancer. Section 4.1 presents the methods of the systematic review. Results of the review are reported in Section 4.2.

4.1 Methods

A registered protocol of this systematic review (CRD42017059561) is available on the PROSPERO website at https://www.crd.york.ac.uk/prospero/display_record.asp?ID=CRD42017059561. The review was conducted following the general principles recommended in CRD's guidance,[25] the PRISMA statement,[26] the NICE Diagnostic Assessment Programme manual[27] and the Cochrane Prognosis Methods Group.[28]

The protocol included a mapping stage, following which minor amendments were made to the inclusion criteria and review methods in consultation with NICE and clinical advisors, in order to focus the evidence review to studies of the highest quality and relevance to the decision problem.

4.1.1. Identification of studies

This systematic review search was an update of a previous systematic review (Ward *et al*., 2013[18]) conducted for NICE Diagnostics Guidance 10 (DG10).[21] The search strategy was adapted to retrieve clinical studies and systematic reviews of five tumour profiling tests: EndoPredict, Oncotype DX, MammaPrint, IHC4 and Prosigna in early breast cancer management.

The search approach involved the following:

- Searching of electronic databases
- Contact with experts in the field
- Scrutiny of bibliographies of retrieved papers
- Identification of relevant studies from the previous review by Ward *et al*., 2013[18] conducted for NICE Diagnostics Guidance 10 (DG10)[21] (see below)
- References included within the evidence dossiers provided by the manufacturers to NICE.

a) Electronic database searches

The search strategy comprised Medical Subject Headings (MeSH) or Emtree Thesauri terms and free-text synonyms for 'breast cancer' combined with the individually named tumour profiling tests. Searches were translated across databases and were not limited by language. Searches for Oncotype DX, MammaPrint, IHC, and Prosigna were limited by publication date from 2011 (the search date in

Ward *et al*., 2013,[18] since these tests were included in this review) whereas no date limits were applied to EndoPredict (as it was not included in the review by Ward *et al*., 2013[18]).

The search strategies are presented in Appendix 1. Literature searching was undertaken in February 2017 in the following electronic databases and trials registries:

- MEDLINE Epub Ahead of Print, In-Process & Other Non-Indexed Citations: Ovid, 1946 to Present
- EMBASE: Ovid, 1974 to 2017 February 24
- Cochrane Database of Systematic Reviews (CDSR): Wiley Interscience, 1996 to present
- Database of Abstracts of Reviews of Effects (DARE): Wiley Interscience, 1995 to 2015 (until close of database)
- Cochrane Central Register of Controlled Trials (CCRCT): Wiley Interscience, 1995 to present
- Health Technology Assessment Database (HTA): Wiley Interscience, 1995 to 2016 (until close of database)
- NHS Economic Evaluation Database (NHS EED): Wiley Interscience, 1995 to 2015 (until close of database)
- Science Citation Index Expanded (SCI-E): Web of Science, 1900 to present
- Conference Proceedings Citation Index – Science (CPCI): Web of Science, 1990 to present
- WHO International Clinical Trials Registry Platform (http://apps.who.int/trialsearch/) [Accessed online 19th January 2017]
- American Society of Clinical Oncology (ASCO): Web of Science
- European Society for Medical Oncology (ESMO): Web of Science

b) Supplementary searches

References of relevant systematic reviews, primary studies and company submissions were checked to identify additional studies.

4.1.2 Inclusion and exclusion criteria

The inclusion and exclusion criteria for selecting studies were as follows:

a) Population and setting

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

In practice, it was anticipated that many potentially relevant studies would include a broader population. Therefore, all relevant studies of early breast cancer were eligible for inclusion. Where

subgroups were reported for the intended population above, these were used in the assessment. Where no subgroups were reported, the study was included and the findings were interpreted with reference to how closely the study population matched the intended population.

The following subgroups were considered within this assessment:

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

This assessment focusses on the use of tumour profiling tests to guide decisions about adjuvant chemotherapy. Use of these tests to guide endocrine therapy decisions, or decisions about neoadjuvant chemotherapy (to shrink the tumour before surgery), were not evaluated.

b) Interventions

The following tumour profiling tests were included:

- EndoPredict and EPClin
- MammaPrint
- Oncotype DX Breast Recurrence Score (RS) and Oncotype DX Breast Recurrence Score-Pathology-Clinical (RSPC)
- Prosigna (or ROR-PT, which is equivalent)
- IHC4 and IHC4+Clinical (IHC4+C).

Commercial versus in silico tests

Studies were included if they assessed the commercially available versions of the tests. For IHC4, as there is no commercially available version of the test, any methodology was included. In addition, some studies used *in silico* (electronic database) versions of tests using publicly-available genetic databases, usually based on whole-genome-expression microarray data. Due to uncertainty about their similarity to the commercially available tests, these studies were included in a separate section of the clinical review. It was beyond the scope of the review to ascertain the quality of the methods used or the degree of overlap between cohorts for these *in silico* studies.

c) Comparators

The comparator for the assessment is standard UK practice for chemotherapy decision-making. This was taken to include: combinations of clinicopathological factors (for example within multivariable models), plus clinicopathological risk tools used in the UK, including PREDICT, the NPI and AOL. The Clinical Treatment Score (CTS), a combination of commonly-used clinicopathological variables, was also included as a comparator even though it is not commonly used in practice as a tool, since it is used in a number of key studies and includes a set of variables which are used in practice.

Other non-UK local or national guidelines such as St Gallen and the National Comprehensive Cancer Network (NCCN) guidelines were excluded where a study also reported comparisons to PREDICT, NPI or AOL, but were included otherwise.

Relevant comparators within individual studies varied according to the study type as follows:

- Studies assessing prognostic performance: no comparator is needed as the aim is to compare outcomes between risks groups for the test being studied
- Studies assessing prediction of chemotherapy benefit: no comparator is needed as the aim is to compare effect of chemotherapy between risks groups for the test being studied.
- Clinical utility studies: relevant comparator is standard clinical practice as above
- Decision impact studies: relevant comparator is standard clinical practice as above (for pre-test chemotherapy decisions).

d) Outcomes and study designs

The clinical effectiveness review considers the clinical effectiveness of the tests in relation to the following broad categories:

- **Analytic validity**, i.e. the ability of the test to accurately and reliably measure the expression of mRNA or proteins by breast cancer tumour cells. Due to time constraints, it was not possible to conduct a full review of analytic validity for all tests. A rapid review of IHC4 will follow as an addendum to this report..
- **Prognostic performance**, i.e. the degree to which the test can accurately predict the risk of an outcome such as disease recurrence and discriminate patients with different outcomes. This is usually assessed by conducting the test on stored tumour samples for which longer-term patient outcome data are available, but where the test did not influence treatment. Study designs include:
    - Reanalysis of randomised controlled trial (RCT) data
    - Analysis of prospective or retrospective cohorts where the test did not influence treatment.

- **Prediction of chemotherapy benefit**, i.e. whether the effect of chemotherapy versus no chemotherapy on patient outcomes differs between test risk groups. This is usually assessed by conducting the test on stored tumour samples for which longer-term outcome data are available. Study designs include:
    - Reanalysis of RCTs in which some patients received chemotherapy and some did not
    - Analysis of prospective or retrospective cohortsin which some patients received chemotherapy and some did not. These could include cohorts where the test did or did not influence practice (the implications for this are discussed within the results).
- **Clinical utility**: This is defined differently throughout the prognostic literature. Here, we define clinical utility studies as those that assess the ability of the test to affect patient outcomes (such as recurrence and survival) though the prospective use of the test to guide treatment decisions (the study may be prospective or retrospective, but use of the test should have been prospective i.e. used in clinical practice rather than conducted on stored tumour samples). Study designs include:
    - RCTs randomising patients to chemotherapy guided by the test or guided by a comparator (e.g. clinical practice).
    - Observational studies reporting clinical outcomes for patients whose treatment was guided by the test. As these studies do not have a comparator, we are primarily interested in outcomes for patients with low risk disease, who as a group have mostly avoided chemotherapy. The observation of good outcomes in this group could, alongside other evidence, support the avoidance of chemotherapy in this group.
- **Decision impact** (i.e. how the tests influence decision making in terms of which patients will be offered chemotherapy). Clincial advice to the EAG suggests chemotherapy rates differ between countries, with lower rates in the UK and Europe compared to the USA. The review therefore included only UK and European studies. Study designs include:
    - Studies assessing change in chemotherapy recommendations and/or decisions before and after use of the test (this design does not include follow-up of clinical outcomes such as recurrence or survival).
- **Health-related quality of life and anxiety**. Study designs include:
    - Studies assessing impact of the test versus usual practice on HRQoL and anxiety
    - Studies assessing HRQoL and anxiety before and after test use.
- **Time to test results**. Studies assessing the time taken to obtain test results.
- **Concordance**: Concordance is defined in this review as the degree to which tests assign the same patients to the same risk groups. They do not report long-term outcomes. A full systematic review of studies which only assess concordance between tests (with no patient

outcome data) was beyond the scope of this assessment. However, the OPTIMA Prelim study was included as a key example of concordance between tests.

Clinical patient outcomes:

Assessment of clinical utility, prognostic ability and prediction of chemotherapy benefit involves making comparisons (between study groups or between risk groups for the test) in terms of clinical patient outcomes. Key clinical outcomes included for this purpose are listed below. Standard definitions for breast cancer outcomes, defined by Hudis *et al*. (2007),[29] are given below, though these are not always consistently or clearly defined in study reports. Within this review, distant recurrence-free survival (DRFS) and distant recurrence-free interval (DRFI) have been combined in some sections where insufficient detail was provided in study reports to distinguish between them.

- distant recurrence/relapse-free survival (DRFS), also referred to as distant metastasis-free survival (DMFS) or distant disease-free survival (DDFS) – events include distant recurrence and death from any cause
- distant recurrence-free interval (DRFI), also referred to as distant metastasis-free interval (DMFI) – events include distant recurrence and death from breast cancer
- recurrence/relapse-free survival (RFS) – events include ipsilateral, locoregional or distant invasive recurrence and death from any cause (not contralateral disease, non-breast cancers, or ductal carcinoma *in situ*, DCIS)
- recurrence/relapse-free interval (RFI) - events include ipsilateral, locoregional or distant recurrence and death from breast cancer (not contralateral disease, non-breast cancers, or ductal carcinoma *in situ*, DCIS)
- invasive disease-free survival (IDFS) – events include ipsilateral, locoregional or distant invasive recurrence, contralateral and non-breast cancers, and death from any cause (not DCIS)
- disease-free survival (DFS) – events include ipsilateral, locoregional or distant recurrence, DCIS, contralateral or non-breast cancers, and death from any cause
- breast cancer-specific survival (BCSS) – events include breast cancer death only
- overall survival (OS) – events include death from any cause only
- disease-related morbidity and mortality
- chemotherapy-related morbidity and mortality.

For the above clinical outcomes, studies were only included if follow-up was at least 5 years for OS and BCSS, or at least 3 years for other outcomes.

The following outcomes were excluded:

- Locoregional recurrence (i.e. within the region of the original tumour); since chemotherapy decisions will mainly impact distant recurrence and survival
- Clinician confidence and patient decisional conflict relating to decision impact of the test (this is beyond the scope of this assessment)
- Prediction of benefit from one type of chemotherapy versus another (the assessment is restricted to benefit of chemotherapy versus no chemotherapy).

Studies not published in English language were included if sufficient PICOS data could be extracted from non-English language full-texts, or from an existing English language abstract. Non-peer-reviewed reports or abstracts were only included if the data were presented in a succinct and accessible manner (e.g. a manuscript prepared for submission to a journal), if sufficient methodological details were reported to allow critical appraisal of the study quality, and if results were reported in sufficient detail.

4.1.3 Study selection process

All records retrieved from the search were exported into a reference management database (EndNote, version X7). After de-duplication, titles/abstracts were assessed for relevance, followed by examination of full texts of potentially includable studies. Study selection was conducted by one reviewer, with discussion between two reviewers for any studies giving rise to uncertainty. A 10% sample was checked by a second reviewer early in the process to ensure, and correct if necessary, mutual understanding of study inclusion.

4.1.4 Data extraction

A data extraction form was constructed in Excel and piloted using two examples of each study design. Data were extracted by one reviewer and checked by a second reviewer. Disagreements were resolved by discussion. Study authors were contacted for any missing data. Where multiple publications related to the same patient cohort, or where pooled analyses were identified, the references selected for inclusion were those which provided the most complete follow-up and the most useful clinical outcomes (DRFS or DRFI were preferred based on clinical advice and use in the health economic model, see Section 5), avoiding double-counting of patients/cohorts where possible. Systematic reviews relevant to the assessment were used to check for additional studies.

4.1.5 Quality assessment

The methodological quality of included studies was assessed using quality assessment tools relevant to the study design. Quality assessment was undertaken by one reviewer and checked by a second reviewer. The quality and design of studies was considered within the narrative synthesis of results.

For clinical utility studies (for which the highest level of evidence is an RCT of the test versus usual practice), quality was assessed using the Cochrane Risk of Bias tool for RCTs.[30]

For studies assessing prognostic ability and prediction of chemotherapy benefit, quality was assessed using relevant criteria selected from the draft Prediction model study Risk Of Bias Assessment Tool (PROBAST) (personal communication, January 2017, Dr Robert Wolff). The PROBAST tool has been developed specifically for use in systematic reviews of prediction models by the Cochrane Prognosis Methods Group[28] but is not yet validated or published. Criteria were selected on the basis of relevance to this review. Table 6 shows the quality criteria used in this assessment and how they were scored.

Table 6: PROBAST quality criteria and scoring

N	Criterion	Scoring
Risk of bias questions		
1	Study design appropriate?	Yes (prognosis): reanalysis of RCT or cohort or nested case control AND patients did not receive chemotherapy
		Yes (predicting chemo benefit): RCT or reanalysis of RCT
		No (prognostic): non-nested case control or case series AND/OR some/all patients had chemotherapy
		No (predicting chemo benefit): patients not randomised to chemotherapy vs. no chemotherapy
2	All eligible patients included?	Yes: all eligible patients from trial or consecutive eligible patients from prospective registry
		No: some eligible patients excluded e.g. not sent for testing; insufficient tissue; test failures; missing data; AND/OR non-prospective registry
		Unclear: if unclear
4	Blinding of test assessors to clinical outcomes?	Yes: blinded
		No: not blinded
		Unclear: if unclear
6	Outcome definition standardised or defined *a priori*?	Yes: reported outcomes which were standardised (e.g. DRFS, OS) or defined *a priori*
		No: outcomes non-standardised and not defined *a priori*
		Unclear if either item unclear
Applicability questions		
3	Patient spectrum	Yes: all patients in scope (ER+, HER2-, LN0-3)

	matches review question?	Mostly: <20% out of scope
		No: >20% out of scope
		Unclear: if unclear
5	Test as per decision problem?	Yes: same as commercially available tests or IHC4 conducted as per Cuzick 2011[24]
		No: different to commercially available tests (eg FFPE vs fresh samples, test methods)

Studies assessing decision impact, analytic validity and HRQoL/anxiety were not quality-assessed due to time constraints.

4.1.6 Data presentation and synthesis

Data were summarised and presented as tabular and narrative syntheses. Meta-analysis was not considered appropriate due to significant heterogeneity between studies. Interpretation of the evidence base was conducted with reference to published hierarchies for predictive studies,[31-33] and with regard to the ability of the study design to adequately address the decision problem. Interpretation of results also considered how closely the study population matched the intended population, the methodological quality of the studies, and the treatment received by patients (in terms of endocrine therapy and chemotherapy).

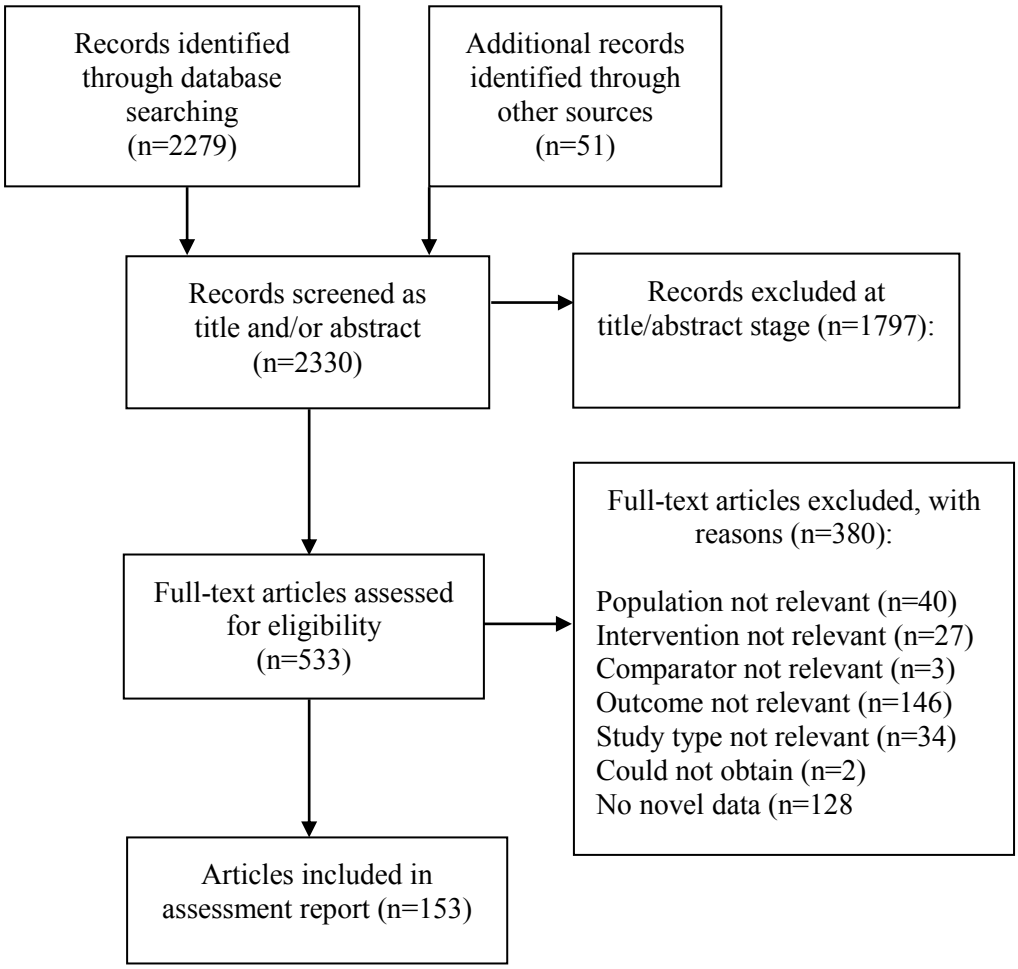
4.2 Results: Overview of main results

Quantity of evidence

The PRISMA flow chart for study selection is shown in Figure 4. The database searches and other sources identified a total of 2330 unique references to screen. Of these, 1797 were excluded at the title/abstract stage and 533 full-text articles were screened, of which 380 were excluded (reasons are listed in Figure 4). Appendix 2 provides a full list of excluded references. In total, 153 references were included in the assessment.

There are numerous TransATAC publications that met the criteria for the review,[24, 34-42] but throughout the report,we present data provided by the TransATAC team as a personal communication to the EAG, which restricts to HR+, HER2-, LN0-3 patients.[43]

Figure 4: PRISMA flow diagram



Overview of results

To orientate the reader to the broad sweep of the evidence and to enable a more informed consideration of the detailed evidence base, we first provide a summary of the results (Section 4.2). The remainder of the review then outlines the evidence base in detail. We have separated the evidence into the following broad categories:

- **Development:** a description of the development of the test. A full review of analytic validity was not possible due to time constraints. A rapid review of IHC4 will follow as an addendum to this report.
- **Prognostic performance:** studies reporting on the ability of the test to predict risk of recurrence and/or survival. The most commonly reported data are Kaplan-Meier estimates of risk of outcome per test risk group and hazard ratios (HRs) between groups, though a small number of studies report C-index data (which in this case is identical to area under the curve (AUC)) and likelihood ratios. In keeping with the majority of studies, we first present unadjusted data, and separately report analyses (usually multivariable Cox proportional

hazard models) which adjust for clinicopathological factors. The C-index is a measure of the goodness of fit of a model with binary outcomes (in this case, it is identical to the AUC). A value <0.5 indicates a poor model, a value of 0.5 indicates the model is no better than chance, a value >0.7 indicates a good model, >0.8 a strong model and a value of 1 indicates a perfect model.[44]

- **Chemotherapy benefit**: Studies in this category compare treatment benefit across risk categories, and most commonly re-analyse RCT data where patients were randomised to chemotherapy or no chemotherapy, and conduct a test for the interaction between treatment and tumour profiling test risk group. The interaction test tells us whether the tumour profiling test is able to predict a differential treatment effect by risk group. We have also included any observational studies which report treatment benefit across risk categories, with or without interaction tests, with appropriate caveats about the possibility of confounding in such studies.
- **Clinical utility**: studies reporting the impact on patient outcomes (such as recurrence and survival) of the prospective use of the test to guide adjuvant chemotherapy treatment decisions. Ideally, such studies would randomise patients to treatment guided by the test or to treatment guided by usual clinical practice. However, given the paucity of RCT evidence, the inherent ethical issues with randomising all patients to chemotherapy and issues with powering such studies, observational studies have also been included in this section.
- **Decision impact:** studies which report the impact of test results on actual chemotherapy decisions or recommendations. Such studies do not report long term follow-up of patients.

There were no data available for clinical utility for Prosigna, EndoPredict or IHC4. Chemotherapy benefit only applies to MammaPrint and Oncotype DX, as these tests claim to be able to identify patients who will benefit from chemotherapy, rather than just those patients who are at high risk of relapse. As such, the clinical review comprises the following main section headings, each with a number of relevant subheadings:

- 4.2 Results: Overview of main results
- 4.3 Results: Oncotype DX
    - 4.3.1 Development: Oncotype DX
    - 4.3.2 Prognostic performance: Oncotype DX
    - 4.3.3 Chemotherapy benefit: Oncotype DX
    - 4.3.4 Clinical utility: Oncotype DX
- 4.4 Results: MammaPrint
    - 4.4.1 Development: MammaPrint
    - 4.4.2 Prognostic performance: MammaPrint

- 4.4.3 Chemotherapy benefit: MammaPrint
- 4.4.4 Clinical utility: MammaPrint

- 4.5 Results: Prosigna
  - 4.5.1 Development: Prosigna
  - 4.5.2 Prognostic performance: Prosigna
- 4.6 Results: EndoPredict and EPClin
  - 4.6.1 Development: EndoPredict and EPClin
  - 4.6.2 Prognostic performance: EndoPredict and EPClin
- 4.7 Results: IHC4 and IHC4+C
  - 4.7.1 Development and Analytic validity: IHC4 and IHC4+C (to follow as Addendum)
  - 4.7.2 Prognostic performance: IHC4 and IHC4+C
- 4.8 Results: All tests compared to each other
  - 4.8.1 Studies reporting more than one test
    - Prognostic performance
  - 4.8.2 Microarray studies
    - Prognostic performance
  - 4.8.3 Concordance
    - OPTIMA Prelim
- 4.9 Results: All tests: Decision impact studies
- 4.10 Results: All tests: Anxiety and health-related quality of life
- 4.11 Results: All tests: Time to test results
- 4.12 Comparison of TransATAC data to other study data

Summary of results

This section provides a summary of results for all included tests, ordered by type of evidence. For the sake of clarity, this section focusses on LN0 and LN+ subgroups only, DRFI/DRFS outcomes and key points. Full descriptions and discussions of the evidence base are reported in Sections 4.3 to 4.10 and should be read in conjunction with this summary to obtain a full understanding. The derivation cohorts are excluded from the summary (i.e. three US cohorts for Oncotype DX,[45] TransATAC[43] for IHC4 and IHC4+C; TransATAC and NSABP B-14 pooled[42] for RSPC; van 't Veer *et al*., 2002[46] for MammaPrint; Van de Vijver *et al.* 2002[47] for Prosigna, Filipits *et al*. 2011[48] for EndoPredict), except in the case of IHC4+C, as only the derivation data reported numerical values.

Risk classification

LN0: Among studies of LN0 patients receiving endocrine monotherapy, percentages categorised as high-risk ranged from 9-33% across all five tests (Table 7): [redacted]-33% for Oncotype DX (4 studies[43, 45, 49-52]); 29% for MammaPrint (1 study[53]); 15-20% for Prosigna/ROR-PT (3 studies[43, 54-56]); [redacted] for EPClin (2 studies[43, 57-59]); [redacted] for IHC4+C (derivation cohort);[43] and not reported for IHC4. Within studies with variable endocrine and chemotherapy use, percentages categorised as high-risk were similar to the above for Oncotype DX (25-28%), but higher for MammaPrint (33 to 73%); this may reflect selection of higher-risk patients for MammaPrint studies (some not ER+, some required chemotherapy).

LN+: Three tests (Prosigna/ROR-PT, EPClin and IHC4+C) categorised far more LN+ than LN0 patients as high-risk among studies of endocrine monotherapy (Table 8): 48-62% for Prosigna/ROR-PT (3 studies[43, 54-56]); [redacted] for EPClin (2 studies[43, 57-59]); and [redacted] for IHC4+C (1 study[43]). Conversely, Oncotype DX categorised similar percentages of LN+ and LN0 patients as high-risk ([redacted] for LN+; 1 study[43]). For MammaPrint, there were no LN+ endocrine monotherapy studies, but in studies with variable endocrine and chemotherapy use, 59-62% were high-risk (2 studies[60, 61]); similar to LN0.

Prognostic performance and additional prognostic value

Oncotype DX: Seven reanalyses of RCTs and four retrospective cohort studies were included (total N=5,156). Generally, similar numbers were high-risk for LN0 and LN+ cohorts, but more were low-risk in LN0, and more intermediate-risk in LN+. Therefore, how many patients would be prescribed chemotherapy would depend on how intermediate patients are handled. 10-year DRFI rates for LN0 low-risk patients were 93%-97% (with endocrine monotherapy), and for intermediate-risk somewhat higher (86%-100%). LN+ patients were generally at higher risk of recurrence than LN0 in both low and intermediate categories (10-year DRFI for LN+ <85% for low and <75% for intermediate). Unadjusted analyses indicated Oncotype DX was prognostic (statistically significant differences between low-risk and high-risk groups) across various recurrence outcomes regardless of lymph node status, though HRs between intermediate-risk and high- or low-risk groups were not always statistically significant. Oncotype DX provided additional prognostic information over most commonly used clinicopathological variables (age, grade, size, nodal status) regardless of lymph node status, and over CTS and NPI in LN0 (but not LN+) patients.

Oncotype DX RSPC: One study[42] derived the RSPC score in a meta-analysis of two RCT datasets (LN0/+; N=1,735), and validated it in another (LN0; N=625). Based on the derivation cohort, the Oncotype DX RSPC algorithm (Oncotype DX plus age, tumour size and grade) appeared to provide

additional prognostic information over Oncotype DX and over clinicopathological variables, and was able to classify more patients into a low-risk category than Oncotype DX whilst maintaining a roughly equivalent rate of distant recurrence in the low-risk group. In the validation cohort, RSPC had prognostic value in univariate analyses (no adjusted analysis was reported. However, RSPC has only been validated in one independent cohort and has not been tested in pre-menopausal or LN+ patients.

MammaPrint: The prognostic value of MammaPrint is based on nine retrospective analyses (total N=1,805), four pooled analysis (N=964; including six of nine series above) and one reanalysis of an RCT (N=538). Studies were variable in terms of nodal status, ER status, and receipt of endocrine and chemotherapy. MammaPrint was statistically significantly prognostic for 10-year DRFS in almost all unadjusted analyses of LN0 and LN+ patients. For LN0 patients, 10-year DRFS/DRFI rates for low-risk patients ranged from 80% to 90% (with varying rates of endocrine and chemotherapy use), while the reanalysis of an RCT reported 10-year DRFS of 93% with endocrine monotherapy and 83% without endocrine or chemotherapy. For LN+ patients, 10-year DRFS rates in low-risk patients ranged from 79% to 91% (with varying rates of endocrine and chemotherapy use). In terms of additional prognostic value, MammaPrint was statistically significantly prognostic for 10-year DRFS/DRFI in multivariable analyses adjusted for clinicopathological risk tools (AOL and NPI) and various combinations of clinicopathological variables in LN0/+ and LN0 cohorts, while adjusted analyses in LN+ cohorts were statistically significant or borderline significant.

Prosigna/ROR-PT: Based on six reanalyses of RCTs and two retrospective analyses of prospective cohorts (total N=9,118), Prosigna/ROR-PT was statistically significantly prognostic for unadjusted analyses of 10-year DRFS/DRFI in LN0 and LN+ patients. The 10-year DRFS/DRFI rates for low-risk patients were 95% to [redacted] in three studies of LN0 patients (endocrine monotherapy), and in LN+ patients these were [redacted] in two studies (endocrine monotherapy) and 92% in one study (all endocrine and chemotherapy). For intermediate-risk, 10-year DRFS/DRFI rates were 87% to 93% for LN0 and [redacted] to 94% for LN+ (endocrine monotherapy). Prosigna/ROR-PT added prognostic information over clinicopathological variables or CTS/CLP/NPI in three studies; this was statistically significant in LN0 patients and either significant or borderline significant in LN+ patients. Oncotype-DX provided additional prognostic information over most commonly used clinicopathological variables (age, grade, size, nodal status) regardless of lymph node status, and over CTS and NPI in LN0 (but not LN+) patients.

Oncotype DX RSPC: One study (Tang *et al*. 2011b)[42] derived the RSPC score in a meta-analysis of NSABP B-14 and TransATAC (LN+/-; N=1735), and validated it in NSABP B-20 (LN0; N=625). Based on the derivation analysis set, the Oncotype DX RSPC algorithm (Oncotype DX plus age, tumour size and grade) appeared to provide additional prognostic information over Oncotype DX and

over clinicopathological variables, and was able to classify more patients into a low-risk category than Oncotype DX whilst maintaining a roughly equivalent rate of distant recurrence in the low-risk group. In the validation cohort, RSPC had prognostic value in a univariate analysis. However, RSPC has only been validated in one independent cohort (unadjusted analysis) and it has not been tested in pre-menopausal or LN+ patients.

EndoPredict and EPClin: Based on three reanalyses of RCTs (total N=3,135) in ER+ HER2- endocrine-treated patients, EPClin was statistically significantly prognostic for unadjusted analyses of 10-year DRFS/DRFI in LN0 and LN+ patients. The 10-year DRFS/DRFI rates for low-risk patients were approximately [redacted] in LN0 and LN+ patients receiving endocrine therapy alone. [redacted], while in two further studies, the EP score added statistically significant information over clinicopathological variables in mixed LN0/LN+ and LN+ patients (no data for EPClin).

IHC4: The IHC4 score has been validated in five reanalyses of RCTs and six retrospective cohort studies (total N=13,434), and provides statistically significant prognostic information consistently in unadjusted analyses in LN+/-, LN0 and LN+ groups. However, most studies used quartiles or tertiles to define risk groups, which are specific to each cohort. Also, many used laboratory methods that differed from the derivation study methodology. Only one validation study[62] used the cut-offs from the original analysis,[24] and provides a second and third validation cohort (BCS and TEAM), but only for the IHC4 component of the test, not including the clinical factors (i.e. not IHC4+C). IHC4 had additional prognostic value over clinicopathological factors in some studies. Test methodologies did not appear to impact on the statistical significance of results, but concerns remain about the conduct of the test in laboratories other than that used to derive the score.

IHC4+C: IHC4+C had prognostic value in unadjusted analyses in the derivation cohort. Additional prognostic value has been reported in the derivation cohort where IHC4+C provided statistically significant information over NPI and CTS in LN0 but not LN+ patients, and in one validation cohort (Nottingham) where statistical significance was maintained after adjustments for CP factors.

Microarray studies: Microarray studies are defined here as those that applied a test algorithm to either *in silico* data (microarray gene expression data held on a database) or to a *de novo* microarray assessment. These studies support conclusions from studies using the commercial versions of the assays in suggesting that Oncotype DX, MammaPrint and EndoPredict can discriminate between high and low-risk patients regardless of LN status.

Outcomes in low-risk and intermediate-risk groups

LN0: Among studies of LN0 patients receiving endocrine monotherapy, the 10-year DRFS/DRFI rates in low-risk groups were similar across all five tests (Table 7): 93% to 97% for Oncotype DX (4 studies[43, 45, 49-52]); 93% for MammaPrint (1 study[53]); 95% to 97% for Prosigna/ROR-PT (3 studies[43, 54-56]); [redacted] to [redacted]% for EPClin (2 studies[43, 57-59]); and [redacted] for IHC4+C (1 study[43]). There were no studies of MammaPrint in this population. Intermediate-risk groups for Oncotype DX, Prosigna/ROR-PT and IHC4+C had worse DRFS/DRFI rates compared to low-risk groups (EPClin and MammaPrint do not have an intermediate-risk group). Many studies of MammaPrint included some ER- patients, did not treat all patients with endocrine therapy, and treated some with chemotherapy; for these studies, 10-year DRFS/DRFI rates in low-risk groups were 80% to 90% (7 studies[47, 53, 63-66]).

LN+: Among studies of LN+ patients receiving endocrine monotherapy (Table 8), 10-year DRFS/DRFI rates in low-risk groups were less favourable for Oncotype DX ([redacted]; 1 study[43]) than for Prosigna/ROR-PT ([redacted] to 100%; 2 studies[43, 54, 55]), EPClin ([redacted]% to 95%; 2 studies[43, 57-59]) or IHC4+C ([redacted]; 1 study[43]). There were no studies of MammaPrint in this population. Intermediate-risk patients had lower DRFS/DRFI than low-risk patients for Oncotype DX [redacted]; Prosigna/ROR-PT [redacted]% to 94% (2 studies[43, 54, 55]) and IHC4+C [redacted]). For MammaPrint, the only LN+ data were in populations which included some ER- patients, did not treat all patients with endocrine therapy, and treated some with chemotherapy; 10-year DRFS/DRFI rates in low-risk groups were 79% to 91% (2 studies[60, 67]).

Chemotherapy benefit

Evidence of chemotherapy benefit was only assessed for Oncotype DX, Oncotype RSPC and MammaPrint. There was no chemotherapy benefit evidence for EndoPredict or EPClin, Prosigna/ROR-PT, IHC4 or IHC4+C.

Oncotype DX and Oncotype RSPC: Analyses of the ability of Oncotype DX to predict benefit from chemotherapy were reported in five studies.[49, 50, 68, 69] Two were reanalyses of RCTs (one LN0,[49, 50] one LN+,[49, 50, 68] total N=1,018) where patients were randomised to endocrine monotherapy, or endocrine therapy plus chemotherapy. Three were observational studies[69-74] (total N~44,000 with some double counting, two LN0,[69, 70, 73, 74] one LN+/-[71, 72]) where patients were treated according to usual practice and their Oncotype DX score. There is some evidence from the two reanalyses of RCTs to suggest that Oncotype DX may predict benefit from chemotherapy, and that benefit from chemotherapy is highest in Oncotype DX high-risk patients. Unadjusted interaction tests between Oncotype DX risk group and chemotherapy benefit were mainly statistically significant. However, the evidence to support Oncotype DX's ability to predict benefit from chemotherapy is weak, possibly

due to insufficient events, and interaction tests adjusted for clinicopathological variables were often non-significant. Also, the RSPC algorithm (Oncotype DX plus age, tumour size and grade) showed a non-significant interaction test between chemotherapy benefit and RSPC risk group, indicating that the incorporation of clinicopathological factors may reduce prediction of chemotherapy benefit. Three observational cohort studies were at high risk of confounding; one reported a statistically significant interaction test but this was only adjusted for limited clinical factors. If predictive ability were assumed, it is unclear below which exact cut-off patients could avoid chemotherapy (though one study suggests this is RS 20), as chemotherapy benefit is uncertain in the intermediate-risk group. Whilst the ongoing RCT TAILOR-X (Trial Assigning IndividuaLized Options for Treatment) will address the issue of whether low and intermediate patients can avoid chemotherapy, it is unclear to what extent it will address the question of whether the test can predict chemotherapy benefit.

MammaPrint: Prediction of chemotherapy benefit for MammaPrint was reported in a pooled analysis of six non-randomised series (N=541, half LN0, half LN1-3) in which patients were treated according to usual practice. The effect of chemotherapy versus no chemotherapy was statistically significant in the MammaPrint high-risk group but not in the low-risk group in unadjusted analyses for 5-year DRFS and BCSS and in adjusted analyses for 5-year BCSS. However, the interaction test between chemotherapy treatment and risk group (for 5 year BCSS) was non-significant (p=0.45). A further pooled analysis of two of the above series, restricted to LN1-3 patients, also reported a statistically non-significant interaction between chemotherapy treatment and risk group for 10-year BCSS (p=0.95). The evidence for the ability of MammaPrint to predict chemotherapy benefit is therefore extremely limited; although unadjusted analyses suggest a greater effect of chemotherapy in high-risk groups, adjusted analyses were only reported for one outcome, and the non-significant interaction tests suggest no statistically significant difference in effect of chemotherapy between risk groups.

Clinical utility

Clinical utility is defined in this assessment as the impact of tests used prospectively in clinical practice on recurrence/survival outcomes. Studies assessing prospective use of tests were only available for Oncotype DX and MammaPrint. There was no clinical utility evidence for EndoPredict or EPClin, Prosigna/ROR-PT, IHC4 or IHC4+C or Oncotype DX RSPC.

Oncotype DX: Without the highest level of evidence (RCT of treatment guided by test vs. treatment guided by usual practice), it is not possible to conclude whether patient outcomes would be affected by use of the test in a clinical setting. In LN0 patients, use of the test in clinical practice appears to result in low rates of chemotherapy use in low-risk patients (2% to 12%), with acceptable outcomes (5-year DRFS/DRFI/IDFS 96% to 99.6%). Rates of chemotherapy use increased with increasing risk category, and were generally higher in LN+ patients; only one study reported 5-year

DRFS/DRFI/IDFS for LN+ patients, which was 97% (7% received chemotherapy). It was not possible to determine whether patients in intermediate- and high-risk categories had better outcomes than low-risk patients as a result of using Oncotype DX due to the observational nature of the studies.

MammaPrint: Two studies reported evidence relating to the clinical utility of MammaPrint. MINDACT is an RCT of MammaPrint versus clinical practice. This study randomised patients with discordant MammaPrint and mAOL risks to chemotherapy or no chemotherapy. For patients who were high-clinical, low-MammaPrint risk, 5-year DMFS was 95.9% with chemotherapy and 94.4% without chemotherapy, an absolute difference of 1.5%. This raises the possibility of avoiding chemotherapy in these patients. In patients who were low-clinical, high-MammaPrint risk, 5-year DMFS was 95.8% with chemotherapy and 95.0% without chemotherapy, an absolute difference of 0.8%. This could be interpreted as showing that MammaPrint may not be useful in this group as it would increase chemotherapy rates without improving outcomes. However, the comparator was mAOL, and it is unclear whether the same would be true for other clinical risk scores.

RASTER is a prospective observational study in which patients were treated according to MammaPrint plus usual clinical practice (LN0) or [redacted]. The 5-year DRFI for LN0 patients was 97.0% for low-risk patients (15% had chemotherapy) and 91.7% for high-risk patients (81% received chemotherapy). [redacted] The DRFI rates in the MammaPrint low-risk group may be considered sufficiently low for these patients to avoid chemotherapy. MammaPrint provided additional prognostic information over AOL and NPI, but not over PREDICT plus. Estimates of prognostic performance between risk groups are likely to be affected by the differing rates of chemotherapy per group, and the fact that chemotherapy use was influenced by MammaPrint.

Decision impact

Decision impact studies assess how decisions to use or not use chemotherapy change pre- and post-use of the test. Only decision impact studies from the UK and Europe were included, since other countries may have very different rates of chemotherapy use. The percentage of patients with any change in treatment recommendation or decision (either to or from chemotherapy) among UK studies was 29% to 49% across four Oncotype studies, 37% in one EndoPredict study, and 27% in one IHC4+C study. Ranges across European (non-UK) studies were 5% to 70% for Oncotype, 38% to 41% for EndoPredict, 14% to 41% for Prosigna and 13% to 51% for MammaPrint. The net change in the percentage of patients with a chemotherapy recommendation or decision (patients changing to chemotherapy minus those changing to no chemotherapy) among UK studies was a reduction of 8% to 23% across four Oncotype studies, an increase of 1% in one EndoPredict study, and a reduction of between 2-26% in one IHC4+C study (unclear due to category definitions). Net changes across

European (non-UK) studies were a reduction of 0% to 64% for Oncotype, reduction of 13% to 26% for EndoPredict, and reduction of 2% to increase of 9% for Prosigna, and reduction of 31% to increase of 8% for MammaPrint.

Concordance

Concordance is defined in this review as the degree to which tests assign the same patients to the same risk groups. They do not report long-term outcomes. A full review of this data was beyond the scope of this review and instead the OPTIMA Prelim study[75] was included as a key example of concordance between tests. OPTIMA Prelim recruited ER+, HER2-, LN1-9 (or LN) with tumour size >30mm) and reported concordance between Oncotype DX, MammaPrint, Prosigna and IHC4. Kappa statistics ranged from 0.33 (Prosigna to IHC4) and 0.53 (Oncotype DX to IHC4 and Prosigna to MammaPrint), indicating modest agreement between tests. Other analyses showed no two tests were more in agreement than others, and that disagreements spanning two risk categories were not uncommon. The authors concluded that whilst tests assigned similar proportions of patients to low/intermediate and high-risk categories, test results for an individual patient could differ markedly depending on which test was used.

Anxiety and HRQoL

Six studies (7 publications)[76-82] reported outcomes relating to anxiety (including worry and distress) and HRQoL. For Oncotype DX (2 studies, N=286),[77, 79] EndoPredict (1 study, N=149)[76] and Prosigna (2 studies N=398)[83, 84] all studies were pre-test/post-test in design, whilst MammaPrint compared patients sub-grouped according to their clinical risk, MammaPrint risk, whether they were assigned to chemotherapy or not, and whether the MammaPrint test result was missing.[81] Across tests, and where reported, state anxiety decreased post-test, and total FACT-G generally stayed the same. However, without a comparator group it is not possible to tell if anxiety would have reduced post-treatment decision regardless of how the decision was made. Emotional and functional wellbeing in FACT-G improved in one study,[84] and FACT-B improved for some subgroups in one study.[81]

Time to test results

One study of 263 US patients reported that the percentage having a delay of at least 42 days from surgery to chemotherapy initiation was 31% for patients for whom an Oncotype test was ordered versus 20% in other patients.

Table 7: Summary of risk categorisation and prognostic and predictive (of chemotherapy benefit) ability across tests: LN0[a]

Test	N studies with DRFS/I	Population	Nodal status	Endo / chemo	% pts per group			% 10yr DRFS/DRFI risk			Significantly prognostic for DRFS/DRFI?	Additional value over CP factors or tests?[a]	Chemo benefit?
					Low	Int	High	Low	Int	High			
LN0, all ET, no CT													
Oncotype DX	4[43, 45, 49-52]	ER+ HER2+/-	LN0	All ET No CT	48 to [redacted]	20 to [redacted]	[redacted] to 33%	93 to 97%	86 to 100%	61 to 77%	Yes (3 of 4 studies, NR in 1)	Yes (three studies)	Weak[c]
Mamma-Print	1[53]	ER+ HER2 NR	LN0	All ET No CT	71%	-	29	93%		85%	NR	NR	
Prosigna / ROR-PT	3[43, 54-56]	Most ER+ HER2-	LN0	All ET No CT	48 to [redacted]	30 to 32%	15 to 20%	95 to 97%	87 to 93%	69 to 85%	Yes (3 of 3 studies)	Yes vs CTS and NPI (2 studies)	NA
EPClin	2[43, 57-59]	ER+ HER2-	LN0	All ET No CT	[redacted] to [redacted]	[redacted]	[redacted] to [redacted]	[redacted] to [redacted]%		[redacted] to [redacted]%	Yes ([redacted] of 2 studies)	[redacted]	NA
IHC4	2 cohorts[62]	ER+ HER2- NR	LNO	All ET No CT	NR	NR	NR	NR	NR	NR	NR	Yes (2 cohorts)	NA
IHC4+C	1[43] (derivation)	ER+ 95% HER2-	LN0	All ET No CT	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NA
LN0, variable ET/CT													
Oncotype DX	2[49, 85]	ER+ HER2+/-	LN0	75-100% ET 79-100% CT	49-51%	21-26%	25-28%	96%[49]	89%[49]	88%[49]	Yes (1 of 1 study)	NR	NR
Oncotype DX RSPC	1[42]	ER+ HER2- NR	LN0	All ET 64% CT	NR	NR	NR	NR	NR	NR	Yes (1 analysis)	Yes (derivation set)	No (1 study)
Mamma-Print	7[47, 53, 63-66, 86d]	70-100% ER+ HER2 NR	LN0	0-25% ET/CT	27 to 67%	-	33 to 73%	80 to 90%		50 to 71%	Yes (4 of 7 studies, 1 not sig, NR in 2)	Yes (pooled study, 2 cohorts, others NR)	Not stat sig (pooled LN0/+)[c]
IHC4	2[87, 88]	ER/HER2 varies		Some ET/CT	Clinical cut-offs not used			NR	NR	NR	Yes (some analyses non-sig)	NR	NA

CP, clinicopathological; CT, chemotherapy; DRFS/I, distant recurrence-free survival/interval; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor-2; LN, number of positive nodes; NR, not reported. [a]For IHC alone, little data by LN status.

[a] Judged via multivariate analyses adjusted for CP factors, change in likelihood ratios, C-index or D-statistics. [b]Sun et al. 2011[89] (China) omitted as much lower DRFS than other studies.

[c] Judged via p values and non-significant interaction tests after adjustments for clinicopathological factors. [d] Where an outlier, Ishitobi 2010 (Japan) omitted due to unknown generalisability

Table 8: Summary of risk categorisation and prognostic and predictive (of chemotherapy benefit) ability across tests: LN+

Test	N studies with DRFS/I	Population	Nodal status	Endo / chemo	% pts per group			% 10yr DRFS/DRFI risk			Significantly prognostic for DRFS/DRFI?	Additional value over CP factors or tests?[a]	Chemo benefit?
					Low	Int	High	Low	Int	High			
LN+, all ET, no CT													
Oncotype DX	1[43]	ER+ HER2-	LN1-3	All ET No CT	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Weak[c]
Prosigna / ROR-PT	3[43, 54-56]	Most ER+ HER2-	LN1-3 (most)	All ET No CT	4% to 25%	27 to 34%	48 to 62%	[redacted] to 100%	[redacted] to 94%	[redacted] to 78%	Yes or borderline (3 studies)	Yes vs CTS, No vs NPI	NR
EPClin	2[43, 57-59]	ER+ HER2-	LN1-3	All ET No CT	[redacted] to [redacted]%	[redacted]	[redacted] to [redacted]	[redacted] to [redacted]	-	[redacted] to [redacted]%	Yes ([redacted] of 2 studies)	[redacted]	NR
IHC4	2 cohorts[62]	ER+ HER2- NR	LN+	All ET No CT	NR			NR			NR	Mixed (1 yes, 1 no)	NR
IHC4+C[a]	1[43]	ER+ HER2-	LN1-3	All ET No CT	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	NR
LN+, variable ET/CT													
Oncotype DX	3[51, 89-91]	ER+ HER2+/-	LN+	74-100% ET/CT	36 to 39%	30 to 34%	30 to 31%	81%[b]	65%[b]	59%[b]	Yes	Yes	NA
Mamma-Print	2[60, 61]	80% ER+ 84% HER2 or NR	LN1-3 LN>3, 26%	Some ET/CT	38 to 41%	-	59 to 62%	79 to 91%		54 to 76%	Yes (2 of 2 studies)	Borderline (1 study)	Not stat sig (pooled LN0/+)[c]
Prosigna / ROR-PT	1[83, 92]	ER+ HER2-	LN>3, 36%	All ET All CT	19%	56%	26%	92%	74%	66%	Yes (1 study)	NR	NR
EPClin	1[83, 92]	ER+ HER2-	LN>3, 36%	All ET All CT	13%	-	87%	100%		72%	Yes (1 study)	NR	NR
IHC4	2[93, 94]	HR+ HER2-	LN+	ET varies 100% CT	Clinical cut-offs not used			No clinical groups			NR	Mixed (1 yes, 1 no)	NR

CP, clinicopathological; CT, chemotherapy; DRFS/I, distant recurrence-free survival/interval; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor-2; LN, number of positive nodes; NR, not reported. [a]For IHC alone, little data by LN status.

[a] Judged via multivariate analyses adjusted for CP factors, change in likelihood ratios, C-index or D-statistics. [b] Sun et al. 2011[89] (China) omitted as much lower DRFS than other studies.

[c] Judged via p values and non-significant interaction tests after adjustments for clinicopathological factors.

4.3 Results: Oncotype DX

4.3.1 Development: Oncotype DX

Oncotype DX was developed through the selection of 250 candidate genes from the published literature, genomic databases, pathway analyses and microarray-based gene expression profiling studies.[45] Three independent breast cancer cohorts (N=447 patients, NSABP-20; Rush University Medical Centre, Chicago, USA; Providence St Joseph's Hospital, Burbank, USA)[95-97] were then used to identify genes that were highly associated with recurrence in all three cohorts, and for which the assay methods performed consistently, and an algorithm derived to fit the data from the three cohorts, using correlational analysis, concordance measurese of accuracy and bootstrap resampling. Data from NSABP-20 were more highly weighted in the derivation set, as the validation set was to be a trial with similar patient characteristics, NSABP-14.

4.3.2 Prognostic performance: Oncotype DX

Study designs and patients: Oncotype DX prognostic performance

Oncotype DX was validated in eleven distinct data sets reported across twelve publications[35, 45, 49, 51, 52, 68, 85, 89-91, 98-100] and a personal communication with the TransATAC team.[43] Study and patient characteristics are presented in Table 9.

Study design: Seven studies were reanalyses of prospectively collected RCT data[35, 43, 45, 49, 51, 68, 90, 91, 98, 99] using archived tissue samples; one of these adopted a nested case-control design.[98, 99] The remaining four data sets were retrospective studies using routinely collected data and archived samples.[52, 85, 89, 100] Data sets ranged from 93[89] to 1065 patients.[51, 90]

Five RCTs were from the USA:

- NSABP-B-14[45] – the National Surgical Adjuvant Breast and Bowel Project (NSABP) which recruited patients between 1982 and 1988 and randomised them to placebo or tamoxifen. Only the tamoxifen arm is included in this analysis. Patients were LN0.
- NSABP-B20[49] – another NSABP trial which recruited patients between 1988 and 1993 and randomised them to either tamoxifen alone, or tamoxifen plus chemotherapy (cyclophosphamide, methotrexate, and fluorouracil (CMF) or methotrexate and fluorouracil (MF). Two analyses are presented, one of the tamoxifen monotherapy arm, which was also as a training set for Oncotype DX, and one of the tamoxifen plus chemotherapy arm, which was not a training set for Oncotype DX. Patients were LN0.
- SWOG-8814[68] – the Southwest Oncology Group trial 8814, which recruited patients between 1989 and 1995 and randomised them to one of three arms: (1) tamoxifen only; (2)

cyclophosphamide, doxorubicin and fluorouracil (CAF) followed by tamoxifen, or (3) CAF with concurrent tamoxifen. Only the tamoxifen arm was included in Albain 2010.[68] Patients were LN+.

- NSABP-28[51, 90] – a third NSABP trial which recruited patients between 1995 and 1998 and randomised them to one of two chemotherapy arms (doxorubicin plus cyclophosphamide (AC; 4 cycles) or four cycles of AC followed by four cycles of paclitaxel). Patients analysed received both endocrine therapy and chemotherapy. Patients were LN+.
- E2197[98, 99] – an Eastern Cooperative Oncology Group (ECOG) trial, which recruited patients between 1997 and 1999 and randomised them to one of two chemotherapy (doxorubicin or docetaxel) plus tamoxifen arms. Patients analysed received both endocrine therapy and chemotherapy. The analysis reported here is a nested case-control study using the trial data. Patients were a mix of LN0/LN+.

The two remaining RCTs were from the UK[35, 43] and France,[91] respectively:

- TransATAC[35, 43] was an international trial, but only UK samples were included in this analysis. The trial evaluated anastrozole, tamoxifen, or the combination of both treatments. Recruitment ended in 2006. Only the tamoxifen arm is included in this analysis. There are numerous TransATAC publications that report data for Oncotpye-DX, but here we present data from a bespoke data request provided by the TransATAC team to the EAG, which restricts to HR+, HER2-, LN0-3 patients.[43]
- PACS01[91] was a French trial which recruited patients between 1997 and 2000 and randomised them to one of two chemotherapy treatment arms. All patients analysed (ER+, HER2-) received chemotherapy and 74% received endocrine therapy (after a protocol amendment, ER+ patients received endocrine therapy). Patients were LN+.

There were four retrospective studies.[52, 85, 89, 100] Importantly, archival tissue samples were analysed and as such patients were not treated according to Oncotype DX scores. Studies in which patients were treated according to test results may be confounded, and are therefore excluded from analysis of prognostic performance, but included in the analysis of clinical utility in Section 4.3.4. One retrospective study[100] was from the USA, while three[52, 85, 89] were from China or Japan:

- Russell *et al*. 2016[100] recruited patients from two hospitals in the USA (University of South Florida and Morton Plan Hospital. The lymph node status, HER2 status and treatments received were not reported.
- Gong *et al*. 2016[85] (China) recruited patients from Sun Yat-sen Memorial Hospital and the Third Hospital of Nanchang City. Three separate cohorts were recruited, but Oncotype DX data were only reported for one cohort, which recruited post-menopausal LN0 patients. A

second cohort reported IHC4 data (see Section 1.2.5.2). Patients received varying levels of endocrine and chemotherapy according to local practice.

- Sun *et al.* 2011[89] (China) recruited patients from the Hospital Affiliated Academy of Military Medical Science, Beijing. Patients were a mix of LN0 and LN+, with over 18% having more than three positive nodes. Patients received varying levels of endocrine and chemotherapy according to local practice.
- Toi *et al.* 2010[52] (Japan) recruited patients diagnosed between 1992 and 1998 who were treated with tamoxifen, but it is unclear whether any were also treated with chemotherapy. Patients were LN0.

Clinical advice received by the EAG suggests that the three studies from China or Japan[52, 85, 89] may be less generalisable to the English context because (a) patients were treated according to usual clinical practice and this may differ in these countries compared to the UK enough to affect prognostic outcomes, and (b) it is possible that people of different ethnicities have different underlying risk profiles and disease natural history. For this reason, data from these studies should be interpreted with caution and with reference to data from studies where the ethnic profile and clinical practice is similar to the UK.

Lymph node status: Amongst the RCT reanalysis studies, TransATAC[35, 43] and E2197[98, 99] recruited patients regardless of lymph node status (E2197 specifically recruited patients with LN1-3 or LN0 with tumour ≥1.1cm); NSABP B-14[45, 51] and NSABP B-20[49] recruited LN0 patients; and SWOG-8814,[68] NSABP B-28[51, 90] and PACS01[91] recruited LN+ patients. Amongst the retrospective studies, two studies recruited LN0 patients[52, 85] and one[89] recruited patients with any LN status, with patients with LN>3 making up 18% of the cohort. Lymph node status was not reported by Russell *et al*. 2016.[100]

Hormone receptor status: All studies recruited either ER+ or HR+ patients.

Menopausal status: Across the eleven data sets, TransATAC and SWOG-8814 recruited only postmenopausal patients.[35, 43, 68] The remainder either did not report the proportion of patients who were post-menopausal,[45, 49, 51, 52, 90, 91, 98, 99] or recruited regardless of menopausal status.[85, 89]

HER2 status: Only TransATAC[35, 43] and Gong *et al*. 2016[85] recruited or reported a subgroup of exclusively HER2- patients. Six studies[45, 49, 51, 52, 90, 91, 100] did not report HER2 status, probably because patients were recruited before this information was routinely collected.

Treatments: Oncotype DX was derived to predict prognosis in patients with HR+ disease who have been treated with endocrine therapy for 5 years. Treatment with chemotherapy, especially if the effect of chemotherapy is differential across risk categories, could potentially reduce the apparent prognostic performance of the test as it could affect event rates. As such, validation cohorts should treat patients with endocrine monotherapy, but not chemotherapy. TransATAC,[35, 43] SWOG-8814 (subgroup 1),[68] NSABP B-14[45, 51] and NSABP B-20[49] (subgroup 1) all treated patients with endocrine monotherapy, whilst E2197[98, 99], SWOG-8814 (subgroup 2, not included here),[68] NSABP B-20[49] (subgroup 2) and NSABP B-28[51, 90] treated all patients with endocrine and chemotherapy. PACS01[91] treated all patients with chemotherapy and 74% with endocrine therapy. Gong *et al*. 2016[85] treated all patients with endocrine therapy and 79% with chemotherapy, whilst Sun *et al* 2011[89] treated 75% with endocrine therapy and 81% with chemotherapy and Toi *et al*. 2010[52] treated all with endocrine therapy but did not report whether patients were also given chemotherapy. Russell *et al*. 2016[100] did not report the proportion of patients receiving chemotherapy or endocrine therapy.

Tests and comparators: Oncotype DX prognostic performance

Two studies did not report how the test was conducted (PACS01 study; Russell *et al*. 2016).[91, 100] In all but three other cases the test was performed on fixed, paraffin embedded tissue by Genomic Health using the commercial Oncotype DX assay. The three exceptions were the two studies from China where the test was not performed by Genomic Health,[85, 89] and Paik *et al.* 2004, as Paik *et al.* 2006 described the assay used in Paik *et al.* 2004 as being "*a preliminary version of the RT-PCR assay (lacking standardized reagents, calibrators, and controls)*". In these three studies, the equivalence of the tests to the commercially offered Oncotype DX assay is unknown.

An analysis of NSABP B-14 included comparison to a "clinical integrator" based on AOL, where the integrator was adjusted to 5-year outcomes rather than the 10 year outcomes used in AOL. The bespoke TransATAC data request provided to the EAG included a comparison of Oncotype DX to three of the tests (IHC4, Prosigna and EndoPredict) and this is presented in Section 4.8.1.

Quality assessment: Oncotype DX prognostic performance

Quality assessment is summarised in Table 10. All studies were validation studies, though a small number of patients included in NSABP B-20 were used in the derivation series for Oncotype DX. Only three studies[35, 43, 45, 68] used an appropriate study design, as eight[49, 51, 85, 89-91, 98-100] included patients who had been treated with chemotherapy or did not report the proportion treated. No studies included all eligible patients and only three[35, 43, 68, 98] stated that they blinded test assessors to patient outcomes. There are concerns about patient spectrum bias in all studies, mainly due to the retrospective nature of the studies and the exclusion of tumour samples with insufficient tissue probably leading to the loss of patients with smaller tumours.

Results: Oncotype DX prognostic performance

Distribution of patients across risk categories

Distributions of patients across risk categories are presented in Table 11 to Table 16. In LN0 cohorts, the proportion of patients ranged from 48%[52] to [redacted] in the low-risk category, from 20%[49, 52] to [redacted] in the intermediate-risk category and [redacted] to 33%[52] in the high-risk category. It is interesting to note that the distribution of patients in the TransATAC analysis [redacted], and that the distribution in the Japanese cohort (Toi *et al*. 2010)[52] indicates more high-risk and fewer low-risk patients that the other LN0 cohorts.

In LN+ cohorts, the proportion of patients ranged from 36%[51, 90] to [redacted] in low-risk patients, from 30%[91] to 34%[51, 90] in intermediate-risk patients and from [redacted] to 32%[68] in high-risk patients. The proportion of low-risk patients was generally lower in LN+ than LN0 cohorts, and the proportion of intermediate- and high-risk patients was generally higher.

Prognostic performance: unadjusted analyses

This section reports unadjusted analyses. Adjusted analyses, which show whether the test has prognostic value over clinicopathological variables, are reported in the section "Additional prognostic value".

DRFS: Table 11 presents DRFS data. One study from China[85] reported 5-year DRFS, with HR for high vs. low-risk of 2.2 (95% CI: 1.11, 4.30, p=0.004) and a C-index (AUC) of 0.685 (95% CI: 0.540, 0.830) indicating the model is better than chance at placing patients into appropriate risk categories.

DRFI: Data relating to DRFI are presented in Table 12. Four studies[35, 43, 45, 49, 52] in LN0 patients receiving 100% endocrine monotherapy reported DRFI, of which three compared DRFI according to RS risk; all showed a statistically significant prognostic effect. For 5-year DFRI, the HR for a 50-point difference in RS was 6.04 (3.88, 9.41, p<0.001) in one study,[45, 51] while in another the HR for high versus low-risk was [redacted].[35, 43] For 10-year DFRI, the HR for high versus low-risk was 3.8 (95% CI: 2.36, 6.1; p<0.001) in one study[45, 51] and [redacted][35, 43] while in a third study the HR for a 50-point difference in RS was 6.20 (95% CI: 2.27, 17.0, p<0.001).[52] Intermediate versus low HRs were lower at [redacted] and 2.21 (1.28, 3.81)[45, 51] at 5 years and [redacted] at 10 years.[43] Across all four studies, estimates of DRFI at 10 years ranged from 93.2%[45] to 96.8%[49] in low-risk patients, from 85.7% [45, 51] to 100%[52] in intermediate-risk patients, and from 60.5%[49] to [redacted] in high-risk patients.

Two studies of LN0 patients[49, 89] who were treated with endocrine therapy and chemotherapy in varying proportions (Table 12) reported 10 year DRFI, with one reporting 5 year DRFI also.[89] Sun *et*

al. 2011[89] (China) reported particularly poor DRFI at both time points in comparison with other studies. DRFI was progressively worse with increasing risk category in both studies (see Table 12) and the difference was statistically significant (p=0.02) in the one study that reported this.[89] In the other study (NSABP B-20),[49] survival in the high-risk group was higher (88.1 (95% CI: 82.0, 94.2)) than in other studies where patients were not treated with chemotherapy.

In LN+ patients (Table 12), only the TransATAC analysis included 100% patients with endocrine monotherapy.[43] In this study, 5-year DRFI was [redacted]

Three LN+ studies[51, 89-91] treated patients with variable endocrine therapy and chemotherapy and each reported statistically significant differences in DRFI between risk groups. For 5-year DRFI, the HR for a 50-point difference in RS was 4.1 (CI: NR, p<0.001) in one study[91] and 4.22 (2.93, 6.07, p<0.001) in another.[51, 90] DRFI rates were generally lower than LN0 groups, again with Sun *et al*. 2011[89] (China) reporting very poor survival rates compared with other studies.

DFS: Table 13 presents DFS data. One study[68] in LN+ patients reported a statistically significant 10-year HR for a 50-point difference in RS (2.64 (95% CI: 1.33, 5.27, p=0.006)) but the assumption of proportional hazards was not met with a 5-10 year HR of 0.86 (95% CI: 0.27, 2.74, p=0.80). One study[100] (in patients of unknown LN status and treatment status) reported statistically significant differences between high- and low-risk patients (p=0.760) but not between high- and intermediate-risk, or low- and intermediate-risk groups (p=0.072 and p=0.760 respectively). Two studies[51, 90, 91] in LN+ patients receiving variable levels of endocrine therapy and chemotherapy reported that RS was statistically significantly prognostic for DRFI (p<0.001 in each case);[51, 90, 91] one reported an HR for a 50-point difference in RS of 3.3 (CI: NR, p<0.001)[91] while the other did not report an HR.[51, 90]

OS and BCSS: Table 14 presents OS and BCSS data. Two studies of LN0 patients treated with endocrine monotherapy reported [redacted].[43, 52] The TransATAC analysis reported [redacted] and the other study reported a statistically significant difference between high and low-risk groups (p=0.008).[52]

The TransATAC study[43] of LN+ patients treated with endocrine monotherapy reported [redacted]

[redacted], whilst Albain *et al*. 2010[68] (LN+) reported an HR for 10-year OS for a 50-point difference in RS of 4.42 (95% CI: 1.96, 9.97, p=0.0006).

In LN+ patients variably treated with endocrine and chemotherapy, one study[91] reported a statistically significant difference in OS (7.7 year median) with an HR for a 50-point difference in RS of 5.0 (CI: NR, p<0.001). Another study reported a statistically significant effect on 10-year OS (p<0.001).[51, 90]

RFI and RFS: Table 15 and Table 16 present RFI and RFS data, respectively. Two studies reported data for these outcomes. Toi *et al*. (Japan)[52] reported a statistically significant difference between high- and low-risk patients for 10-year RFI and RS (both p<0.05). The E2197 analysis[98, 99] reported very similar rates of 5- and 10-year RFI across subgroups of LN0, LN+ and LN+/- patients who were all treated with endocrine and chemotherapy; survival was progressively worse with increasing risk category but no significance tests were reported (the C-index (AUC) for 5-year RFI was 0.69).

Additional prognostic value

This section reports adjusted analyses, which indicate the additional prognostic value of IHC4 over clinicopathological factors. The clinicopathological factors adjusted for vary from study to study, and are detailed in the footnotes to the tables.

Table 17 presents data relating to the additional prognostic value of Onctoype-Dx RS over clinicopathological variables. One study (E2197)[98, 99] reported RFI for a mixed cohort of LN+/- patients (Table 17). For RFI, HRs for a 50-point differerence in RS (adjusted for number of positive nodes, tumour size, age, HER2 status and grade) were borderline statistically significant at 5 years (2.12; 95% CI: 0.97, 4.65, p=0.06) and 10 years (2.27; 95% CI: 1.04, 4.97). However, in a subgroup of HER2- patients, the adjusted HR for a 50-point differerence in RS was not statistically significant (data NR).

Two studies (NSABP B-14 and the Japanese study)[45, 52] reported analyses of LN0 patients who received endocrine monotherapy. Both reported analyses adjusted for clinicopathological variables. HRs for a 50-point differerence in RS were statistically significant in all DRFI and RFI analyses,[45, 52] with a statistically significant increase in likelihood ratio χ^2 (p<0.001) over age and tumour size alone, and over age, tumour size, tumour grade, HER2 amplification, ER and PR.[45] HRs for a 50-point differerence in RS adjusted for age and tumour size were not statistically significant for RFS and OS in one study.[52]

In a study of LN0 patients[89] some of whom had endocrine and/or chemotherapy the HR for DRFS, for a 1-point difference in RS, was 1.03 (95% CI: 1.01, 1.06, p=0.017), but it was unclear if all

clinicopathological variables listed were included in the model (age, tumour size, nodal status, ER, PR, HER2, endocrine therapy, chemotherapy, St Gallen), or just endocrine therapy and chemotherapy.

Three studies assessed LN+ patients, some or all of whom were treated with endocrine and chemotherapy. HRs for Oncotype DX RS adjusted for clinicopathological variables (see footnote to Table 17) were statistically significant in all three studies[51, 89-91] for outcomes including DRFI, DRFS, DFS and OS; only one reported an HR, which was for a 1-point difference in RS (1.03 (95% CI: 1.00, 1.07), p=0.039).[89] Notably, of these three studies, only Sun *et al*. (2011) adjusted for ER, PR and HER2.[89]

Oncotype DX versus Adjuvant! Online

Two studies (E2197 and NSABP-B-14)[45, 98, 99] compared Oncotype Dx RS with AOL (Table 18). The E2197[98, 99] study (LN0/+, 100% endocrine and chemotherapy) compared Oncotype DX against a model (the "clinical integrator") based on AOL, where the integrator was adjusted to 5-year outcomes rather than AOL's 10 year outcomes. For RFI, based on the C-indexes (AUC) reported (Oncotype DX 0.69; Integrator 0.61; p-values NR) and on HRs (Oncotype DX HR for 50-point difference: 2.51 (95% CI: 1.71, 3.70; p<0.001); integrator HR: 1.51 (95% CI: 1.07, 2.13; p=0.02)), the integrator performed less well than Oncotype DX (statistical significance NR). Analyses (not in Table 18) where patients were sub-grouped by the integrator or RS risk groups, and the other test applied to the patients in that risk group, showed that both tests provided additional prognostic information over the other.

The NSABP B-14 analysis[45] of LN0 patients treated with endocrine monotherapy showed that Oncotype DX was statistically significantly prognostic for DRFI when adjusted for AOL (HR for 50-point difference 2.83 (95% CI: 1.91, 4.18, p<0.001). In addition, AOL was statistically significantly prognostic for DRFI when adjusted for Oncotype DX (HR 1.93; 95% CI: 1.27, 2.91, p=0.002) (Table 18). When clinical variables were added into the model, the HR for AOL was no longer statistically significant (HR 0.86 (95% CI: 0.45, 1.62, p=0.636)) whereas that for Oncotype DX was (HR 2.37 (95% CI: 1.58, 3.55, p<0.001)).

Oncotype DX versus CTS and NPI

The TransATAC analysis[43] reports a reduced dataset of patients where data for all four in-scope tests are available. Additional prognostic value over NPI or the Clinical Treatment Score (CTS, a combination of nodal status, tumour size, grade, age and treatment) was assessed via increases in likelihood ratio χ^2 for 10-year DRFI, for Oncotype DX plus NPI or CTS, over NPI or CTS alone (Table 18). Increases in likelihood ratio χ^2 were [redacted]

[redacted]

Prognostic performance: Oncotype RSPC

The Oncotype RSPC algorithm includes Oncotype RS plus age, tumour size and grade.[42] Table 19 presents data relating to Oncotype RSPC. One study (Tang *et al*. 2011b)[42] derived the RSPC score in a meta-analysis of NSABP B-14 and TransATAC (LN+/- patients, 100% endocrine monotherapy), and performed a limited validation in NSABP B-20 (LN0 patients, 100% treated with endocrine therapy; 64% also with chemotherapy).

Derivation: In the derivation cohort, both RSPC and Oncotype DX RS had statistically significantly ($p<0.001$) worsening 10-year DRFI rates as test scores increased (HR/CI NR). However, DRFI rates were not significantly different between RSPC and RS within each risk group (respectively, 93.5% vs. 94.1%, p=0.68 for low-risk; 82.4% vs. 86.2%, p=0.27 for intermediate-risk; and 73.8% vs. 70.5%, p=0.42 for high-risk. RSPC was able to reclassify RS intermediate patients as 16.9% (n=46) high-risk RSPC and 55.1% (n=150) low-risk RSPC; RS low-risk patients as 1.9% (n=15) high-risk RSPC and 8.9% (n=70) intermediate-risk RSPC; and RS high-risk patients as 28.6% (n=NR) intermediate-risk RSPC. The increase in likelihood ratio χ^2 for 10-year DRFI was 76.9 ($p<0.001$) for RSPC over RS, and 45.4 ($p<0.001$) for RSPC over grade, tumour size and age.

Vlaidation: Only HRs were reported for the validation cohort (NSABP B-20), and this was 2.43 ($p<0.001$) for RSPC and 2.22 ($p<0.001$) for RS.

Further data relating to the RSPC were reported in the TransATAC data request. However, as the original derivation cohort includes TransATAC patients, these data are not included in this analysis. They are included in the section on multiple tests (Section 4.8.1).

Discussion: Oncotype DX and RSPC prognostic performance

Oncotype DX

Oncotype DX was validated in eleven distinct datasets. Seven were re-analyses of RCTs, [35, 45, 49, 51, 68, 90, 91, 98, 99, 101] where three treated patients with endocrine monotherapy (one study recruited mixed lymph node status,[35, 101] one recruited LN0 patients[45] and one recruited LN+ patients[68]); one treated some patients with endocrine monotherapy and some with endocrine and chemotherapy (LN0 patients[49]); two treated all patients with endocrine therapy and chemotherapy (one study recruited mixed lymph node status,[98, 99] one recruited LN+ patients[51, 90]); and one treated all patients with chemotherapy and 74% with endocrine therapy (LN+ patients).[91] The remaining four were

retrospective studies where patients were treated according to usual practice (without Oncotype DX) with varying levels of endocrine therapy and chemotherapy. [52, 85, 89, 100] The total number of patients included was 5156.

The quality of the studies overall was poor to moderate according to the criteria used, with particular concerns about differences in endocrine and chemotherapy treatments given to patients, blinding of test assessors to patient outcomes, and the potential for attrition of patients with small tumours, due to insufficient tumour sample being available to run the test. Only 4 studies[45, 49, 51, 68, 101] appropriately treated patients with endocrine monotherapy, and it was not always clear if this was for at least 5 years. The remaining cohorts were confounded by under-treatment with endocrine therapy and/or treatment with chemotherapy, both of which can affect recurrence and may alter the observed HRs for outcomes between high- and low-risk groups. A lack of blinding is likely to have a low impact as Oncotype DX is an objective test. The potential loss of patients with small tumours is of unknown concern, as it is unknown whether Oncotype DX would have a different prognostic performance in these patients.

The proportion of patients classified as low-risk ranged from 48% to [redacted] in LN0 cohorts and was generally lower, ranging from 36% to [redacted], in LN+ cohorts. The proportion of patients who were classified as intermediate risk ranged from 20% to [redacted] in LN0 cohorts, and was generally higher in LN+ cohorts, ranging from 30% to 34%. The proportion of patients who were classified as high risk ranged from [redacted] to 33% in LN0 patients and was similar in LN+ patients, ranging from [redacted] to 32%. The number of patients who are likely to be prescribed chemotherapy on the basis of their test result will depend on how intermediate-risk patients are handled and whether they would be handled the same in LN0 and LN+ groups.

Data on discrimination largely comprised recurrence/survival per risk group and HRs between risk groups. 10 year DRFI in low-risk LN0 patients treated with endocrine monotherapy ranged from 93% to 97% (4 studies), [45, 49, 52, 101] and was similar where 100% patients received endocrine therapy and chemotherapy (96%, 1 study).[49] LN+ patients had much lower 10 year DRFI (82% (1 study) [101] with endocrine monotherapy; 81% (1 study) [51 90] where 100% patients received endocrine therapy and chemotherapy. 10 year DRFI in LN0 intermediate-risk patients treated with endocrine monotherapy ranged from 86% to 100% (4 studies),[45, 49, 52, 101] and was similar where 100% patients received endocrine therapy and chemotherapy (89%, 1 study).[49] LN+ intermediate-risk patients had much lower10 year DRFI; [redacted] (1 study) [101] with endocrine monotherapy; 65% (1 study) [51 90] where 100% patients received endocrine therapy and chemotherapy. 10 year DRFI in LN0 high-risk patients treated with endocrine monotherapy ranged from 61% to 77% (4 studies),[45, 49, 52, 101] and was similar where 100% patients received endocrine therapy and chemotherapy (88%, 1 study).[49] LN+ high-risk

patients had much lower 10 year DRFI; [redacted] (1 study) [101] with endocrine monotherapy; 59% (1 study) [51 90] where 100% patients received endocrine therapy and chemotherapy. All the DRFI rates in this paragraph exclude one study from China, which appeared an outlier with very low DRFI rates (Sun et al. 2011).[89] The study from Japan[52] also reported some unusual results in that intermediate-risk patients had better DRFI than low-risk patients (e.g. 10 year DRFI 97% and 100% respectively). It is unclear if, for both of these studies, [52, 89] the unusual results are due to small sample sizes (N=98 and N=200), differences in treatment practices, or differences in ethnicity.

Despite confounding from treatment, the studies reporting prognostic performance data reported largely statistically significant differences between high-risk and low-risk groups whether through HRs or through analyses of event rates per group, and this was the case regardless of lymph node status. However, differences between the intermediate-risk group and the high- or low-risk groups were not always statistically significant, particularly in the LN+ population.

Two studies reported a C-index (AUC). One study[98, 99] was in LN+/- patients and the other was in LN0 patients.[85] In both cases the C-index was 0.69, which indicated that the model was better than chance at placing patients into appropriate risk categories and nearly reaches the 0.7 cut-off for a "good" test.[44]

The analyses which reported multivariable Cox models that were adjusted for clinicopathological variables generally indicated that the prognostic performance of Oncotype DX had additional benefit over these factors, as HRs remained significant in most analyses (the exception being RFS and OS analyses by Toi *et al*. 2010).[52] This was consistent regardless of lymph node status and variables adjusted for, which included age, tumour size and LN status (where applicable) in all cases, and grade in most (Toi *et al.* 2010 and Sun *et al*. 2011 being the exceptions).[52, 89] However, covariates included in multivariate analyses varied, and it is not clear if all important covariates were included in all analyses.

In addition, in comparison to other clinicopathological tests, the likelihood ratio χ^2 for Oncotype DX when added to a model containing CTS or NPI was [redacted] for LN0 patients in the TransATAC dataset, while [redacted] for LN+ patients.[101] Compared with AOL, and to a model based on AOL but with 5-year outcomes, Oncotype DX also appeared to provide additional prognostic information.

Oncotype DX RSPC

One study (Tang *et al*. 2011b)[42] derived the RSPC score in a meta-analysis of NSABP B-14 and TransATAC (LN+/-; N=1735), and performed a limited validation in NSABP B-20 (LN0; N=625). Based on the derivation analysis set, the Oncotype DX RSPC algorithm (Oncotype DX plus age, tumour size and grade) appeared to provide additional prognostic information over Oncotype DX and over clinicopathological variables, and was able to classify more patients into a low-risk category than Oncotype DX whilst maintaining a roughly equivalent rate of distant recurrence in the low-risk group. In the validation cohort, RSPC had prognostic value in a univariate analysis. However, RSPC has only had limited validation in one independent cohort and it has not been tested in pre-menopausal or LN+ patients.

Conclusion: Oncotype DX and RSPC prognostic performance

Seven re-analyses of RCTs and four retrospective cohort studies were included with a total of 5156 patients. The generalisability of the evidence base to the decision problem is uncertain due to the loss of patients with insufficient tumour sample available to be tested. Generally, when comparing LN0 to LN+ patients, similar numbers were at high risk, but more were at low risk in LN0 cohorts, and more at intermediate risk in LN+ cohorts. How many patients would be prescribed chemotherapy would depend on how intermediate patients are handled. 10 year DRFI rates suggest patients in the LN0 low-risk group are at very low risk of recurrence (10 year DRFI range 93%-97%) in the absence of chemotherapy), and patients in the intermediate risk group at somewhat higher risk (10 year DRFI range 86%-100%). LN+ patients were generally at higher risk of recurrence than LN0 in both low and intermediate categories (10 year DRFI <85% and <75% respectively). Unadjusted analyses indicated Oncotype DX had prognostic power (statistically significant differences between low-risk and high-risk groups) across various recurrence outcomes, regardless of lymph node status. HRs between intermediate-risk group and the high- or low-risk groups were not always statistically significant. Oncotype DX provided additional prognostic information over most commonly used clinicopathological variables (age, grade, size, nodal status) regardless of lymph node status, and over CTS and NPI in LN0 (but not LN+) patients. On the basis of proportions classified as low-risk and DRFI rates, RSPC may outperform Oncotype DX in LN0 patients, but this data is from the derivation cohort, with only limited validation data from one independent cohort, and it has not been tested in pre-menopausal or LN+ patients.

Table 9: Study and patient characteristics: Oncotype DX prognostic performance

Reference; N	Cohorts	Country	Study design	Details of test	Cut-offs	Population	Nodal status	Endo / chemo
Reanalyses of RCTs: LN status mixed								
100% ET monotherapy								
Sestak 2017 (data request)[43] Dowsett 2010[35 a] [redacted]	TransATAC	UK	Reanalysis of prospective trial (RCT); archive tissue	FFPE Genomic Health	18-30	100% HR+ 100% HER2- Postmenopausal 100% Female	[redacted]	100% ET monotherapy
Variable ET&CT								
Goldstein 2008 (5 year) ; [98]Sparano 2012[99] (10-year) N=465	E2197 (ECOG trial)	USA	Nested Case-Control from prospective RCT; archive tissue	FFPE Genomic Health	18-30	100% HR+ 44% HER2- (34.1% unknown), Meno NR 100% Female If LN0, tumour ≥1.1cm	LN0, 56.5% LN1-3, 43.5%	100% ET & CT 40% aromatase inhibitor
Reanalyses of RCTs: LN0 studies								
100% ET monotherapy								
Paik 2004;[45] Wolmark 2016[51] N= 668	NSABP B-14	USA	Reanalysis of prospective trial (RCT); archive tissue	FFPE Genomic Health[b]	18-30	100% ER+ HER2+/-, % NR Meno NR Female NR	LN0	100% ET monotherapy
Paik 2006[49] N= 651	NSABP B-20	USA	Reanalysis of prospective trial (RCT); archive tissue	FFPE Genomic Health	18-30	100% ER+ HER2+/-, % NR Meno NR Female 100%	LN0	1) 100% ET monotherapy (N=227) 2) 100% ET + 100% CT (N=424)
Reanalyses of RCTs: LN+ studies								
100% ET monotherapy								
Albain 2010[68] N=148 (tamoxifen monotherapy subgroup)	SWOG-8814	USA	Reanalysis of prospective trial (RCT); archive tissue	FFPE Genomic Health	18-30	100% HR+ 91% HER2- Postmenopausal 100% Female	LN+, 100% LN>3, 37%	100% ET monotherapy
100% CT&ET								

Reference; N	Cohorts	Country	Study design	Details of test	Cut-offs	Population	Nodal status	Endo / chemo
Wolmark 2016[51 c] Mamounas 2012[90 d] N=1065	NSABP B-28 (Also reports NSABP B-14, listed here under Paik 2004)	USA	Reanalysis of prospective trial (RCT); RS available	FFPE Genomic Health (Assumed for B-28)	18-30	100% ER+ HER2 NR Meno NR Female NR	LN+	100% CT & ET
Variable ET&CT								
Penault-Llorca 2014[91] N=530	PACS01	France	Reanalysis of prospective trial (RCT); unclear if archive tissue	NR	NR	100% HR+ HER2 NR Meno NR Female NR	LN+	100% CT 74.2% ET
Retrospective studies								
Gong 2016[85] O-DX subgroup N=153	SYSMH; CCSYU; 3rdHNC	China	Retrospective reanalysis of routinely collected data; archive tissue	FFPE Multiplex branched-DNA liquid chip technology Surexam, Guangzhou, China	NR, assume 18-30	100% HR+ 100% HER2- 61% post meno % female NR non-metastatic	LN0	100% ET; 79% CT
Russell 2016[100] N=135	University of South Florida; Morton Plan Hospital	USA	Observational study (not treated according to O-DX)	NR	NR	100% ER+ HER2- NR Meno NR Female NR	NR	NR – usual practice guided by MMP
Sun 2011[89] N=93	Hospital Affiliated Academy of Military Medical Science, Beijing	China	Retrospective reanalysis of routinely collected data; consecutive	FFPE qRT-PCR (not Genomic Health)	18-30	100% HR+ 86% HER2- (7.5% unclear) 82.6% Premeno 100% Female	LN+/- LN0, 61.3% LN1-3, 19.4% LN>3, 18.3%	75.3% ET 80.6% CT
Toi 2010[52] N=200	8 Japanese hospitals (unnamed)	Japan	Retrospective reanalysis of routinely collected data; archive tissue	FFPE Genomic Health	18-30	100% ER+ HER2 NR Meno NR % Female NR T1-T2	LN0	100% ET % CT NR

Reference; N	Cohorts	Country	Study design	Details of test	Cut-offs	Population	Nodal status	Endo / chemo
Oncotype DX RSPC data								
Tang 2011b[42] **B-14:** n=647 **TransATAC:** n=1088 **B-20:** n=625	NSABP B-14 & TransATAC meta-analysis NSABP B-20	NSABP: USA TransATAC: UK	Reanalysis of prospective trials (RCT); archive tissue	FFPE Genomic Health	**RSPC:** 12% - 20% **RS:** 18-30	100% ER+ **B-14:** HER2+/-, % NR **TransATAC:** HER2+/- **B-20:** HER2+/-, % NR	**B-14:** LN0 **TransATAC:** LN+/- **B-20:** LN0	**B-14:** 100% ET **TransATAC:** 100% ET **B-20:** 36% ET; 64% CT&ET

O-DX, Oncotype DX; MMP, MammaPrint; FFPE, formalin fixed, paraffin embedded tumour samples; SYSMH, Sun Yat-sen Memorial Hospital; CCSYSU, Cancer Centre of Sun Yat-sen University; 3rdHNC, Third Hospital of Nanchang City

[a] TransATAC is reported across several publications, each with a different aim and/or reporting results of different tests. Data was also made available to the EAG via NICE [redacted]; [b] from Paik 2006, about Paik 2004 "a preliminary version of the RT-PCR assay (lacking standardized reagents, calibrators, and controls)"; [c] Note data for B-14 also reported in this article, but reported here under Paik 2004;[d] Note a second abstract (Mamounas 2012)[102] presented data for the same cohort, but split by chemotherapy treatment group, and has been excluded as it added no new data to Mamounas 2012[90]

Table 10: Quality assessment of Oncotype DX prognostic performance studies

Reference: N	Cohorts	Derivation or validation?	Study design appropriate?	All eligible patients included?	Blinding (of test assessors to outcomes)	Definition of outcome standardised or *a priori*?	Applicability: Patient Spectrum	Applicability: Test as per decision problem?
Albain 2010[68]	SWOG-8814	V	Y, reanalysis of endocrine only arm of RCT	N InT; TF	Y	Y	No, InT, TF, >20% LN>3. However, adjustments applied in several analyses	Y
Goldstein 2008 (5 year);[98] Sparano 2012[99] (10-year)	E2197 (ECOG trial)	V	N- authors identify possible bias; all CT	UC	Y	Y	No, >20% HER2+	Y
Gong 2016[85]	SYSMH; CCSYU; 3rdHNC	V	N, cohort study, some CT	N InT; MD	UC	Y	N, InT, MD	N – Oncotype DX algorithm, but used Surexam, Guangzhou, China assay.
Paik 2004[45]	NSABP B-14	V	Y, reanalysis of RCT; endocrine only	N InT	UC	Y	N, InT, %HER2- NR	UC
Paik 2006[49]	NSABP B-20	D (ET arm) V (ET&CT arm)	N, reanalysis of RCT; some CT	N InT	UC	Y	N, InT, %HER2- NR	Y
Penault-Llorca 2014[91]	PACS01	V	N, reanalysis of RCT; some CT	N InT	UC	Y	N InT	UC[a]
Russell 2016[100]	University of South Florida; Morton Plan Hospital	V	N, cohort study, usual practice (some CT)	N InT, SfT	UC	Y	N InT	Y
Sun 2011[89] N=93	Hospital Affiliated Academy of Military Medical Science (HAAMMS), Beijing	V	N, cohort study (retrospective) some CT	N InT; MD	UC	Y	N InT, MD, 18% LN>3	N Oncotype DX algorithm, but assay not Genomic Health
Toi 2010[52]	8 Japanese hospitals (unnamed)	V	UC, cohort study (retrospective), %CT NR	N InT; MD; FT	UC	Y	N InT, MD, FT, HER2 NR	Y

Reference: N	Cohorts	Derivation or validation?	Study design appropriate?	All eligible patients included?	Blinding (of test assessors to outcomes)	Definition of outcome standardised or *a priori*?	Applicability: Patient Spectrum	Applicability: Test as per decision problem?
Wolmark 2016[51] Mamounas 2012[90] N=1065	NSABP B-28	V	N, reanalysis of RCT; all CT	N InT; FT	UC	Y	N InT; FT	Y
Sestak 2017 (data request)[43] Dowsett 2010[35]	TransATAC	V	Y, reanalysis of RCT, ET monotherapy	N InT; FT	Y	Y	N InT, FT	Y
RSPC								
Tang 2011b[42]	NSABP B-14 & TransATAC meta-analysed NSABP B-20	D, V	Y, reanalysis of RCT N, B-20 some CT	N, InT; ER+ by RS[b]	UC	Y	UC %HER- NR	Y

InT, insufficient tissue; TF, test failure; MD, missing data; ER, oestrogen receptor status; RS, recurrence score; SfT, sent for test; SYSMH, Sun Yat-sen Memorial Hospital; CCSYSU, Cancer Centre of Sun Yat-sen University; 3rdHNC, Third Hospital of Nanchang City

[a] In this analysis, patients were classed as ER+ using RS rather than histology, which does not reflect clinical practice as patients would be selected for RS testing using histology

[b] from Paik 2006, about Paik 2004, "a preliminary version of the RT-PCR assay (lacking standardized reagents, calibrators, and controls)" suggests the assay used was somewhat different to the commercial version now available.

Table 11: Oncotype DX prognostic performance, DRFS

Reference; N	Cohorts	Population	Nodal status	ET/CT	% pts per group			% DRFS risk: 0-5 yr			% DRFS risk: 0-10 yr			DRFS: HR (95% CI) (unless otherwise stated)
					Low	Inter	High	Low	Inter	High	Low	Inter	High	0-5 yr
LN0, variable ET&CT														
Gong 2016[85] N=153	SYSMH; CCSYU; 3rdHNC	100% HR+ 100% HER2-	1) LN0	1)100% ET; 79% CT	49	26	25							**High vs. low:** 2.2, (1.11, 4.30, p=0.004) **High vs. Inter:** 1.9, (0.55, 6.47, p=0.108) **Inter vs. low:** 1.0, (0.67, 1.52, p=0.953) **C-index (AUC)** 0.685 (95% CI: 0.540, 0.830)

DRFS, distant recurrence-free survival; SYSMH, Sun Yat-sen Memorial Hospital; CCSYSU, Cancer Centre of Sun Yat-sen University; 3rdHNC, Third Hospital of Nanchang City; N, number of patient; ET, endocrine therapy; pts, patients; CT, chemotherapy; Inter, intermediate group; yr, year; HR Hazard ratio; CI, confidence interval; HR+, hormone receptor positive; LN, lymph node; AUC, area under the curve;

Table 12: Oncotype DX prognostic performance, DRFI

Reference; N	Cohorts	Population	Nodal status	ET/CT	Cut-offs	% pts per group			% DRFI risk: 0-5 yr (95% CI)			% DRFI risk: 0-10 yr (95% CI)			DRFI: HR (95% CI)	
						Low	Inter	High	Low	Inter	High	Low	Inter	High	0-5	0-10 yr
LN0/+																
Variable ET & CT																
Sun 2011[89] N=93[a]	HAAMMS	100% HR+ 86% HER2-	LN+/-	LN+/- 75.3% ET 80.6% CT	18-30	37	31	32	5.5yr median FU **Low vs High:** p<0.001 **Inter vs High:** p=0.003						5.5yr median FU: RS 50 pt difference: 2.35 (1.58, 3.49), p<0.001	
LN0																
100% ET monotherapy																
Paik 2004,[45] Tang 2011a Wolmark 2016[51] N= 668	NSABP B-14	100% ER+ HER2+/-, % NR	LN0	100% ET	18-30	51	22	27	97.9 (95.6, 99.0)[b]	90.8 (84.7, 94.5)[b]	77.9 (71.1, 83.4)[b]	93.2 (90.4, 96.0)	85.7 (79.7, 91.7)	69.5 (62.6, 76.4)	RS 50-point difference: 6.04 (3.88, 9.41), p<0.001[b]	**Inter vs Low:** 2.21 (1.28, 3.81)
									Log rank p<0.001[b]			p<0.001 high vs low				**High vs Low:** 3.8 (2.36, 6.1) p<0.001
												5-10 yr 95.2 (92.1, 97.2)[b]	**5-10 yr** 94.4 (88.0, 97.5)[b]	**5-10 yr** 89.2 (82.4, 93.4)[b]	**5-10 year:** RS 50 point difference: 1.55 (0.81, 2.97), p=0.20[b]	
												Log rank p=0.06** high vs low				
												5-15 yr 93.3 (89.6, 95.7)[b]	**5-15 yr** 88.1 (79.9, 93.1)[b]	**5-15 yr** 86.4 (79.0, 91.3)[b]		
Toi 2010[52] N=200	8 Japanese hospitals	100% ER+ HER2 NR	LN0	100% ET	18-30	48	20	33				96.7 (90.0, 98.9)	100 (NR)	75.2 (62.2 to 84.3)		50-point increase: 6.20 (2.27, 17.0), p<0.001
												p<0.001 log rank (low vs. high)				

Reference; N	Cohorts	Population	Nodal status	ET/CT	Cut-offs	% pts per group			% DRFI risk: 0-5 yr (95% CI)			% DRFI risk: 0-10 yr (95% CI)			DRFI: HR (95% CI)	
						Low	Inter	High	Low	Inter	High	Low	Inter	High	0-5	0-10 yr
Sestak 2017 (data request)[43] Dowsett 2010[35] [redacted]	TransATAC	100% HR+ 100% HER2- Postmeno	LN0	100% ET	18-30	[redacted]										
Paik 2006[49] N=227	NSABP B-20	100% ER+ HER2+/-	LN0	100% ET	18-30	59	20	21				96.8 (93.7, 99.9)	90.9 (82.5, 99.4)	60.5 (46.2, 74.8)		
100% ET+CT																
Paik 2006[49] N=424	NSABP B-20	100% ER+ HER2+/-	LN0	100% ET&CT	18-30	51	21	28				95.6 (92.7, 98.6)	89.1 (82.4, 95.9)	88.1 (82.0, 94.2)		
Variable ET&CT																
Sun 2011[89] N=57 [a]	HAAMMS	100% HR+ 86% HER2-	LN0	75.3% ET 80.6% CT	18-30	-	-	-	84.4 (77.2, 91.6)	72.6 (62.1, 83.1)	41.7 (27.5, 55.9)	57.9 (41.4, 74.4)	43.0 (23.7, 62.3)	20.8 (4.4, 37.2)		
												p=0.02				
LN+																
100% ET monotherapy																
Sestak 2017 (data request)[43] Dowsett 2010[35] [redacted]	TransATAC	100% HR+ 100% HER2- Postmeno	LN1-3	100% ET	18-30	[redacted]										
Variable ET&CT																
Wolmark 2016[51] Mamounas 2012[90] N=1065	NSABP-28	100% ER+ HER2 NR	LN+	100% CT & ET	18-30	36	34	30	91.6 (88.3, 94.0)	81.2 (76.8, 84.9)	70.3 (64.9, 75.1)	80.9 (76.4, 84.6) [c]	64.9 (59.6, 69.7) [c]	55.8 (50.0, 61.2) [c]	RS 50-point difference: 4.22 (2.93, 6.07), p<0.001	

Reference; N	Cohorts	Population	Nodal status	ET/CT	Cut-offs	% pts per group			% DRFI risk: 0-5 yr (95% CI)			% DRFI risk: 0-10 yr (95% CI)			DRFI: HR (95% CI)	
						Low	Inter	High	Low	Inter	High	Low	Inter	High	0-5	0-10 yr
									Log rank p<0.001			p<0.001				
												5-10 yr 88.3 (84.3, 91.4)	**5-10 yr** 79.9 (74.7, 84.2)	**5-10 yr** 79.3 (73.1, 84.3)	**5-10 yr:** RS 50-point difference: 1.66 (1.05, 2.61), p=0.04	
												Log rank p=0.02				
Penault-Llorca 2014[91] N=530	PACS01	100% HR+	LN+	100% CT 74.2% ET	NR	39	30	31	93.7 (89.4, 96.3)	87.3 (81.0, 91.6)	69.3 (61.5, 75.8)	p<0.001			7.7yr median FU, RS 50 point difference: 4.1 (CI NR), p<0.001	
Sun 2011[89] N=35 [a]	HAAMMS	100% HR+ 86% HER2-	LN+	LN+/- 75.3% ET 80.6% CT	18-30				62.5 (45.4, 79.6)	66.7 (51.0, 82.4)	16.7 (7.9, 25.5)	62.5 (45.4, 79.6)	33.3 (8.5, 58.1)	16.7 (7.9, 25.5)		
												p=0.038				

HAAMMS, Hospital Affiliated Academy of Military Medical Science, Beijing; N, number of patient; ET, endocrine therapy; pts, patients; CT, chemotherapy; Inter, intermediate group; yr, year; HR Hazard ratio; CI, confidence interval; HR+, hormone receptor positive; LN, lymph node; AUC, area under the curve.

[a] Outcome described as DRFS, but the definition is for DRFI as it excludes contralateral disease, loco-regional relapse, other primary cancers and non-breast cancer deaths; [b] from Wolmark 2016; [c] this data is from Mamounas 2012. The same data is reported in the Company Submission as DRFS. As DRFI is defined and reported in Wolmark 2016, we have assumed Mamounas 2012 is correct in calling this DRFI.

Table 13: Oncotype DX prognostic performance, DFS[a]

Reference; N	Cohorts	Population	Nodal status	ET/CT	% pts per group			% DFS risk: 0-5 yr			% DFS risk: 0-10 yr			DFS: HR (95% CI)	
					Low	Inter	High	Low	Inter	High	Low	Inter	High	0-10 yr	Other
LN+, 100% ET monotherapy															
Albain 2010[68] N=148	SWOG-8814	100% HR+ 91% HER2- Postmeno	LN+, 100% LN>3, 37%	100% ET	37	31	32	-	-	-	60	49	43	RS 50 point difference: 2.64 (1.33, 5.27, p=0.006)	Assumption of proportional hazards not met (p=0.0016) **0-5 years** HR 5.55 (2.32, 3.28, p=0.0002) **5-10 years** HR 0.86 (0.27, 2.74, p=0.80)
LN status NR, ET & CT NR															
Russell 2016[100] N=135	University of South Florida; Morton Plan Hospital	100% ER+ HER2- NR Meno NR Female NR	NR	NR – usual practice guided by MMP	53	26	21							4.5 yr median FU: Mantel-Cox Log Rank Inter vs low: p=0.760 High vs low: p=0.036 Inter vs high: p=0.072	
LN+, variable ET&CT															
Penault-Llorca 2014[91] N=530	PACS01	100% HR+	LN+	100% CT 74.2% ET	39	30	31	90.8 (86.0, 94.1)	84.9 (78.3, 89.6)	64.6 (56.7, 71.4)				7.7yr median FU, RS 50 point difference: 3.3 (CI NR), p<0.001	
								p<0.001							
Wolmark 2016[51] Mamounas 2012[90] N=1065	NSABP-28	100% ER+ HER2 NR Meno NR Female NR	LN+	100% CT & ET	36	34	30				75.8 (71.1, 79.8)	57.0 (51.6, 61.9)	48.0 (42.3, 53.4)		
											p<0.001				

[a]DFS, disease-free survival (definition unclear for all studies); N, number of patient; ET, endocrine therapy; CT, chemotherapy; pts, patients; Inter, intermediate group; yr, year; HR Hazard ratio; CI, confidence interval; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; Fu, follow-up; RS, Oncotype DX recurrence score

Table 14: Oncotype DX prognostic performance, OS & BCSS

Reference; N	Cohorts	Population	Nodal status	ET/CT	outcome	% pts per group			% risk: 0-5 yr (95% CI)			% risk: 0-10 yr (95% CI)			OS: HR (95% CI)		
						Low	Inter	High	Low	Inter	High	Low	Inter	High	0-5 years	0-10 years	Other
LN0/+, variable ET&CT																	
Sun 2011[89] N=93	HAAMMS	100% HR+ 86% HER2- (7.5% unclear)	LN+/-	75.3% ET 80.6% CT	BCSS[a]	37	31	32	RS as categorical or continuous variable p=0.553.								
LN0, 100% ET monotherapy																	
Toi 2010[52] N=200	8 Japanese hospitals (unnamed)	100% ER+ HER2 NR	LN0	100% ET	OS	48	20	33				93.6 (86.4, 97.1)	97.4 (83.2, 99.6)	80.9 (68.7, 88.7)			
												p=0.008 log rank test high vs low					
Sestak 2017 (data request)[43] Dowsett 2010[35] [redacted]	TransATAC	100% HR+ 100% HER2- Postmeno	LN0	100% ET	OS	[redacted]									[redacted]		
LN+, 100% ET monotherapy																	
Albain 2010[68] N=148	SWOG-8814	100% HR+ 91% HER2- Postmeno	LN+, 100% LN>3, 37%	100% ET	OS	37	31	32				77	68	51		RS 50 point difference: 4.42 (1.96, 9.97, p=0.0006)	RS risk categories: log rank p=0.003 Proportional hazards not met

Reference; N	Cohorts	Population	Nodal status	ET/CT	outcome	% pts per group			% risk: 0-5 yr (95% CI)			% risk: 0-10 yr (95% CI)			OS: HR (95% CI)		
						Low	Inter	High	Low	Inter	High	Low	Inter	High	0-5 years	0-10 years	Other
Sestak 2017 (data request)[43] Dowsett 2010[35] [redacted]	TransATAC	100% HR+ 100% HER2- Postmeno	LN1-3	100% ET	OS	[redacted]			[redacted]			[redacted]			[redacted]	[redacted]	
LN+, variable ET&CT																	
Penault-Llorca 2014[91] N=530	PACS01	100% HR+	LN+	100% CT 74.2% ET	OS	39	30	31	99.0 (96.2, 99.8)	95.6 (90.9, 97.9)	85.6 (79.1, 90.2)					7.7yr median FU, RS 50 point difference: 5.0 (CI NR), p<0.001	
									p<0.001								
Wolmark 2016[51] Mamounas 2012[90] N=1065	NSABP-28	100% ER+ HER2 NR Meno NR Female NR	LN+	100% CT & ET	OS	36	34	30				90.0 (86.4, 92.6)	74.7 (69.8, 78.9)	63.0 (57.4, 68.2)			
												p<0.001					

HAAMMS, Hospital Affiliated Academy of Military Medical Science, Beijing; N, number of patient; ET, endocrine therapy; CT, chemotherapy; pts, patients; Inter, intermediate group; yr, year; HR Hazard ratio; CI, confidence interval; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; Fu, follow-up; RS, Oncotype DX recurrence score;

[a] Called "overall survival" in the publication, but defined as only breast-cancer deaths

Table 15: Oncotype DX prognostic performance, RFI

Reference; N	Cohorts	Population	Nodal status	ET/CT	% pts per group			% RFI risk: 0-5 yr			% RFI risk: 0-10 yr			Other
					Low	Inter	High	Low	Inter	High	Low	Inter	High	
LN0, 100% ET monotherapy														
Toi 2010[52] N=200	8 Japanese hospitals (unnamed)	100% ER+ HER2 NR Meno NR % Female NR T1-T2	LN0	100% ET	48	20	33				94.5 (87.2, 97.7)	97.5 (83.5, 99.6)	75.4 (62.4, 84.4)	
											High vs Low: p<0.05			
LN0, 100% ET&CT														
Goldstein 2008;[98] Sparano 2012[99] N=233	E2197 (ECOG trial)	100% HR+ 44% HER2- Pre/post-meno	LN0	100% ET&CT	-	-	-	96[a]	86[a]	87[a]	93[a]	76[a]	81[a]	
LN+/-, 100% ET&CT														
Goldstein 2008;[98] Sparano 2012[99] N=465	E2197 (ECOG trial)	100% HR+ 44% HER2- Pre/post-meno	LN0, 56.5% LN1-3 43.5%	100% ET&CT	46	30	24	96[a]	87[a]	83[a]	92[a]	77[a]	75[a]	C-index (AUC) 0.69 at 0-5yr
LN+, 100% ET&CT														
Goldstein 2008;[98] Sparano 2012[99] N=232	E2197 (ECOG trial)	100% HR+ 44% HER2- Pre/post-meno	LN1 (N=123)	100% ET&CT	-	-	-	98[a]	90[a]	82[a]	93.5[a]	85[a]	62.5[a]	
			LN2-3 (N=109)		-	-	-	92[a]	84[a]	67[a]	88[a]	76[a]	63[a]	

RFI, recurrence-free interval; N, number of patient; ET, endocrine therapy; CT, chemotherapy; pts, patients; Inter, intermediate group; yr, year; HR Hazard ratio; CI, confidence interval; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; AUC, area under the curve;

[a] Read off graph, RFI from recurrence rates

Table 16: Oncotype DX prognostic performance, RFS

Reference; N	Cohorts		Population	Nodal status	ET/CT	% pts per group			% RFS risk: 0-10 yr			HR
						Low	Inter	High	Low	Inter	High	10 year
	LN0, 100% ET monotherapy											
Toi 2010[52] N=200	8 Japanese hospitals		100% ER+ HER2 NR	LN0	100% ET	48	20	33	90.4 (82.4, 94.9)	94.9 (81.2, 98.7)	76.6 (64.1, 85.2)	
									High vs Low: p<0.05			
	RFS, relapse-free survival (events include locoregional or distant recurrence or death from any cause; censored are contralateral disease, new cancer, deaths before recurrence)											

Table 17: Oncotype DX, additional prognostic value over clinicopathological factors

Reference; N	Cohorts	Population	Nodal status	ET/CT	Cut-off	Outcome	Test or comparator[a]	Multivariable Cox model (adjusted for CP factors[a]): HR (95% CI) unless otherwise stated	
LN+/- 100% ET & CT									
Goldstein 2008;[98] Sparano 2012[99] N=465	E2197 (ECOG trial)	100% HR+ 44% HER2- Pre/post-meno	LN0, 56.5% LN1-3 43.5%	All ET & CT 40% aromatase inhibitor		**RFI**	**O-DX vs CP factors[a]**	**5 year** [98] HR (RS 50 point difference): 2.12 (0.97, 4.65, p=0.06)[b] 3.13, 1.60, 6.14; p=0.0009 [b]	**10 year[99]** RS 50 point difference: 2.27 (1.04, 4.97)
LN0, 100% ET monotherapy									
Paik 2004,[45] Tang 2011a[50] N= 668	NSABP B-14	100% ER+ HER2+/-, % NR	LN0	100% ET	**18-30**	**DRFI**	**O-DX vs CP factors[a]**	**Increase in likelihood ratio χ^2 over clinical factors[a] or comparator**	**Multivariable Cox model (adjusted for CP factors[a]): HR (95% CI)**
								Cox model 1: 33.7, p<0.001 **Cox model 2:** 15.2, p<0.001 **Cox model 3[c]:** NR	**RS 50 point difference:** **Cox model 1:** 3.21 (2.23, 4.61, p<0.001) **Cox model 2:** 2.81 (1.70, 4.64, p<0.001) **Cox model 3****:** 2.34 (1.56, 3.5), p<0.001
Toi 2010[52] N=200	8 Japanese hospitals (unnamed)	100% ER+ HER2 NR Meno NR % Female NR T1-T2	LN0	100% ET	18-30	**DRFI**	**O-DX vs CP factors[a]**	RS 50 point difference: 6.03 (2.17, 16.7), p<0.001	
						RFI	**O-DX vs CP factors[a]**	RS 50 point difference: 3.38 (1.32, 8.69)	
						RFS	**O-DX vs CP factors[a]**	RS 50 point difference: 2.09 (0.84, 5.20)	
						OS	**O-DX vs CP factors[a]**	RS 50 point difference: 2.67 (0.93, 7.62)	
LN0, variable ET & CT									
Sun 2011[89] N=57	HAAMMS	100% HR+ 86% HER2- (7.5% unclear)	LN0	75.3% ET 80.6% CT	18-30	**DRFS**	**RS (not Genomic Health) vs CP factors[a]**	RS 1-point difference: 1.03 (1.01, 1.06), p=0.017	
LN+, variable ET & CT									

Reference; N	Cohorts	Population	Nodal status	ET/CT	Cut-off	Outcome	Test or comparator[a]	Multivariable Cox model (adjusted for CP factors[a]): HR (95% CI) unless otherwise stated
Penault-Llorca 2014[91] N=530	PACS01	100% HR+	LN+	100% CT 74.2% ET	**NR**	**NR, assume DRFI**	**O-DX vs CP factors[a]**	p<0.001
Sun 2011[89] N=35	HAAMMS	100% HR+ 86% HER2- (7.5% unclear)	LN+	75.3% ET 80.6% CT	18-30	**DRFS**	**RS (not Genomic Health)**	RS 1-point difference: 1.03 (1.00, 1.07), p=0.039
Wolmark 2016[51] Mamounas 2012[90] N=1065	NSABP-28	100% ER+ HER2 NR Meno NR Female NR	LN+	100% CT & ET	18-30	**DFS, DRFI, OS**	**O-DX vs CP factors[a]**	p<0.001

N, number of patient; ET, endocrine therapy; CT, chemotherapy; pts, patients; Inter, intermediate group; yr, year; HR Hazard ratio; CI, confidence interval; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; RS, Oncotype DX recurrence score.

[a] adjusted for: Goldstein 2008: number positive nodes, tumour size, age, HER2 status, grade; Paik 2004: Cox model 1 adjusted for age and tumour size; Cox model 2 adjusted for age, tumour size, tumour grade, HER2 amplification, amounts of oestrogen and progesterone-receptor protein; Cox model 3 adjusted for age, tumour size, grade; Penault-Llorca 2014: treatment, age, tumor size & grade, number of + nodes, surgery type and Ki67 status; Sun 2011, unclear if all CP factors kept in the analysis: age, tumour size, nodal status, ER, PR, HER2, ET, CT, ST Gallen, RS; Toi 2010: Adjusted for age and clinical tumour size; Wolmark 2016/Mamounas 2012: does not specify which covariates were included for which outcomes, but selected from treatment, age, tumour size, tumour grade, number of + nodes and type of surgery; [b] first Cox model used centrally determined disease grade, second Cox model used locally determined disease grade; [c] reported in Tang 2011a

Table 18: Oncotype DX, additional prognostic value over comparators

Reference; N	Cohorts	Population	Nodal status	ET/CT	Cut-off	Outcome	Test or comparator[a]	Outcomes
LN0/+ 100% ET & CT								
Sestak 2017 (data request)[43] Dowsett 2010[35] [redacted]	TransATAC	100% HR+ 100% HER2- Postmeno	LN+/-	100% ET	**18-30**	**DRFI**		**Increase in likelihood ratio χ^2 over comparator**
							O-DX vs CTS	[redacted]
							O-DX vs NPI	[redacted]
Goldstein 2008;[98] Sparano 2012[99] N=465	E2197 (ECOG trial)	100% HR+ 44% HER2- Pre/post-meno	LN0, 56.5% LN1-3 43.5%	All ET & CT 40% aromatase inhibitor	**18-30**	**RFI**	**O-DX vs Integrator based on AOL[b]**	**5 yr:[98]** **For 50-point difference using central grade:** **RS HR:** 2.51 (95% CI: 1.71. 3.70, p<0.001) **Integrator HR:** 1.51 (95% CI: 1.07, 2.13, p=0.02) **For 50-point difference using local grade:** RS HR: 2.64 (95% CI: 1.80, 3.87; p<0.001) Integrator HR: 1.34 (95% CI: 0.94, 1.91, p=0.11) Interaction term was not significant indicating effect of RS is largely independent of the level of the integrator.
		100% HR+ 100% HER2-						**5 yr:[98] C-index (AUC)** **RS:** 0.69 **Integrator (central grade):** 0.61 **Integrator (local grade):** 0.56
LN0, 100% ET monotherapy								
Paik 2004,[45] Tang 2011a[50] N= 668	NSABP B-14	100% ER+ HER2+/-, % NR	LN0	100% ET	**18-30**	**DRFI**		**Multivariable Cox model (adjusted for CP factors[a]): HR (95% CI)**
							O-DX vs AOL[b]	**Cox model 4 (only AOL and RS):** **AOL:** 1.93 (1.27, 2.91), p=0.002, **RS 50 point difference:** 2.83 (1.91, 4.18), p<0.001
							O-DX vs AOL& CP factors[b]	**Cox model 5 (AOL, RS, age, tumour size, grade):** **AOL:** 0.86 (0.45, 1.62), p=0.636 **RS:** 2.37 (1.58, 3.55), p<0.001
Sestak 2017 (data request)[43] Dowsett 2010[35] [redacted]	TransATAC	100% HR+ 100% HER2- Postmeno	LN0	100% ET	**18-30**	**DRFI**	**O-DX vs CTS**	**Increase in likelihood ratio χ^2 over comparator** [redacted]
							O-DX vs NPI	[redacted]

Reference; N	Cohorts	Population	Nodal status	ET/CT	Cut-off	Outcome	Test or comparator[a]	Outcomes	
LN+									
Sestak 2017 (data request)[43] Dowsett 2010[35] [redacted]	TransATAC	100% HR+ 100% HER2- Postmeno	LN+	100% ET	**18-30**	**DRFI**	**O-DX vs CTS**	**Increase in likelihood ratio χ^2 over comparator**	
								[redacted]	[redacted]
							O-DX vs NPI	[redacted]	[redacted]

N, number of patient; ET, endocrine therapy; CT, chemotherapy; pts, patients; Inter, intermediate group; yr, year; HR Hazard ratio; CI, confidence interval; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node;DRFI, Distant recurrence free interval; RFI, recurrence free interval; Fu, follow-up; RS, Oncotype DX recurrence score; AUC, area under the curve; AOL, Adjuvant! Online; CTS, clinical treatment score; NPI, Nottingham Prognostic Index; O-DX, Oncotype DX

[b]In this analysis the Cox model only included O-DX and an integrator based on AOL, where the integrator was adjusted to 5-year outcomes rather than AOL's 10 year outcomes.

Table 19: Oncotype DX RSPC, discrimination, reclassification and additional prognostic value

Reference; N	Cohorts	Population	Nodal status	ET/CT	Cut-off	Outcome	Test	% pts per group			Discrimination				Reclassification	Additional prognostic value
											% 10yr DRFI (95% CI)			HR, p-value		
								L	I	H	L	I	H			Difference in likelihood ratio χ^2
Node-negative and node-positive																
Tang 2011b[42] **B-14:** n=647 **TransATAC:** n=1088 **B-20:** n=625	NSABP B-14 & Trans-ATAC meta-analysed	ER+ HER2+/-, %NR	LN0 (B-14); LN+/- (Trans-ATAC)	100% ET	12% - 20% risk	DRFI 10yr	RSPC	64	18	18	93.5 (91.5, 95.5)	82.4 (77.1, 87.7)	73.8 (68.4, 79.2)	HR/CI NR, p<0.001 with increasing risk group	RSPC vs RS: DRFI risks not significantly different between RS and RSPC within each risk group (p=0.68, p=0.27 and p=0.42 for low-, inter- and high-risk groups) RS int-risk pts (n=272): 16.9% high-risk RSPC 55.1% low-risk RSPC RS low-risk pts (n=783): 1.9% high-risk RSPC 8.9% inter-risk RSPC RS high-risk pts (n NR): 28.6% inter-risk RSPC Other differences NR	**RSPC vs O-DX RS:** 76.9, p<0.001 **RSPC vs grade, tumour size, age:** 45.4, p<0.001
							RS	54	27	19	94.1 (92.2, 96.0)	86.2 (81.9, 90.5)	70.5 (63.4, 76.5)	HR/CI NR, p<0.001 with increasing risk group		
	NSABP B-20	ER+ HER2+/-, %NR	LN0	100% ET; 64% CT			RSPC							**RSPC:** 2.43, p<0.001 **RS**: 2.22, p<0.001		

DRFI, distant recurrence free interval; N, number of patient; ET, endocrine therapy; CT, chemotherapy; pts, patients; L, low risk group; I, intermediate risk group; H, high risk group; yr, year; HR Hazard ratio; CI, confidence interval; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; RS, Oncotype DX recurrence score; RSPC, algorithm including recurrence score with clinical and pathological factors; NR, not reported;

4.3.3 Chemotherapy benefit: Oncotype DX

Five data sets, reported across eleven published references[42, 49, 50, 68-70, 72, 74, 103-105] and one AIC manuscript[71] (Table 20), have conducted analyses that assess the ability of Oncotype DX to predict benefit of chemotherapy. Chemotherapy benefit relates to the ability of the test to predict which patients will respond to chemotherapy, and can be assessed by considering whether the effect of chemotherapy versus no chemotherapy on patient outcomes differs according to the test score, e.g. by comparing HRs or p-values between risk groups. Formal assessments of chemotherapy benefit include interaction tests which assess whether the difference is statistically significant.

Study designs: Oncotype DX chemotherapy benefit

Two data sets[42, 49, 50, 68] were re-analyses of RCTs, which provide evidence relating to the extent of any interaction between the effect of chemotherapy and Oncotype DX on outcome (i.e. whether the result of the test is able to predict a differential treatment effect).

Albain *et al*. 2010[68] conducted a re-analysis of the Southwest Oncology Group (SWOG)-8814 study, a Phase 3, open-label, parallel-group RCT. Two arms of the trial were reanalysed: the tamoxifen only arm and the tamoxifen plus cyclophosphamide, doxorubicin and fluorouracil (CAF-T) arm.

Paik *et al*. 2006,[49] Tang *et al*. 2011a[50] and Tang *et al.* 2011b[42] re-analysed the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-20 trial in which patients were randomised to tamoxifen alone, or to tamoxifen plus clyclophosphamide, methotrexate and fluorouracil (CMF-T), or to tamoxifen plus methotrexate and fluorouracil (MF-T). It should be noted that some of the patients of the B-20 trial were used to derive the Oncotype DX score.[49] Tang *et al*. 2011b[42] derived the prognostic Oncotype DX RSPC algorithm using the TransATAC and NSABP B-14 data sets, and then tested the ability of the RSPC to predict benefit from chemotherapy in the NSABP B-20 data set.

The remaining three data sets (MD Anderson Center,[69, 70] Clalit Health Services[71, 72] and SEER registry)[74, 103] were retrospective observational studies where patients were treated according to routine practice and their Oncotype DX score.

Patients: Oncotype DX chemotherapy benefit

The RCTs comprised one data set in LN+[68] and one in LN0[42, 49, 50] patients; however, neither data set matched the decision problem exactly in other respects.

The SWOG-8814[68] data set comprised all HR+, LN+ patients with 38.1% having four or more positive lymph nodes. All patients were post-menopausal and 12% were HER2+. A total of 367 (40%) out of the 927 patients recruited to the original trial were included in the analysis, with attrition

due to missing samples, insufficient tissue and test failures. Analyses in this study were adjusted for LN1-3 and ≥4.

The NSABP B-20[42, 49, 50] data set comprised ER+, LN0 patients, with an unreported percentage being HER2+. A total of 651 (28%) out of the 2363 patients recruited to the original trial were included in the analysis, with attrition due to missing clinical variables, missing samples, insufficient tissue, and test failures.

For the derivation of RSPC a further 26 patients were excluded from NSABP B-20 as, for this analysis, ER positivity was scored based on the results of the Oncotype DX ER gene expression rather than the clinicopathological ER score, leaving 625 patients for analysis.[42]

Of the observational studies, one data set originated from the MD Anderson Cancer Centre[69, 70] in the USA (n=1424), and reported a retrospective analysis of HR+, HER2-, LN0 patients in one report[69] and of a subset of these patients with Stage 1 disease in a previous report[70]. The aim of the former was to report the survival in patients treated with and without chemotherapy with an RS of 11 to 25, although several other analyses were also reported, and the latter reported exploratory analyses by tumour size subgroups.

The second data set originated from Clalit Health Services, in Israel (n=627) , and reported an analysis of ER+, HER2- patients, [redacted][71, 105] ([redacted]) and with LN1-3.[72] The aim of the study was to report outcomes in patients who underwent RS testing.

The third data set was a retrospective analysis of the SEER registry in the USA,[74, 103] and aimed to determine BCSS by baseline RS scores and clinical covariates, but also reported a test for the interaction between RS and treatment on BCSS for HR+, HER2-, LN0 patients (n=40,134).

Quality assessment: Oncotype DX chemotherapy benefit

Table 21 presents the quality assessment of the included studies. The two reanalyses of RCTs[42, 49, 50, 68] were at some risk of bias, largely because of patient spectrum bias, where those individuals excluded because of insufficient tissue may be systematically different to the included patients and no attempt was made to account for missing data. Other sources of bias arising from the analysis of the data include not accounting for stratification factors used in the randomisation of patients to treatment, excluding potentially relevant prognostic variables and treatment effect modifiers, and not considering higher order and non-linear terms in the Cox regression. Blinding of test assessors to clinical outcomes was only conducted in Albain *et al*. 2010.[68]

The three observational studies[69-72, 74, 103, 105] are limited by their non-randomised design, whereby patients who received chemotherapy are likely to be systematically different in terms of known (and potentially unknown) prognostic variables (e.g. age) and treatment effect modifiers to those who did not, leading to a high risk of confounding. They also only recruited patients for whom an Oncotype DX test had been ordered and it is unclear how this may have affected the patient spectrum and generalisability to the decision problem. Two studies did, however, blind the test assessors to the long-term outcomes.[69-72]

Results: Oncotype DX chemotherapy benefit

Re-analysis of RCTs

Table 22 presents data from RCTs relating to the ability of Oncotype DX to predict benefit from chemotherapy.

DFS: Albain *et al.* 2010 reported 5 and 10 year DFS and Tang *et al*. 2011a report 10 year DFS.[42, 49, 50, 68] 10-year HRs for the effect of chemotherapy compared with no chemotherapy showed a progressively greater effect on DFS when moving from low-risk to high-risk Oncotype DX categories in both studies (see Table 22) but only the high-risk group in Tang *et al*. 2011a (HR 0.41 (95% CI: 0.23, 0.71); unadjusted)[50] and Albain *et al*. 2010 (HR 0·59 (95% CI: 0·35, 1·01); log-rank p-value=0.033; adjusted for the number of positive nodes) were statistically significant.[68] Formally, the test for the interaction between treatment and RS risk group test was not statistically significant in Tang *et al*. 2011a (p=0.082).[50] Albain *et al.* 2010 assessed the effect of RS on the continuous scale and its interaction with treatment adjusted for the number of positive nodes and found the interaction to be borderline statistically non-significant (p=0.053).[68] However, Albain *et al*. 2010[68] also found that the effect of recurrence score on treatment varied over time and that recurrence score is a treatment effect modifier in the first 5 years (interaction p-value=0.029) but not after 5 years (interaction p-value=0.580). Within the first 5 years, they performed an additional Cox regression adjusting for age, ethnic origin, tumour size, progesterone status, grade, P53 and HER2, treatment, continuous recurrence score and the interaction between continuous recurrence score and treatment, and found that the interaction remained statistically significant (p-value not presented). Notably, this analysis omitted ER status, and a futher analysis with adjustment for ER status only (by Allred-scoring) was not statistically significant (p=0.15).

DRFI: This was the primary outcome in Tang *et al* 2011a,[50] but was not reported by Albain *et al*. 2010,[68] where an exploratory analysis of BCSS was presented instead. For DRFI in Tang *et al*. 2011a,[50] HRs for no chemotherapy compared with chemotherapy showed a similar trend as DFS, with the Oncotype DX high-risk category showing a statistically significant effect of chemotherapy (HR 0.26 (95% CI: 0.13, 0.53); unadjusted); the test of the interaction between treatment and recurrence

score was also statistically significant (p=0.031).[50] Paik *et al*. 2006[49] performed several Cox regressions adjusting for age, tumour size, ER, PR, tumour grade, recurrence score as a continuous variable, treatment and the interaction between treatment and recurrence score (interaction p-values 0.035 to 0.068); thus, there is weak evidence for an interaction between treatment and continuous recurrence score.

Tang *et al.* 2011a[50] also reported the effect of chemotherapy by AOL risk groups in patients with RS scores and reported a test for the interaction between treatment and AOL risk group (p=0.99), indicating that it was unable to predict the benefit of chemotherapy; HRs were low-risk 0.58 (95% CI: 0.23, 1.42); intermediate-risk 0.54 (95% CI: 0.20, 1.46); high-risk 0.53 (95% CI: 0.25, 1.1)). In an additional analysis of 1952 patients from B-20 with tumour grade, the test for the interaction between treatment and AOL risk group was statistically non-significant (p=0.219). However, although the effects of treatment were similar in patients at intermediate- and high-risk by AOL, there was evidence of no effect of treatment in patients at low-risk; HRs low-risk 0.92 (95% CI: 0.53, 1.62); intermediate-risk 0.52 (95% CI: 0.29, 0.93); high-risk 0.53 (95% CI: 0.36, 0.77).

Breast Cancer Specific Survival: BCSS also showed a statistically significant effect in the high-risk group in Albain *et al.* 2010 (p=0.033; adjusted for the number of positive nodes), although no interaction test was reported and data was not reported for intermediate and low risk patients.[68]

Overall survival: HRs were reported for both data sets for chemotherapy compared with no chemotherapy in low-, intermediate- and high-risk groups (see Table 22). HRs showed the greatest effect of chemotherapy in the high-risk groups; the HR was statistically significant in Tang *et al*. (HR 0.31 (95% CI: 0.16, 0.60); unadjusted) [50] and borderline statistically significant in Albain *et al.* 2010 (HR 0·56 (95% CI: 0·31, 1·02), p=0·057; adjusted for the number of positive nodes).[68] In Tang *et al*,[50] the test for the interaction between treatment and recurrence score (i.e. low-, intermediate- and high-risk) was statistically significant (p=0.011). Albain *et al*. 2010 assessed the effect of RS on the continuous scale and its interaction with treatment adjusted for the number of positive nodes and found the interaction with treatment statistically significant over 10 years (p=0.026) and within the first 5 years (p=0.016).

Tang 2011a[50] also reported the effect of chemotherapy by AOL risk groups in patients with RS scores and reported a test for the interaction between treatment and AOL risk group (p=0.311). In an additional analysis of 1952 patients from B-20 with tumour grade, the test for the interaction between treatment and AOL risk group was significant (p=0.009); HRs low-risk 1.26 (95% CI: 0.81, 1.95); intermediate-risk 0.53 (95% CI: 0.31, 0.9); high-risk 0.57 (95% CI: 0.40< 0.82).

Whilst the results from Tang *et al*. 2011a suggest that Oncotype DX is better at identifying individuals who would benefit from chemotherapy than AOL, the authors did not provide a formal comparison of the performance of the models and the relative benefit of Oncotype DX over AOL remains unclear.

Cut-off below which chemotherapy has no benefit: Albain *et al.* 2010 suggested that within the first 5 years, the effect of chemotherapy on DFS was clinically equivalent to the effect of no chemotherapy for recurrence scores up to about 20 but that chemotherapy performed better at higher scores. Paik *et al*. 2006[49] explored the effect of treatment, Oncotype DX score as a continuous variable and their interaction on distant recurrence but were unable to estimate the cut-off below which there was no benefit from chemotherapy as chemotherapy provided a benefit at all risk scores.

Observational studies

Data relating to the ability of Oncotype DX to predict benefit from chemotherapy from observational studies is presented in Table 23. These studies are at high risk from confounding.

DRFS, IDFS, RFS and BCSS: The MD Anderson study reported DRFS[69, 70] (using Cox regression by risk group adjusted for treatment, age at diagnosis, tumour size, grade, histologic subtype, Ki-67 expression, LVI, type of surgery and endocrine therapy at both the 18-30 RS cut-off and the 11-25 RS cut off). The Clalit Health study reported DR (using Cox regression adjusted for treatment, RS risk group, age, tumour size and histologic grade, although statistically non-significant covariates were excluded from the final model).[71, 72, 105] Neither of the authors included terms in their models to assess the interaction between the effect of chemotherapy and Oncotype DX risk group.

The Clalit Health Services study reported a subgroup of patients with one micro metastasis up to 3 lymph node metastases (LN1micro to LN3),[72] and a subgroup of patients [redacted] (n=270/627 and [redacted] had micro metastases, respectively). In the LN1micro to LN3 group, rates of DR and BC death for chemotherapy-treated and untreated patients were reported as exploratory analyses in patients with Oncotype DX RS scores 18-30 and scores 11-25 only. Statistical tests were not conducted, but for both endpoints, those treated with chemotherapy had more favourable results compared with those not treated with chemotherapy, and this was more evident in the subgroup of patients with Oncotype DX RS scores 18-30 (DRFS 97.8% compared with 90.4%; BCSS 98.9% compared with 96.3%, respectively) than in the group with score 11-25 (DR 97.3% compared with 95.9%; BC death 100% compared with 98.8%). [redacted]

[redacted]

The MD Anderson study[69, 70] presented Kaplan-Meier survival functions by risk group and 5-year DRFS, IDFS, RFS and OS rates for LN0 patients only. At both RS cut offs, event rates were too few in the low-risk categories to allow an analysis. Kaplan Meier survival functions indicated no difference between chemotherapy and no chemotherapy for any outcome and unadjusted log-rank tests were not statistically significant. The observed event rates were similar or worse in chemotherapy treated patients in the intermediate RS category (11-25). Analyses using the 11-25 RS cut off reported HRs>1 for the effect of chemotherapy in the intermediate-risk group, and HRs<1 for the effect of chemotherapy in the high-risk group, across all outcomes, although p-values were not statistically significant. Analyses using the 18-30 RS cut-off reported HRs <1 in all risk categories (except the RS <18 risk group, where the HR was 1.09 (95% CI: 0.14, 8.62, p=0.938), though HRs were closer to 1 in the intermediate-risk groups than in the high-risk groups. P-values were non-significant and no tests for the interactions between treatment and RS were reported. Results are presented in Table 23.

A further analysis, unadjusted for potential prognostic variables and treatment effect modifiers, was conducted which split the Stage 1 disease patients in the intermediate-risk group (RS 18-30) by tumour size, and found the effect of chemotherapy versus no chemotherapy (HR not reported) was statistically significant in the pT1c (tumour size >10mm, log rank test p=0.02) patients, but not in pT1b (tumour size >5mm, ≤10mm, log-rank test p=0.752) patients. However, the direction of effect was not clear because of conflicting statements within the published report.[70]

The SEER registry study[74, 103] used Cox regression adjusted for treatment, age, tumour size, and recurrence score risk group with and without terms for the interaction between treatment and recurrence score risk group. They found that the association between RS and BCSS remained prognostic, but was attenuated for those with chemotherapy compared to those reported as having no chemotherapy or unknown treatment (interaction p=0.03). They also fitted recurrence score as a continuous variable, although no details were provided of the extent of the interaction with treatment.

One further study (Sparano 2012; ECOG trial E2197)[99] noted that their data were consistent with previous reports indicating greater chemotherapy treatment effect for high RS (RS>20), based on the levelling off of a plot (see source paper),[99] but offered no formal analysis.

Results: RSPC

RSPC was derived in the TransATAC and NSABP B-14 data sets[42] and is based on the Oncotype DX score with the addition of clinicopathological variables (namely RS using a natural cubic spline with 2 degrees-of-freedom with knots at 5, 18 and 50; age; tumour size and grade; nodal status; and hormonal treatment) formally incorporated. Data are available only in LN0 patients. The prognostic ability of RSPC is reported in Section 4.3.2. In the same publication,[42] the NSABP B-20 data set was used to assess the scores' abilities to predict chemotherapy benefit based on 625 (26%) of 2,362 randomised individuals who had available tumour blocks, Oncotype DX ER expression ≥6.5 and complete information on tumour grade and size, and age. Whilst there was a weak statistically significant interaction between treatment effect and Oncotype DX RS risk score (p=0.037) with a standardised HR of 0.66 (95% CI: 0.44, 0.97), there was insufficient evidence of an interaction between treatment and RSPC risk score (p=0.10) with a standardised HR of 0.65 (95% CI: 0.39, 1.09) (data not tabulated).

Discussion: Oncotype DX and RSPC chemotherapy benefit

Analyses relating to the ability of Oncotype DX to predict benefit from chemotherapy were reported in five studies. [49, 50, 68, 69] Two were re-analyses of RCTs (total N=1018, NSABP B-20 study in LN0,[49, 50] and SWOG-8814 study in LN+ patients[49, 50, 68]) where patients were randomised to endocrine monotherapy, or endocrine therapy plus chemotherapy. Three were observational studies [69-74] (total N~44,000 with some double counting, two LN0,[69, 70, 73, 74] one mixed LN+/-[71, 72]) where patients were treated according to usual practice and their RS.

From the re-analyses of RCTs, based on the HRs for chemotherapy vs. no chemotherapy between Oncotype DX risk categories, the greatest benefit of chemotherapy appears to be for patients in the high-risk category and the HRs appear to be greater in the LN0 population (for high-risk patients, HRs for DFS for chemotherapy versus no chemotherapy were 0.41 (95% CI: 0.23, 0.71) in LN0 patients[49] and 0.59 (95% CI: 0.35, 1.01)[68] in LN+). Unadjusted interaction tests were statistically significant for 10 year DRFI and OS in NSABP B-20 (LN0) (p=0.031 and p=0.011 respectively),[49, 50] and in SWOG-8814 (LN+) for 5 year DFS and OS (p=0.029 and p=0.016 respectively),[68] whereas interaction tests for 10 year DFS (NSABP B-20, p=0.082)[49, 50] and 5-10 year DFS and OS (SWOG-8814, p=0.58 and p=0.87 respectively)[68] were not statistically significant. Adjusted interaction tests were not always statistically significant. Adjustments for age, tumour size, ER, PR and tumour grade gave a range of p=0.035 to 0.068 in NSABP B-20 (LN0),[49, 50] whilst analyses adjusted for age, ethnic origin, tumour

size, progesterone status, grade, P53, and HER2 remained statistically significant (p-values not reported), whilst an adjustment for ER status resulted in a non-significant interaction (p=0.15) in SWOG-8814 (LN+).[68] The Oncotype DX cut-off below which chemotherapy could be avoided was reported to be approximately 20 in SWOG-8814,[68] but NSABP B-20 authors could not determine a cut-off as there was no point below which chemotherapy did not confer an advantage.[49, 50]

Overall the evidence for the prediction of chemotherapy benefit by Oncotype DX from the reanalyses of RCTs was weak since some interaction tests were not statistically significant, possibly due to insufficient events, and could be spurious as a consequence of omitting potentially important covariates from the statistical models. It was not clear whether all relevant clinicopathological variables were included in a single model for either study (e.g. ER status was omitted from the adjusted analyses in SWOG-8814;[68] analyses in NSABP B-20 appeared to only include each covariate separately),[49, 50]or whether all stratification factors used in randomising patients to treatment were included as well. Categorising the continuous Oncotype RS score into risk groups may lead to loss of information and has the potential to create spurious interactions between RS and chemotherapy benefit due to imbalances in clinicopathological variables between risk groups, especially if these are not adjusted for. Authors rarely provided information on model comparison or considered inclusion of non-linear or higher order covariates. Other potential biases in the reanalyses of RCTs included attrition of samples; exclusion of patients due to missing data for covariates; and inclusion of HER2+ patients (who are out of scope for this assessment).

From the three observational cohort studies,[69-74, 105] evidence was mixed and at high risk from confounding, since patients who received chemotherapy were likely to be at higher risk than patients who did not. Only one study reported an interaction test, and this was statistically significant (p=0.03), but only adjusted for grade, tumour size, age and race (omitting ER and PR).[73, 74] The other two studies only reported HRs for chemotherapy versus no chemotherapy in intermediate- (MD Anderson and Clalit Health studies)[69-72, 105] and high risk patients (MD Anderson),[69, 70] and these were statistically non-significant, even after adjustment for confounders in one study. [69, 70]

RSPC was derived in TransATAC and NSABP B-14,[42] and validated in NSABP B-20.[42] An interaction test was non-significant (p=0.10),[42] suggesting that the interaction between treatment effect and RS risk group may be confounded by clinicopathological variables.

In practice, it is unlikely that chemotherapy decisions would be made on Oncotype DX scores independent of clinicopathological variables. Evidence relating to the ability of the test to predict chemotherapy benefit over and above routinely collected clinicopathological variables was provided in both RCT data sets in the adjusted interaction tests. [49, 50, 68] Interestingly, Tang *et al*. 2011a[50] tested

the ability of AOL to predict benefit from chemotherapy in a large cohort of 1952 patients, and found it to have predictive ability for OS. However, the inclusion of clinicopathological variables alongside RS in the RSPC algorithm resulted in a loss of predictive ability (p=0.10), suggesting that the interaction between treatment effect and RS risk group may be spurious and explainable by confounding from clinicopathological variables.[50]

Conclusion: Oncotype DX and RSPC chemotherapy benefit

In conclusion, there is some evidence from two reanalyses of RCTs to suggest that Oncotype DX may predict benefit from chemotherapy, and that benefit from chemotherapy is highest in Oncotype DX high-risk patients. Unadjusted interaction tests between Oncotype DX risk group and chemotherapy benefit were mainly statistically significant. However, the evidence to support Oncotype DX's ability to predict benefit from chemotherapy is weak, possibly due to insufficient events, and interaction tests adjusted for clinicopathological variables were often non-significant. Also, the RSPC algorithm (Oncotype plus age, tumour size and grade) showed a non-significant interaction test between chemotherapy benefit and RSPC risk group. Three observational cohort studies were at high risk of confounding; one reported a statistically significant interaction test but this was only adjusted for limited factors. If predictive ability were assumed, it is unclear below which exact cut-off patients could avoid chemotherapy (though one study suggests this is RS 20), as chemotherapy benefit is uncertain in the intermediate-risk group. Whilst TAILOR-X will address the issue of whether low and intermediate patients can avoid chemotherapy, it is unclear to what extent it will address the question of whether the test can predict chemotherapy benefit.

Table 20: Study and patient characteristics: Oncotype DX and RSPC for chemotherapy benefit

Reference; N	Cohorts	Country	Study design	Details of test	Cut-offs	Population	Nodal status	Endo / chemo
Reanalysis of RCT – Oncotype DX								
Albain 2010[68] N=367	SWOG-8814	USA	Reanalysis of prospective trial (RCT); archive tissue	FFPE Genomic Health	18-30	100% HR+ 12% HER2+ Postmenopausal 100% Female	LN+, 100% LN>3, 38%	100% ET monothearpy
Paik 2006[49] Tang 2011a[50] N= 651	NSABP B-20	USA	Reanalysis of prospective trial (RCT); archive tissue	FFPE Genomic Health	18-30	100% ER+ % NR HER2+/- Meno NR Female 100%	LN0	1) 100% ET monotherapy (N=227) 2) 100% ET + 100% CT (N=424)
Observational studies – Oncotype DX								
Barcenas 2017[69] Le Du 2015[69, 70] N=1424	MD Anderson Centre	USA	Retrospective cohort study	NR	11-25	100% HR 100% HER2- 67% postmeno 99% female Had O-DX test	LN0	91% ET 22% CT Treated according to usual practice with O-DX test
Stemmer 2017[71] Stemmer 2016[72] Stemmer 2016[105] [redacted] 2) N=627	Clalit Health Services	Israel	Retrospective cohort study	NR	[redacted]18-30	100% ER+ 100% HER2- Meno NR Had O-DX test	1) [redacted] 2)LNmic-LN3	Treated according to usual practice with O-DX test [redacted] 2) %ET NR 27% CT
Petkov 2016[73] Roberts 2016[74] Roberts2017[104]N=40,134	SEER registry	USA	Retrospective cohort study	NR Genomic health	18-30	100% HR+ 100% HER2- 40-85 years old Unclear if only those with O-DX test	LN0	ET NR CT 23% Treated according to usual practice with O-DX test
Reanalysis of RCT – RSPC								
Tang 2011b[42] **B-20:** n=625	NSABP B-20	USA	Reanalysis of prospective trials (RCT); archive tissue	FFPE Genomic Health	**RSPC:** 12% - 20%	100% ER+ HER2+/-, % NR	**B-20:** LN0	**B-20:** 36% ET; 64% CT&ET

CT, chemotherapy; ET, endocrine therapy; pts, patients; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; RS or O-DX, Oncotype DX recurrence score; FFPE, formalin fixed paraffin embedded; postmeno, postmenopausal; Meno, menopausal status; RSPC, recurrence score- clinical-pathological score

Table 21: Quality assessment of studies reporting the ability of Oncotype DX and RSPC to predict chemotherapy responsiveness

Author, Year	Cohort name	Derivation or validation?	Study design appropriate?	All eligible patients included?	Blinding (of test assessors to outcomes)	Definition of outcome standardised or a priori?	Applicability: Patient Spectrum	Applicability : Test as per decision problem?
O-DX								
Albain 2010[68]	SWOG-8814	V	Y, R-RCT	N InT, TF	Y	Y	N: >20% >LN3+[a]	Y
Paik 2006[49] Tang 2011a[50]	NSABP B-20	V **	Y, R-RCT	N InT, <5% cancer cells, MS	UC	Y	UC, % HER2+ NR	Y
Barcenas 2017[69] Le Du 2015[69, 70]	MD Anderson Cancer Centre	V	N, not RCT	N, SFT	Y	Y	No, SFT	Y
Stemmer 2017[71] Stemmer 2016[72] Stemmer 2016[105]	Clalit Health Services[71, 72]	V	N, not RCT	N, SFT	Y	Y	No, SFT	Y
Petkov 2016[73] Roberts 2016[74] Roberts2017[104]	SEER registry[74, 103]	V	N, not RCT	N, SFT	N	Y	No, SFT	Y
O-DX RSPC								
Tang 2011b[42]	NSABP B-20 cohort	D & V of RSPC[b]	Y, R-RCT	N Pts ER+ by RS only; MS	UC	Y	Unclear - % HER2+ NR	Y

Y, Yes; N, No; UC, unclear; R-RCT, Reanalysis of RCT; InT, insufficient tissue; TF, test failure; MS, missing samples; D, Development; V, validation; SFT, only those sent for test included;
[a] Most/all analyses adjusted for number of positive nodes (1 to 3 and 4 or more); [b]used some of O-DX derivation sample

Table 22: The prediction of chemotherapy responsiveness by Oncotype DX – Reanalyses of RCT data

Author, year	Cohort	Outcome	Prediction of chemotherapy benefit: HR for no CT vs CT (95% CI), unless stated otherwise			Interaction tests and other data	Adjusted interaction tests
			Low: RS <18	**Intermediate: RS 18-30**	**High: RS >30**		
Paik 2006[49, 50] Tang 2011a[36]	NSABP B-20 (USA) ER+, LN0, HER2+/- N=651 (227 ET; 424 ET+CT)	**DFS 10 yr**	0.91 (0.57, 1.45)	0.79 (0.43, 1.47)	**0.41 (0.23, 0.71)**	p=0.082	Interaction test for AOL p=0.357; In a larger cohort (n=1952) AOL p=0.099
		DRFI 10 yr	1.31 (0.46, 3.78), p=0.61 % DRF (95% CI) TM 96.8 (93.7, 99.9) TM+C 95.6 (92.7, 98.6) Difference in DRFI at 10 yrs -1.1%	0.61 (0.24, 1.59), p=0.39 % DRF (95% CI) TM 90.9 (82.5, 99.4) TM+C 89.1 (82.4, 95.9) Difference in DRFI at 10 yrs -1.8%	**0.26 (0.13, 0.53), p<0.001** % DRF (95% CI) TM 60.5 (46.2, 74.8) TM+C 88.1 (82.0, 94.2) Difference in DRFI at 10 yrs 27.6%	Likelihood ratio test for interaction (categorical RS) **p=0.031**[a] RS as continuous function: CT benefit increased as RS increased; clear cut-off below which there is no benefit could not be accurately defined.	Interaction tests for O-DX RS adjusted for age, tumour size, ER, PR and tumour grade in different models giving range of p=0.035 to 0.068. Interaction test for AOL p=0.99; In a larger cohort (n=1952) AOL p=0.219
		OS 10 yr	1.37 (0.63, 3.01)	0.94 (0.4, 2.25)	**0.31 (0.16, 0.60)**	**p=0.011**	Interaction test for AOL, p=0.311; In a larger cohort (n=1952) AOL **p=0.009**
Albain 2010[68]	SWOG-8814 (USA) HR+, LN+, HER2+/- N=367	**DFS 10 yr**	1·02 (0·54, 1·93)[b] SLR p=0·97	0·72 (0·39, 1·31)[b] SLR p=0·48	0·59 (0·35, 1·01)[b] SLR **p=0·033**[c]	Interaction (linear RS)[b] All years: p=0.053 0-5 years: **p=0.029**[b] 5-10 years: p=0.58[b]	Interaction test for O-DX RS significant (p NR) after adjustment for age, ethnic origin, tumour size, progesterone status, grade, P53, and HER2 by TAB250. Interaction for O-DX RS non-significant after adjustment for Allred-scored ER status (p=0·15).
			Whole sample p=0·054[b]				
		DFS 0-5 yr	1·34 (0·47, 3·82)[b]	0·95 (0·43, 2·14)[b]	0·59 (0·32, 1·11)[b]		
		DFS 5-10 yr	0·88 (0·38, 1·92)[b]	0·52 (0·21, 1·27)[b]	0·60 (0·22, 1·62)[b]		
		BCSS	NR	NR	TM 54% TM+C 73% **p=0.033**[c]		

		OS **10 yr**	1·18 (0·55,2·54, p=0·68)[b] SLR p=0.63	0·84 (0·40, 1·78, p=0·65) [b] SLR p=0.85	0·56 (0·31, 1·02, p=0·057) [b] SLR **p=0.027**	Interaction (linear RS) [b] All yrs: **p=0.026** 0-5 yrs: p=0.016 5-10 yrs: p=0.87	

CT, chemotherapy; DFS, disease free survival; OS, overall survival; DRFI, distant recurrence free interval; CI, confidence interval; HR, Hazard Ratio; CT, chemotherapy; TM, tamoxifen monotherapy; TM+C, Tamoxifen plus chemotherapy; SLR, stratified log rank test; AOL, Adjuvant Online; O-DX RS or RS; Oncotype DX recurrence score

[a] reported in Paik *et al*. 2006[49] as p=0.038; reported in Tang *et al*. 2011b[42] as p=0.037 for the standardised HR;

[b] Adjusted for number of positive nodes (1 to 3 and 4 or more); [c] Unclear why the HR 95% CI does not indicate statistical significance but the p-value does, possibly due to use of log-rank test.

Table 23: The prediction of chemotherapy responsiveness by Oncotype DX – Observational studies

Author, year	Study	Outcome	Cut-off	Prediction of chemotherapy benefit: CT vs no CT			Additional predictive value Adjusted HR, CT vs no CT(95% CI)		
				Low RS	Intermediate RS	High RS	Low RS	Intermediate RS	High RS
Barcenas 2017[69] Le Du 2015[70] Median FU 58 months HR+, HER2-, LN0, Stage I-II, had O-DX test All risk groups, all years N =1424 Diagnosed 2005 to 2011 and included in K-M analysis: Intermediate RS: N=547 High RS: N=142	MD Anderson Centre USA	**5 yr DRFS**	11-25	NR	**Actuarial 5-yr rate (% (95% CI)):** CT, 96 (87, 99); No CT, 96 (94, 98) **HR:** NR, p=0.97	**HR:** NR, p=0.74	NR	1.25 (0.32, 4.92, p=0.746) [a]	0.67 (0.16, 2.78, p=0.584) [a]
			18-30	NR	NR	NR	NR, too few events	0.80 (0.23, 2.71, p=0.716)	0.32 (0.07, 1.47, p=0.143)
		5 yr DRFS[b]	18-30	NR	**Stage 1 disease, Intermediate-risk (RS 18-30) only** [b] (HRs NR) pT1a (n=13), NR pT1b (n=95) p=0.752 pT1c (n=246) **p=0.020**	NR	NR	NR	NR
		5 yr IDFS	11-25	NR	**Actuarial 5-yr rate (% (95% CI)):** CT, 89 (80, 94); No CT, 93 (90, 95) **HR** NR, p=0.35	**HR:** NR, p=0.56	NR	1.64 (0.73, 3.71, p=0.233) [a]	0.67 (0.21, 2.07, p=0.483) [a]
			18-30	NR	NR	NR	1.09 (0.14, 8.62, p=0.938)	0.78 (0.34, 1.80, p=0.571)	0.50 (0.13, 2.02, p=0.334)
		5 yr RFS	11-25	NR	**Actuarial 5-yr rate (% (95% CI)):** CT, 95 (86, 98); No CT,96 (94, 98) **HR**: NR, p=0.75	**HR:** NR, p=0.94		1.46 (0.41, 5.23, p=0.564) [a]	0.78 (0.17, 3.52, p=0.748 [a]
			18-30	NR	NR	NR	NR, too few events	0.98 (0.32, 3.06, p=0.975)	NR
		5 yr OS[c]	11-25	NR	**Actuarial 5-yr rate (% (95% CI)):** CT, 98 (91, 99); No CT, 98 (96, 99) **HR**: NR, p=0.91	**HR:** NR, p=0.18	NR	2.19 (0.44, 11.0, p=0.340) [a]	0.28 (0.04, 2.05, p=0.209) [a]
			18-30	NR	NR	NR	NR, too few events	0.86 (0.15, 4.91, p=0.861)	0.13 (0.01, 1.30, p=0.082)
Stemmer 2017[71] Stemmer 2016[72] Stemmer 2016[105]	Clalit Health Services[71, 72] Israel	**5 yr DRFI[d]**	18-30	NR	**LN1micro-LN3: % DRF** CT (40%): 97.8%; No CT (60%): 90.4%	NR	NR		
			11-25	NR	**LN1micro-LN3: % DRF** CT (18%): 97.3%; No CT (82%): 95.9%	NR	NR		

Author, year	Study	Outcome	Cut-off	Prediction of chemotherapy benefit: CT vs no CT			Additional predictive value Adjusted HR, CT vs no CT(95% CI)		
				Low RS	Intermediate RS	High RS	Low RS	Intermediate RS	High RS
Median follow-up 6 years ER+, HER2-, had O-DX test [redacted] LN1micro-LN3, N=627[e]			[redacted]	[redacted]	[redacted]	[redacted]	NR		
			[redacted]	[redacted]	[redacted]	[redacted]	NR		
					[redacted]		NR		
		5 year BCSS[d]	18-30		**LN1micro-LN3** CT (40%): 98.9%; No CT (60%): 96.3%		NR		
			11-25		**LN1micro-LN3** CT (18%): 100%; No CT (82%): 98.8%		NR		
Petkov 2016[73] Roberts 2016[74] Roberts2017[104] FU 38 months HR+, HER2-, LN0[f]	SEER registry USA	**Acutarial 5 year BCSS[c]**	18-30	NR			Multivariable model[g] including chemotherapy treatment: RS remained significantly prognostic for 5-year BCSS for both chemo-treated and untreated (or unknown) patients, but strength of association between RS categories and BCSS attenuated for those with chemotherapy reported as "yes" (p=0.03 for covariate-adjusted interaction), versus those reported as "no/unknown". Similar analysis with RS as continuous variable also significant both with and		

Author, year	Study	Outcome	Cut-off	Prediction of chemotherapy benefit: CT vs no CT			Additional predictive value Adjusted HR, CT vs no CT(95% CI)		
				Low RS	Intermediate RS	High RS	Low RS	Intermediate RS	High RS
N=40,134							without adjustment for covariates (p<0.001 for both)		

N, number of patient; CT, chemotherapy; yr, year; HR Hazard ratio; CI, confidence interval; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; RS, Oncotype DX recurrence score; K-M, Kaplan Meier; DRFS, Distant recurrence free survival; IDFS, invasive disease free survival; RFS, relapse free survival; OS, overall survival; DR, distant recurrence; BCSS, breast cancer specific survival

a Adjusted for age at diagnosis, tumour size, grade, histologic subtype, LVI, type of surgery, and endocrine therapy. Covariates producing unstable estimates were removed. Ki-67 was removed due to too many missing values; b Data from Le Du 2015,97 where only Stage I disease patients were included. 17 intermediate patients also in TAILORx study; c OS & BCSS data does not meet our inclusion criteria as follow-up was <5 years; d Converted to DRFI from DR; converted to BCSS from BCSM; e Note overlap between LN0-1micro and LN1micro-LN3 analyses; f HR+ by O-DX and by IHC; HER2 status by O-DX; g adjusted for grade, tumour size, age, race.

4.3.4 Clinical Utility: Oncotype DX

In this review, clinical utility relates to the impact of the prospective use of the test on patient outcomes such as survival and recurrence. The ideal study design would be an RCT where patients are randomised to treatment guided by the test or treatment according to usual practice. Additional study designs for clinical utility are observational cohorts (either prospective or retrospective) where patients received the test prospectively in clinical practice, and data are available for both the test results and clinical outcomes. These observational designs are at higher risk of bias from confounding.

Five data sets reported across nine published references[70, 72, 104-110] and one AIC manuscript[71] reported evidence relating to the clinical utility of Oncotype DX and met the inclusion criteria for the review. One further study[74, 103, 104] did not meet the inclusion criteria for the review in that the follow up was less than 5 years (for outcome BCSS). We have presented data relating to this study as it was the only identified study presenting subgroup analyses for micrometastases and by race, both of which were subgroups specified in the NICE scope[22] and for which there are very limited data.

Study design and chemotherapy rates: Oncotype DX clinical utility

Study characteristics are presented in Table 24. Two studies had a prospective trial design.[106-109] Only one study, the Trial Assigning Individualized Options for Treatment (TAILORx),[**106**] randomises patients to treatment guided by the test or treatment according to usual practice. This study aims to assess the clinical utility of Oncotype DX. Women with RS<11 were assigned to endocrine therapy alone, while women with RS 11-25 were randomised to either endocrine therapy plus chemotherapy or endocrine therapy alone. As of July 2017, this study had only reported results for the low-risk (RS<11) group (n=1626). Data for this group are effectively prospective observational data.

The West German Study Group Plan B (WSG Plan B)[107-109, 111] trial (n=3198) is also a prospective RCT, but does not aim to assess the clinical utility of Oncotype DX, as it randomises patients with RS≥12 to two different sorts of chemotherapy. However, a translational research aim was to assess the risk of recurrence in patients with RS <12 who were not treated with adjuvant chemotherapy. This group is again effectively a prospective observational cohort.

Three studies had an observational design and were retrospective analyses of routinely collected data at three centres or areas: MD Anderson Cancer Centre in the USA (n=1030),[70] Clalit Health Services[71, 72, 105] in Israel (n=2010 LNmic-LN3; n=627 LN0-LNmic), and the Memorial Sloan Kettering Centre in the USA (n=1406).[110] In all cases, treatment was given according to routine clinical practice, including the Oncotype DX RS, which resulted in differing levels of chemotherapy being prescribed per risk group and per study. Chemotherapy ranged from 1%[105] to 12%[110] in low RS

groups (RS <18), from [redacted][71] to 43%[70] in the intermediate-risk group (RS 18-30) and [redacted] to 90%[70] in the high-risk group.

The study that did not meet the inclusion criteria (due to insufficient follow-up length) was of a similar design to the other retrospective analyses, and was based on the prospectively maintained SEER (Surveillance, Epidemiology, and End Results) database and Genomic Health's clinical laboratory database.[74, 103] Chemotherapy rates for low (RS <18), intermediate (RS 18-30) and high (RS>30) risk patients were 7%, 34% and 69% in lymph node negative patients, respectively, and somewhat higher at 23%, 47% and 75% in lymph node positive patients, respectively.

Patients: Oncotype DX clinical utility

Prospective trials: Both trials[106-109] recruited HR+, HER2- patients, but TAILORx recruited LN0 patients with tumours sized 1.1 to 5cm (or 0.6 to 1.0cm in intermediate or high-risk tumours), whilst WSG Plan B recruited clinically high-risk (pT1-T4c; LN+ (or LN0 with a risk factor (CpT2, grade 2/3, high uPA/PAI-1, <35 years, or HR-negative))) patients with 0 to three positive lymph nodes.

Observational studies: All three data sets[70-72, 105, 110] recruited ER+, HER2- patients and only recruited patients who had had an Oncotype DX test. It was not always clear how (or even whether) patients were selected for the test, and how this may have affected the patient spectrum. The MD Anderson study recruited only Stage 1 patients,[70] the Memorial Sloan Kettering study recruited Stage 1 and 2 patients[110] and the Clalit Health Services study did not restrict by stage of disease.[72] The MD Anderson and Memorial Sloan Kettering studies recruited only patients with no or micro lymph node metastases (LN0-LNmic).[70, 110] The Clalit Health Services reported two subgroups across two publications:[72, 105] patients with LN0-LNmic[105] and patients with micro metastases or between one and three lymph node metastases (LNmic – LN3).[72, 105] [redacted]

The study that did not meet the inclusion criteria (SEER database)[74, 103] recruited patients with LN0 to LN3, and subgrouped patients according to age (40-85 years), lymph node status (LN0, LNmic-LN3, LNmic alone) and race (black, white, other).

Quality assessment: Oncotype DX clinical utility

The highest level of evidence for clinical utility is an RCT of treatment guided by the test versus treatment guided according to usual practice. Assessment with the Cochrane risk of bias tool for RCTs indicates all studies are of poor quality to meet this aim (Table 25).

Results: Oncotype DX clinical utility

Data relating to the clinical utility of Oncotype DX are presented in Table 26. Whilst all studies report data relating to recurrence or survival, differences in cut off points (RS<11, <12 and <18), patient populations (clinically high-risk, LN0, LN+), treatment regimens (some patients had chemotherapy in some studies) and outcome measures (DRFS, DFS, DRFI, BCSS, OS) precluded a meaningful meta-analysis.

Whilst two studies use RCT datasets, neither presents data for the test versus usual practice. As such, the evidence base is exclusively single-armed in nature and cannot address the question of whether the test can improve patient outcomes compared to usual practice. It can, however, reveal something about the ability of the test to identify a group at very low risk of recurrence who could avoid chemotherapy. Data relating to risk in intermediate and high-risk categories are, without a no-test comparator arm, difficult to interpret in the context of clinical utility. The results presented here are therefore divided into two subsections:

- Outcomes in low-risk patients: Assessing the ability of the test to identify a group of patients at low-risk of recurrence who can avoid chemotherapy
- Outcomes in intermediate- and high-risk patients treated according to clinical practice: Observational data relating to clinical outcomes in these patients.

A further section relating to protocol-defined subgroups then follows:

- Outcomes in protocol-defined subgroups.

Outcomes in low-risk patients

DRFS: The TAILORx trial[106] and the MD Anderson observational study[70] reported 5-year DRFS in low-risk patients. DRFS appears very low for patients with RS<11 (99.3%)[106] but somewhat higher when the cut point is increased to RS<18 (95.9%)[70] even though this study included only Stage 1 patients. The difference could indicate that the cut-point of RS<11 gives better outcomes; however, it could be due to differences in patient populations e.g. nodal status (LN0 versus LN0-LNmic respectively) or level of chemotherapy use (0% versus 6.4% respectively).

DRFI: The Clalit Health[71, 72] and the Memorial Sloan Kettering[110] observational studies reported DR rates at 5 years, which have been converted into 5-year DRFI (proportion free of distant recurrence, not including death, at 5 years) for ease of comparison with other outcomes.

In both studies,[71, 72, 105, 110] a subset of patients received chemotherapy in all risk groups (Table 26). In the LN0-LNmic group, 5-year DRFI in the low-risk group (RS<18) was similar in both studies, at 99.5% (95% CI: 98.4, 99.8)[105] and 99.6% (95% CI: NR)[110] respectively, although chemotherapy rates

were somewhat different at 1% and 12%, respectively. In the LNmic-LN3 group, reported for the Clalit Health study only, DRFI in the low-risk group (RS<18) was lower at 96.8% (95% CI: NR).[72]

For LN0-LNmic patients, a lower cut point for low-risk patients (RS<11) was reported in the Memorial Sloan Kettering study[110] and the proportion of patients free from distant recurrence at 5 years was higher compared to RS<18, at 99.9% (95% CI: NR). A similar result was reported for this cut off in the [redacted]. For LNmic-LN3 patients, the lower cut point of RS<11 surprisingly resulted in a DRFI of 95.1% (95% CI: NR), which was slightly lower than for RS<18 (96.8%; 95% CI: NR).[72]

IDFS: The WSG Plan B study[107-109] reported 5-year IDFS, at cut points RS<12 for low-risk, as 94.2% (95% CI: 91.2, 97.3). TAILORx[106] reported IDFS for low-risk (RS<11) patients as 93.8% (95% CI: 92.4, 94.9%).

BCSS/OS: OS was reported in the TAILORx study,[106] and BCSS (converted from breast cancer death rates) was reported in the Clalit Health study for both subgroups ([redacted] and LNmic-LN3)[71, 72, 112] and for the SEER registry.[74, 103] OS was reported in the WSG Plan B study,[107-109] but follow up was less than 5 years and the data were not extracted. 5-year OS in TAILORx[106] was 98.0% (95% CI: 97.1, 98.6%) for patients with RS<11. In the Clalit Health study, LN0-1mic with RS<18, BCSS was 99.9% (95% CI: 99.0, 100.0%).[72, 105] [redacted] For the LNmic-LN3 subgroup of the Clalit Health study,[72] BCSS was 98% in RS<11 patients and 99.1% in RS<18 patients.

Outcomes in intermediate and high-risk patients

DRFS: The MD Anderson study[70] also reported 5-year DRFS for the high-risk group. This was 76.4% (95% CI: 59.2, 87.1%). The difference between risk groups was statistically significant in an unadjusted analysis (p<0.0001) and non-significant in a multivariable analysis (p=0.083 for high vs. low; p=0.066 for intermediate vs. low).

DRFI: Data on intermediate and high-risk groups were reported in the Clalit Health study for both LN0-1mic[72, 105] and for the LNmic-LN3[72] groups. DRFI decreased with increasing risk group in both subgroups, and [redacted] but not reported for the LNmic-LN3 subgroup. The LNmic-LN3 subgroup had [redacted] 5-year DRFI in all risk groups (DRFI RS<18 96.8%, RS18-30 93.4%, RS>30 83.6%) compared with [redacted]

[redacted] but again, surprisingly, DRFI was lower in the RS<11 analyses than the RS<18 (Table 26).

IDFS: The WSG Plan B study[107-109] reported 5-year IDFS, at cut points 12-25 for intermediate-risk and >25 for high-risk. These were 94% and 84% respectively, with p<0.001 between groups (multivariable p=0.001). Tailor X[106] reported IDFS for low-risk (RS<11) patients as 93.8% (95% CI: 92.4, 94.9%).

BCSS/OS: OS was not reported for the intermediate- and high-risk groups in TAILORx.[106] 5-year BCSS for intermediate- and high-risk groups in [redacted] and 97.4% (95% CI: NR) and 86.9% (95% CI: NR) in the LNmic-LN3 subgroups (p-value not reported) of the Clalit Health Services study.[72]

Outcomes in protocol-defined subgroups

Micrometastases: The NICE scope lists micrometastases as a subgroup of interest to the assessment. Only one study that met the inclusion criteria for the review reported data for patients with micrometastases separately (Clalit Health Services),[72] and as such an additional study (SEER database)[74, 103] that followed up patients for <5 years and reported actuarial 5 year BCSS was included.

In the Clalit Health Services LNmic-LN3 analysis,[72] 5-year DRFI was generally higher in the LNmic group compared to the LN1mic LN-3 group, for example, for low-risk patients (RS<18) DRFI was 99.3% (95% CI: NR) and 96.8% (95% CI: NR) respectively. However, BCSS was very similar in each group at 99.3% (95% CI: NR) and 99.1% (95% CI: NR), respectively.

The SEER registry data[74, 103, 104] reported subgroups of LN0 (ages 40-84 years), LN1-LN3 (all ages) and LNmic (all ages). Actuarial 5 year BCSS for low-risk patients (RS<18) were similar at 99.6% (95% CI: 99.4%, 99.7%), 98.9% (95% CI: 97.4, 99.6%) and 99.4% (95% CI: 97.4, 99.9%), respectively (though data for micrometastases is from a later publication with more patients).[104] Data were also similar across subgroups within the intermediate group (LN0 98.6, LN+ 97.7) and high-risk group (LN0 95.6, LN+ 85.7). There was a statistically significant difference between groups for LN0 (p<0.001, unadjusted and multivariable) and LN+ patients (p<0.001 for unadjusted; not reported for multivariable; Table 26).

Race: The NICE scope lists race as a subgroup of interest to the assessment. Only the SEER registry data[74, 103] (which followed up patients for <5 years and reported actuarial 5 year BCSS) reported an

analysis by race, whereby patients were categorised as white, black or other. Data were reported for LN0 and LN1-3 patients separately, and showed generally similar rates across race categories, within risk categories (Table 26).

Clinical utility Oncotype summary

In this review, clinical utility relates to the impact of the prospective use of the test on patient outcomes such as survival and recurrence. The evidence base included two prospective cohorts (within the RCTs TAILORx[106] and WSG PlanB),[108, 109, 111] three retrospective cohorts where patients were treated using Oncotype DX to guide their treatment,[70-72, 105, 110] and one further retrospective cohort that did not meet the inclusion criteria for the review due to insufficient follow-up length (which was included as it provided subgroup analyses for micrometastases and by race, both of which were subgroups specified in the NICE scope[22] and for which there are very limited data).[73, 74] The total number of patients included in these analyses is ~54,000 (some double counting from Clalit Health Services cohort).[71, 72, 105]

Chemotherapy rates in low-risk (RS<18) ranged from 2% to 12% (4 studies) in LN0 patients, and from 7% to 23% (2 studies) in LN+ patients. In intermediate (RS 18-30) patients, chemotherapy rates ranged from 25% to 43% (3 studies) in LN0 patients and from 40% to 47% (2 studies) in LN+ patients. These data perhaps indicate that lymph node status was considered in treatment decisions, though no formal comparison has been made. In high risk patients, chemotherapy rates were similar in LN0 (90% and 88%) and LN+ patients (90%).

Studies generally reported different outcomes (5 year DRFS (n=2), DRFI (n=2), IDFS (n=1), BCSS (n=2) and OS (n=1)), making comparisons across studies difficult. For outcomes including recurrence (DRFS, DRFI and IDFS) RS<18 low-risk patients had outcomes ranging from 96% (5-year DRFI) to 99.6% (5 year DRFI) (LN0) and 97% (LN+; 5-year DRFI) whilst RS<11 low-risk patients had outcomes ranging from 94% (5-year IDFS) to 99.9% (5-year DRFI) (LN0) and 95% (LN+; 5-year DFRI). Clinical advice to the EAG suggests that these levels of recurrence are acceptable in a low-risk population.

It was beyond the scope of the assessment to determine whether the newer cut points used in TAILORx (RS11-25) should be used, or whether the original cut points of RS 18-30 would be preferable. Data relating to this is summarised in the results, and the general observation can be made that whilst use of lower cut-points may result in better outcomes in the low-risk group (though data is mixed on this point), it would also result in fewer patients being classified as low-risk.

It was not possible to draw any conclusions as to whether patients in intermediate and high-risk categories had better outcomes as a result of using Onoctype DX to guide treatment as there were no comparator (no-Oncotype DX) groups.

The data on micrometastases are difficult to interpret as there is no analysis that reports all nodal statuses in the same patient group (i.e. LN0, LNmic, LN1-3) to allow a comparison. The analyses that have been done show that the trend for worse outcomes with increasing risk group holds true in this group.[71, 72, 74, 103]

The data relating to the performance of the test in patients of different races showed that whilst BCSS survival differed similarly according to risk categories in all race categories.[74, 103]

Conclusions

Without the highest level of evidence, it is not possible to conclude whether patient outcomes would be affected by use of the test in a clinical setting. In LN0 patients, use of the test in clinical practice appears to result in low rates of chemotherapy use in low-risk patients (2% to 12%), with acceptable outcomes (DRFS/DRFI/IDFS 96% to 99.6%). Rates of chemotherapy use increased with increasing risk category, and were generally higher in LN+ patients; only one study reported DRFS/DRFI/IDFS for LN+ patients, which was 97% (7% received chemotherapy). It was not possible to draw any conclusions as to whether patients in intermediate and high-risk categories had better outcomes as a result of using Onoctype DX due to the observational nature of the studies.

Table 24: Clinical utility studies: Oncotype DX

Reference; N	Cohorts	Country	Study design	Details of test	Cut-offs	Population	Nodal status	Endo / chemo
Sparano 2015[106] LN0, N=1626	TAILORx	USA	Prospective cohort (within an RCT)	FFPE Genomic Health	RS<11 pts only	100% HR+ 100% HER2- 70% postmeno 100% female Tumour size 1.1 to 5cm, or 0.6-1.0cm with inter/high grade, indicated for CT[a]	LN0	100% ET 100% CT
Le Du 2015[70] N=1030	MD Anderson	USA	Retrospective cohort study	NR	11-25	100% ER+ 100% HER2- 64% postmeno 100% female Stage I disease Had O-DX test	LN0/LNmic	98% ET 27% CT Treated according to usual practice with O-DX
Nitz 2017 [108, 109, 111] N=2642	WSG PlanB	Germany	Prospective cohort (within an RCT)	NR Genomic Health	12-25	100% HR+ 100% HER2- Pre/post meno 100% female High clinical risk [d]	LN0-3 LN0 58.8% LN1-3 41.2%	Treated according to RS: RS<12 endo only RS≥12, chemo + endo[e]
[redacted] Stemmer 2016[72] Stemmer 2016[105] [redacted] LN1mic – LN3, N=627[72]	Clalit Health Services	Israel	Retrospective cohort study	NR	[redacted]18-30	100% ER+ 100% HER2- Meno NR Had O-DX test	1) [redacted] 2)LNmic-LN3	Treated according to usual practice with O-DX test [redacted] 2) % ET NR 27% CT
Wen 2017[110] N=1406	Memorial Sloan Kettering	USA	Retrospective cohort study	NR	RS <18 pts only Cut point RS 11	100% HR+ 100% HER2- 64% postmeno 99.9% female All pts tumour >0.5cm routinely tested and some <0.5cm RS<18 only	LN0-mic	Treated according to usual practice with O-DX test 97% ET 12% CT

Petkov 2016[73] Roberts 2016[74] 1) LN0, all ages N=40,134 2) LNmic-LN3, all ages, N =4,691	SEER registry	USA	Retrospective cohort study	NR Genomic health	18-30	100% HR+ 100% HER2- 40-85 years old Unclear if only those with O-DX test	1) LN0 2)LNmic-LN3	Treated according to usual practice with O-DX test 1) ET NR 23% CT 2) ET NR 35% CT
N, number of patient; CT, chemotherapy; ET, endocrine therapy; FFPE, formalin fixed paraffin embedded; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; RS, Oncotype DX recurrence; mic, micrometastases; NR, not reported [a] indicated for CT by NCCN guidelines; [d] HER2-negativity; pT1-T4c; LN+ [or LN0 with a risk factor (CpT2, grade 2/3, high uPA/PAI-1, <35 years, or HR-negative)][111] [e] patients were treated according to Oncotype DX score, with those with RS<12 receiving ET only, and those with RS≥12 receiving CT+ET;								

Table 25: Quality assessment of clinical utility studies: Oncotype DX

	Random sequence generation	Allocation concealment	Blinding (participants and personnel)	Blinding (outcome assessment)	Incomplete outcome data	Selective reporting
TAILORx[106]	High	High	High	Low	High	Unclear
MD Anderson Le Du 2015[70]	High	High	High	Low	High	Unclear
WSG PlanB[107-109]	High	High	High	Low	High	Unclear
Clalit Health Services[redacted]	High	High	High	Low	High	Unclear
Memorial Sloan Kettering[110]	High	High	High	Low	High	Unclear
SEER registry[74, 103]	High	High	High	Low	High	Unclear
High/low/unclear relates to risk of bias on each criterion						

Table 26: Clinical utility results: Oncotype DX

Study	Study design	Patients	Subgroup, N	Treatment	Cut off[a]	Low-risk: % risk of outcome (95% CI)	Intermediate-risk: % risk of outcome (95% CI)	High-risk: % risk of outcome (95% CI)	Comparison	Adjusted HR (95% CI)[b]
DRFS – 5 year										
TAILORx Sparano 2015[106]	P (obvs arm of RCT)	HR+, HER2-, Tumour size[c]	LN0, N=1626	100% Endocrine therapy	<11	99.3 (98.7, 99.6)	NA	NA	NA	
MD Anderson Le Du 2015[70]	R	ER+, HER2-, Stage 1, had O-DX test	LN0/LNmic, N=1030	Chemo per group: RS <18: 6.4% RS 18-30: 42.7% RS >30:89.8%	18-30	95.9 (93.0, 97.6) [d]	NR	76.4 (59.2, 87.1) [d]	**p<0.0001**	High vs low: 2.20 (0.90, 5.36), p=0.083 Int vs low: 1.88 (0.96, 3.68), p=0.066 High vs int: 1.17 (0.54, 2.51), p=0.690
DRFI– 5 year										
Clalit Health Services [redacted] Stemmer 2016[72] Stemmer 2016[105]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
			LN0-1mic, N= 1,594[105]	RS<18: 1% RS18-30: 26% RS>30: 89%	18-30	99.5 (98.4, 99.8)	98.8 (97.2, 99.4)	93.1 (87.1, 96.3)	NR	
			[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	R		LN1mic – LN3, N=627[72]	Chemo per group: RS<18: 7% RS18-30: 40% RS>30: 90%	18-30	96.8 (NR)	93.4 (NR)	83.6 (NR)	NR	
			LN1mic, N =270[72]			99.3 (NR)	89.2 (NR)	80.6 (NR)		
			LN1mic – LN3, N =627[72]	RS<11: 7% RS11-25: 18% RS >25: 81%	11-25	95.1 (NR)	96.1 (NR)	86.8 (NR)		
			LN1mic, N=270[72]			97.8 (NR)	95.9 (NR)	83.9 (NR)		

Study	Study design	Patients	Subgroup, N	Treatment	Cut off[a]	Low-risk: % risk of outcome (95% CI)	Intermediate-risk: % risk of outcome (95% CI)	High-risk: % risk of outcome (95% CI)	Comparison	Adjusted HR (95% CI)[b]
Memorial Sloan Kettering Wen 2017[110]	R	ER+, HER2-, Stage 1&2, had O-DX test, low RS	LN0 or micro, N=1406	Chemo: RS<18: 12%	<18	99.6%[e]	NA	NA	NR	
					<11	99.9%[e]	NA	NA		
					11 to 17	99.7%[e]	NA	NA		
IDFS– 5 year										
WSG PlanB[f] Nitz 2017 [108, 109, 111]	P	Clinically high-risk,[g] HR+, HER2- patients	LN0-3[107, 108], N=2642	RS<12 endo only; RS≥12, chemo + endo	12-25	94.2 (91.2, 97.3) [h, i]	94.3 (92.8, 95.8) [h]	84.2 (80.6, 87.8) [h]	HR=2.33 (1.73, 3.14), p<0.001 [j]	For continuous score (100-75th vs 0-25th percentiles): 1.73 (1.21, 2.47), p=0.001
TAILORx Sparano 2015[106]	P (RCT)	HR+, HER2-, Tumour size,[c]	LN0, N=1626	100% Endocrine therapy	<11	93.8 (92.4, 94.9)	NA	NA	NA	
BCSS– 5 year										
Clalit Health Services [redacted] Stemmer 2016[72] Stemmer 2016[105] Company Submission[113]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	R		LN0-1mic, N= 1,594[105]	RS<18: 1% RS18-30: 26% RS>30: 89%	18-30	99.9 (99.0 , 100.0)	98.5 (97.1, 99.2)	90.6 (84.5, 94.4)	NR	
	R		LN1mic – LN3, N =627[72]	Clinical practice, including O-DX RS. Chemo per group: RS<18: 7% RS18-30: 40% RS>30: 90%	18-30	99.1 (NR)	97.4 (NR)	86.9 (NR)	NR	
			Ln1mic, N =270[72]		18-30	99.3 (NR)	96.8 (NR)	83.9 (NR)		
			LN1mic – LN3, N =627[72]	RS<11: 7% RS11-25: 18% RS >25: 81%	11-25	98.0 (NR)	99.0 (NR)	90.4 (NR)		
			Ln1mic, N=270[72]			97.8 (NR)	98.8 (NR)	89.3 (NR)		
OS– 5 year										
TAILORx Sparano 2015[106]	P (RCT)	HR+, HER2-, Tumour size[c] N=1629	LN0	100% Endocrine therapy	<11	98.0 (97.1, 98.6)	NA	NA	NA	
BCSS with less than 5 years follow-up (Does not meet inclusion criteria for review).[k] Actuarial 5 year BCSS[l] NOTE: LNmic data has unclear follow-up length.										

Study	Study design	Patients	Subgroup, N	Treatment	Cut off[a]	Low-risk: % risk of outcome (95% CI)	Intermediate-risk: % risk of outcome (95% CI)	High-risk: % risk of outcome (95% CI)	Comparison	Adjusted HR (95% CI)[b]
SEER registry Petkov 2016[73] Roberts 2016[74]	R	HR+, HER2-[m]	LN0, all ages N=40,134	CT per group: LN0, all ages: NR	11-25	99.6 (99.4, 99.8)	99.3 (99.2, 99.4)	96.4 (95.6, 97.0)	p<0.001	
			LN0, 40-84 years of age, N =38,568	LN0, 40-85 years: RS <18: 7% RS 18-30: 34% RS >25: 69%	18-30	99.6 (99.4, 99.7)	98.6 (98.3, 98.9)	95.6 (94.4, 96.6)	Int vs low: HR 3.1 (2.3, 4.3) High vs low: HR 11.0 (7.8, 15.5) All: p<0.001	Int vs low: HR 3.0 (2.1, 4.2) High vs low: HR 7.8 (5.3, 11.6) All: p<0.001
			LNmic, N =2820[104]	NR		98.9 (97.4, 99.6)	99.1 (97.9, 99.6)	84 (74.1, 90.4)	NR	
			LNmic-LN3, all ages, N =4691	LN1-3: <18: 23% 18-30: 47% >25: 75%		99.0 (98.0, 99.5)[n]	97.7 (95.9, 98.7)	85.7 (76.2, 91.6)	p<0.001	
			LN0, Black, N =2,890			99.2 (0.28)	98.2 (0.58)	94.3 (2.17)	p<0.0001	
			LN0, White, N =33,684			99.6 (0.07)	98.6 (0.15)	95.6 (0.61)	p<0.0001	
			LN0, Other race, N =3,321			99.8 (0.15)	99.2 (0.36)	95.3 (1.89)	p<0.0001	
			LN1-3, Black, N =328			99.4 (0.56)	98.9 (1.12)	91.3 (8.31)	p=0.4117	
			LN1-3, White, N =4,021			99.0 (0.39)	97.6 (0.75)	84.1 (4.21)	p<0.0001	
			LN1-3, Other race, N =320			98.5 (1.53)	99.1 (0.92)	100 (0)	p=0.8427	

Study	Study design	Patients	Subgroup, N	Treatment	Cut off[a]	Low-risk: % risk of outcome (95% CI)	Intermediate-risk: % risk of outcome (95% CI)	High-risk: % risk of outcome (95% CI)	Comparison	Adjusted HR (95% CI)[b]
P, Prospective; obvs, observational; RCT, randomised controlled trial; R, Retrospective; HR Hazard ratio; CI, confidence interval; RS, Oncotype DX recurrence score; WSG, West German Study Group; FU, follow-up; RS, recurrence score; DFS, disease free survival; DRFI, distant recrurrence free interval; DRFS, distant recurrence free survival; IDFS, invasive disease free survival; BCSS, breast cancer specific survival; OS, overall survival; NA, not applicable, NR, not reported; N, number of patients; EBC, early breast cancer; LN, lymph nodes; DDFS, distant disease free survival; O-DX, Oncotype DX; chemo-endo, combination therapy of chemotherapy and endocrine therapy; KM, Kaplan Meier; BCD, Breast cancer death; FFR, freedom from recurrence of breast cancer; OS, overall survival; DRFI, distant recurrence free interval (excludes death from other causes); Mic, micrometastases a Range of intermediate group, e.g. where cut points were <18, 18-30 and >30 for low, intermediate or high, this is shown as 18-30; [b] Adjustments: **Le Du 2015**: Not reported; **Stemmer**: age, tumour size, grade; **Nitz 2017**: nodal status, tumour size, grade, ER, PR, Ki67, IHC4; **SEER registry:** age, grade, tumour size, race; c Tumour size 1.1 to 5cm or 0.6 to 1.0 in intermediate or high-risk tumours; [d] Median FU 58 months; [e] median follow-up 46 months; [f] Overall survival data not presented here as follow-up <5 years. Nitz *et al* 2017[111] published after searches, but only added 95% CIs to data already available from conference abstracts; [g] HER2-negativity; pT1-T4c; LN+ [or LN0 with a risk factor (CpT2, grade 2/3, high uPA/PAI-1, <35 years, or HR-negative)]; [h]55 month FU[107, 108]; [i] this data for 348/404 with RS<12, in whom CT was omitted after a protocol amendment;[j] assume low-risk vs intermediate/high-risk [k] Clinical advice to the EAG stated that for survival outcomes, a minimum of 5 year data was required. This data is presented as an exception as it is included in Genomic Health's submission, and it presents data on micro-metastases and by number of lymph nodes, and analyses relating to race; [l] follow-up 38 months; [m] HR+ by O-DX and by IHC; HER2 status by O-DX; [n] 98.7% (95% CI 97.1 to 99.4) for LN1-LN3									

4.4 Results: MammaPrint

4.4.1 Development: MammaPrint

Derivation of the 70-gene signature used a case-control design with 78 node-negative (LN0) patients aged under 55 years: 34 patients with and 44 patients without distant metastases within 5 years (van 't Veer *et al*., 2002).[114] Validation in an additional 19 patients is described within the same article; these patients were also young and LN0, 12 with and 7 without distant metastases within 5 years. Derivation of the 70-gene signature used a DNA microarray containing approximately 25,000 genes.

The first main validation study of the 70-gene signature used a retrospective consecutive series of 295 patients (151 LN0 and 144 LN+ patients) from the Netherlands Cancer Institute (NKI), described by van de Vijver *et al*.[47] (2002). Of these, 61 patients (21%) were also part of the derivation set.[114] Again a 25,000-gene microarray was used to identify the 70-gene signature. This study showed that the signature was significantly prognostic for 5-year DMFS and OS in LN0 and LN+ patients. Updated results for this cohort have since been reported and are presented below.

Threshold: MammaPrint

MammaPrint classifies patients as low-risk (good prognosis) and high-risk (poor prognosis). Correlation coefficients are calculated for the expression level of the 70 genes between individual patients and an "average" good prognosis profile based on the derivation study by van 't Veer *et al*..[114] In the first version of MammaPrint, samples were classified as low-risk if the correlation coefficient was greater than 0.4 and high-risk if less than 0.4 (van 't Veer *et al*.).[114] In a later version of MammaPrint, this threshold was mathematically adjusted to 0 so that low-risk samples are greater than 0 and high-risk samples are ≤0. Both thresholds are the same apart from the adjustment to zero. The same threshold is used in all clinical studies (personal communication with manufacturer).

Prognostic performance in derivation and first validation cohorts

In the derivation cohort (n=78),[114] the test incorrectly identified 3/34 patients who recurred as good prognosis and 12/44 patients who did not recur as poor prognosis. The initial validation cohort in the same article (n=19)[114] incorrectly identified 2/19 patients (whether these were recurrences or non-recurrences was not reported). A multivariable logistic regression analysis that included "classical prognostic factors" (variables not reported) reported an odds ratio for distant metastasis of 18 (95% CI 3.3 to 94) for low- compared with high-risk patients in the derivation cohort (n=78), and a likelihood ratio p-value of 0.0001, though it was unclear whether the patients included were from the derivation cohort or the validation cohort.

Equivalence of different test methods: MammaPrint

Following development of the MammaPrint mini-array specific to the 70 genes, Glas *et al*.[115] (2006) demonstrated that the 70-gene MammaPrint microarray provided very similar results to the 25,000-gene microarray. Within the 78 patients from the derivation set,[114] risk group classification was very similar between the 25,000-gene array and the MammaPrint 70-gene array (Pearson correlation 0.92). For 145 of 151 LN0 patients from the van de Vijver *et al.* validation study,[47] HRs for low vs. high-risk for DMFS over all follow-up were very similar for the two array types: HR 5.5 (95% CI: 2.5, 12.2) for the 25,000-gene array, and HR 5.6 (95% CI: 2.4, 7.3) for the 70-gene array.[115]

Beumer *et al*. (2016)[116] showed that fresh-frozen and FFPE paired samples give very similar results (Pearson correlation 0.93); that the MammaPrint 70-gene mini-array and whole-genome 25,000-gene array give near-perfect correlation (Pearson correlation 0.99); that samples repeated over 10 years give an overall reproducibility of 97%; and that precision and repeatability (using repeated measurements) are both 98% overall.

4.4.2 Prognostic performance: MammaPrint

Study designs and patients: MammaPrint prognostic performance

Several publications describe validation of the prognostic value of MammaPrint. Many include overlapping cohorts of patients, sometimes pooled with other cohorts, sometimes focussing on patient subgroups (e.g. ER+ or LN0/LN+), sometimes updating the data with longer follow-up, and reporting a range of different outcomes. Therefore, it should be noted that there is some overlap between patient cohorts within the references included here. Table 27 shows both the study reference(s) (column 1) and the cohort(s) (column 2) used for each analysis.

Prognostic data on MammaPrint mainly consists of retrospective analyses of consecutive patient series, many from the Netherlands plus some from other countries. The main nine cohorts are listed below (and in Table 27). Five cohorts consisted of LN0 patients,[63-66, 86] one of LN+,[60] and three included a mix of LN0 and LN+ patients.[47, 117, 118] Three cohorts did not receive adjuvant chemotherapy,[64, 65, 117] while in the other six a subset received chemotherapy,[47, 60, 63, 66, 86, 118] though treatment was not influenced by the MammaPrint test since this was performed later on stored tumour samples. In the majority of these series, around 70-80% of patients were ER+, while HER2 was not well reported (Table 27). The nine cohorts are:

- van de Vijver 2002:[47] Retrospective analysis of consecutive series from the Netherlands Cancer Institute (NKI, 1984-95), age ≤52 years; 51% LN0. Updated data are presented in subsequent articles, the most recent being Drukker *et al*.[67] (2014) Independent data from the same centre are reported in Mook 2010[65] (ages 55-71, LN0) and Bueno-de-Mesquita 2009,[63] and there may be some overlap with Mook 2009[60] (1994-2001) and Kok 2009 (1985-94).[117]

- Bueno-de-Mesquita 2009:[63] Retrospective analysis of consecutive series from two Dutch hospitals (NKI and Reinier de Graaf Hospital, 1996-99), all LN0.
- Mook 2010:[65] Retrospective analysis of consecutive series from NKI (1984-96), age 55-71 years, all LN0.
- Mook 2009:[60] Retrospective analysis of consecutive series from NKI and Italy (1994-2001), all LN+ (LN1-3).
- Kok 2009/2012:[117] Retrospective analysis of consecutive series from NKI (1985-94), 82% LN+.
- Buyse 2006[64] (TRANSBIG): Retrospective cohort from the UK, France and Sweden (1980-1999); all LN0.
- Yao 2015:[118] Retrospective analysis of consecutive series from two US centres (1992-2010); mix of LN0 and LN+.
- Wittner 2008:[66] Retrospective analysis of consecutive series from one US centre (1985-1997); all LN0.
- Ishitobi 2010:[86] Retrospective analysis of cases from Osaka Medical Centre, Japan (1998-2001); all LN0.

In addition, there is one retrospective analysis of an RCT:

- Stockholm Tamoxifen (STO-3) trial (Esserman 2016,[119] Lindstrom 2015,[120] van 't Veer 2017,[53] company submission[121]): LN0 patients receiving no chemotherapy.

A number of additional analyses pooled data on patients with specific characteristics from two or more of the above cohorts, as follows:

- Mook *et al.* (2010)[122] pooled 964 patients from seven series and reports prognostic performance;[47, 60, 63-65, 117, 123] patients are a mix of LN0 and LN+, and the analysis is restricted to T1 patients (tumour ≤2cm) which means that a higher proportion of patients are ER+ than in the original analyses. The analysis included six series in which MammaPrint did not influence treatment, plus one study (RASTER)[123-125] in which patients were treated according to usual practice plus MammaPrint.
- Knauer *et al*.[126] (2010) pooled 541 patients from six of seven series above (LN0 or LN1-3) and reports whether MammaPrint predicts benefit from chemotherapy (Section 4.4.3). Again, this analysis included the RASTER observational study.[123-125]
- Bueno-de-Mesquita *et al*.[127] (2011) pooled 139 ER+ LN0 untreated patients from two series[47, 63]
- Beumer *et al*.[128] (2016) pooled patients with lobular breast cancer from five series.[47, 117, 123, 129]

Tests and comparators: MammaPrint prognostic performance

All prognostic studies used the MammaPrint 70-gene microarray. The majority used frozen tumour samples, while FFPE samples were used in the STO-3 trial,[53, 119, 120] and both frozen and FFPE samples were used in the USA series (Yao *et al*., 2015[118]) (Table 27). Patients were categorised as low-risk (or good prognosis) and high-risk (or poor prognosis).

None of the MammaPrint analyses included other in-scope tests (except for some of the whole-transcriptome microarray studies; see Section 4.8.2). Comparators for prognostic studies included AOL and NPI.

Quality assessment: MammaPrint prognostic performance

All data sets included for prognostic performance were validation studies (Table 28), though the Van de Vijver 2002[47] cohort included a small proportion of patients from the derivation set (Van 't Veer *et al*. 2002).[114] Most analyses excluded some patients recruited to the original trial or cohort, or this was unclear. Blinding of test assessors to outcomes was reported in around half the studies. Outcomes did not always match standardised defintions; several described analyses of distant metastases but were not clear whether all deaths and breast cancer deaths were counted as events or were censored, which makes it difficult to know whether the analyses were of DRFS or DRFI.[53, 63, 64, 86, 126-128] As noted above, many studies were retrospective analyses of patient series of whom some received chemotherapy in accordance with usual practice; the corresponding different levels of chemotherapy use in the high- and low-risk groups may confound results. Additionally, retrospective selection of cohorts who did (or did not) have chemotherapy may introduce spectrum bias since these patients may be systematically different to the whole population. In addition, many studies included a proportion of patients who were out of scope (ER- and/or HER2+ and/or >3 positive nodes).

Results: MammaPrint prognostic performance

Prognostic data for MammaPrint is provided in Table 29 to Table 32.

Distribution of patients by risk group

For LN0 patients, the percentage of patients categorised as low-risk varied widely: 20% to 71% across seven analyses[53, 63-67, 86, 119] (Table 29). For LN+ patients, 38% and 41% were categorised as low-risk in two analyses.[60, 67] A further analysis of LN0 patients showed that, of those who were low clinical risk (via three tools: AOL, NPI and St Gallen), 77% were MammaPrint low-risk; conversely, of those at high clinical risk, only 27% were MammaPrint low-risk.[127]

Prognostic performance: unadjusted analyses

This section reports unadjusted analyses. Adjusted analyses, which show whether the test has prognostic value over clinicopathological variables, are reported in the section "Additional prognostic value"

Mix of LN0/+ patients with varying endocrine and chemotherapy use: Two unadjusted analyses pooled six or seven European validation series; both showed MammaPrint to be significantly prognostic for DRFS/DRFI and BCSS. Mook *et al*.[122] (2010) pooled 964 patients from seven series[47, 60, 63-65, 117, 123] (84% ER+, varying levels of chemotherapy and endocrine therapy). MammaPrint was significantly prognostic for 10-year DRFS (HR 2.70 (95% CI 1.88 to 3.88, p<.0001); Table 29) and BCSS (HR 4.22 (95% CI 2.70 to 6.60, p<0.001); Table 31), with 10-year DRFS rates in the low-risk group of 87% at 10 years (Table 29). Knauer *et al*.[126] (2010) pooled 541 patients from six of these series (restricted to LN0-3 patients, all had endocrine therapy and 42% chemotherapy). MammaPrint was again significantly prognostic for 5-year DRFS and BCSS, with 95% DRFS in the low-risk group at 5 years (no data for later follow-up). Separate results for ER+ patients from three of the above series were reported by Kok *et al*. (2012);[117] MammaPrint was significantly prognostic for 10-year BCSS among patients pooled from two series[47, 65] (all ER+, 91% LN0, no adjuvant treatment, HR 4.52 (95% CI 2.01 to 10.2, p<0.001)) and also from NKI patients[117] (all ER+, 82% LN+, all endocrine-treated, HR 2.78 (95% CI 1.30 to 5.94, p=0.008)) (Table 31).

In terms of longer follow-up, 25-year follow-up[67] of the initial van de Vijver (2002)[47] cohort (51% LN0; 37% had chemotherapy and 14% endocrine therapy) reported that MammaPrint was statistically significantly prognostic for unadjusted analyses of DRFS for the whole 0-25 year period (HR 3.1 (95% CI 2.02 to 4.86, p<0.001)); however, most of this difference was seen in the first 5 years (HR 9.6, 95% CI 4.2 to 22.1), with subsequent individual 5-year bands from 5-10 years to 20-25 years not showing a statistically significant difference in DRFS between risk groups (Table 29). Results for OS showed a similar pattern, with a statistically significant prognostic effect for years 0-5 and 0-25 (p<0.0001); there was also a statistically significant difference in years 5-10 for OS (p=not reported; Table 30). A separate USA series (Yao *et al*. 2015,[118] 72% LN0, 43% had chemotherapy and 87% endocrine therapy) also showed statistically significant prognostic ability for DRFS at 10 years (HR 2.91 (95% CI 0.97 to 8.68), p=0.045, Table 29) with DRFS rates in the low-risk group of 96% at 10 years; results were similar (low-risk 10-year DRFS 98%) in a subset with no chemotherapy.

LN0: Four of five retrospective LN0 cohorts (all having varying levels of endocrine and chemotherapy) assessing the prognostic ability of MammaPrint reported statistically significant prognostic performance in unadjusted analyses.[47, 63-65] The exception was one study of 100 US patients (Wittner, 2008)[66] in which MammaPrint was not statistically significantly prognostic for

DRFI (p=0.330 at 10 years; HR NR). In the van de Vijver 2002[47] cohort (age ≤52 years), MammaPrint was statistically significantly prognostic for DRFS (Table 29) and OS (Table 30) over years 0-10[63] and years 0-25[67] (HRs range from 4.6 to 10.7, all p<0.001). In the Bueno-de-Mesquita 2009 cohort (age <55 years),[63] MammaPrint was statistically significantly prognostic for DRFS (HR 5.7 (95% CI 1.6 to 20, p=0.007)) and OS (HR 3.4 (95% CI 1.2 to 9.6, p=0.021)) at 5 years. In Mook 2010 (age 55-71 years),[65] MammaPrint was statistically significantly prognostic for 5-year DRFS (4.6 (95% CI 1.8 to 12.0, p=0.01) and BCSS (HR 19.1 (95%CI 2.5 to 148, p=0.005), though 10 year outcome data were available but no statistical significance levels were reported (Table 29; Table 31). In TRANSBIG (Buyse 2006[64]), for all follow-up (median 13.6 years), MammaPrint was statistically significantly prognostic for DRFI (HR 2.32 (95% CI 1.35 to 4.00, p=0.002), OS (HR 2.79 (95% CI 1.60 to 4.87, p<0.001) and BCSS (HR 1.50 (95% CI 1.04 to 2.16, p=0.032). In addition, the STO-3 trial (van 't Veer 2017[53]) reported 10-year DRFS rates (93% in low-risk; 85% in high-risk; Table 29) but no statistical significance levels were reported.[53] [redacted] 5-year DRFS was also reported for a Japanese cohort (Ishitobi *et al*., 2010[86]), with 5-year DRFS of 100% for low-risk patients and 94% for high-risk; however, no statistical significance levels were reported (Table 29).

Patient outcomes may vary by receipt of chemotherapy and endocrine therapy. In low-risk patients, 10-year DRFS rates were 88% in a pooled analysis of patients receiving no chemotherapy or endocrine therapy from the van de Vijver[47] and Bueno-de-Mesquita[63] cohorts (Bueno-de-Mesquita 2011[127]); 86% in van de Vijver 2002[63] (4% chemotherapy, 4% endocrine therapy); 80% in Mook 2010[65] (no chemotherapy, 18% endocrine therapy); and in the STO-3 trial (van 't Veer 2017,[53] ER+ patients), 10-year DRFS was 93% with endocrine monotherapy and 83% without endocrine or chemotherapy, while 10-year DRFI was 90% in TRANSBIG (no chemotherapy or endocrine therapy).[64]

Three LN0 cohorts included comparisons to clinical risk tools (AOL and NPI), which appeared to have less prognostic value than MammaPrint, though there were no comparisons available for some in-scope comparators (such as PREDICT or modified AOL). NPI was statistically significantly prognostic for 10-year DRFS and OS (both p<0.001)) in the van de Vijver 2002 cohort,[47, 63] but was not statistically significantly prognostic for 5-year DRFS (p=0.14)) and borderline non-significant for 5-year OS (p=0.053) in the Bueno-de-Mesquita 2009 cohort,[63] and was statistically significantly prognostic for DRFI (p=0.043) but not OS (p=0.092) or DFS (p=0.58) in TRANSBIG[64] (all follow-up; Table 29, Table 30 and Table 31). AOL was statistically significantly prognostic for 10-year OS (p=0.017) but not DRFS (p=0.14) in the van de Vijver 2002 cohort,[47, 63] but was not statistically

significantly prognostic for 5-year DRFS (p=0.14) or OS (p=0.22) in the Bueno-de-Mesquita 2009 cohort,[63] nor for DRFI (p=0.092), OS (p=0.085) or BCSS (p=0.092) in TRANSBIG.[64]

LN+: Two cohorts reported separate results for LN+ patients, both with varying endocrine and chemotherapy use; both showed statistically significant prognostic performance of MammaPrint.[47, 60] In the van de Vijver 2002[47] cohort (in which a quarter had more than 3 positive nodes), MammaPrint was statistically significantly prognostic for DRFS (HR 2.24 (95% CI 1.25 to 4.00, p=0.01)) and OS (HR 1.83 (95% CI 1.07 to 3.11, p=0.03) over years 0-25[67] and for 10-year BCSS[60] (HR 6.60 (95% CI 1.97 to 22.10, p=0.002)) (Table 29, Table 30 and Table 31). In the Mook 2009[60] cohort (all LN1-3), MammaPrint was again statistically significantly prognostic for DRFS (HR 4.13 (95% CI 1.72 to 9.96), p=0.002), OS (HR 5.40 (95% CI 2.11 to 13.80, p<0.001)) and BCSS (HR 5.70 (95% CI 2.01 to 16.23, p=0.001)) over 0-10 years. In both cohorts, some patients received chemotherapy, though results remained statistically significant in a subgroup of patients not receiving chemotherapy in Mook 2009[60] (only reported for BCSS, HR 7.33 (95% CI 1.61 to 33.49, p=0.01); Table 31). In low-risk patients, 10-year DRFS rates were 79% in van de Vijver 2002[63] (rates of adjuvant treatment not reported) and 91% in Mook 2009[60] (56% chemotherapy, 73% endocrine therapy).

Low or high clinical risk: Patients at low- or high-risk via three clinical tools (AOL, NPI and St Gallen) were assessed in a pooled analysis of LN0 untreated patients from two series[47, 63] (Bueno-de-Mesquita *et al.*,[127] 2011; Table 32). Patients with all-low clinical risk according to all three clinical tools showed a statistically significant prognostic effect of MammaPrint on 10-year OS (HR NR, p=0.016) but not DRFI (HR NR, p=0.19), though 10-year DRFI was numerically more favourable in the MammaPrint low-risk group (87%) than in the high-risk group (70%). Patients with all-high clinical risk did not show a statistically significant effect on either OS (HR NR, p=0.17) or DRFI (HR NR, p=0.19), and had relatively poor 10-year DRFI even in the MammaPrint low-risk group (77%) though this was numerically more favourable than in the high-risk group (45%). In a separate analysis, LN+ patients (LN1-3) at high clinical risk via AOL in Mook 2009[60] showed a statistically significant prognostic effect of MammaPrint on 10-year BCSS (HR 4.12 (95%IC 1.45 to 11.76, p=0.008)); Table 32). Statistical significance levels in this analysis may have been affected by the small sample sizes per subgroup.

Lobular breast cancer: A pooled analysis of patients with invasive lobular breast cancer from five series[47, 117, 123, 129] (Beumer *et al.*,[128] 2016) showed that MammaPrint was statistically significantly prognostic for 10-year DRFS (HR : 3.31 (95%CI 1.79 to 6.12, p<0.001)) and OS (HR 3.58 (95% CI 1.84 to 6.95, p<0.001)) in all patients (34% LN+) and in a sub-analysis of LN0 patients (DRFS HR 7.81 (95% CI 2.89 to 21.07, p<0.001); OS HR 7.47 (95% CI 2.58 to 21.58, p<0.001)), Table 29 and Table 30).

Additional prognostic value

This section reports adjusted analyses, which indicate the additional prognostic value of IHC4 over clinicopathological factors. The clinicopathological factors adjusted for vary from study to study, and are detailed in the footnotes to the tables.

Among mixed LN0/+ cohorts, the van de Vijver 2002 cohort reported that MammaPrint was statistically significantly prognostic for 10-year DRFS (HR 4.6 (95% CI 2.3–9.2, p<0.001) in a multivariable analysis which included age, lymph node status, tumour size, grade, vascular invasion, ER status, surgery type, chemotherapy and endocrine therapy. In the pooled analysis of seven series by Mook *et al*.[122] (2010), which incorporated some or all of the van de Vijver 2002[47] cohort, MammaPrint was also statistically significantly prognostic for 10-year DRFS (HR 2.43 (95% CI 1.56 to 3.77, p<0.001) and BCSS in a multivariable analysis adjusted for age, tumour size, nodal status, grade, ER, HER2, surgery, endocrine therapy and chemotherapy (p<0.001; Table 33 and Table 35). However, in the USA series (Yao *et al*. 2015,[118]), MammaPrint prognostic value for 10-year DRFS was borderline statistically significant in the unadjusted analysis (p=0.045, Table 33) and borderline non-statistically significant in a multivariable analysis (HR 3.01 (95% CI 0.88 to 10.33, p=0.08, Table 33).

Among LN0 patients, MammaPrint remained statistically significantly prognostic for distant recurrence when adjusted for either AOL or NPI in three cohorts: for 10-year DRFI in van de Vijver 2002 (p=0.001),[47, 63, 64] for 5-year DRFI in Bueno-de-Mesquita 2009[63] (p=0.02), and for DRFI (all follow-up) in TRANSBIG[64] (p=not reported) (Table 33). C-indices (reported as AUC) were reported by Bueno-de-Mesquita 2009[63] for both cohorts (Bueno-de-Mesquita[63]; van de Vijver 2002[47]) and showed a higher value (0.75 (95% CI 0.61 to 0.89) and 0.76 (95% CI 0.68 to 0.85) respectively) for MammaPrint and clinicopathological factors (age, tumour size, grade, ER, PR, HER2) than for either the factors on their own, or MammaPrint on its own, though differences were not statistically compared (Table 33). For OS (Table 34), MammaPrint remained statistically significantly prognostic in van de Vijver 2002[47, 63, 64] at 10-year when adjusted for AOL or NPI (p<0.001), in TRANSBIG[64] (all follow-up) when adjusted for AOL or NPI (p=not reported), and in Bueno-de-Mesquita 2009[63] at 5-year when adjusted for AOL (p=0.044), but not NPI (p=0.086). C-indicies reported by Bueno-de-Mesquita 2009[63] for OS showed the same trends as for DRFI (data not shown). For other outcomes, MammaPrint remained statistically significantly prognostic for 5-year BCSS in Mook 2010[65] when adjusted for AOL (p=0.01) and for DFS (all follow-up) in van de Vijver 2002[47, 64] when adjusted for AOL (HR 4.80 (95%CI 2.37 to 9.71, p not reported)) but not for DFS in TRANSBIG[64] when adjusted for AOL or NPI (p=not reported) (Table 35).

Among LN+ patients, MammaPrint was statistically significantly prognostic for 10-year BCSS (HR 7.17 (95% CI 1.81 to 28.43, p=0.005), Table 35) but borderline significant for 10-year DRFS (2.99 (95% CI 0.996 to 8.99, p=0.051), Table 33) in Mook 2009[60] when adjusted for age, tumour size, nodal status, grade, ER, HER2, surgery, endocrine therapy, chemotherapy. MammaPrint was borderline non-statistically significantly prognostic for 10-year BCSS in van de Vijver 2002[47, 60] (HR 3.63 (95% CI 0.88 to14.96, p=0.07)) when adjusted for the same variables.

Among lobular breast cancer patients, MammaPrint was statistically significantly prognostic for 10-year DRFS (p=0.037 in all patients and p=0.001 in LN0; Table 33) but not statistically significant for 10-year OS (p=0.070 in all patients and p=0.008 in LN0; Table 34) when adjusted for age, nodal status, grade, ER, HER2, and chemotherapy.

Discussion: MammaPrint prognostic performance

Prognostic value of MammaPrint is mainly based on nine small retrospective analyses of consecutive patient series: four from the Netherlands, two multi-European, two from the USA and one from Japan (total N=1,805).[47, 60, 61, 63-66, 86, 117, 130] These cohorts cover a mix of LN0 and LN+ patients, with variable proportions receiving endocrine therapy and chemotherapy; in most studies around 70-80% were ER+, while HER2 status was not well reported. Four analyses pooling some of the above cohorts[122, 126, 127, 131] are also included due to their focus on specific subgroups. In addition, there was one reanalysis of an RCT (Swedish STO-3 trial; N=538), of which a subgroup had endocrine monotherapy.[53, 119, 120, 132] Most analyses excluded some patients from the original cohort, some because of insufficient tumour sample, which may introduce bias due to attrition of patients with smaller tumours.

The percentage of patients categorised as low-risk ranged from 20% to 71%, and high-risk from 29% to 80%, across seven analyses of LN0 patients.[53, 61, 63-66, 86] In two analyses of LN+ patients,[60, 61] percentage categorised as low-risk was 38% and 41%, while percentage high-risk was 59% and 62%.

Among LN0/+ studies, a pooled unadjusted analysis of patients from seven series (N=964; one-third had endocrine and one-quarter chemotherapy) showed that MammaPrint was statistically significantly prognostic for 10-year DRFS, with 10-year DRFS of 87% in low-risk patients.[122] In terms of longer follow-up, MammaPrint was statistically significantly prognostic in an unadjusted analysis of DRFS over 0-25 years[61] in a LN0/LN+ cohort;[47] most of this difference occurred in the first 5 years.

Among LN0 patients, in the only reanalysis of an RCT (STO-3 trial), patients receiving endocrine monotherapy had a 10-year DRFS of 93%, while in those without endocrine or chemotherapy it was 83%; there were no statistical comparisons between MammaPrint risk groups.[53] Four of five

retrospective LN0 cohorts reported statistically significant prognostic performance of MammaPrint for 10-year DRFS/DRFI, based on unadjusted HRs between risk groups.[47, 63-65] The 10-year DRFS/DRFI rates in low-risk patients ranged from 80% to 90% across three analyses (with varying rates of endocrine and chemotherapy use).[47, 64, 65] Three of the LN0 cohorts included comparisons to AOL and NPI, which appeared to have less prognostic value than MammaPrint, though no statistical comparisons were reported. There were no comparisons to other risk tools such as PREDICT or mAOL. Among LN+ patients, two cohorts reported statistically significant prognostic performance of MammaPrint based on unadjusted HRs between risk groups, with 10-year DRFS rates in low-risk patients of 79% and 91% (with varying rates of endocrine and chemotherapy use).[60, 63]

Several studies reported adjusted analyses relating to the additional prognostic value of MammaPrint over existing clinicopathological risk scores and clinicopathological variables. A pooled analysis of LN0/LN+ patients from seven series[122] showed that MammaPrint was statistically significantly prognostic for 10-year DRFS in a multivariable analysis adjusting for clinicopathological variables. Among LN0 patients, MammaPrint was statistically significantly prognostic for DRFI when adjusted for AOL or NPI in three cohorts.[47, 63, 64] C-indices from two of these cohorts showed higher values when MammaPrint was included alongside clinicopathological factors than for either alone, though differences were not statistically compared.[47, 63] In one analysis of LN+ patients, MammaPrint was borderline statistically significant for 10-year DRFS and statistically significantly prognostic for 10-year BCSS,[60] though in another[47] BCSS at 10 years was borderline non-statistically significant.

Conclusions: MammaPrint prognostic performance

The prognostic value of MammaPrint is based on nine retrospective analyses, four pooled analysis (including six of the nine retrospective series and one prospective series) and one reanalysis of an RCT. Studies were variable in terms of nodal status, ER status, and receipt of endocrine and chemotherapy. The percentage of LN0 patients categorised as low-risk ranged from 20% to 71%%, and high-risk from 29% to 80%. In LN+ patients, the percentage low-risk was 38% to 41%, and percentage high-risk 59% to 62%. MammaPrint was statistically significantly prognostic for 10-year DRFS in almost all unadjusted analyses of LN0 and LN+ patients (as well as in pooled analyses). For LN0 patients, 10-year DRFS/DRFI rates for low-risk patients ranged from 80% to 90% (with varying rates of endocrine and chemotherapy use), while the reanalysis of an RCT reported 10-year DRFS of 93% with endocrine monotherapy and 83% without endocrine or chemotherapy. Interestingly, although on the whole MammaPrint low-risk 10-year DRFS rates are lower than for the other in-scope tests, the 93% figure for patients having endocrine monotherapy is more in line with other tests and may better reflect the population used in studies of other tests (ER+, endocrine monotherapy). For LN+ patients, 10-year DRFS rates in low-risk patients ranged from 79% to 91% (with varying rates of endocrine and chemotherapy use). In terms of additional prognostic value, MammaPrint was

statistically significantly prognostic for 10-year DRFS/DRFI in multivariable analyses adjusted for clinicopathological risk tools (AOL and NPI) and various combinations of clinicopathological variables in LN0/+ and LN0 cohorts, while adjusted analyses in LN+ cohorts were statistically significant or borderline significant.

Table 27: Characteristics of prognostic studies: MammaPrint

Reference(s)	Cohort(s)	N	Country	Study design	Test	Test details	Cut-offs	Population	Nodal status	ET / CT
Pooled analyses of patient cohorts: LN status mixed										
Variable ET&CT										
Beumer 2016[128] **Lobular cancer**	Lobular cancers, 5 pooled series: - van de Vijver 2002[47] - Bueno-de-Mesquita 2007[123] (RASTER) - Kok 2012[117] - Michaut 2016[129] (RATHER; NKI, UK) - North Shore & Fox Chase, US	217	Neths, US, UK	Pooled cohorts	**MMP**	Sample type NR MMP microarray	Low, high; details NR	Invasive lobular breast cancer 94% ER+ 92% HER2- % female NR	LN0, 66% LN1-3, 24% LN>3, 9%	59% ET (low 58%, high 62%) 22% CT (low 19%, high 33%)
Knauer 2010[126]	Pooled 6 series: - van de Vijver 2002[47] - Bueno-de-Mesquita 2009[63] - Mook 2009[60] (LN1-3) - Mook 2010[65] (age 55-71) - Bueno-de-Mesquita 2007[123] (RASTER) - Kok (personal com.)	541	Various	Pooling of 6 consecutive cohorts	**MMP**	Frozen MMP microarray	Low, high (details NR)	90% ER+ 89% HER2- Pre/post-meno % female NR pT1-3	LN0, 49% LN1-3, 51%	All ET 42% CT
Mook 2010[122]	Pooled 7 series: - van de Vijver 2002[47] (NKI 84-95) - Bueno-de-Mesquita 2009[63] (NKI+RdGG) - Mook 2009[60] (LN1-3, NKI+Italy) - Mook 2010[65] (age 55-71, NKI) - Bueno-de-Mesquita 2007[123] (RASTER) - Kok 2012[117] (NKI 1985-94) - Buyse 2006[64] (TRANSBIG)	964	Various	Pooling of 7 consecutive cohorts	**MMP**	Frozen MMP microarray	Low (good), high (poor); details NR	84% ER+ 68% HER2- (23% missing) Pre/post-meno % female NR pT1 (≤2cm)	LN0, 72% LN+, 27% (% LN>3 NR)	32% ET (low 27%, high 38%) 22% CT (low 10%, high 37%)
No ET&CT										
Kok 2012[117]	Pooled 2 series: - van de Vijver 2002[47] - Mook 2010[65] age 55-71	100 + 51	Neths	Two pooled cohorts	**MMP**	Frozen MMP microarray	Low (good), high (poor); details NR	All ER+ HER2 NR Pre/post-meno % female NR	LN0, 91% LN1-3, 7% LN>3, 2%	No ET/CT

Reference(s)	Cohort(s)	N	Country	Study design	Test	Test details	Cut-offs	Population	Nodal status	ET / CT
Retrospective studies: LN status mixed										
100% ET monotherapy										
Kok 2012[117]	Kok 2009[133] (NKI 1985-94)	121	Neths	1 cohort	**MMP**	Frozen MMP microarray	Low (good), high (poor); details NR	All ER+ HER2 NR Pre/post-meno % female NR	LN0, 18% LN1-3, 65% LN>3, 18%	All ET, no CT
Variable ET&CT										
Drukker 2014[67] van de Vijver 2002[47]	- van de Vijver 2002[47]	295	Neths	Retrospective, consecutive	**MMP**	Frozen MMP microarray	Low >0.4, high <0.4	77% ER+ HER2 NR Age ≤52 100% female	LN0, 51% LN1-3, 36% LN>3, 13%	14% ET (low 15%, high 13%) 37% CT (low 38%, high 37%)
Yao 2015[118]	NorthShore University Health System & Fox Chase Cancer Center (1992-2010)	373 (all) 238 (subgrp)	USA	Retrospective, consecutive	**MMP**	Frozen or FFPE MMP microarray	Low, high; details NR	**All:** 74% ER+ 83% HER2- Stage 1-2b **Subgrp:** All HR+ all HER2- 100% female	LN0, 72% LN1-3, 25% LN>3, 5%	**Subgrp:** 87% ET (low 92%, high 79%) 43% CT (low 37%, high 53%)
Reanalyses of RCTs: LN0										
100% ET monotherapy OR No ET&CT										
van 't Veer 2017[53] Esserman 2016[119] Lindstrom 2015[120] [redacted]	Stockholm Tamoxifen (STO-3) trial: ER+ subgroup	538	Sweden	Reanalysis of RCT	**MMP**	FFPE MMP microarray	Low >0, high <0	All ER+ 96% HER2- Post-meno % female NR Tumours <30mm	LN0	Analysis 1: All ET, no CT Analysis 2: No ET, no CT
Pooled analyses of patient cohorts: LN0										
No ET&CT										
Bueno-de-Mesquita 2011[127]	Pooled 2 series: - van de Vijver 2002 (NKI, 84-95)[47] - Bueno-de-Mesquita 2009[63] (NKI 96-99)	186	Neths	Pooling of 2 cohorts to form 1 consecutive series	**MMP**	Frozen MMP microarray	Low (good), high (poor); details NR	76% ER+ 76% HER2- Pre/post-meno 100% female	LN0	No ET No CT

Reference(s)	Cohort(s)	N	Country	Study design	Test	Test details	Cut-offs	Population	Nodal status	ET / CT
Retrospective studies: LN0										
Variable ET&CT										
Bueno-de-Mesquita 2009[63]	1) Bueno-de-Mesquita 2009[63] (NKI+RdGG 1996-99) 2) van de Vijver 2002[47]	1) 123 2) 151	Neths	Retrospective, consecutive	**MMP**	Frozen MMP microarray	Low (good), high (poor); details NR	1) 76% ER+ 93% HER2- pT1–2, <55 yr 2) 72% ER+ HER2 NR, pT1–2, age ≤52	LN0	1) 22% ET (low 28%, high 15%); 25% CT (low 16%, high 36%) 2) 4% ET (low 5%, high 3%); 4% CT (low 3%, high 4%)
Buyse 2006[64]	1) TRANSBIG (1980-1999)[64] 2) van de Vijver 2002[47]	1) 302 2) 151	1) France, Sweden, UK 2) Neths	Retrospective cohorts	**MMP**	Frozen MMP microarray	Correlation coeff. low >0.4, high <0.4	1) 70% ER+ HER2 NR, <61yr T1-2 (≤5cm) % female NR 2) 72% ER+ HER2 NR, pT1–2, age ≤52	LN0	1) No ET/CT 2) Some ET/CT
Ishitobi 2010[86]	Osaka Medical Centre (1998-2001)	102	Japan	Retrospective analysis of cases	**MMP**	Frozen MMP microarray	Good (low, if above threshold) or poor (high)	51% ER+ HER2 NR ≤70yrs, T1-3 100% female	LN0	73% ET (low 85%, high 70%) 28% CT (low 10%, high 33%)
Mook 2010[65]	NKI 1984-96[65] (55-71yr)	148	Neths	Retrospective, consecutive	**MMP**	Frozen MMP microarray	Low (good), high (poor); details NR	78% ER+ HER2 NR Post-meno, T1-2 100% female	LN0	18% ET No CT
Wittner 2008[66] N=100	Massachusetts General Hospital (1985-1997)	100	USA	Retrospective, consecutive	**MMP**	Frozen MMP microarray	Low (good) >0.4, high (poor) <0.4	80% ER+ HER2 NR Pre/post-meno 100% female	LN0	24% ET 21% CT

Reference(s)	Cohort(s)	N	Country	Study design	Test	Test details	Cut-offs	Population	Nodal status	ET / CT
Retrospective studies: LN+										
Variable ET&CT										
Mook 2009[60]	1) NKI+Italy 1994-2001[60] 2) van de Vijver 2002[47]	1) 241 2) 106	Neths, Italy	Retrospective, consecutive	**MMP**	Frozen MMP microarray	Low (good), high (poor); details NR	1) 79% ER+ 84% HER2- 2) 82% ER+ 84% HER2- All: Pre/post-meno, age ≤70 % female NR T1-3	LN1-3 (inc. micromets)	1) 73% ET (low 82%, high 65%); 56% CT (low 41%, high 67%) 2) 23% ET (low 26%, high 21%); 70% CT (low 77%, high 65%)

CT, chemotherapy; ER, oestrogen receptor; ET, endocrine therapy; FFPE, formalin-fixed paraffin-embedded; HER2, human epidermal growth factor receptor 2; HR+, hormone-receptor positive; LN, number of positive nodes; MMP, MammaPrint; NKI, Netherlands Cancer Institute; NR, not reported; qRT-PCR, quantitative reverse transcription polymerase chain reaction; RASTER, MicroarRAy PrognoSTics in Breast CancER study; RATHER, RAtional THerapy for breast cancer study; RdGG, Reinier de Graaf Hospital; R-RCT, reanalysis of RCT

Table 28: Quality assessment of prognostic studies: MammaPrint

Reference(s); N	Cohort(s)	Derivation or validation?	Study design appropriate?	All eligible patients included?	Blinding (of test assessors to outcomes)?	Outcome definition standardised or *a priori*?	Applicability: Patient Spectrum	Applicability: Test as per decision problem?
Beumer 2016[128] N=217 Lobular cancer	Lobular cancers, 5 **pooled** series[47, 117, 123, 129]	V	N, cohorts, some CT	UC	UC	Y	Most (6% ER-, 8% HER2+, 9% LN>3)	Y
Bueno-de-Mesquita 2011[127] N=139	**Pooled** 2 series: van de Vijver 2002;[47] Bueno-de-Mesquita 2009[63]	V[a]	Y, consecutive cohorts, no CT	UC	Y	Y	Most (all ER+, all LN0, 86% HER2-)	Y
Bueno-de-Mesquita 2009[63] N=123+151	1) Bueno-de-Mesquita 2009[63] 2) van de Vijver 2002[47]	V[a]	N, consecutive cohorts, some CT	N InT	Y	Y	N (24%+7% ER-, 7% HER2 or NR)	Y
Buyse 2006[64] N=302+151	1) TRANSBIG[64] 2) van de Vijver 2002[47]	V[a]	Y, retrospective cohort, no CT	N RNA qual, missing data	UC	Y	N (ER- 30%, HER2 NR)	Y
Drukker 2014[67] van de Vijver 2002[47] N=295	- van de Vijver 2002[47]	V[a] (21% also in derivation set)	N, retrospective, some CT	Y	UC	Y	N (23% ER-, HER2 NR, 13% LN>3)	Y
van 't Veer 2017[53] Esserman 2016[119] Lindstrom 2015[120]	Stockholm Tamoxifen (STO-3) trial: ER+ subgroup	V	Y, reanalysis of RCT, no CT	N InT, TF	UC	Y	Most (HER2 NR)	Y
Ishitobi 2010[86]	Osaka Medical Centre	V	N, case series, some CT	N Lack of RNA, TF	Y	Y	N (49% ER-, HER2 NR)	Y
Knauer 2010[126] N=541	**Pooled** 6 series[47, 60, 63, 65, 117, 123]	V[a]	N, cohorts, some CT	UC	Y	Y	Most (10% ER-, 11% HER2+)	Y
Kok 2012[117] 1) N=121 2) N=100+51	1) Kok 2009[133] 2) **Pooled** 2 series: - van de Vijver 2002[47] - Mook 2010[65] (55-71)	V[a]	Y, consecutive cohorts, no CT	UC	UC	Y	Most (HER2 NR; LN>3 18% (1) and 2% (2))	Y

Reference(s); N	Cohort(s)	Derivation or validation?	Study design appropriate?	All eligible patients included?	Blinding (of test assessors to outcomes)?	Outcome definition standardised or *a priori*?	Applicability: Patient Spectrum	Applicability: Test as per decision problem?
Mook 2010[122] N=964	**Pooled** 7 series[47, 60, 63-65, 117, 123]	V[a]	N, cohorts, some CT	N TF, MD	Y	Y	UC (16% ER-, 9% HER2+, 23% HER2 unknown, LN>3 % NR)	Y
Mook 2010[65] N=148	NKI 1984-96[65]	V	Y, consecutive cohort, no CT	N InT, RNA qual, MD	Y	Y	N (22% ER-, HER2 NR)	Y
Mook 2009[60] N=241+106	1) NKI+Italy[60] 2) van de Vijver 2002[47]	V[a]	N, retrospective, 56% + 70% CT	N InT, RNA qual	Y	Y	N (21%+18% ER-, 16% HER2+)	Y
Wittner 2008[66] N=100	Massachusetts, USA	V	N, retrospective, some CT	UC	UC	Y	N (20% ER-, HER2 NR)	Y
Yao 2015[118] N=238	NorthShore & Fox Chase	V	N, retrospective, some CT	UC	Y	Y	Most (for HR+ HER2- subgroup; LN NR)	Y

Y, yes; N, no; UC, unclear
D, Development; InT, insufficient tissue; MD, missing data; MS, missing samples; LN, number of positive nodes; qRT-PCR, quantitative reverse transcription polymerase chain reaction; R-RCT, reanalysis of RCT; TF, test failure; V, validation
[a]van de Vijver 2002[47] included 61 patients from the derivation set

Table 29: Prognostic performance of MammaPrint: distant recurrence-free survival/interval (DRFS/DRFI)

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Test or comp.	% pts per grp		% DRFS/DRFI risk : 0-5 yr		% DRFS/DRFI risk: 0-10 yr		DRFS/DRFI: HR (95% CI)
						Low	High	Low	High	Low	High	
Pooled analyses of patient cohorts: LN status mixed												
Variable ET&CT												
Beumer 2016[128] N=217 **Lobular cancer**	Lobular cancers, 5 pooled series[47, 117, 123, 129]	94% ER+ 92% HER2-	LN0, 66% LN+, 34%	59% ET 22% CT	**MMP**	76	24	-	-	-	-	**0-10yr:** 3.31 (1.79, 6.12), p<0.001
		93% ER+ 93% HER2-	LN0	51% ET 12% CT	**MMP**	82	18	-	-	-	-	**0-10yr:** 7.81 (2.89, 21.07), p<0.001
Knauer 2010[126] N=541	Pooled 6 series[47, 60, 63, 65, 117, 123]	90% ER+, 89% HER2-	LN0, 49% LN1-3, 51%	All ET 42% CT	**MMP**	47	53	95	82	-	-	**0-5 yr:** 3.88 (1.99, 7.58), p<0.01
Mook 2010[122] N=964	Pooled 7 series[47, 60, 63-65, 117, 123]	84% ER+ 68% HER2-	LN0, 72% LN+, 27%	32% ET 22% CT	**MMP**	54	46	95	80	87	72	**0-10 yr:** 2.70 (1.88, 3.88), p<0.001
	T1 only	N=552		No ET No CT	**MMP**			96	78	86	70	**0-10 yr:** 2.90 (1.83, 4.79), p<0.001
Retrospective studies: LN status mixed												
Variable ET&CT												
Drukker 2014[67] N=295	- van de Vijver 2002[47]	77% ER+ HER2 NR	LN0, 51% LN1-3, 36% LN>3, 13%	14% ET 37% CT	**MMP**	39	61	94.7	58.5	82.0	50.0	**0-5 yr:** 9.6 (4.2, 22.1), p=NR **5-10yr:** 1.1 (0.5, 2.5), p=NR **10-15yr:** 1.2 (0.2, 6.0), p=NR **15-20yr:** 1.1 (0.1, 17.9), p=NR **20-25yr:** 0.3 (0, 2.9), p=NR **0-25yr:** 3.1 (2.02, 4.86), p<0.0001
Yao 2015[118] N=238	NorthShore & Fox Chase, USA	All HR+ All HER2-	LN0, 72% LN+, 28%	87% ET 43% CT	**MMP**	60	40	-	-	96	87	**0-10 yr:** 2.91 (0.97, 8.68), p=0.045
		HR+/-	LN+/-	No CT	**MMP**	61	39	-	-	98	85	-
Reanalyses of RCTs: LN0												
100% ET monotherapy												
van't Veer 2017[53] Esserman[119] ET: N=281	STO-3 trial: ER+ analysis	All ER+ HER2 NR	LN0	All ET No CT	**MMP**	71	29	-	-	93	85	-

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Test or comp.	% pts per grp		% DRFS/DRFI risk : 0-5 yr		% DRFS/DRFI risk: 0-10 yr		DRFS/DRFI: HR (95% CI)
						Low	High	Low	High	Low	High	
No ET&CT												
van't Veer 2017[53] Esserman[119] No ET: N=257	STO-3 trial: ER+ analysis	All ER+ HER2 NR	LN0	No ET No CT	**MMP**	67	33	-	-	83	70	-
Pooled analyses of patient cohorts: LN0												
No ET&CT												
Bueno-de-Mesquita 2011[127]	Pooled[47, 63] N=186	76% ER+ 76% HER2-	LN0	No ET/CT	**MMP**	45	55	-	-	88	55	-
Retrospective studies: LN0												
Variable ET&CT												
Bueno-de-Mesquita 2009[63] N=123	- Bueno-de-Mesquita 2009[63]	76% ER+ 93% HER2-	LN0	22% ET 25% CT	**MMP**	52	48	98	78	-	-	**0-5 yr:** 5.7 (1.6, 20), p=0.007
					AOL	-	-	-	-	-	-	**0-5 yr:** 4.6 (0.61, 35.1), p=0.14
					NPI	-	-	-	-	-	-	**0-5 yr:** 2.2 (0.78, 6.5), p=0.14
Buyse 2006[64] Co. submission[121]	- TRANSBIG[64] N=302	70% ER+ HER2 NR	LN0	No ET/CT	**MMP**	37	63	-	-	90	71	**DRFI all FU (med 13.6yr)** 2.32 (1.35, 4.00), p=0.002
					AOL							1.68 (0.92, 3.07), p=0.092
					NPI							1.65 (1.02, 2.66), p=0.043
Drukker 2014[67] Bueno-de-Mesquita 2009[63] N=151	- van de Vijver 2002[47]	72% ER+ HER2 NR	LN0	4% ET 4% CT	**MMP**	40	60	94.9[67]	52.4[67]	86[63]	50[63]	**0-10 yr:** 5.5 (2.5, 12), p<0.001[63] **0-25yr:** 4.57 (2.31, 9.04); p<0.0001[67]
					AOL	-	-	-	-	-	-	**0-10 yr:** 1.7 (0.84, 3.6), p=0.14
					NPI	-	-	-	-	-	-	**0-10 yr:** 3.1 (1.6, 5.9), p<0.001
Ishitobi 2010[86] N=102	Osaka Medical Centre	51% ER+ HER2 NR	LN0	73% ET 28% CT	**MMP**	20	80	100	94	-	-	-
Mook 2010[65] N=148	- NKI 1984-96[65] (55-71yr)	78% ER+ HER2 NR	LN0	18% ET No CT	**MMP**	61	39	93	72	80	67	**0-5 yr:** 4.6 (1.8, 12.0), p=0.001 **0-10 yr:** Data per group but p-values NR
Wittner 2008[66] N=100	Massachusetts USA	80% ER+ HER2 NR	LN0	24% ET 21% CT	**MMP**	27	73	-	-	-	-	**DRFI:** **0-5 yr:** PPV=12%, NPV=100%, p=0.192 **0-10 yr:** PPV=14%, NPV=100%, p=0.330
Retrospective studies: LN+												
Variable ET&CT												
Drukker 2014[67] N=144	- van de Vijver 2002[47]	ER+/-, HER2 NR	LN1-3, 74% LN>3, 26%	Some ET Some CT	**MMP**	38	62	94.5	64.7	78.6	54.3	**0-25yr:** 2.24 (1.25, 4.00); p=0.01

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Test or comp.	% pts per grp		% DRFS/DRFI risk : 0-5 yr		% DRFS/DRFI risk: 0-10 yr		DRFS/DRFI: HR (95% CI)
						Low	High	Low	High	Low	High	
Mook 2009[60] N=241	- NKI+Italy[60]	79% ER+, 84% HER2-	LN1-3	73% ET 56% CT	**MMP**	41	59	98	80	91	76	**0-10yr:** 4.13 (1.72, 9.96), p=0.002

-, not reported; AOL, Adjuvant! Online; CI, confidence interval; comp, comparator; CT, chemotherapy; DMFS, distant metastasis-free survival; DRFI, distant recurrence free interval; DRFS, distant recurrence free survival; ER, oestrogen receptor; ET, endocrine therapy; FU, follow-up; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; LN, number of positive nodes; MMP, MammaPrint; NPI, Nottingham Prognostic Index; NPV, negative predictive value; NR, not reported; PPV, positive predictive value; St G, St Gallen; TTDM, time to distant metastasis.

Table 30: Prognostic performance of MammaPrint: Overall survival

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Test or comp.	% pts per group		% OS risk: 0-5 yr		% OS risk: 0-10 yr		OS: HR (95% CI) (unless stated otherwise)
						Low	High	Low	High	Low	High	
Pooled analyses of patient cohorts: LN status mixed												
Variable ET&CT												
Beumer 2016[128] N=217 **Lobular cancer**	Lobular cancers, 5 pooled series[47, 117, 123, 129]	94% ER+ 92% HER2-	LN0, 66% LN+, 34%	59% ET 22% CT	**MMP**	76	24	-	-	-	-	**0-10yr:** 3.58 (1.84, 6.95), p<0.001
		93% ER+ 93% HER2-	LN0	51% ET 12% CT	**MMP**	82	18	-	-	-	-	**0-10yr:** 7.47 (2.58, 21.58), p<0.001
Retrospective studies: LN status mixed												
Variable ET&CT												
Drukker 2014[67] N=295	- van de Vijver 2002[47]	77% ER+ HER2 NR	LN0, 51% LN1-3, 36% LN>3, 13%	14% ET 37% CT	**MMP**	39	61	97.4	74.0	92.8	55.7	**0-5yr:** 11.3 (3.5, 36.4), p=NR **5-10yr:** 6.1 (2.4, 15.6), p=NR **10-15yr:** 1.5 (0.6, 3.5), p=NR **15-20yr:** 0.6 (0.2, 1.7), p=NR **20-25yr:** 0.2 (0, 2.1), p=NR **0-25yr:** 2.9 (1.90, 4.28), p<0.0001
Pooled analyses of patient cohorts: LN0												
No ET&CT												
Bueno-de-Mesquita 2011[127]	Pooled[47, 63] N=186	76% ER+ 76% HER2-	LN0	No ET/CT	**MMP**	45	55	-	-	91	56	
Retrospective studies: LN0												
Variable ET&CT												
Bueno-de-Mesquita 2009[63] N=123	- Bueno-de-Mesquita 2009[63]	76% ER+ 93% HER2-	LN0	22% ET 25% CT	**MMP**	52	48	97	82	-	-	**0-5yr:** 3.4 (1.2, 9.6), p=0.021
					AOL	-	-	-	-	-	-	**0-5yr:** 2.5 (0.59, 11), p=0.22
					NPI	-	-	-	-	-	-	**0-5yr:** 2.8 (0.99. 7.8), p=0.053
Buyse 2006[64] N=302	- TRANSBIG[64]	70% ER+ HER2 NR	LN0	No ET/CT	**MMP**	37	63	-	-	-	-	**All (med 13.6yr):** 2.79 (1.60, 4.87), p<0.001 **C-index (AUC)** 0.648
					AOL							**All (med 13.6yr):** 1.67 (0.93, 2.98), p=0.085 **C-index (AUC)** 0.576
					NPI							**All (med 13.6yr):** 1.49 (0.94, 2.36), p=0.092

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Test or comp.	% pts per group		% OS risk: 0-5 yr		% OS risk: 0-10 yr		OS: HR (95% CI) (unless stated otherwise)
						Low	High	Low	High	Low	High	
Drukker 2014[67] Bueno-de-Mesquita 2009[63] N=151	- van de Vijver 2002[47]	72% ER+ HER2 NR	LN0	4% ET 4% CT	**MMP**	40	60	96.7[67]	71.1[67]	94[63]	51[63]	**0-10yr:** 10.7 (3.9., 30), p<0.001[63] **0-25yr:** 4.73 (2.46, 9.07); p<0.0001[67]
					AOL	-	-	-	-	-	-	**0-10yr:** 2.8 (1.2, 6.6), p=0.017[63]
					NPI	-	-	-	-	-	-	**0-10yr:** 3.4 (1.8, 6.6), p<0.001[63]
Retrospective studies: LN+												
Variable ET&CT												
Drukker 2014[67] N=144	- van de Vijver 2002[47]	ER+/- HER2 NR	LN1-3, 74% LN>3, 26%	Some ET/CT	**MMP**	38	62	98.2	76.9	92.5	58.7	**0-5yr and 0-10yr:** HRs not reported **0-25yr:** 1.83 (1.07, 3.11); p=0.03
Mook 2009[60] N=241	- NKI+Italy[60]	79% ER+, 84% HER2-	LN1-3	73% ET 56% CT	**MMP**	41	59	-	-	-	-	**0-10yr:** 5.40 (2.11, 13.80), p<0.001

-, not reported; AOL, Adjuvant! Online; CI, confidence interval; comp, comparator; CT, chemotherapy; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; LN, number of positive nodes; MMP, MammaPrint; NPI, Nottingham Prognostic Index; OS, overall survival.

Table 31: Prognostic performance of MammaPrint: Other outcomes

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Outcome	Test or comp.	% pts per group		% risk of outcome: 0-5yr		% risk of outcome: 0-10yr		HR (95% CI)
							Low	High	Low	High	Low	High	
Pooled analyses of patient cohorts: LN status mixed													
Variable ET&CT													
Knauer 2010[126] N=541	Pooled 6 series[47, 60, 63, 65, 117, 123]	90% ER+ 89% HER2-	LN0, 49% LN1-3, 51%	All ET 42% CT	**BCSS**	**MMP**	47	53	97	87	-	-	**0-5 yr:** 4.81 (1.98, 11.67), p<0.01
Mook 2010[122] N=964	Pooled 7 series[47, 60, 63-65, 117, 123]	84% ER+ 68% HER2-	LN0, 72% LN+, 27%	32% ET 22% CT	**BCSS**	**MMP**	54	46	99	88	91	72	**0-10 yr:** 4.22 (2.70, 6.60), p<0.001
		N=552	LN+/-	No ET/CT					99	85	91	69	**0-10 yr:** 4.67 (2.67, 8.18), p<0.001
No ET&CT													
Kok 2012[117] N=100+51	Pooled 2 series: van de Vijver[47] + Mook 2010[65]	ER+ HER2 NR	LN0, 91% LN+, 9%	No ET/CT	**BCSS**	**MMP**	56	44	97.6	80.9	90.2	63.3	**0-10 yr:** 4.52 (2.01, 10.2), p<0.001
Retrospective studies: LN status mixed													
100% ET monotherapy													
Kok 2012[117] N=121	NKI 1985-94[133]	ER+ HER2 NR	LN0, 18% LN+, 82%	All ET No CT	**BCSS**	**MMP**	69	31	96.2	72.5	80.6	63.4	**0-10 yr:** 2.78 (1.30, 5.94), p=0.008
Reanalyses of RCTs: LN0													
Variable ET&CT													
Esserman[119] [redacted]	STO-3 trial: [redacted]	All ER+ HER2 NR	LN0	[redacted]	**BCSS**	[redacted]	[redacted]	[redacted]	-	-	-	-	[redacted]
Retrospective studies: LN0													
Variable ET&CT													
Mook 2010[65] N=148	- NKI 1984-96[65] (55-71yr)	78% ER+ HER2 NR	LN0	18% ET No CT	**BCSS**	**MMP**	61	39	99	80	90	69	**0-5 yr:** 19.1 (2.5, 148), p=0.005 **0-10 yr:** 3.9 (CI NR), p=NR
						AOL	50	50	-	-	-	-	**0-5 yr:** 5.3 (CI NR) **0-10 yr:** 6.2 (2.1, 18.0), p=0.001
No ET&CT													
Buyse 2006[64] N=302	- TRANSBIG[64]	70% ER+ HER2 NR	LN0	No ET/CT	**DFS**	**MMP**	37	63	-	-	-	-	**All FU (med 13.6yr):** 1.50 (1.04, 2.16), p=0.032

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Outcome	Test or comp.	% pts per group		% risk of outcome: 0-5yr		% risk of outcome: 0-10yr		HR (95% CI)
							Low	High	Low	High	Low	High	
						AOL							1.30 (0.86, 1.95), p=0.21
						NPI							1.10 (0.78, 1.56), p=0.58
Retrospective studies: LN+													
Variable ET&CT													
Mook 2009[60] N=241	- NKI+Italy[60]	79% ER+ 84% HER2-	LN1-3	73% ET 56% CT	**BCSS**	**MMP**	41	59	99	88	96	76	**0-10 yr:** 5.70 (2.01, 16.23), p=0.001
		All ER+ N=191	LN1-3	Some ET/CT			NR	NR	-	-	-	-	**0-10 yr:** 9.75 (2.26, 42.01), p=0.002
		N=101	LN1-3	No CT			NR	NR	-	-	-	-	**0-10 yr:** 7.33 (1.61, 33.49), p=0.01
		N=166	LN1-3	All ET			NR	NR	-	-	-	-	**0-10 yr:** 3.63 (1.21, 10.94), p=0.02
van de Vijver 2002;[47] Mook 2009[60] N=106	- van de Vijver 2002[47]	82% ER+ 84% HER2-	LN1-3	23% ET 70% CT	**BCSS**	**MMP**	41	59	-	-	98	64	**0-10 yr:** 6.60 (1.97, 22.10), p=0.002

-, not reported; AOL, Adjuvant! Online; BCSS, breast cancer-specific survival; CI, confidence interval; CT, chemotherapy; DFS, disease-free survival; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; LN, number of positive nodes; MMP, MammaPrint.

Table 32: Prognostic performance of MammaPrint for patients at high or low clinical risk

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Outcome	Test or comp.	% pts per grp		% risk of outcomes: 0-5 yr		% risk of outcomes: 0-10 yr		HR (95% CI)
							Low	High	Low	High	Low	High	
Low-risk via AOL+NPI+St Gallen (LN0)													
Bueno-de-Mesquita 2011[127]	Pooled[47, 63] N=139	ER+, Low clin[b]	LN0	No ET/CT	**DRFI**	**MMP**	77	23	-	-	87	70	**0-10yr:** HR NR, p=0.19
					OS	**MMP**	77	23	-	-	100	86	**0-10yr:** HR NR, p=0.016
Discordant risk via AOL+NPI+St Gallen (LN0)													
Bueno-de-Mesquita 2011[127]	Pooled[47, 63] N=139	ER+, Discordant clin[b]	LN0	No ET/CT	**DRFI**	**MMP**	66	34	-	-	91	63	**0-10yr:** HR NR, p=0.004
					OS	**MMP**	66	34	-	-	88	58	**0-10yr:** HR NR, p=0.06
High-risk via AOL+NPI+St Gallen (LN0)													
Bueno-de-Mesquita 2011[127]	Pooled[47, 63] N=139	ER+, High clin[b]	LN0	No ET/CT	**DRFI**	**MMP**	27	73	-	-	77	45	**0-10yr:** HR NR, p=0.19
					OS	**MMP**	27	73	-	-	77	53	**0-10yr:** HR NR, p=0.17
High-risk via AOL (LN+)													
Mook 2009[60] N=209	- NKI+Italy[60]	High-risk AOL	LN1-3	Some ET Some CT	**BCSS**	**MMP**	NR	NR	-	-	94	76	**0-10 yr:** 4.12 (1.45, 11.76), p=0.008

-, not reported; AOL, Adjuvant! Online; CI, confidence interval; comp, comparator; CT, chemotherapy; DMFS, distant metastasis-free survival; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; LN, number of positive nodes; MMP, MammaPrint; NPI, Nottingham Prognostic Index; NR, not reported; St G, St Gallen.

Table 33: Additional prognostic value for DRFS/DRFI: MammaPrint

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Test or comparator[a]	C-index (AUC) (95% CI)	Increase in LR χ^2 over CP factors[a]	Multivariable model adj. for CP factors[a], AOL[b] or NPI[c]: HR (95% CI)
Pooled analyses of patient cohorts: LN status mixed								
Variable ET&CT								
Beumer 2016[128] N=217 **Lobular cancer**	Lobular cancers, 5 pooled series[47, 117, 123, 129]	94% ER+ 92% HER2-	LN0, 66% LN+, 34%	59% ET 22% CT	**MMP**			**10yr:** 2.08 (1.05, 4.14), p=0.037[a]
		93% ER+ 93% HER2-	LN0	51% ET 12% CT	**MMP**			**10yr:** 6.40 (2.14, 19.17), p=0.001[a]
Mook 2010[122] N=941	Pooled 7 series[47, 60, 63-65, 117, 123]	84% ER+ 68% HER2-	LN0, 72% LN+, 27%	32% ET 22% CT	**MMP**			**10yr:** 2.43 (1.56. 3.77), p<0.001[a]
		All ER+ (n=788)	LN+/-	Some ET/CT	**MMP**			**10yr:** 2.51 (1.60, 3.95), p<0.001[a]
		N=552	LN+/-	No ET/CT	**MMP**			**10yr:** 2.54 (1.49, 4.34), p=0.001[a]
Retrospective studies: LN status mixed								
Variable ET&CT								
van de Vijver[47] N=295	- van de Vijver[47]	77% ER+ HER2 NR	LN0, 51% LN+, 49%	14% ET 37% CT	**MMP**			**10yr:** 4.6 (2.3–9.2), p<0.001[a]
Yao 2015[118] N=373	NorthShore & Fox Chase	74% ER+ 83% HER2-	LN0, 72% LN+, 28%	65% ET 58% CT	**MMP**			**10yr:** 3.01 (0.88, 10.33), p=0.08[a]
Pooled analyses of patient cohorts: LN0								
No ET&CT								
Bueno-de-Mesquita 2011[127] N=186	Pooled[47, 63]	76% ER+ 76% HER2-	LN0	No ET/CT	**MMP**		Change log likelihood p<0.001	
Retrospective studies: LN0								
Variable ET&CT								
Bueno-de-Mesquita 2009[63] N=123	- Bueno-de-Mesquita[63]	76% ER+ 93% HER2-	LN0	22% ET 25% CT	**MMP**	CP: 0.66 (0.50 to 0.82) CP+MMP: 0.75 (0.61 to 0.89) MMP: 0.69 (0.56 to 0.82)	Change log likelihood 5.5, p=0.023	**5yr:** 4.8 (1.3, 17), p=0.018[b] 5.4 (1.4, 21), p=0.015[c]
van de Vijver 2002;[47] Bueno-de-Mesquita 2009;[63] Buyse 2006[64] N=151	- van de Vijver 2002[47]	72% ER+ HER2 NR	LN0	4% ET 4% CT	**MMP**	CP: 0.70 (0.61 to 0.79) CP+MMP: 0.76 (0.68 to 0.85) MMP: 0.68 (0.60 to 0.77)	Change log likelihood 15.8, p<0.01	**10yr:**[63] 5.3 (2.4, 12), p<0.001[b] 4.3 (1.8, 10), p=0.001[c] **All FU (med 6.7yr):** [64] 6.07 (2.64, 13.98)[b]

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Test or comparator[a]	C-index (AUC) (95% CI)	Increase in LR χ^2 over CP factors[a]	Multivariable model adj. for CP factors[a], AOL[b] or NPI[c]: HR (95% CI)
No ET&CT								
Buyse 2006[64] N=302	- TRANSBIG[64]	70% ER+ HER2 NR	LN0	No ET/CT	**MMP**			**5yr:** 4.68 (CI NR)[b] **10yr:** 3.5 (CI NR)[b] **All FU (med 13.6yr):** 2.13 (1.19, 3.82);[b] 2.15 (1.19 , 3.92)[c]
Retrospective studies: LN+								
Variable ET&CT								
Mook 2009[60] N=241	- NKI+Italy[60]	79% ER+ 84% HER2-	LN1-3	73% ET 56% CT	**MMP**			**10yr:** 2.99 (0.996, 8.99), p=0.051 [a]

AOL, Adjuvant! Online; CI, confidence interval; CP, clinical/pathological; CT, chemotherapy; DRFI, distant metastasis-free interval; DMFS, distant metastasis-free survival; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; LN, number of positive nodes; LR, likelihood ratio; MMP, MammaPrint; NPI, Nottingham Prognostic Index; TTDM, time to distant metastasis.
[a]Adjusted for: van de Vijver 2002;[47]age, lymph node status, tumour size, grade, vascular invasion, ER status, Surgery type, chemotherapy and endocrine therapy; Mook 2009 (LN1-3) + Mook 2010 (pooled): age, tumour size, nodal status, grade, ER, HER2, surgery, endocrine therapy, chemotherapy; Bueno-de-Mesquita 2011+2009: age, tumour size, grade, ER, PR, HER2; Yao 2015: age, tumour size, grade, ER, HER2; Beumer 2016: age, nodal status, grade, ER, HER2, chemotherapy (similar results when only adjusting for CP factors associated with MMP outcome). [b]Adjusted for AOL. [c]Adjusted for NPI.

Table 34: Additional prognostic value for overall survival: MammaPrint

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Test or comparator[a]	Likelihood ratio χ^2	Increase in LR χ^2 over CP factors[a]	Multivariable model adj. for CP factors[a], AOL[b] or NPI[c] : HR (95% CI)
Pooled analyses of patient cohorts: LN status mixed								
Variable ET&CT								
Beumer 2016[128] N=217 **Lobular cancer**	Lobular cancers, 5 pooled series[47, 117, 123, 129]	94% ER+ 92% HER2-	LN0, 66% LN+, 34%	59% ET 22% CT	**MMP**			**10yr:** 2.02 (0.94, 4.30), p=0.070[a]
		93% ER+ 93% HER2-	LN0	51% ET 12% CT	**MMP**			**10yr:** 5.10 (1.52, 17.17), p=0.008[a]
Pooled analyses of patient cohorts: LN0								
No ET&CT								
Bueno-de-Mesquita 2011[127] N=186	Pooled[47, 63]	76% ER+ 76% HER2-	LN0	No ET/CT	**MMP**		Change log likelihood p=0.005	
Retrospective studies: LN0								
Variable ET&CT								
Bueno-de-Mesquita 2009[63] N=123	- Bueno-de-Mesquita[63]	76% ER+ 93% HER2-	LN0	22% ET 25% CT				**5yr:** 3.0 (1.0, 8.9), p=0.044[b] 2.7 (0.87, 8.1), p=0.086[c]
van de Vijver 2002;[47] Bueno-de-Mesquita 2009;[63] Buyse 2006[64] N=151	- van de Vijver 2002[47]	72% ER+ HER2 NR	LN0	4% ET 4% CT	**MMP**		Change log likelihood 19.7, p<0.01	**10yr:**[63] 9.6 (3.4, 27), p<0.001[b] 8.5 (2.9, 25), p<0.001[c] **All FU (med 6.7yr):** [64] 17.46 (4.12, 74.00)[b]
No ET&CT								
Buyse 2006[64] N=302	- TRANSBIG[64]	70% ER+ HER2 NR	LN0	No ET/CT	**MMP**			**5yr:** 16.99 (CI NR)[b] **10yr:** 3.46 (CI NR)[b] **All (med 13.6yr):** 2.63 (1.45, 4.79);[b] 2.89 (1.58, 5.29)[c]

AOL, Adjuvant! Online; CI, confidence interval; CP, clinical/pathological; CT, chemotherapy; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; LN, number of positive nodes; LR, likelihood ratio; MMP, MammaPrint; NPI, Nottingham Prognostic Index; OS, overall survival.
[a]Adjusted for: Bueno-de-Mesquita 2011+2009: age, tumour size, grade, ER, PR, HER2; Beumer 2016: age, nodal status, grade, ER, HER2, chemotherapy (similar results when only adjusting for CP factors associated with MMP outcome). [b]Adjusted for AOL. [c]Adjusted for NPI.

Table 35: Additional prognostic value for other outcomes: MammaPrint

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Outcome	Test or comparator[a]	Likelihood ratio χ^2	Increase in LR χ^2 over CP factors[a]	Multivariable model adj. for CP factors[a], AOL[b] or NPI[c]: HR (95% CI)
Pooled analyses of patient cohorts: LN status mixed									
Variable ET&CT									
Mook 2010[122] N=964	Pooled 7 series[47, 60, 63-65, 117, 123]	84% ER+ 68% HER2-	LN0, 72% LN+, 27%	32% ET 22% CT	**BCSS 10yr**	**MMP**			HR 3.25 (1.92, 5.51), p<0.001[a]
		All ER+ (n=788)	LN+/-		**BCSS 10yr**	**MMP**			3.43 (1.98, 5.95), p<0.001[a]
			LN+/-	No ET/CT	**BCSS 10yr**	**MMP**			3.47 (1.83, 6.60), p<0.001[a]
No ET&CT									
Kok 2012[117] N=100+51	Pooled 2 series: van de Vijver[47] + Mook 2010[65]	ER+ HER2 NR	LN0, 91% LN+, 9%	No ET/CT	**BCSS 10yr**	**MMP**			2.56 (0.91, 7.17), p=0.074
Retrospective studies: LN status mixed									
100% ET monotherapy									
Kok 2012[117] N=121	NKI 1985-94[133]	ER+ HER2 NR	LN0, 18% LN+, 82%	All ET No CT	**BCSS 10yr**	**MMP**			1.88 (0.77, 4.61), p=0.17
Retrospective studies: LN0									
Variable ET&CT									
Mook 2010[65] N=148	- NKI 1984-96[65] (55-71yr)	78% ER+ HER2 NR	LN0	18% ET No CT	**BCSS**	**MMP**			**5yr:** 14.4 (1.7, 122), p=0.01[b] **10yr:** 2.2 (CI NR)[b]
van de Vijver 2002;[47] Bueno-de-Mesquita 2009;[63] Buyse 2006[64]	- van de Vijver 2002[47] N=151	72% ER+ HER2 NR	LN0	4% ET 4% CT	**DFS**	**MMP**			**All FU (med 6.7yr):** [64] 4.80 (2.37, 9.71)[b]
No ET&CT									
Buyse 2006[64] N=302	- TRANSBIG[64]	70% ER+ HER2 NR	LN0	No ET/CT	**DFS**	**MMP**			**5yr:** 2.16 (CI NR)[b] **10yr:** 1.66 (CI NR)[b] **All (med 13.6yr):** 1.36 (0.91, 2.03);[b] 1.45 (0.97, 2.16)[c]

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Outcome	Test or comparator[a]	Likelihood ratio χ^2	Increase in LR χ^2 over CP factors[a]	Multivariable model adj. for CP factors[a], AOL[b] or NPI[c]: HR (95% CI)
Retrospective studies: LN+									
Variable ET&CT									
Mook 2009[60] N=241	- NKI+Italy[60]	79% ER+ 84% HER2-	LN1-3	73% ET 56% CT	**BCSS 10yr**	**MMP**			7.17 (1.81, 28.43), p=0.005[a]
van de Vijver 2002;[47] Mook 2009[60] N=106	- van de Vijver 2002[47]	82% ER+ 84% HER2-	LN1-3	23% ET 70% CT	**BCSS 10yr**	**MMP**			3.63 (0.88, 14.96), p=0.07[a]

AOL, Adjuvant! Online; BCSS, breast cancer-specific survival; CI, confidence interval; CP, clinical/pathological; CT, chemotherapy; DFS, disease-free survival; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; LN, number of positive nodes; LR, likelihood ratio; MMP, MammaPrint; NPI, Nottingham Prognostic Index.
[a]Adjusted for: Mook 2009 (LN1-3) + Mook 2010 (pooled): age, tumour size, nodal status, grade, ER, HER2, surgery, endocrine therapy, chemotherapy. [b]Adjusted for AOL. [c]Adjusted for NPI.

4.4.3 Chemotherapy benefit: MammaPrint

Study designs and patients

Two references have reported the ability of MammaPrint to predict the benefit of chemotherapy, i.e. whether the relative effect of chemotherapy differs between MammaPrint risk groups. The article by Knauer *et al*.[126] (2010) reported a pooled analysis of 541 patients, of whom 100% received endocrine therapy and 42% recevied chemotherapy, from six consecutive patient series as detailed in Table 36. Overall, 90% were ER+, 89% HER2-, and half were LN0 while half had 1-3 positive nodes (LN1-3). This publication did not report separate analyses for LN0 and LN+ groups.

Additionally, the article by Mook *et al*.[60] (2009) reported a pooled analysis of two of the six patient series from Knauer *et al*.[126] (Table 36), with an extended follow-up (10 years), but restricted to LN1-3 patients (including micrometastases).

Quality assessment

Table 37 presents the quality assessment of studies assessing MammaPrint prediction of chemotherapy benefit. There were no reanalyses of RCTs assessing chemotherapy benefit. Both studies used pooled retrospective cohorts, where patients were treated according to usual practice (in addition, one of the six cohorts in Knauer[126] was the prospective RASTER study[123] where patients were treated according to usual practice plus MammaPrint). As such, those who received chemotherapy are likely to be systematically different in terms of known (and potentially unknown) prognostic and clinical factors (e.g. age, nodal status) to those who did not, leading to a high risk of confounding. Both studies blinded the test assessors to clinical outcomes, and both used standard outcome definitions. Both studies included a proportion of patients outside the scope (ER- and/or HER2+).

Results

The pooled analysis of six consecutive series by Knauer *et al*.[126] (2010) reported that at 5 years, there was a statistically significant effect of chemotherapy in the MammaPrint high-risk group but no statistically significant effect in the low-risk group, though HRs favoured chemotherapy in both groups (Table 38). Unadjusted HRs for DRFS (for no chemotherapy vs. chemotherapy) were 0.26 (95% CI: 0.03, 2.02, p=0.20) in the low-risk group and 0.35 (95% CI: 0.17, 0.71, p<0.01) in the high-risk group, while unadjusted HRs for BCSS were 0.58 (95% CI: 0.07, 4.98, p=0.62) in the low-risk group and 0.21 (95% CI: 0.07, 0.59, p<0.01) in the high-risk group. Multivarible analyses of the effect of chemotherapy on 5 year BCSS were again statistically significant in the high-risk group (HR 0.21, 95% CI: 0.06, 0.80, p=0.02) but not the low-risk group (HR not estimable, p=0.98) (Table 38). However, the interaction test for chemotherapy treatment and risk group was not statistically significant (p=0.45; the interaction test appears to relate to 5-year BCSS as opposed to DRFS but this

is unclear in the publication). This indicates that the effect of chemotherapy versus no chemotherapy on 5-year BCSS was not statistically significantly different between risk groups. It is unclear whether this interaction test relates to the adjusted or unadjusted analysis.

For the two pooled LNmicro-3 cohorts reported by Mook *et al*., 2009[60] (these were subsets of two of the six cohorts pooled in Knauer *et al.*[126]), the only evidence relating to prediction of chemotherapy benefit was a test of the interaction between chemotherapy treatment and risk group (within a multivariable analysis of 10-year BCSS), which was not statistically significant (p=0.95, Table 38).

Discussion: MammaPrint chemotherapy benefit

Prediction of chemotherapy benefit for MammaPrint was reported within a pooled analysis of 541 patients across six patient series (half LN0, half LN1-3).[126] The effect of chemotherapy versus no chemotherapy on 5-year DRFS and BCSS was statistically significant in the MammaPrint high-risk group but not in the low-risk group in unadjusted analyses for 5-year DRFS and BCSS and in analyses for 5-year BCSS adjusted for clinicopathological variables (not reported for DRFS). However, the interaction test for chemotherapy treatment and risk group (for 5 year BCSS) was non-significant (p=0.45). A further pooled analysis of two of the above series, with follow-up to 10 years but restricted to LN1-3 patients, also reported a statistically non-significant interaction between chemotherapy treatment and risk group for 10-year BCSS (p=0.95).[60]

Both studies used pooled retrospective cohorts where patients were treated according to usual practice (or usual practice plus MammaPrint within RASTER,[123] one of the six pooled cohorts). As such, those who received chemotherapy are likely to be systematically different in terms of known (and unknown) prognostic and clinical factors to those who did not, leading to a high risk of confounding. In the analysis of six series,[126] it was unclear whether the interaction test was unadjusted or adjusted, and if so for which factors. In the analysis of LN1-3 patients from two series,[60] the interaction test was conducted within a multivariable analysis adjusted for clinicopathological variables.

Conclusions: MammaPrint chemotherapy benefit

Prediction of chemotherapy benefit for MammaPrint was reported within a pooled analysis of 541 patients within six non-randomised patient series (half LN0, half LN1-3) in which patients were treated according to usual practice. The effect of chemotherapy versus no chemotherapy on 5-year DRFS and BCSS was statistically significant in the MammaPrint high-risk group but not in the low-risk group in unadjusted analyses for 5-year DRFS and BCSS and in adjusted analyses for 5-year BCSS. However, the interaction test for chemotherapy treatment and risk group (for 5 year BCSS) was non-significant (p=0.45). A further pooled analysis of two of the above series, restricted to LN1-3 patients, also reported a statistically non-significant interaction between chemotherapy treatment and

risk group for 10-year BCSS (p=0.95). The evidence for the ability of MammaPrint to predict chemotherapy benefit is extremely limited; although unadjusted analyses suggest a greater effect of chemotherapy in high-risk groups, adjusted analyses were only reported for one outcome, and the non-significant interaction tests suggest there was no statistically significant difference in effect of chemotherapy between risk groups.

Table 36: Characteristics of chemotherapy benefit studies: MammaPrint

Reference(s)	Cohort(s)	N	Country	Study design	Test	Details of test	Cut-offs	Population	Nodal status	ET / CT
Knauer 2010[126]	Pooled 6 series: - van de Vijver 2002[47] - Bueno-de-Mesquita 2009[63] - Mook 2009[60] (LN1-3) - Mook 2010[65] (age 55-71) - Bueno-de-Mesquita 2007[123] (RASTER) - Kok (personal com.)	541	Various	Pooling of 6 consecutive cohorts	**MMP**	Frozen MMP microarray	Low, high (details NR)	90% ER+ 89% HER2- Pre/post-meno % female NR pT1-3	LN0, 49% LN1-3, 51%	All ET 42% CT
Mook 2009[60]	1) NKI+Italy 1994-2001[60] 2) van de Vijver 2002[47]	1) 241 2) 106	Neths, Italy	Retrospective, consecutive	**MMP**	Frozen MMP microarray	Low (good), high (poor); details NR	1) 79% ER+ 84% HER2- 2) 82% ER+ 84% HER2- All: Pre/post-meno, age ≤70 % female NR	LN1-3 (inc. micromets)	1) 73% ET (low 82%, high 65%); 56% CT (low 41%, high 67%) 2) 23% ET (low 26%, high 21%); 70% CT (low 77%, high 65%)

CT, chemotherapy; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; HR+, hormone-receptor positive; LN, number of positive nodes; MMP, MammaPrint; NKI, Netherlands Cancer Institute; NR, not reported; qRT-PCR, quantitative reverse transcription polymerase chain reaction; RASTER, MicroarRAy PrognoSTics in Breast CancER study; RdGG, Reinier de Graaf Hospital

Table 37: Quality assessment of studies predicting chemotherapy responsiveness: MammaPrint

Reference(s)	Cohorts	Derivation or validation?	Study design appropriate?	All eligible patients included?	Blinding (of test assessors to outcomes)	Definition of outcome standardised or a priori?	Applicability: Patient Spectrum	Applicability: Test as per decision problem?
Knauer 2010[126]	Pooled 6 series: - NKI, van de Vijver 2002[47] - Bueno-de-Mesquita 2009[63] - Mook 2009[60] (LN1-3), NKI+EIO (Italy) - Mook 2010[65] (age 55-71) - Bueno-de-Mesquita 2007[123] (RASTER) - Kok (personal com.)	V	N, not RCT data, pooled cohorts,	NR	Y	Y	Most (10% ER-, 11% HER2+)	Y
Mook 2009[60] (LN1-3)	- NKI+EIO (Italy)[60] - NKI, van de Vijver 2002[47]	V	N, not RCT data,retrospective cohort,	N InT, RNA quality	Y	Y	N (21%+18% ER-, 16% HER2+)	Y

Y, Yes; N, No; UC, unclear; R-RCT, Reanalysis of RCT; InT, insufficient tissue; TF, test failure; MS, missing samples; D, Development; V, validation; CT, chemotherapy; NKI, Netherlands Cancer Institute-Antoni van Leeuwenhoek hospital, Amsterdam; European Institute of Oncology, Italy

Table 38: Prediction of chemotherapy responsiveness: MammaPrint

Reference(s); Design; Country	Cohorts	Population	% patients in each risk group		Outcome	Low-risk			High-risk			Adjusted HRs[a]	Interaction tests
			Low	High		No CT: % risk	CT: % risk	HR (95% CI)	No CT: % risk	CT: % risk	HR (95% CI)		
Knauer 2010[126] Pooled cohorts N=541	Pooled 6 series: - NKI, van de Vijver 2002[47] - Bueno-de-Mesquita 2009[63] - Mook 2009[60] (LN1-3), NKI+EIO (Italy) - Mook 2010[65] (age 55-71) - Bueno-de-Mesquita 2007[123] (RASTER) - Kok (personal com.)	90% ER+ 89% HER2- All ET 42% CT LN0, 49% LN1-3, 51%	47	53	**DRFS 5 yr**	93	99	0.26 (0.03, 2.02), p=0.20	76	88	0.35 (0.17, 0.71), p<0.01	NR	NR
					BCSS 5 yr	97	99	0.58 (0.07, 4.98), p=0.62	81	94	0.21 (0.07, 0.59), p<0.01	No CT vs CT[a] (95% CI): **Low:** not estimable, p=0.98 **High:** 0.21 (0.06, 0.80), p=0.02	Interaction[b] (risk group+CT): p=0.45
Mook 2009[60] (LNmicro-3) Retrospective N=347	1. NKI+EIO (Italy)[60] (n=241) 2. NKI, van de Vijver 2002[47] (n=106)	79/82% ER+ 84% HER2- 73/23% ET 56/70% CT LNmicro-3	41	59	**BCSS 10yr**	NR	NR	NR	NR	NR	NR		Interaction (risk group+chemo, series 1+2 pooled, multivariable[a]): p=0.95

BCSS, breast cancer-specific survival; DRFS, distant recurrence-free survival; CI, confidence interval; HR, Hazard Ratio; ET, endocrine therapy; CT, chemotherapy; NR, not reported; micro, micrometastases;
[a]Adjusted for: Knauer 2010: age, tumour size, nodal status, grade, ER, PR, endocrine therapy, chemotherapy; Mook 2010: age, tumour size, nodal status, grade, ER, HER2, surgery, endocrine therapy, chemotherapy. [b]Unclear whether interaction test in Knauer 2010 relates to adjusted or unadjusted analysis.

4.4.4 Clinical Utility: MammaPrint

Overview

Two studies reported evidence relating to clinical utility of MammaPrint (the impact of prospective use of the test on clinical outcomes). MINDACT is an RCT of MammaPrint versus clinical practice.[134] RASTER[123-125] is a prospective observational study in which patients were treated according to usual practice plus MammaPrint. As these two studies are very different in design, they are reported separately below.

Clinical utility RCT: MINDACT

Study design

MINDACT (Cardoso *et al*., 2016)[134] is a partially-randomised prospective study of MammaPrint versus clinical practice. Patients with discordant risk scores (high/low or low/high) via MammaPrint and modified AOL (mAOL; included HER2 status) were randomised to chemotherapy or no chemotherapy; in other words, discordant-risk patients were randomised to treatment determined by MammaPrint or treatment determined by mAOL.

Patients with concordant risk were not randomised, but were followed as prospective cohorts. High/high-risk patients (via both MammaPrint and mAOL) were all recommended to receive chemotherapy, while low/low-risk patients were all recommended no chemotherapy.

The primary aim was to determine whether patients who were high-clinical and low-MammaPrint risk could potentially avoid chemotherapy (i.e. to compare outcomes for patients randomised to chemotherapy or no chemotherapy in this group). Results were also presented for low-clinical high-MammaPrint patients. Secondary analyses included an analysis of discordant patients according to treatment group (chemotherapy versus no chemotherapy), as well as for all patients when chemotherapy was recommended according to clinical risk or to MammaPrint risk. The percentage of patients assigned to chemotherapy with each strategy was also reported.

Patients and tests

MINDACT enrolled 6693 patients from nine European countries (Table 39). Of these, using ITT analyses, 2634 (39%) were low clinical, low MammaPrint risk and were assigned to no chemotherapy; 1873 (28%) were high clinical, high MammaPrint risk and were assigned to chemotherapy; 1497 (22%) were high clinical, low MammaPrint risk and were randomised to chemotherapy or no chemotherapy; and 690 (10%) were low clinical, high MammaPrint risk and were again randomised.

Of all 6693 patients, 88% were hormone-receptor-positive (HR+) and 90% HER2-. In terms of nodal status, overall 79% were LN0 and 21% LN1-3. However, this varied by group: in the discordant groups, only 52% were LN0 among high clinical, low MammaPrint patients, while 98% were LN0 among low clinical, high MammaPrint patients; in the concordant groups 94% were LN0 in the low-risk concordant group and 74% were LN0 in the high-risk concordant group.

Frozen tumour samples were used, and the MammaPrint 70-gene test was conducted using an FDA-approved MammaPrint whole-transcriptome microarray. Cut-offs were not reported, but were assumed by the EAG to be the same as in previous studies.

Quality assessment

Discordant-risk patients were randomised centrally and randomisation was stratified by institution, risk group, ER, PR, nodal status, age, HER2, axillary treatment, and type of surgery; hence, randomisation sequence and allocation concealment were judged to be low risk of bias. No details of blinding were reported (Table 40).

Intention-to-treat (ITT) and per-protocol analyses were reported. Some patients did not adhere to their recommended chemotherapy or no chemotherapy allocation. Other patients had a change in clinical risk group due to initial incorrect reporting of clinical characteristics, or a change in MammaPrint risk group due to a change in the RNA-extraction solution which affected the calculation of risk group. For ITT, patients were analysed in their originally-allocated clinical/MammaPrint risk groups and in their randomised treatment groups. Per protocol analysis excluded patients who were ineligible, or were non-adherent to chemotherapy recommendations, or had a change in their clinical or MammaPrint risk group. This report uses ITT results (where available).

Results

Adherence to recommended treatment

In the discordant-risk groups, overall adherence to chemotherapy assignment was 86%. Among high clinical, low MammaPrint risk patients, adherence was 85% for chemotherapy and 89% for no-chemotherapy. Among low clinical, high MammaPrint risk patients, adherence was 80% for chemotherapy and 88% for no-chemotherapy. However, results presented here are for the ITT analyses which analyse patients within their allocated groups regardless of adherence.

High clinical, low MammaPrint group

The primary aim was to assess whether patients who were high-clinical (mAOL) but low-MammaPrint risk could avoid chemotherapy, i.e. whether outcomes were similar for chemotherapy versus no chemotherapy. In this group (N=1497; 52% LN0), using ITT analyses, 5-year DMFS was

95.9% (95% CI: 94.0, 97.2) with chemotherapy and 94.4% (95% CI: 92.3, 95.9) without chemotherapy, an absolute difference of 1.5% favouring chemotherapy, though the HR was not statistically significant (adjusted HR 0.78, 95% CI 0.50, 1.21, p=0.267). Similar differences between chemotherapy and no chemotherapy were reported for 5-year DMFI, DFS and OS, as well as among both LN0 and LN1-3 patients and a LN0 HR+ HER2- subgroup (Table 41).

This finding was interpreted by the authors as showing little difference in outcomes for chemotherapy versus no chemotherapy, implying that patients who were high-clinical but low-MammaPrint risk could potentially avoid chemotherapy. Statistically, this met the primary objective in that the lower bound of the 95% CI for 5-year DMFS in the no-chemotherapy group was at least 92% (this lower bound was 92.3% in the ITT analysis and 92.5% in the per protocol analysis).

Low clinical, high MammaPrint group

Results were also presented for the low-clinical (mAOL) high-MammaPrint risk group (Table 42). Among these patients (N=690; 98% LN0), again using ITT data, 5-year DMFS was 95.8% (95% CI: 92.9, 97.6) with chemotherapy and 95.0% (95% CI: 91.8, 97.0) without chemotherapy, an absolute difference of 0.8% (adjusted HR 1.17, 95% CI: 0.59, 2.28, p=0.657). This finding, though again showing little difference in outcomes between chemotherapy and no chemotherapy, has quite a different interpretation. Given that low clinical risk patients could be assumed (in general) to not be recommended chemotherapy in current practice, these results imply that low-clinical risk patients with a high-risk MammaPrint result still have little benefit from chemotherapy, implying that MammaPrint should not be used to guide treatment in low clinical risk patients as it would result in patients receiving chemotherapy but not gaining any benefit.

Non-randomised concordant-risk groups

In terms of outcomes for the non-randomised groups, patients with low/low-risk (recommended no chemotherapy) had a 5-year DMFS of 97.6% (95% CI: 96.9, 98.1), i.e. slightly more favourable than the discordant groups. Conversely, patients with high/high-risk (recommended chemotherapy) had a 5-year DMFS of 90.6 (95% CI: 89.0, 92.0), i.e. slightly less favourable than the discordant groups. Results for DFS and OS followed a similar pattern (Table 43).

Estimated outcomes according to clinical and MammaPrint treatment strategies

Results were also reported for analyses, firstly assuming that chemotherapy recommendations were determined by clinical risk, and secondly by MammaPrint risk (Table 44). Both these analysis included all concordant-risk patients (low/low, recommended no chemotherapy, and high/high, recommended chemotherapy). Of the discordant-risk patients, the clinical strategy only included the clinical high, MammaPrint low patients who were randomised to chemotherapy and the clinical low,

MammaPrint high patients who were randomised to no chemotherapy (and vice versa for the MammaPrint strategy; see Table 44). Since half of randomised patients were excluded from each analysis, the remaining discordant patients were double-weighted; the outcomes are therefore described as "estimated".

The 5-year DMFS for both strategies were very similar: 95.0% for the clinical strategy and 94.7% for the MammaPrint strategy (95% CIs not reported). This was interpreted as the MammaPrint strategy leading to little difference in outcomes even though fewer patients had chemotherapy (see below). However, any potential difference between treatment according to the MammaPrint or clinical strategy in the discordant group could be considered to be "diluted" by the concordant-risk groups who had the same treatment and outcomes with either strategy. This analysis also assumes that in the MammaPrint strategy, all patients would be treated according to MammaPrint, whereas the results above indicate this may not be justified for low-clinical high-MammaPrint patients.

Reclassification of patients via clinical or MammaPrint risk (and implications for chemotherapy)

Of all 6693 patients, 3356 (50%) overall were high clinical risk via mAOL, while 2398 (36%) were high MammaPrint risk (Table 45). Therefore, overall, 14% fewer (958/6693) were categorised as high-risk via MammaPrint than mAOL. Of those at high clinical risk, 46% (1550/3356) could be reclassified to low-risk by MammaPrint.

Multivariable analysis

In a multivariable analysis adjusted for chemotherapy use, clinical risk, and patient and tumour characteristics, MammaPrint low/high-risk grouping was statistically significantly associated with 5-year DMFS (HR for high vs low-risk 2.41, 95% CI: 1.79, 3.26, p<0.001).

Discussion: RCT of clinical utility for MammaPrint (MINDACT)

One RCT assessed the clinical utility of MammaPrint. In MINDACT (total N=6693),[134] patients with discordant risk scores via MammaPrint and mAOL were randomised to chemotherapy or no chemotherapy, while patients with concordant high-risk were recommended chemotherapy and those with concordant low-risk were recommended no chemotherapy. The primary aim was to determine whether patients who were high-clinical but low-MammaPrint risk could avoid chemotherapy. In this group (N=1550; 52% LN0), 5-year DMFS was 95.9% (95% CI: 94.0, 97.2) with chemotherapy and 94.4% (95% CI: 92.3, 95.9) without chemotherapy, an absolute difference of 1.5% (adjusted HR 0.78, 95% CI 0.50, 1.21, p=0.267). This finding was interpreted by the authors as suggesting that these patients could avoid chemotherapy. Clinical advice to the EAG suggests that chemotherapy would usually only be indicated where it is likely to provide an absolute improvement in 5-year DRFS of 2%-3%, which suggest that it may be reasonable to withhold chemotherapy in patients with high-

clinical low-MammaPrint risk given the above absolute difference in 5-year DRFS of 1.5% for chemotherapy vs. no chemotherapy.

In patients who were low-clinical but high-MammaPrint risk (N=592; 98% LN0), 5-year DMFS was 95.8% (95% CI: 92.9, 97.6) with chemotherapy and 95.0% (95% CI: 91.8, 97.0) without chemotherapy, an absolute difference of 0.8% (adjusted HR 1.17, 95% CI: 0.59, 2.28, p=0.657). This finding could be interpreted as showing that use of MammaPrint in low clinical risk patients could lead to more patients being prescribed chemotherapy, but not receiving a survival benefit from treatment. Additional analyses assessed strategies in which chemotherapy recommendations for all patients were determined by either clinical risk or MammaPrint risk. These included concordant (non-randomised) and discordant (randomised) patients who had treatment that matched either their clinical risk (treatment determined by clinical risk group) or MammaPrint risk (treatment determined by MammaPrint risk group). The 5-year DMFS was very similar: 95.0% for clinical strategy and 94.7% for MammaPrint strategy. This was interpreted as the MammaPrint strategy leading to little difference in outcomes while sparing many patients from chemotherapy (of those at high clinical risk, 46% were MammaPrint low-risk and could potentially be spared chemotherapy). Given the results in the low clinical risk group (where treatment according to MammaPrint risk groups would result in more patients receiving chemotherapy but with no DMFS advantage), the most advantageous strategy may be to only test clinical high-risk patients with MammaPrint. However, the comparator in this study was mAOL, and it is unclear whether the same would be true for other clinical risk scores.

Conclusions: RCT of clinical utility for MammaPrint (MINDACT)

MINDACT randomised patients with discordant MammaPrint and mAOL risks to chemotherapy or no chemotherapy. For patients who were high-clinical, low-MammaPrint risk, 5-year DMFS was 95.9% with chemotherapy and 94.4% without chemotherapy, an absolute difference of 1.5%. This raises the possibility of avoiding chemotherapy in these patients. In patients who were low-clinical, high-MammaPrint risk, 5-year DMFS was 95.8% with chemotherapy and 95.0% without chemotherapy, an absolute difference of 0.8%. This could be interpreted as showing that MammaPrint may not be useful in this group as it would increase chemotherapy rates without improving outcomes. However, the comparator was mAOL, and it is unclear whether the same would be true for other clinical risk scores.

Table 39: Study and patient characteristics: MINDACT (clinical utility RCT)

Reference	Cohort; N		Country	Study design	Details of test	Cut-offs	Population	Nodal status	Endo / chemo
Cardoso 2016[134]	MINDACT RCT	6693 total (see below)	9 European countries	RCT and prospective cohort (see below)	Frozen MMP whole-transcriptome microarray	Low, high (no details)	88% ER+/PR+ 90% HER2-	[a]LN0, 79% LN1-3, 21%	Some ET (% NR) CT according to clinical / MMP risk
		1497 high clin, low MMP		RCT			98% HR+ 92% HER2-	LN0, 52% LN1-3, 48%	Randomised to CT or no CT
		690 low clin, high MMP		RCT			90% HR+ 88% HER2-	LN0, 98% LN1-3, 2%	Randomised to CT or no CT
		2634 low clin, low MMP		Prospective cohort			100% HR+ 96% HER2-	LN0, 94% LN1-3, 6%	No CT recommended
		1873 high clin, high MMP		Prospective cohort			62% HR+ 81% HER2-	LN0, 74% LN1-3, 26%	CT recommended

-, not reported; CT, chemotherapy; ER, oestrogen receptor; ET, endocrine therapy; FFPE, formalin-fixed paraffin-embedded; HER2, human epidermal growth factor receptor 2; LN, number of positive nodes; MMP, MammaPrint; NR, not reported.
[a]Micrometastases 0.2-2mm considered LN+; isolated tumour cells considered LN0

Table 40: Quality assessment: MINDACT (clinical utility RCT)

	Random sequence generation	Allocation concealment	Blinding (participants and personnel)	Blinding (outcome assessment)	Incomplete outcome data	Selective reporting
Cardoso 2016[134] MINDACT RCT	Low (stratified)	Low (centrally randomised)	Unclear	Unclear	Low	Low

High/low/unclear relates to risk of bias on each criterion.

Table 41: Clinical utility of MammaPrint (MINDACT): High clinical, low MMP group (ITT[a])

Study	Subgroup	N	Population	Nodal status	Outcome	No CT: % risk of outcome (95% CI)	CT: % risk of outcome (95% CI)	HR adj[b] (95% CI)	Absolute difference (95% CI)
Node-negative and node-positive									
Cardoso 2016[134]	High mAOL, low MMP	1497	98% HR+ 92% HER2-	LN0, 52% LN1-3, 48%	**DMFS 5yr**	94.4 (92.3, 95.9)	95.9 (94.0, 97.2)	0.78 (0.50, 1.21), p=0.267	1.5%
					DRFI 5yr	95.3 (93.4, 96.6)	96.6 (94.8, 97.8)	0.76 (0.47, 1.22), p=0.253	1.3%
					DFS 5yr	90.1 (87.5, 92.1)	92.9 (90.5, 94.7)	0.71 (0.50, 1.01), p=0.055	2.8%
					OS 5yr	97.0 (95.4, 98.1)	98.4 (97.0, 99.1)	0.69 (0.35, 1.35), p=0.278	1.4%
Node-negative									
Cardoso 2016[134]	High mAOL, low MMP	787	-	LN0	**DMFS 5yr**	93.2 (90.1, 95.4)	95.7 (93.0, 97.4)	0.69 (0.39, 1.21), p=0.193	2.5%
		699	All HR+ All HER2-	LN0	**DMFS 5yr**	93.9 (90.6, 96.1)	95.5 (92.5, 97.3)	0.80 (0.44, 1.45), p=0.456	1.6%
Node-positive									
Cardoso 2016[134]	High mAOL, low MMP	709	-	LN1-3	**DMFS 5yr**	95.6 (92.7, 97.4)	96.3 (93.1, 98.1)	0.88 (0.42, 1.82), p=0.724	0.7%

-, not reported; CI, confidence interval; CT, chemotherapy; DFS, disease-free survival; DMFS, distant metastasis-free survival; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; HR+, hormone receptor-positive; LN, number of positive nodes; mAOL, modified Adjuvant! Online; MMP, MammaPrint; OS, overall survival.

[a]ITT analysis includes initially-allocated risk groups and treatment assignment, irrespective of adherence to treatment

[b]HRs adjusted for institution, risk group, ER, PR, nodal status, age, HER2, axillary treatment, surgery; HR below 0 favours CT

Table 42: Clinical utility of MammaPrint (MINDACT): Low clinical, high MMP group (ITT[a])

Study	Subgroup	N	Population	Nodal status	Outcome	No CT: % risk of outcome (95% CI)	CT: % risk of outcome (95% CI)	HR adj[b] (95% CI)	Absolute difference (95% CI)
Node-negative and node-positive									
Cardoso 2016[134]	Low mAOL, high MMP	690	90% HR+ 88% HER2-	LN0, 98% LN1-3, 2%	**DMFS 5yr**	95.0 (91.8, 97.0)	95.8 (92.9, 97.6)	1.17 (0.59, 2.28), p=0.657	0.8%
					DRFI 5yr	95.6 (92.5, 97.5)	98.1 (95.7, 99.1)	0.63 (0.27, 1.47), p=0.282	2.5%
					DFS 5yr	90.1 (86.1, 93.0)	92.1 (88.3, 94.6)	0.87 (0.53, 1.45), p=0.603	2.0%
					OS 5yr	97.8 (95.5, 99.0)	97.1 (94.5, 98.5)	1.28 (0.54, 3.02), p=0.578	-0.7%
Node-negative									
Cardoso 2016[134]	Low mAOL, high MMP	635	-	LN0	**DMFS 5yr**	95.1 (91.9, 97.1)	96.0 (93.1, 97.7)	1.09 (0.54, 2.19), p=0.815	0.9%
		534	All HR+ All HER2-	LN0	**DMFS 5yr**	95.5 (91.6, 97.6)	95.1 (91.5, 97.2)	1.45 (0.68, 3.08), p=0.333	-0.4%
Node-positive									
Cardoso 2016[134]	Low mAOL, high MMP	- (N too small)	-	LN1-3	**DMFS 5yr**	-	-	-	-

-, not reported; CI, confidence interval; CT, chemotherapy; DFS, disease-free survival; DMFS, distant metastasis-free survival; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; HR+, hormone receptor-positive; LN, number of positive nodes; mAOL, modified Adjuvant! Online; MMP, MammaPrint; OS, overall survival.

[a]ITT analysis includes initially-allocated risk groups and treatment assignment, irrespective of adherence to treatment

[b]HRs adjusted for institution, risk group, ER, PR, nodal status, age, HER2, axillary treatment, surgery; HR below 0 favours CT

Table 43: Clinical utility of MammaPrint (MINDACT): Outcomes for non-randomised groups[a]

Study	Subgroup	N	Population	Nodal status	Outcome	No CT: % risk (95% CI)	CT: % risk (95% CI)
Low clinical, low MMP (node-negative and node-positive)							
Cardoso 2016[134]	Low mAOL, low MMP	2745	100% HR+ 96% HER2-	LN0, 94% LN1-3, 6%	**DMFS 5yr**	97.6 (96.9, 98.1)	NA
					DFS 5yr	92.8 (91.7, 93.7)	NA
					OS 5yr	98.4 (97.8, 98.9)	NA
High clinical, high MMP (node-negative and node-positive)							
Cardoso 2016[134]	High mAOL, high MMP	1806	62% HR+ 81% HER2-	LN0, 74% LN1-3, 26%	**DMFS 5yr**	NA	90.6 (89.0, 92.0)
					DFS 5yr	NA	85.3 (83.4, 87.0)
					OS 5yr	NA	94.7 (93.4, 95.7)

CI, confidence interval; CT, chemotherapy; DFS, disease-free survival; DMFS, distant metastasis-free survival; HER2, human epidermal growth factor receptor 2; HR+, hormone receptor-positive; LN, number of positive nodes; mAOL, modified Adjuvant! Online; MMP, MammaPrint; NA, not available; OS, overall survival.

[a]Analysis for "corrected risk" groups, since a minority of patients initially allocated to incorrect clinical or MMP risk groups

Table 44: Clinical utility of MammaPrint (MINDACT): Estimated outcomes according to clinical and MMP treatment strategies (ITT)

Study	Subgroup	Patients & treatment	N	DMFS 5yr: % estimated risk
Cardoso 2016[134]	Clinical strategy: CT recommended according to clinical (mAOL) risk	Clin low MMP low: no CT Clin low MMP high: no CT Clin high MMP low: CT Clin high MM high: CT (Excludes: Clin low MMP high: CT) (Excludes: Clin high MMP low: no CT)	6698 (discordant patients double weighted since under-represented)	95.0 (CI NR)
	MMP strategy: CT recommended according to MMP risk	Clin low MMP low: no CT Clin low MMP high: CT Clin high MMP low: no CT Clin high MM high: CT (Excludes: Clin low MMP high: no CT) (Excludes: Clin high MMP low: CT)	6690 (discordant patients double weighted since under-represented)	94.7 (CI NR)
CI, confidence interval; CT, chemotherapy; DMFS, distant metastasis-free survival; mAOL, modified Adjuvant! Online; MMP, MammaPrint; NR, not reported.				

Table 45: Clinical utility of MammaPrint (MINDACT): Reclassification of patients via clinical (mAOL) or MMP risk

Study	Subgroup	High clinical risk, N (%)	High MMP risk, N (%)	Difference in N (%) categorised high-risk overall	Of high clinical risk, N (%) reclassified to low-risk by MMP
Cardoso 2016[134]	All patients (N=6693)	3356 (50%)	2398 (36%)	958/6693 (14% fewer) high-risk by MMP	1550/3356 (46%)

Clinical utility observational study: RASTER

Study design

RASTER (Drukker *et al*., 2013;[124] Drukker *et al*., 2014;[125] Bueno-de-Mesquita *et al*., 2007;[123] [redacted]) is a prospective observational study in which LN0 patients in the Netherlands were treated according to MammaPrint plus usual clinical practice. The aims were to assess the impact of MammaPrint on treatment decisions and to prospectively record outcomes for patients categorised as high or low-risk via MammaPrint, via clinical risk tools, and for various combinations of MammaPrint risk and clinical risk. [redacted]

In the prospective observational study of LN0 patients, receipt of chemotherapy was guided by MammaPrint in combination with the Dutch Institute of Healthcare Improvement (CBO) guidelines of 2004[137] and clinician and patient preference. As such, estimates of prognostic performance (HRs between groups; c-indices) are confounded by the differing rates of chemotherapy in different risk groups (usually more chemotherapy in the high-risk group compared with the low-risk group). Estimates of the impact of the test on clinical outcomes (DRFI, DRFS and OS rates) and chemotherapy use for MammaPrint reflect the use of MammaPrint in routine clinical practice in conjunction with the CBO guidelines, rather than MammaPrint on its own. Conversely, estimates for other risk tools (NPI, Predict, AOL) are confounded by differential rates of chemotherapy in each risk group, and cannot be used to estimate the impact of those tests on clinical outcomes, but can provide some estimate of prognostic performance, albeit confounded by treatment.

Patients and tests

RASTER assessed prospective use of MammaPrint in 427 LN0 patients, age <61 years, of whom 80% were ER+ and 84% HER2- (Table 46).[124, 125] [redacted] Frozen tumour samples were used[124] ([redacted]). The MammaPrint 70-gene microarray was used, stating that cut-offs were the same as in previous studies.[124]

Quality assessment

Since RASTER was not an RCT, it was judged to be at a high risk of bias using standard RCT criteria (Table 47).

Results for LN0 patients

Results for MammaPrint (in conjunction with CBO guidelines and patient/clinician preference):

Of all 427 LN0 patients in RASTER, MammaPrint was low-risk in 51% (of whom 15% received chemotherapy) and high-risk in 49% (of whom 81% received chemotherapy). At 5 years, DRFI was

97.0% for low-risk and 91.7% for high-risk (p=0.03 between groups, HR NR; Table 48).[124, 125] [redacted] 5-year overall survival was not statistically significantly different between MammaPrint groups (p=0.35, HR NR; Table 49).[124, 125]

Results for clinical risk tools: MammaPrint results were compared against various clinical risk tools applied retrospectively to the data (Table 48 and Table 49). Both NPI and PREDICT Plus categorised approximately the same number of patients into the high-risk groups (42% and 47% respectively) as did MammaPrint (49%), and chemotherapy rates in high-risk groups for NPI and PREDICT Plus (84% and 78% respectively) were similar to MammaPrint (81%). Likewise, 5-year DRFI rates in the low-risk groups for NPI and PREDICT Plus (96.7% and 96.8% respectively) were similar to MammaPrint (97.0%), and likewise 5-year DRFI rates in the high-risk groups for NPI and PREDICT Plus (91.3% and 91.7% respectively) were similar to MammaPrint (91.7%). Both NPI and PREDICT Plus showed a significant difference between groups (p=0.03 and p=0.004).[124, 125]

Conversely, AOL categorised more patients as high-risk (69%) than did MammaPrint, NPI or PREDICT Plus, and high-risk AOL patients had a lower chemotherapy rate (60%). 5-year DRFI was similar for the low-risk group (96.7%) but not so much reduced in the high-risk group (93.4%) as for MammaPrint, NPI or PREDICT Plus, and the difference between groups for AOL was not statistically significant (p=0.24; Table 48).[124, 125] [redacted]

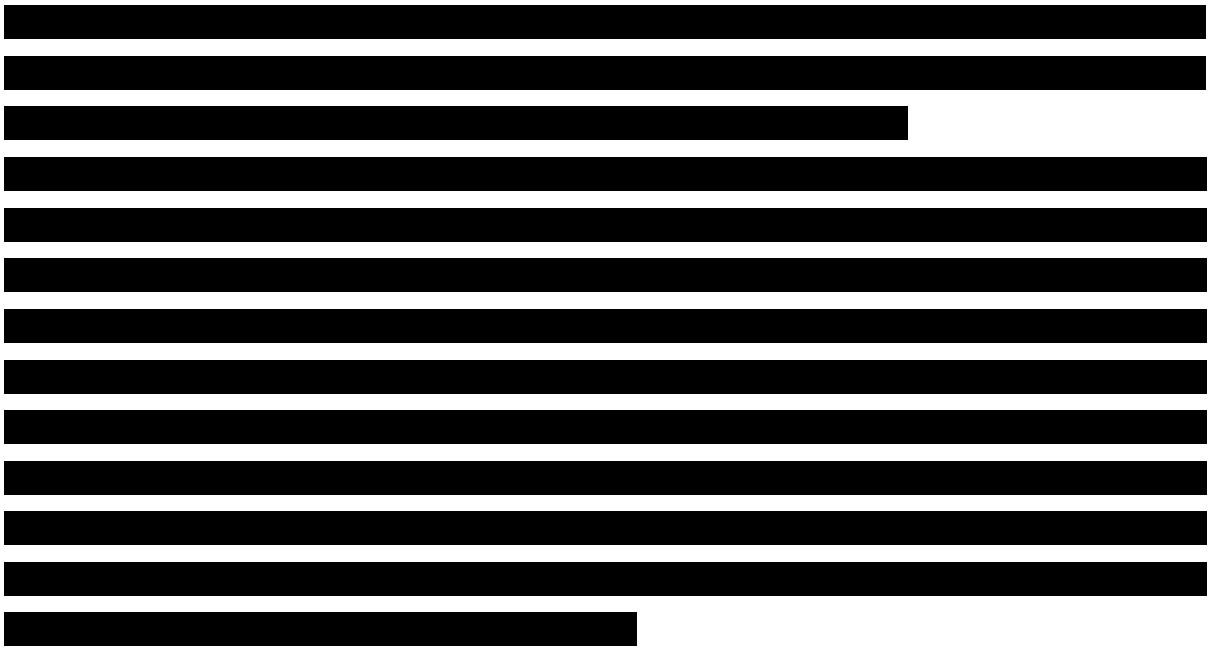
MammaPrint results for patients at high/low clinical risk: Also presented were results by MammaPrint risk group for patients at a high or low clinical risk. The high clinical risk group is particularly of interest to determine whether patients with a high-clinical low-MammaPrint result could safely avoid chemotherapy. Within patients who were high-risk via NPI or PREDICT Plus [redacted], 25% of each (same percentage in all cases) were MammaPrint low-risk (of whom 41-57% received chemotherapy) while 75% were MammaPrint high-risk (of whom 91-93% received chemotherapy). Within NPI and PREDICT Plus high-risk patients, 5-year DRFI for MammaPrint low-risk was 95.5% and 93.9%, while for MammaPrint high-risk it was 89.9% and 91.0%, respectively (Table 48; no p-values reported).[124, 125]

Conversely, AOL categorised more patients as high-risk than did NPI or PREDICT Plus. Of these, a higher proportion fell into the MammaPrint low-risk group (42%), in which chemotherapy rates were lower (24%) and 5-year DRFI higher (98.4%). Of 117 AOL-high-risk patients who received no chemotherapy, 80% were MammaPrint low-risk, and 5-year DRFI for these MammaPrint low-risk patients was 98.9%.[124, 125] However, no such data are reported for NPI or PREDICT Plus, which categorise fewer patients as high-risk.

Of patients at low clinical risk, 5-year DRFI for MammaPrint low-risk patients ranged from 95.3% to 98.0% (Table 48),[124, 125] whilst for MammaPrint high-risk patients 5-year DRFI ranged from 93.9% to 100%, though it should be noted that high-risk patients had more chemotherapy (57-59%) than low-risk patients (3-8%).

Additional prognostic value of MammaPrint:

Table 50 shows C-indexes (AUC) for clinical risk tools alone and in addition to MammaPrint. The addition of MammaPrint to AOL or NPI statistically significantly increased the C-index (AUC) (p=0.03 and p=0.05 respectively), while the addition of MammaPrint to PREDICT Plus did not statistically significantly increase the C-index (AUC) (p=0.27; Table 50).[125]



Discussion: Observational study of clinical utility for MammaPrint (RASTER)

One observational study assessed the clinical utility of MammaPrint. RASTER is a prospective observational study in which 427 LN0 patients[123-125] in the Netherlands were treated according to MammaPrint plus usual clinical practice. Among LN0 patients, 51% were categorised as low-risk. The 5-year DRFI was 97.0% for low-risk patients (15% had chemotherapy) and 91.7% for high-risk

patients (81% received chemotherapy); p=0.03 between groups (HR NR).[124, 125] [redacted]

The addition of MammaPrint to retrospectively-applied AOL, NPI or PREDICT Plus gave a C-index (AUC) which was statistically significantly greater than that for AOL or NPI alone, but not statistically significantly greater than for PREDICT Plus alone.[125] NPI and PREDICT Plus were similar to MammaPrint in terms of proportion categorised into each risk group, chemotherapy rates per risk group, and DRFI rates per risk group, while AOL categorised more as high-risk and high-risk patients had lower chemotherapy rates and better outcomes. [redacted]

Conclusions: Observational study of clinical utility for MammaPrint (RASTER)

RASTER is a prospective observational study in which patients were treated according to MammaPrint plus usual clinical practice practice (LN0) or [redacted]. The 5-year DRFI for LN0 patients was 97.0% for low-risk patients (15% had chemotherapy) and 91.7% for high-risk patients (81% received chemotherapy). [redacted] The DRFI rates in the MammaPrint low-risk group may be considered sufficiently low for these patients to avoid chemotherapy. MammaPrint provided additional prognostic information over AOL and NPI, but not over PREDICT plus. Estimates of prognostic performance between risk groups are likely to be affected by the differing rates of chemotherapy per group, and the fact that chemotherapy use was influenced by MammaPrint.

Table 46: Study and patient characteristics: RASTER (clinical utility observational study)

Reference	Cohort; N		Country	Study design	Details of test	Cut-offs	Population	Nodal status	Endo / chemo
Drukker 2013[124] Drukker 2014[125] Bueno-de-Mesquita 2007[123] [redacted]	RASTER node-negative 16 community hospitals	427	Neths	Prospective observational; treatment influenced by MMP result	Frozen MMP microarray	Low (good), high (poor); cut-offs as in previous studies	80% ER+ 84% HER2- Age <61 100% female	LN0	43% ET (low 27%, high 59%) 47% CT (low 15%, high 81%)
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

-, not reported; CT, chemotherapy; ER, oestrogen receptor; ET, endocrine therapy; FFPE, formalin-fixed paraffin-embedded; HER2, human epidermal growth factor receptor 2; LN, number of positive nodes; MMP, MammaPrint; NR, not reported; RASTER, MicroarRAy PrognoSTics in Breast CancER study.

Table 47: Quality assessment: RASTER (clinical utility observational study)

	Random sequence generation	Allocation concealment	Blinding (participants and personnel)	Blinding (outcome assessment)	Incomplete outcome data	Selective reporting
Drukker 2013[124] Drukker 2014[125] Bueno-de-Mesquita 2007[123] [redacted] Cohort study	High	High	High	UC	High	Unclear

High/low/unclear relates to risk of bias on each criterion.
RASTER, MicroarRAy PrognoSTics in Breast CancER study.

Table 48: Clinical utility of MammaPrint (RASTER study): DRFI[a] in node-negative patients

Study	Subgroup N	Population ET/CT	Nodal status	Outcome	Test or comparator	% pts per grp		% CT per grp		% DRFIrisk: 0-5yr		% DRFI risk: 0-10yr		HR (95% CI), p-value	
						Low	High	Low	High	Low	High	Low	High	0-5 yr	0-10 yr
Node-negative															
RASTER LN0[124, 125, 135]	All patients N=427	80% ER+ 84% HER2- 43% ET 47% CT	LN0	**DRFI**	**MMP**	51	49	15	81	97.0	91.7	[redacted]	[redacted]	p=0.03	[redacted]
					AOL	31	69	18	60	96.7	93.4	-	-	p=0.24	-
					NPI	58	42	21	84	96.7	91.3	-	-	p=0.03	-
					PREDICT Plus	53	47	20	78	96.8	91.7	-	-	p=0.004	-
					mAOL	[redacted]	[redacted]	-	-	-	-	[redacted]	[redacted]	-	[redacted]
ER+ patients															
RASTER LN0[121, 135]	ER+ patients N=342	All ER+ HER2 NR	LN0	**DRFI**	**MMP**	[redacted]	[redacted]	-	-	-	-	[redacted]	[redacted]	-	[redacted]
					mAOL	-	-	-	-	-	-	[redacted]	[redacted]	-	-
High clinical risk															
RASTER LN0[124, 125, 135]	AOL high; N=295	ER+/- HER2+/-	LN0	**DRFI**	**MMP**	42	58	24	87	98.4	89.8	-	-	-	-
	NPI high; N=179					25	75	57	93	95.5	89.9	-	-	-	-
	PREDICT Plus high; N=199					25	75	41	91	93.9	91.0	-	-	-	-
	mAOL high; N=183					[redacted]	[redacted]	-	-	-	-	[redacted]	[redacted]	-	-
High clinical risk, untreated															
RASTER LN0[124, 125]	AOL high; no CT (N=117)	ER+/- HER2+/-	LN0	**DRFI**	**MMP**	80	20	0	0	98.9	-	-	-	-	-
	AOL high; no ET/CT (N=75)					93	7	0	0	100.0	-	-	-	-	-
Low clinical risk															
RASTER LN0[124, 125, 135]	AOL low; N=132	ER+/- HER2+/-	LN0	**DRFI**	**MMP**	72	28	3	57	95.3	100.0	-	-	-	-
	NPI low; N=248					71	29	5	59	97.4	95.3	-	-	-	-
	PREDICT Plus low; N=228					75	25	8	57	98.0	93.9	-	-	-	-
	mAOL low; N=NR					-	-	-	-	-	-	[redacted]	[redacted]	-	-
Low clinical risk, untreated															
RASTER	AOL low; no CT (N=108)	ER+/-	LN0	**DRFI**	**MMP**	85	15	0	0	95.1	-	-	-	-	-

Study	Subgroup N	Population ET/CT	Nodal status	Outcome	Test or comparator	% pts per grp		% CT per grp		% DRFIrisk: 0-5yr		% DRFI risk: 0-10yr		HR (95% CI), p-value	
						Low	High	Low	High	Low	High	Low	High	0-5 yr	0-10 yr
LN0[124, 125]	AOL low; no ET/CT (N=93)	HER2+/-				95	5	0	0	95.0	-	-	-	-	-

-, not reported; AOL, Adjuvant! Online; CT, chemotherapy; DRFI, distant recurrence free interval; DRFS, distant recurrence free survival; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; LN, number of positive nodes; mAOL, modified AOL (includes HER2); MMP, MammaPrint; NPI, Nottingham Prognostic Index; NR, not reported; RASTER, MicroarRAy PrognoSTics in Breast CancER study.

[a]In RASTER, definition of DRFI includes DR and BC death as events, which is more similar to definitions of DRFS/DMFS in most studies in this review.

Table 49: Clinical utility of MammaPrint (RASTER study): overall survival in node-negative patients

Study	Subgroup N	Population	Nodal status	Endo / chemo	Outcome	Test or comparator	% pts per group		% CT per group		% OS risk per group		HR (95% CI), p-value
							Low	High	Low	High	Low	High	0-5 yr
Node-negative													
RASTER[124, 125, 135]	All patients N=427	80% ER+ 84% HER2-	LN0	43% ET 47% CT	**OS 5yr**	**MMP**	51	49	15	81	98.3	96.9	p=0.35
						AOL	31	69	18	60	100.0	96.5	p=0.02

AOL, Adjuvant! Online; CT, chemotherapy; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; LN, number of positive nodes; MMP, MammaPrint; OS, overall survival; RASTER, MicroarRAy PrognoSTics in Breast CancER study.

Table 50: Clinical utility of MammaPrint (RASTER study): Additional prognostic value in node-negative patients

Study	Subgroup N	Population	Nodal status	Endo / chemo	Outcome	Test or comparator	% pts per group		% CT per group		C-index (AUC)	Increase in C-index (AUC) over CP factors
							Low	High	Low	High		
Node-negative												
RASTER[124, 125, 135]	All patients N=427	80% ER+ 84% HER2-	LN0	43% ET 47% CT	**DRFI 5yr**	**MMP**	51	49	15	81	-	
						AOL	31	69	18	60	0.532	
						AOL+MMP	-	-	-	-	0.619	p=0.03
						NPI	58	42	21	84	0.591	
						NPI+MMP	-	-	-	-	0.638	p=0.05
						PREDICT Plus	53	47	20	78	0.627	
						PREDICT Plus +MMP	-	-	-	-	0.662	p=0.27

-, not reported; AOL, Adjuvant! Online; CP, clinical/pathological; CT, chemotherapy; DRFI, distant recurrence-free interval; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; LN, number of positive nodes; MMP, MammaPrint; NPI, Nottingham Prognostic Index; RASTER, MicroarRAy PrognoSTics in Breast CancER study.

Table 51: Clinical utility of MammaPrint (RASTER study): DRFI[a] in node-positive patients

Study	Subgroup N	Population ET/CT	Nodal status	Outcome	Test or comparator	% pts per grp		% CT per grp		% DRFI risk: 0-5yr		% DRFIrisk: 0-10yr		HR (95% CI), p-value	
						Low	High	Low	High	Low	High	Low	High	0-5 yr	0-10 yr
Node-positive															
[redacted]															
High clinical risk															
[redacted]															
[redacted]															

4.5 Results: Prosigna

Prosigna is based on a Risk Of Recurrence (ROR) score called ROR-PT, which incorporates the PAM50 gene signature, a weighting for a proliferation score (P, a subset of the 50 genes) and information on tumour size (T). Nodal status is then used when converting the score into a risk category. The commercial Prosigna test uses the nCounter system to analyse ROR-PT. Other research-based versions of ROR-PT exist, for example using qRT-PCR. This assessment includes all studies assessing ROR-PT, whether or not they use the formal Prosigna test. However, studies assessing other versions of the ROR score, such as ROR-S (subtype) or ROR-T/ROR-C (subtype and tumour size) or ROR-P (subtype and proliferation score), are excluded. Studies assessing ROR-PT via whole-transcriptome microarray (in silico studies) are summarised in Section 4.8.2.

Within this section, the test is referred to as ROR-PT since this covers both Prosigna and other versions of ROR-PT that do not use the nCounter system (but are equivalent to Prosigna in terms of incorporation of PAM50 gene signature and clinical factors).

4.5.1 Development: Prosigna

The PAM50 gene signature was developed and validated by Parker *et al*.[138] (2009) using microarray and qRT-PCR. Risk of recurrence (ROR) models were trained on 141 node-negative (LN0), untreated patients from the Netherlands Cancer Institute (NKI; van de Vijver, 2002),[47] which was also part of the first validation cohort for MammaPrint. ROR models tested included ROR-S and ROR-T. Validation in untreated LN0 patients showed that both ROR-S and ROR-T statistically significantly improved prognosis over clinico-pathologic variables, and that ROR-T statistically significantly improved prognosis over ROR-S. This study is not discussed further as it did not include ROR-PT.

Use of Prosigna (ROR-PT) via the nCounter system was developed and validated by in the British Columbia cohort by Wallden *et al*. (2015), which is included in this section.[139]

4.5.2 Prognostic performance: Prosigna

Study designs: Prosigna prognostic performance

Eight data sets were used to assess the prognostic performance of ROR-PT (Table 52). These included six reanalyses of RCTs (TransATAC,[36, 43] ABCSG-8,[54, 55] CALGB 9741,[140] NCIC MA.21,[141] GEICAM 9906[83, 92] and NCIC MA.12[142]) and two retrospective analyses of prospective cohorts (the Danish Breast Cancer Cooperative Group [DBCG] cohort[56, 143-145] and two analyses of the British Columbia cohort[139, 146]).

Patients: Prosigna prognostic performance

Two analyses of RCTs (TransATAC[36, 43] and ABCSG-8[54, 55]) included patients who were ER+ HER2-, a mix of LN0 and LN+, and received only endocrine treatment (no chemotherapy). Conversely, the

other four analyses of RCTs[83, 92, 140-142] included higher-risk patients who received adjuvant chemotherapy; more patients in these trials were node-positive (LN+), and not all were ER+ HER2- (Table 52).

[redacted]

The two retrospective analyses of prospective cohorts[56, 139, 143-146] included patients who were mostly ER+ HER2-, a mix of LN0 and LN+, and received only endocrine treatment (no chemotherapy).

Tests and comparators: Prosigna prognostic performance

Four analyses of RCTs[36, 43, 54, 55, 140, 141] and two analyses of prospective cohorts[56, 139, 143-145] measured ROR-PT using the nCounter device, while two analyses of RCTs[83, 92, 142] and one of a prospective cohort[146] used qRT-PCR (Table 52). The cut-offs used to define risk groups varied across studies, while some analyses assessed ROR-PT as a continuous score (see Table 52 for details).

Some data sets were also used to evaluate other in-scope tests as follows (see Section 4.8.1 on comparing tests). TransATAC was used to evaluate Oncotype DX, EndoPredict and IHC4. The GEICAM 9906 analysis,[83, 92] as well as a pooled analysis of ABCSG-6 and 8,[57-59] were used to evaluate EndoPredict.

Quality assessment: Prosigna prognostic performance

All data sets reported here were validation studies (Table 53). All analyses excluded some patients recruited to the original trial or cohort. Blinding of test assessors to outcomes was reported in five analyses. All used standardised outcomes.

Results: Prosigna prognostic performance

Table 54 to Table 58 present the data for all patients (LN0 or LN+) and separate data for LN0 and LN+ patients.

Distribution of patients by risk group

Some studies reported the percentages of patients categorised into each risk group by ROR-PT (Table 54). For LN0 patients, the percentages categorised as low-risk were reported in two analyses: [redacted] in TransATAC[43] and 48% in ABCSG-8.[54, 55] Among LN+ patients, far fewer patients were categorised

as low-risk: [redacted] in TransATAC;[43] 4% in ABCSG-8;[54, 55] 19% in GEICAM 9906;[83, 92] and 25% in DBCG.[56] The percentage of patients categorised as intermediate-risk was [redacted][43] and 32%[54, 55] in LN0 patients and ranged from 27% to 56% in LN+ patients.[43, 54-56, 83, 92]

Prognostic performance: unadjusted analyses

This section reports unadjusted analyses. Adjusted analyses, which show whether the test has prognostic value over clinicopathological variables, are reported in the section "Additional prognostic value".

For LN0 patients, ROR-PT was statistically significantly prognostic for DRFS/DFMS/DRFI in all three data sets (TransATAC,[43] ABCSG-8,[54, 55] DBCG[56]), with the proportion of patients having 10-year DRFS/DFMS/DRFI in the low-risk groups being [redacted] (TransATAC[43]), 96.5% (ABCSG-8[54, 55]) and 95.1% (DBCG[56]). HRs and p-values between groups are reported in many differing formats and timepoints so are summarised in Table 54 rather than in the text. ROR-PT was also statistically significantly prognostic for late (5-15-year) recurrence in the one study reporting this (Table 54).[54, 55]

For LN+ patients, ROR-PT was statistically significantly prognostic for 10-year DRFS/DFMS/DRFI in all four data sets (TransATAC,[43] ABCSG-8,[54, 55] DBCG[56] and GEICAM 9906[83, 92]), with the proportion of patients having 10-year DRFS/DFMS/DRFI in the low-risk groups being [redacted] (TransATAC[43]), 100.0% (ABCSG-8[54, 55]) and 92% (GEICAM 9906[83, 92]), or 95.1% in the combined low/intermediate-risk groups (DBCG[56]). ROR-PT was also statistically significantly prognostic for late (5-10-year) recurrence in the two studies reporting this (Table 54).[54-56, 144]

In terms of other outcomes (Table 55 and Table 56), ROR-PT was statistically [redacted] for 10-year overall survival in LN0 and LN+ patients in TransATAC;[43] for relapse-free survival (RFS) and breast cancer specific survival in LN0 patients in the British Columbia cohort;[146] and for RFS in CALGB 9741;[140] but not for RFS in NCIC MA.21.[141] ROR-PT was also statistically significantly prognostic in both pre- and post-menopausal patients (CALGB 9741[140]) and in ductal and lobular breast cancer patients (DBCG, Laenkholm *et al*., 2016[145]).

Additional prognostic value

This section reports adjusted analyses, which indicate the additional prognostic value of IHC4 over clinicopathological factors. The clinicopathological factors adjusted for vary from study to study, and are detailed in the footnotes to the tables.

Likelihood ratios: The TransATAC analysis[43] reports a reduced dataset of patients where data for all four in-scope tests are available. Additional prognostic value was assessed via increases in likelihood

ratio χ^2 for 10-year DRFI, for ROR-PT plus NPI or CTS, over NPI or CTS alone (Table 57). Increases in likelihood ratio χ^2 were [redacted] (Table 57). In ABCSG-8,[54] likelihood ratios also showed a statistically significant increase for ROR-PT over the Clinical Linear Predictor (same variables as CTS) in LN0 patients (p<0.0001) and LN+ patients (p=0.0002). Similar results were found for other outcomes (Table 58).

C-indexes (AUC): In ABCSG-8,[54] C-indexes were numerically higher for ROR-PT than for the Clinical Linear Predictor in both LN0 and LN+ patients, but statistical significance levels were not reported. Similarly in the British Columbia analysis by Wallden *et al.* 2015,[139] C-indexes were higher for ROR-PT than for AOL or IHC4+tumour size in LN0 patients, but statistical significance levels were not reported (Table 57).

Multivariable Cox models: ABCSG-8[54] and DBCG[56, 143] used multivariable analyses to show that Prosigna was an independent prognostic parameter for 10-year DMFS/DRFS after adjustment for clinical factors across a mix of nodal status (Table 57).

Discussion: Prosigna prognostic performance

Prognostic value of Prosigna (or ROR-PT assessed via any method) was based on six reanalyses of RCTs[36, 54, 55, 83, 92, 101, 140, 142, 147] and retrospective analyses of two prospective cohorts[56, 143-146, 148] (total N=9,118). Two of the RCTs (TransATAC[36, 101] and ABCSG-8;[54, 55] total N=2,252) and the two retrospective analyses (total N=3,508) included patients who were all/mostly ER+ HER2- and received endocrine monotherapy. The other four RCTs[83, 92, 140, 142, 147] (total N=3,358) included higher-risk patients (not restricted to ER+ HER2-, higher proportion LN+) and all received chemotherapy. All excluded some originally-recruited patients, sometimes due to insufficient tumour sample which may introduce bias due to attrition of patients with smaller tumours.

In two studies of LN0 patients,[54, 55, 101] the percentage of patients categorised as Prosigna/ROR-PT low-risk ranged from 48% to [redacted], the percentage intermediate-risk from [redacted] to 32%, and the percentage high-risk from [redacted] to 20%. Across four studies of LN+ patients,[54-56, 83, 92, 101] the percentage low-risk ranged from 4% to 25%, the percentage intermediate-risk from 27% to 56%, and the percentage high-risk from 26% to 62%. The number of patients who are likely to be prescribed chemotherapy on the basis of their test result will depend on how intermediate-risk patients are handled and whether they would be handled the same in LN0 and LN+ groups.

Prosigna/ROR-PT was statistically significantly prognostic for 10-year DRFS/DRFI in all unadjusted analyses of LN0 and LN+ patients. In Prosigna/ROR-PT low-risk groups, 10-year DRFS/DRFI rates for LN0 patients ranged from 95% to [redacted] (three studies of endocrine monotherapy),[54-56, 101] while for LN+ patients these were [redacted] (in two studies of endocrine monotherapy)[54, 55, 101] and 92% (in one study where all received endocrine and chemotherapy).[83, 92] In intermediate-risk groups, 10-year DRFS/DRFI rates for LN0 patients were [redacted] to 93% (endocrine monotherapy),[54, 55, 101] and for LN+ patients were [redacted] to 94% (endocrine monotherapy)[54, 55, 101] and 74% (endocrine and chemotherapy).[83, 92] Use of chemotherapy could potentially influence patient outcomes in either direction: negatively since trials of chemotherapy may have selected higher-risk patients than trials not assessing chemotherapy, or positively due to the effect of chemotherapy.

In terms of additional prognostic value, multivariable analyses of two datasets[54, 56, 143] showed that ROR-PT was an independent prognostic parameter for 10-year DMFS/DRFS after adjustment for clinicopathological variables across LN0/LN+ and LN+ patients. Two studies reported an [redacted] in likelihood ratio χ^2 for ROR-PT plus CTS/CLP/NPI over CTS/CLP/NPI alone; this [redacted] was statistically significant in LN0 and LN+ patients in ABCSG-8,[54] [redacted].[101]

Conclusions: Prosigna prognostic performance

Based on six reanalyses of RCTs and two retrospective analyses of prospective cohorts, Prosigna/ROR-PT was statistically significantly prognostic for unadjusted analyses of 10-year DRFS/DRFI in LN0 and LN+ patients. Among LN0 patients, approximately 50% were categorised as low-risk, 30% as intermediate-risk and [redacted] to 20% as high-risk. Among LN+ patients, 4% to 25% were low-risk, 27% to 56% intermediate-risk, and 26% to 62% high-risk. The 10-year DRFS/DRFI rates for low-risk patients were 95% to [redacted] in three studies of LN0 patients (all endocrine only), and in LN+ patients these were [redacted] in two studies (endocrine therapy only) and 92% in one study (all endocrine and chemotherapy). ROR-PT added prognostic information over clinicopathological variables or CTS/CLP/NPI in three studies; this was statistically significant in LN0 patients and either significant or borderline significant in LN+ patients.

Table 52: Characteristics of prognostic studies: Prosigna

Reference(s)	Cohort(s)	N pts	Country	Study design	Test	Details of test	Cut-offs	Other tests	Population	Nodal status	Endo / chemo
Reanalyses of RCTs: LN status mixed											
100% ET monotherapy											
Sestak 2017 (data request)[43] Dowsett 2013[36]	**TransATAC**	[redacted]	UK	R-RCT	ROR-PT	FFPE nCounter	LN0: 40; 60 LN1-3: 15; 40	O-DX EPClin IHC4+C	ER+ HER2- Postmeno 100% female	LN0, [redacted] LN1-3, [redacted]	All ET No CT
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8**	1397	Austria	R-RCT	ROR-PT	FFPE nCounter	LN0: 40; 60 LN1-3: 15; 40 LN>3: all high	EPClin (ABCSG-6+8)	ER+ HER2- Postmeno 100% female	LN0, 71%[b] LN1-3, 26%[b] LN>3, 3%[b]	All ET No CT
Variable ET&CT											
Chia 2012[142]	**NCIC MA.12**	398	Canada	R-RCT	ROR-PT	FFPE qRT-PCR	Continuous score		73% HR+ HER2 NR Premeno 100% female Stage I-III	LN0, 25% LN1-3, 55% LN>3, 20%	Some ET (% NR) All CT
Liu 2015[141]	**NCIC MA.21**	1094	Canada + USA	R-RCT	ROR-PT	FFPE nCounter	LN0: 40; 60 LN1-3: 15; 40 LN>3: all high + cont. score		58% ER+ 71% HER2- 31% postmeno 100% female	LN0, 30% LN1-3, 42% LN>3, 28%	58% ET All CT
Reanalyses of RCTs: LN+											
Variable ET&CT											
Liu 2016[140]	**CALGB 9741** (Alliance)	1311	USA	R-RCT	ROR-PT	FFPE nCounter	Continuous score		64% ER+ HER2 NR 51% postmeno 100% female	All LN+ (1-5 nodes, % NR)	ET NR All CT
100% ET&CT											
Martin 2016,[83] 2014[92]	**GEICAM 9906**	555	Spain	R-RCT	ROR-PT (research-based)	qRT-PCR then microarray	LN+: 18; 65	EP; EPClin	ER+ HER2- 46% postmeno Stage II-III 100% female	All LN+ LN1-3, 64% LN>3, 36%	All ET All CT

Reference(s)	Cohort(s)	N pts	Country	Study design	Test	Details of test	Cut-offs	Other tests	Population	Nodal status	Endo / chemo
Retrospective studies: LN status mixed											
100% ET monotherapy											
Ejlertsen 2015[143]; Laenkholm 2015[56], 2015[144], 2016[145]	**DBCG** 2000-2003	2722	Denmark	Retro. analysis of prosp. cohort	ROR-PT	FFPE nCounter	LN0: 40; 60 LN1-3: low 0-40; high >40		HR+ HER2 NR Postmeno 100% female	LN0, 46% LN1-3, 54%	All ET No CT
Nielsen 2010[146]	**British Columbia** 1986-1992	786	Canada	Retro. analysis of prosp. cohort	ROR-PT	FFPE qRT-PCR	Continuous score? (unclear)		ER+ 89% HER2- 96% postmeno 100% F	LN0, 28% LN1-3, 46% LN>3, 19% Missing, 7%	All ET No CT
Retrospective studies: LN0											
100% ET monotherapy											
Wallden 2015[139]	**British Columbia** (years NR)	232	Canada	Retro. analysis of prosp. cohort	ROR-PT	FFPE nCounter	Continuous score		ER+ 91% HER2- 94% postmeno (% female NR)	All LN0	All ET No CT

ABCSG, Austrian Breast and Colorectal Cancer Study Group; AC/T, doxorubicin, cyclophosphamide + paclitaxel; CEF, dose-intense cyclophosphamide, epirubicin + flurouracil; CT, chemotherapy; DBCG, Danish Breast Cancer Cooperative Group; EC/T, dose-dense, dose-intense epirubicin, cyclophosphamide + paclitaxel; CEF, cyclophosphamide, epirubicin and fluorouracil; CMF, cyclophosphamide, methotrexate and fluorouracil; DC, doxorubicin and cyclophosphamide; ER, oestrogen receptor; ET, endocrine therapy; FEC, 5-Fluorouracil, epirubicin, and cyclophosphamide; FEC-P, FEC + paclitaxel; FFPE, formalin-fixed paraffin-embedded; HER2, human epidermal growth factor receptor 2; HR+, hormone-receptor positive; LN, number of positive nodes; NR, not reported; prosp, prospective; qRT-PCR, quantitative reverse transcription polymerase chain reaction; R-RCT, reanalysis of RCT; retro, retrospective.

[redacted]

[b]Nodal status for all 1478 patients; NR for 1397 who were HER2-

Table 53: Quality assessment of prognostic studies: Prosigna

Reference(s)	Cohort(s)	Derivation or validation?	Study design appropriate?	All eligible patients included?	Blinding (of test assessors to outcomes)	Outcome definition standardised *or a priori*?	Applicability: Patient Spectrum	Applicability: Test as per decision problem?
Sestak 2017 (data request),[43] Dowsett 2013[36]	**TransATAC**	V	Y, R-RCT, no chemo	N (InT, MS, TF)	Y	Y	Y	Y
Chia 2012[142]	**NCIC MA.12**	V	N, R-RCT, adj chemo	N (InT, MS, TF)	UC	Y	N (27% HR-/unknown, HER2 NR, 20% LN>3)	N (qRT-PCR, continuous score)
Ejlertsen 2015[143]; Laenkholm 2015[56], 2015[144], 2016[145]	**DBCG**	V	Y, prospective cohort, no chemo	N (reason NR)	UC	Y	UC (HER2 NR)	Y
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8**	V	Y, R-RCT, no chemo	N (InT, MS, TF)	Y	Y	Y (for subgroup analysis)	Y
Liu 2016[140]	**CALGB 9741**	V	N, R-RCT, adj chemo	N (InT, MS, TF)	Y	Y	N (36% ER-, HER2 NR, LN>3 NR)	N (continuous score)
Liu 2015[141]	**NCIC MA.21**	V	N, R-RCT, adj chemo	N (InT, MS, TF)	Y	Y	N (42% ER-, 29% HER2+ / unknown, 28% LN>3)	Y
Martin 2016,[83] 2014[92]	**GEICAM 9906**	V	N, R-RCT, adj chemo	N (reason NR)	Y	Y	N (36% LN>3)	N, Prosigna via qRT-PCR then microarray
Nielsen 2010[146]	**British Columbia**	V	Y, prospective cohort, no chemo	N (InT, TF)	UC	Y	Most (11% HER2+ / missing; 19% LN>3)	No - qRT-PCR, continuous score? (unclear)
Wallden 2015[139]	**British Columbia**	V D (nCounter)	Y, prospective cohort, no chemo	N (InT, TF)	UC	Y	Most (9% HER2+ / missing)	No - continuous score

Y, yes; N, no; UC, unclear

ABCSG, Austrian Breast and Colorectal Cancer Study Group; D, Development; DBCG, Danish Breast Cancer Cooperative Group; InT, insufficient tissue; LN, number of positive nodes; MS, missing samples; qRT-PCR, quantitative reverse transcription polymerase chain reaction; R-RCT, reanalysis of RCT; TF, test failure; V, validation.

Table 54: Prognostic performance of Prosigna: distant recurrence-free survival (DRFS/DRFI)[a]

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Endo / chemo	Test	% pts per group			% DRFS/DRFI risk: 0-5 yr			% DRFS/DRFI risk: 0-10 yr			DMFS/DRFS[a]: HR (95% CI)
						Low	Int	High	Low	Int	High	Low	Int	High	
LN status mixed															
100% ET monotherapy															
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT; UK	ER+ HER2- N=[redacted]	LN0, [redacted] LN1-3, [redacted]	All ET No CT	**ROR-PT** nCounter	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8** R-RCT; Austria	ER+ HER2- N=1397	LN0, 71%[b] LN1-3, 26%[b] LN>3, 3%[b]	All ET No CT	**ROR-PT** nCounter	35	32	33	-	-	-	96.6	91.1	79.9	**5-15yr: L vs I:** 3.74 (NR), p=0.002[c]. **L vs H:** 6.90 (3.08, 15.45), p<0.001[c]
Laenkholm 2015[56], 2015[144]	**DBCG** Cohort; Denmark	HR+ HER2 NR N=2722	LN0, 46% LN1-3, 54%	All ET No CT	**ROR-PT**	27	29	44	-	-	-	95.7	-	79.2	**5-10yr: L vs I:** NR, p=0.0074 I vs H: NR, p=0.0091
LN0															
100% ET monotherapy															
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT; UK	ER+ HER2- N=[redacted]	LN0	All ET No CT	**ROR-PT** nCounter	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8** R-RCT; Austria	ER+ HER2- N=984	LN0	All ET No CT	**ROR-PT** nCounter	48	32	20	-	-	-	96.5	90.0	84.7	**5-15yr: L vs I:** 4.03 (NR), p=0.002[c]. **L vs H:** 4.74 (1.89, 11.87), p<0.001[c]
Laenkholm 2015[56]	**DBCG** Cohort; Denmark	HR+ HER2 NR N=1256	LN0	All ET No CT	**ROR-PT**	NR	NR	NR	-	-	-	95.1	92.7	81.5	**0-10 yr: L vs I:** NR, p=0.1543 **I vs H:** NR, p<0.0001
LN+															
100% ET monotherapy															
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT; UK	ER+ HER2- N=[redacted]	LN1-3	All ET No CT	**ROR-PT** nCounter	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Endo / chemo	Test	% pts per group			% DRFS/DRFI risk: 0-5 yr			% DRFS/DRFI risk: 0-10 yr			DMFS/DRFS[a]: HR (95% CI)
						Low	Int	High	Low	Int	High	Low	Int	High	
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8** R-RCT; Austria	ER+ HER2- N=413	LN1-3, 89%[b] LN>3, 11%[b]	All ET No CT	**ROR-PT** nCounter	4	34	62	-	-	-	100	93.6	76.2	**5-15yr: L vs I or L vs H:** no events. **I vs H:** 3.15 (1.20, 8.24), p=0.020[c]
Laenkholm 2015[56]	**DBCG** Cohort; Denmark	HR+ HER2 NR N=1466	LN1-3	All ET No CT	**ROR-PT**	25	27	48	-	-	-	95.2		78.1	**0-10 yr: L/I vs H:** NR, p<0.0001
100% CT&ET															
Martin 2016,[83] 2014[92]	**GEICAM 9906** R-RCT; Spain	ER+ HER2- N=536	LN1-3, 64% LN>3, 36%	All ET **All CT**	**ROR-PT (research)**	19	56	26	-	-	-	92	74	66	**0-10 yr: L vs I:** 4.4 (NR) **L vs H:** 5.8 (NR), p<0.0001

-, not reported; ABCSG, Austrian Breast and Colorectal Cancer Study Group; CI, confidence interval; CT, chemotherapy; DBCG, Danish Breast Cancer Cooperative Group; DMFS, distant metastasis-free survival; DRFS, distant recurrence free survival; ER, oestrogen receptor; ET, endocrine therapy; H, high; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; HR+, hormone-receptor positive; I/int, intermediate; L, low; LN, number of positive nodes; R-RCT, reanalysis of RCT.

[a]DMFS (GEICAM, ABSCG); DRFI (TransATAC); [b]Nodal status for all patients; NR for HER2- subgroup; [redacted]; [c]5-15 yr in ABCSG-8 analysis of Prosigna

Table 55: Prognostic performance of Prosigna: Overall survival

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Endo / chemo	Test	% pts per group			% OS risk: 0-5yr			% OS risk: 0-10yr			OS: HR (95% CI)
						Low	Int	High	Low	Int	High	Low	Int	High	
LN status mixed															
100% ET monotherapy															
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT; UK	ER+ HER2- N=[redacted]	LN0 [redacted] LN1-3, [redacted]	All ET No CT	**ROR-PT** nCounter	[redacted]									[redacted]
LN0															
100% ET monotherapy															
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT; UK	ER+ HER2- N=[redacted]	LN0	All ET No CT	**ROR-PT** nCounter	[redacted]									[redacted]
LN+															
100% ET monotherapy															
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT; UK	ER+ HER2- N=[redacted]	LN1-3	All ET No CT	**ROR-PT** nCounter	[redacted]									[redacted]

-, not reported; CI, confidence interval; CT, chemotherapy; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; int, intermediate; LN, number of positive nodes; OS, overall survival; R-RCT, reanalysis of RCT.

[redacted]

Table 56: Prognostic performance of Prosigna: Other outcomes

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Outcome	Endo / chemo	Test	% pts per group			% risk of outcome per group			HR (95% CI)
							Low	Int	High	Low	Int	High	
LN status mixed													
Variable CT&ET													
Liu 2015[141]	**NCIC MA.21** R-RCT; Canada+USA	58% ER+, 71% HER2- N=1094	LN0, 30% LN1-3, 42% LN>3, 28%	**RFS 8yr**	58% ET **All CT**	**ROR-PT** nCounter	3	18	79	-	-	-	Low/int vs high: 1.27 (0.83. 1.95), p=0.275
LN0													
100% ET monotherapy													
Nielsen 2010[146]	**British Columbia** Cohort; Canada	ER+, 89% HER2- N=222	LN0	**BCSS 10+yr**	All ET No CT	**ROR-PT** qRT-PCR	-	-	-	-	-	-	Between groups: p=0.026 (cut-points unclear)
				RFS 10+yr	All ET No CT	**ROR-PT** qRT-PCR	-	-	-	-	-	-	Between groups: p=0.009 (cut-points unclear)
LN+													
Variable CT&ET													
Liu 2016[140]	**CALGB 9741** R-RCT; USA	64% ER+, HER2 NR N=1311	All LN+ (1-5 nodes, % NR)	**RFS 12.5yr**	ET NR **All CT**	**ROR-PT** nCounter	N/A	N/A	N/A	N/A	N/A	N/A	Per 10-unit change: 1.12 (1.07, 1.18), p<0.0001

-, not reported; BCSS, breast cancer-specific survival; CI, confidence interval; CT, chemotherapy; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; int, intermediate; LN, number of positive nodes; RFS, relapse-free survival (locoregional or distant); R-RCT, reanalysis of RCT.

Table 57: Additional prognostic value for DRFI/DRFS: Prosigna

Reference(s)	Cohort(s)	Population	Nodal status	Endo / chemo	Outcome	Test or comparator[a]	Likelihood ratio χ^2	Increase in LR χ^2 over CP[a]	C-index (AUC)	Increase in C-index (AUC) over CP[a]	Multivariable model (adj. for CP factors[a]): HR (95% CI)
LN status mixed											
100% ET monotherapy											
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT	ER+ HER2- N=[redacted]	LN0, [redacted] LN1-3, [redacted]	All ET No CT	**DRFI 10yr**	**ROR-PT** nCounter	[redacted]	[redacted]			
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8** R-RCT	ER+ HER2- N=1397	LN0, 71%[b] LN1-3, 26%[b] LN>3, 3%[b]	All ET No CT	**DRFS 10yr**	**ROR-PT** nCounter		29.94 (p<0.0001)	0.720	NR	**L vs I:** 2.15 (1.21, 3.81), p=0.009; **L vs H:** 4.26 (2.44, 7.43), p<0.0001
						CLP			0.688		
Laenkholm 2015[56]	**DBCG** Cohort	HR+ HER2 NR N=2722	LN0, 46% LN1-3, 54%	All ET No CT	**DRFS 10yr**	**ROR-PT** nCounter		p<0.0001			HR (20-point change in ROR): 1.7 (1.5, 1.9)
LN0											
100% ET monotherapy											
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT	ER+ HER2- N=[redacted]	LN0	All ET No CT	**DRFI 10yr**	**ROR-PT** nCounter	[redacted]	[redacted]			
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8** R-RCT	ER+ HER2- N=984	LN0	All ET No CT	**DRFS 10yr**	**ROR-PT** nCounter		Over CLP: 20.32 (p<0.0001)	0.692	NR	
						CLP			0.639		
Wallden 2015[139]	**British Columbia** Cohort	ER+, 91% HER2- N=232	LN0	All ET No CT	**DRFS (time NR)**	**ROR-PT** nCounter			0.675[c]	NR	
						AOL			0.587[c]	NR	
						IHC-T			0.590[c]	NR	

Reference(s)	Cohort(s)	Population	Nodal status	Endo / chemo	Outcome	Test or comparator[a]	Likelihood ratio χ^2	Increase in LR χ^2 over CP[a]	C-index (AUC)	Increase in C-index (AUC) over CP[a]	Multivariable model (adj. for CP factors[a]): HR (95% CI)
LN+											
100% ET monotherapy											
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT; UK	ER+ HER2- N=[redacted]	LN1-3	All ET No CT	**DRFI 10yr**	**ROR-PT** nCounter	[redacted]	[redacted]			
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8** R-RCT	ER+ HER2- N=413	LN1-3, 89%[b] LN>3, 11%[b]	All ET No CT	**DRFS 10yr**	**ROR-PT** nCounter		Over CLP: 17.45 (p=0.0002)	0.743	NR	
						CLP			0.667		
Ejlertsen 2015[143]	**DBCG** Cohort	HR+ HER2 NR N=1466	LN1-3	All ET No CT	**DMFS 10yr**	**ROR-PT** nCounter					N1+ p<0.0001, N2+ p=0.0001; N3+: p=0.008
100% CT&ET											
Martin 2016,[83] 2014[92]	**GEICAM 9906** R-RCT	ER+ HER2- N=536	LN1-3, 64% LN>3, 36%	All ET **All CT**	**DMFS 10yr**	**ROR-PT** qRT-PCR (research)			0.644	Adding ROR-PT to EP-clin + CP: p=0.567	

ABCSG, Austrian Breast and Colorectal Cancer Study Group; AOL, Adjuvant! Online; CI, confidence interval; CLP, Clinical Linear Predictor; CP, clinical/pathological; CT, chemotherapy; CTS, Clinical Treatment Score; DBCG, Danish Breast Cancer Cooperative Group; DMFS, distant metastasis-free survival; DRFS, distant recurrence free survival; ER, oestrogen receptor; ET, endocrine therapy; H, high; HER2, human epidermal growth factor receptor-2; HR, hazard ratio; I, intermediate; IHC-T, IHC-4 + tumour size; L, low; LN, number of positive nodes; LR, likelihood ratio; NR, not reported; R-RCT, reanalysis of RCT.

[a]CP factors (ABSCG) = age, grade, nodal status, tumour size, Ki67. CP factors (GEICAM) = age, grade, nodal status, tumour size, treatment, ER, PR, Ki67. CTS (TransATAC) and CLP (ABCSG-8) = age, grade, nodal status, tumour size, treatment. CP factors (DBCG): not reported which; [b]Nodal status for all patients, NR for HER2- subgroup; [c]Estimated from graph. [redacted]

Table 58: Additional prognostic value for other outcomes: Prosigna

Reference(s)	Cohort(s)	Population	Nodal status	Endo / chemo	Outcome	Test or comparator[a]	C-index (AUC)	Increase in C-index (AUC) over CP[a]	Multivariable model (adj. for CP factors[a]): HR (95% CI)
LN status mixed									
Variable CT&ET									
Chia 2012[142]	**NCIC MA.12** R-RCT	73% HR+, N=398	LN0, 25% LN+, 75%	Some ET All CT	**OS 10 yr**	**ROR-PT** qRT-PCR	0.611		
					DFS 10yr	**ROR-PT** qRT-PCR	0.576		
Liu 2015[141]	**NCIC MA.21** R-RCT	58% ER+, 71% HER2- N=1094	LN0, 30% LN1-3, 42% LN>3, 28%	58% ET All CT	**RFS 8yr**	**ROR-PT** nCounter			**L/I vs H:** 1.98 (0.53, 7.45), p=0.311; **HR (cont score):** 1.01 (1.00, 1.02), p=0.029
LN0									
100% ET monotherapy									
Nielsen 2010[146]	**British Columbia** Cohort	ER+, 89% HER2- N=222	LN0	All ET No CT	**BCSS >10yr**	**ROR-PT** qRT-PCR	0.69	p=0.002 vs AOL p=0.033 vs IHC-T	
						AOL	0.56		
						IHC-T	0.63		
					RFS >10yr	**ROR-PT** qRT-PCR	0.67	p=0.001 vs AOL p=0.047 vs IHC-T	
						AOL	0.57		
						IHC-T	0.62		
Wallden 2015[139]	**British Columbia** Cohort	ER+, 91% HER2- N=232	LN0	All ET No CT	**BCSS (time NR)**	**ROR-PT** nCounter	0.672[b]		
						AOL	0.565[b]		
						IHC-T	0.560[b]		
LN+									
100% ET monotherapy									
Nielsen 2010[146]	**British Columbia** Cohort	ER+, 89% HER2- N=511	LN1-3, 70% LN>3, 30%	All ET No CT	**BCSS >10yr**	**ROR-PT** qRT-PCR	0.62	p=0.59 vs AOL p=0.30 vs IHC-T	
						AOL	0.63		
						IHC-T	0.61		
					RFS >10yr	**ROR-PT** qRT-PCR	0.60	p=0.72 vs AOL p=0.31 vs IHC-T	
						AOL	0.61		

Reference(s)	Cohort(s)	Population	Nodal status	Endo / chemo	Outcome	Test or comparator[a]	C-index (AUC)	Increase in C-index (AUC) over CP[a]	Multivariable model (adj. for CP factors[a]): HR (95% CI)
						IHC-T	0.59		

BCSS, breast cancer-specific survival; CI, confidence interval; CP, clinical/pathological; CT, chemotherapy; DFS, disease-free survival; ER, oestrogen receptor; H, high; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; HR+, hormone receptor-positive; I, intermediate; L, low; LN, number of positive nodes; LR, likelihood ratio; NR, not reported; OS, overall survival; RFS, relapse-free survival (locoregional or distant); R-RCT, reanalysis of RCT.

[a]CP factors (ABSCG) = age, grade, nodal status, tumour size, Ki67. CP factors (GEICAM) = age, grade, nodal status, tumour size, treatment, ER, PR, Ki67. CTS (TransATAC) and CLP (ABCSG-8) = age, grade, nodal status, tumour size, treatment. CLP . CP factors (MA.21): not reported which. [b]Estimated from graph.

4.6 Results: EndoPredict and EPClin

4.6.1 Development: EndoPredict and EPClin

EndoPredict and EPClin risk scores were trained on 964 ER+ HER2- endocrine-treated samples (65% node-negative) from a range of sources (Filipits *et al*., 2011).[48] EndoPredict generates an EP score based on the gene signature alone. The EPClin score is calculated from the EP score plus information on tumour size and nodal status.

4.6.2 Prognostic performance: EndoPredict and EPClin

Study designs: EndoPredict and EPClin

Three data sets, all re-analyses of RCTs, have been used to validate the prognostic performance of EndoPredict (Table 59). Analysis of UK-based patients from the TransATAC trial was reported by Buus *et al*. (2016)[34] and updated data for 878 patients (used in this report) were provided via personal communication with the TransATAC team (Sestak, 2017).[43] Analysis of 1702 patients pooled from the Austrian Breast and Colorectal Cancer Study Group (ABCSG)-6 and ABCSG-8 trials was reported by Dubsky *et al* (2013a and 2013b) plus subgroup analyses provided to NICE by Myriad Genetics.[48, 57-59] Finally, 555 patients from the Spanish GEICAM 9906 trial were analysed by Martin *et al*. (2014, 2016).[83, 92]

Patients: EndoPredict and EPClin

All three data sets either consisted of, or had analyses available for, ER+, HER2- patients. In terms of nodal status, two of the three data sets included LN0 patients (TransATAC[34, 43] and ABCSG-6+8[57-59]). All three data sets included LN+ patients, all of whom had 1-3 positive nodes (LN1-3) except in GEICAM 9906[83, 92] in which 36% had >3 positive nodes. Patients in all three analyses received 5 years of endocrine therapy. Patients in the GEICAM 9906 analysis[83, 92] also received adjuvant chemotherapy, while those in the other two analyses did not.

For TransATAC, two sets of data were presented in the analysis reported to the EAG via NICE.[43] The "full dataset" refers to data on all [redacted] patients with EndoPredict data available, while the "reduced dataset" refers to [redacted] patients with data for all four in-scope tests analysed in TransATAC. In this report, data for the "full dataset" is used where available; if not available than the "reduced dataset" is used. [redacted]

Tests and comparators: EndoPredict and EPClin

All three data sets assessed the tests as marketed (though in TransATAC[34] a correction factor was applied to account for differences in RNA extraction methods), using qRT-PCR and standard cut-offs for risk groups (5 for EndoPredict and 3.3 for EPClin). The three data sets were also used to evaluate

other in-scope tests as follows (see Section 4.8.1 on comparing tests). TransATAC was used to evaluate Oncotype DX, Prosigna and IHC4+C.[24, 36, 38, 39] GEICAM 9906 was used to evaluate a "research-based" version of PAM50 ROR-PT. [83, 92] ABCSG-8 (but not ABCSG-6) was used to evaluate Prosigna.[54, 55]

Quality assessment: EndoPredict and EPClin

All three data sets were validation studies and re-analyses of RCTs (Table 60). All analyses excluded some original trial patients (or this was unclear). Blinding of test assessors to outcomes was reported in two analyses.[34, 83, 92] All used standardised outcomes.

Results: EndoPredict and EPClin

Table 61, Table 62 and Table 63 present the data for all patients (mix of LN0 and LN+) and separate data for LN0 and LN+ patients.

Distribution of patients by risk group

The percentage of LN0 patients categorised as EPClin low-risk was [redacted] in TransATAC[43] and [redacted] in ABCSG-6+8.[57-59] Far fewer LN+ patients were categorised as EPClin low-risk: [redacted] in TransATAC,[43] [redacted] in ABCSG-6+8[57-59] and 13% in GEICAM 9906[83, 92] (Table 61).

Prognostic performance: unadjusted analyses

This section reports unadjusted analyses. Adjusted analyses, which show whether the test has prognostic value over clinicopathological variables, are reported in the section "Additional prognostic value"

LN0: Both analyses of LN0 patients (TransATAC[43] and ABCSG-6+8[57-59]) showed that EPClin was statistically significantly prognostic for 10-year DRFS/DRFI. The proportion of patients with 10-year DRFS/DRFI in the EPClin low-risk groups was [redacted] in TransATAC[43] and [redacted] in ABCSG-6+8[57-59] (Table 61). HRs for the low vs. high-risk groups were [redacted] in TransATAC[43] and [redacted] in ABCSG-6+8.[57-59] EndoPredict and EPClin remained statistically significantly prognostic for DRFI during both early (0-5-year) and late (5-10-year) follow-up in ABCSG-6+8.[58]

In terms of overall survival, EPClin was [redacted] for 10-year overall survival in the one study of LN0 patients reporting this outcome (TransATAC,[43] Table 62).

LN+: [redacted] analyses of LN+ patients showed that EPClin was statistically significantly prognostic for 10-year DMFS/DRFS/DRFI. The proportion of patients with 10-year DMFS/DRFS/DRFI in the

EPClin low-risk groups was [redacted] in TransATAC;[43] [redacted] in ABCSG-6+8;[57-59] and 100% in GEICAM 9906[83, 92] (Table 61). HRs for the low vs. high-risk groups were [redacted] in TransATAC;[43] [redacted] in ABCSG-6+8;[57-59] and for GEICAM not estimable since there were no events in the low-risk group (p<0.0001).[83, 92] EPClin was also statistically significantly prognostic for 10-year overall survival in GEICAM,[83, 92] [redacted] TransATAC[43] (Table 62). However, as noted above, only a relatively small proportion of LN+ patients were classed as low-risk (13% to [redacted] across studies) [43, 57-59, 83, 92]

Comparison to guidelines: In the ABCSG-6+8 analysis,[57] the hazard ratio for 10-year DRFI for low vs. intermediate/high-risk groups across all patients (two-thirds LN0) was higher for EPClin (HR 5.11, 95% CI: 3.48, 7.51, p<0.001) than when classifying patients as low/high risk according to any of three clinical guidelines: NCCN 2007 (HR 2.16, p=0.119), St Gallen 2011 (HR 2.78, p<0.001) or German S3 2008 guidelines (HR 2.20, p=0.014).

Patients at high clinical risk: The ABCSG-6+8 analysis[57] also reported results for patients classed as high or high/intermediate-risk via the three clinical guidelines: NCCN 2007, St Gallen 2011, and German S3 guidelines 2008. Around 60% were categorised as low-risk via EPClin. EPClin was statistically significantly prognostic for 10-year DRFI in these high-clinical-risk patients (Table 61).

Additional prognostic value

This section reports adjusted analyses, which indicate the additional prognostic value of IHC4 over clinicopathological factors. The clinicopathological factors adjusted for vary from study to study, and are detailed in the footnotes to the tables.

Likelihood ratios: The TransATAC analysis[43] reports a reduced dataset of patients where data for all four in-scope tests are available. Additional prognostic value was assessed via increases in likelihood ratio χ^2 for 10-year DRFI, for EPClin plus NPI or CTS, over NPI or CTS alone (Table 63). Increases in likelihood ratio χ^2 were [redacted].

C-indexes (AUC): In LN+ patients in GEICAM 9906, adding EndoPredict to a combination of clinicopathological variables increased the C-index from 0.654 to 0.672 (p=0.0018), while EPClin gave a higher C-index of 0.693 (p=NR; Table 63).[92] In ABCSG-6+8 (two-thirds LN0), the C-index was only reported for years 5-10 (no data for years 0-5).[58] In this period, the C-index increased when adding EndoPredict to a combination of clinical variables or to AOL (both p<0.001; Table 63).

Multivariable Cox models: Both ABCSG-6+8[57-59] (mix of LN0/LN+) and GEICAM 9906[83, 92] (LN+) used multivariable analyses to show that EndoPredict (no data reported for EPClin) was an independent prognostic variable for 10-year DMFS/DRFI after adjustment for clinical variables ($p<0.001$;[57-59] $p=0.003$;[83, 92] Table 63).

Discussion: EndoPredict and EPClin prognostic performance

The prognostic value of EPClin was based on three reanalyses of RCTs (all ER+ HER2-, total N=3,135).[34, 57-59, 83, 92, 101] Two reported on LN0 patients (total N=1,836)[34, 57-59, 101] and all three on LN+ patients (total N=1,201; two of three restricted to LN1-3). Patients received endocrine monotherapy in two trials[34, 57-59, 101] and all patients received endocrine and chemotherapy in the GEICAM trial.[83, 92] All excluded some original trial patients (or this was unclear), sometimes due to insufficient tumour sample which may introduce bias due to attrition of patients with smaller tumours.

The percentage of patients categorised as EPClin low-risk in LN0 patients (two studies)[34, 57-59, 101] was [redacted] and [redacted], and the percentage high-risk was [redacted] and [redacted]. For LN+ patients (three studies),[34, 43, 57-59, 83, 92] the percentage categorised as low-risk ranged from 13% to [redacted] and the percentage high-risk from [redacted] to 87%. EPClin was statistically significantly prognostic for DRFS/DRFI for all unadjusted analyses at 10 years (and most analyses at 5 years) in LN0 and LN+ patients,[34, 57-59, 83, 92, 101] and in one analysis of patients at high clinical risk.[57] Rates of 10-year DRFS/DRFI in EPClin low-risk groups were [redacted] to [redacted] in LN0 patients (two studies),[57-59, 101] and in LN+ patients ranged from [redacted] (two studies with only endocrine therapy)[57-59, 101] to 100% (one study using endocrine and chemotherapy).[83, 92] Use of chemotherapy in the GEICAM study[83, 92] could influence patient outcomes in either direction: negatively due to potential selection of higher-risk patients, or positively due to the effect of chemotherapy.

In terms of additional prognostic value, TransATAC reported [redacted].[101] Two further studies reported that the EndoPredict EP score added statistically significant information over clinicopathological variables in LN+ and mixed LN0/LN+ patients (based on multivariable analyses and differences in C-index (AUC) for 10-year DMFS/DRFI); however neither reported the additional prognostic value of EPClin.[57-59, 83, 92]

Conclusions: EndoPredict and EPClin prognostic performance

Based on three reanalyses of RCTs (total N=3,135) in ER+ HER2- endocrine-treated patients, EPClin was statistically significantly prognostic for unadjusted analyses of 10-year DRFS/DRFI in LN0 and LN+ patients. The percentage of patients categorised as EPClin low-risk ranged from [redacted] to [redacted] for LN0 patients and 13% to [redacted] for LN+ patients. The 10-year DRFS/DRFI rates for low-risk patients

were approximately [redacted] in LN0 and LN+ patients receiving endocrine therapy alone. [redacted], while in two further studies the EP score added statistically significant information over clinicopathological variables in mixed LN0/LN+ and LN+ patients (no data for EPClin). There was no evidence relating to chemotherapy benefit or clinical utility for EndoPredict or EPClin.

Table 59: Characteristics of prognostic studies: EndoPredict and EPClin

Reference(s)	Cohort(s)	N pts	Country	Study design	Test	Details of test	Cut-offs	Other tests	Population	Nodal status	Endo / chemo
Reanalyses of RCTs: LN status mixed											
100% ET monotherapy											
Sestak 2017 (data request),[43] Buus 2016[34]	**TransATAC**	[redacted]	UK	R-RCT	EPClin	FFPE qRT-PCR, Sividon	3.3	O-DX ROR-PT IHC4+C	ER+ HER2- Postmeno 100% female	LN0, [redacted] LN1-3, [redacted]	All ET 5yr No CT
Dubsky 2013a,[57] 2013b,[58] Myriad[59]	**ABCSG-6+8**	1702 (all) [redacted] (LN0-3)	Austria	R-RCT	EP EPClin	FFPE qRT-PCR	5 3.3	ROR-PT (ABSCG-8)	ER+ HER2- Postmeno Stage I-II 100% female	LN0, 68% LN1-3, 27% LN>3, 5%	All ET 5yr No CT
Reanalyses of RCTs: LN+											
100% CT&ET											
Martin 2016,[83] 2014[92]	**GEICAM 9906**	555	Spain	R-RCT	EP EPClin	FFPE qRT-PCR	5 3.3	ROR-PT	ER+ HER2- 46% postmeno Stage II-III 100% female	All N+ LN1-3, 64% LN>3, 36%	All ET 5yr All CT

ABCSG, Austrian Breast and Colorectal Cancer Study Group; CT, chemotherapy; ER, oestrogen receptor; ET; endocrine therapy; FFPE, formalin-fixed paraffin-embedded; HER2, human epidermal growth factor receptor 2; LN, number of positive nodes; qRT-PCR, quantitative reverse transcription polymerase chain reaction; R-RCT, reanalysis of RCT [redacted]

Table 60: Quality assessment of prognostic studies: EndoPredict and EPClin

Reference(s)	Cohort(s)	Derivation or validation?	Study design appropriate?	All eligible patients included?	Blinding (of test assessors to outcomes)?	Outcome definition standardised or *a priori*?	Applicability: Patient Spectrum	Applicability: Test as per decision problem?
Sestak 2017 (data request),[43] Buus 2016[34]	**TransATAC**	V	Y, R-RCT, no chemo	N, InT, MS, TP	Y	Y	Y	Y
Dubsky 2013a,[57] 2013b,[58] Myriad[59]	**ABCSG-6+8**	V	Y, R-RCT, no chemo	UC	UC	Y	Y (for subgroup analysis of LN0-3)	Y
Martin 2016,[83] 2014[92]	**GEICAM 9906**	V	N, R-RCT, adj chemo	N (reason NR)	Y	Y	N (36% LN>3)	N, Prosigna via qRT-PCR then microarray

Y, yes; N, no; UC, unclear
ABCSG, Austrian Breast and Colorectal Cancer Study Group; D, Development; InT, insufficient tissue; MS, missing samples; LN, number of positive nodes; qRT-PCR, quantitative reverse transcription polymerase chain reaction; R-RCT, reanalysis of RCT; TF, test failure; V, validation

Table 61: Prognostic performance of EndoPredict and EPClin: distant recurrence-free survival (DRFS/DRFI)[a]

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Endo / chemo	Test or comparator	% pts per group		% DRFS/DRFI risk: 0-5 yr		% DRFS/DRFI risk: 0-10 yr		DMFS/DRFS/DRFI[a]: HR (95% CI)
						Low	High	Low	High	Low	High	
Reanalyses of RCTs: LN status mixed												
100% ET monotherapy												
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT; UK	ER+ HER2- [redacted]	LN0, [redacted] LN1-3, [redacted]	All ET No CT	**EPClin**	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Dubsky 2013a,[57] 2013b,[58] Myriad[59]	**ABCSG-6+8** R-RCT; Austria	ER+ HER2- N=1702	LN0, 68% LN1-3, 27% LN>3, 5%	All ET No CT	**EP**	49	51	-	-	-	-	**0-5 yr:** 2.80 (1.81, 4.34), p<0.001 **5-10 yr:** 3.28 (1.48, 7.24), p=0.002
					EPClin	63	37	-	-	95.3	-	**0-5 yr:** 4.82 (3.12, 7.44), p<0.001 **0-10 yr:** 5.11 (3.48, 7.51), p<0.001 **5-10 yr:** 6.25 (2.72, 14.36), p<0.001
		[redacted]	LN0, [redacted] LN1-3, [redacted]	All ET No CT	**EPClin**	[redacted]	[redacted]	-	-	[redacted]	[redacted]	[redacted]
Reanalyses of RCTs: LN0												
100% ET monotherapy												
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT; UK	ER+ HER2- [redacted]	LN0	All ET No CT	**EPClin**	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Dubsky 2013a,[57] 2013b,[58] Myriad[59]	**ABCSG-6+8** R-RCT; Austria	ER+ HER2- [redacted]	LN0	All ET No CT	**EPClin**	[redacted]	[redacted]	-	-	[redacted]	[redacted]	[redacted]
Reanalyses of RCTs: LN+												
100% ET monotherapy												
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT; UK	ER+ HER2- [redacted]	LN1-3	All ET No CT	**EPClin**	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Dubsky 2013a,[57] 2013b,[58] Myriad[59]	**ABCSG-6+8** R-RCT: Austria	ER+ HER2- [redacted]	LN1-3	All ET No CT	**EPClin**	[redacted]	[redacted]	-	-	[redacted]	[redacted]	[redacted]

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Endo / chemo	Test or comparator	% pts per group		% DRFS/DRFI risk: 0-5 yr		% DRFS/DRFI risk: 0-10 yr		DMFS/DRFS/DRFI[a]: HR (95% CI)
						Low	High	Low	High	Low	High	
100% CT&ET												
Martin 2016,[83] 2014[92]	**GEICAM 9906** R-RCT; Spain	ER+ HER2- N=555	LN1-3, 64% LN>3, 36%	All ET All CT	**EP**	25	75	-	-	93	70	**0-10 yr:** 4.8, (2.5, 9.6, p<0.0001)
					EPClin	13	87	-	-	100	72	**0-10 yr:** Not estimable, p<0.0001
	Premeno, N=300	NR	LN1-3, 64% LN>3, 36%		**EP**	24	76			93	67	**0-10 yr:** 6.7 (2.4, 18.3, p<0.0001)
	Postmeno, N=255	NR			EP	27	73			92	74	**0-10 yr:** 3.3 (1.3, 8.5, p=0.0069)
	Premeno, N=300	NR	LN1-3, 64% LN>3, 36%		**EPClin**	12	88			100	70	**0-10 yr:** HR NR, p=0.0006
	Postmeno, N=255	NR			**EPClin**	13	87			100	76	**0-10 yr:** HR NR, p=0.0023
High/intermediate-risk via clinical guidelines (LN0/+)												
Dubsky 2013a[57]	**ABCSG-6+8** R-RCT; Austria	**NCCN** N=1603	LN+/- (% NR)	All ET No CT	**EPClin**	61	39	-	-	95	77	**0-10 yr:** 5.09 (3.42, 7.58), p<0.001
		St Gallen N=1358			**EPClin**	58	42	-	-	95	75	**0-10 yr:** 5.18 (3.38, 7.93), p<0.001
		S3 N=1454			**EPClin**	58	42	-	-	95	76	**0-10 yr:** 5.60 (3.64, 8.61), p<0.001

-, not reported; ABCSG, Austrian Breast and Colorectal Cancer Study Group; CI, confidence interval; CT, chemotherapy; DMFS, distant metastasis-free survival; DRFI, distant recurrence free interval; DRFS, distant recurrence free survival; ET; endocrine therapy; ER, oestrogen receptor; HR, hazard ratio; HER2, human epidermal growth factor receptor 2; LN, number of positive nodes; R-RCT, reanalysis of RCT.
[a]DMFS (GEICAM, ABSCG for ROR-PT); DRFI (TransATAC); DRFI (ABSCG for EPClin)

Table 62: Prognostic performance of EndoPredict and EPClin: overall survival

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Endo / chemo	Test	% pts per group		% OS risk: 0-5 yr		% OS risk: 0-10 yr		OS: HR (95% CI) 0-5 yr
						Low	High	Low	High	Low	High	
Reanalyses of RCTs: LN status mixed												
100% ET monotherapy												
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT; UK	ER+ HER2- [redacted]	LN0, [redacted] LN1-3, [redacted]	All ET No CT	**EPClin**	[redacted]						[redacted]
Reanalyses of RCTs: LN0												
100% ET monotherapy												
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT; UK	ER+ HER2- [redacted]	LN0	All ET No CT	**EPClin**	[redacted]		[redacted]		[redacted]		[redacted]
Reanalyses of RCTs: LN+												
100% ET monotherapy												
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT; UK	ER+ HER2- [redacted]	LN1-3	All ET No CT	**EPClin**	[redacted]		[redacted]		[redacted]		[redacted]
100% CT&ET												
Martin 2016,[83] 2014[92]	**GEICAM 9906** R-RCT; Spain	ER+ HER2- N=536	LN1-3, 64% LN>3, 36%	All ET All CT	**EP**	25	75	-	-	92	67[a]	**0-10 yr:** 3.9 (2.0, 7.5), p<0.0001
					EPClin	13	87	-	-	99 [a]	69 [a]	**0-10 yr:** 19.4 (2.7, 138.7), p<0.0001

-, not reported; ABCSG, Austrian Breast and Colorectal Cancer Study Group; CI, confidence interval; CT, chemotherapy; ET, endocrine therapy; ER, oestrogen receptor; HR, hazard ratio; HER2, human epidermal growth factor receptor 2; LN, number of positive nodes; OS, overall survival; R-RCT, reanalysis of RCT.
[a] Estimated off graph
[redacted]

Table 63 Additional prognostic value for DRFI/DMFS: EndoPredict and EPClin

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Endo / chemo	Outcome	Test or comparator [a]	Likelihood ratio χ^2	Increase in LR χ^2 over CP factors	C-index (AUC)	Increase in C-index (AUC) over CP factors[a]	Multivariable model (adj. for CP factors[a]): HR (95% CI)
Reanalyses of RCTs: LN status mixed											
100% ET monotherapy											
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT; UK	ER+ HER2- [redacted]	LN0, [redacted] LN1-3, [redacted]	All ET No CT	**DRFI 10yr**	**EPClin**	[redacted]	[redacted]			
Dubsky 2013a,[57] 2013b[58]	**ABCSG-6+8** R-RCT; Austria	ER+ HER2- N=1702	LN0, 68% LN1-3, 27% LN>3, 5%	All ET No CT	**DMFS 0-5yr**	**EP**					1.20 (1.10, 1.31), p<0.001
						EPClin					
					DMFS 5-10yr	**EP**					1.28 (1.10, 1.48), p=0.001
						EPClin			0.786		
						EP vs AOL			0.765	p<0.001	
						EP vs CP [a]			0.716	p<0.001	
						AOL			0.674		
						CP factors [a]			0.644		
Reanalyses of RCTs: LN0											
100% ET monotherapy											
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT; UK	ER+ HER2- [redacted]	LN0	All ET No CT	**DRFI 5 years**	**EPClin**	[redacted]	[redacted]			
					DRFI 10yr		[redacted]	[redacted]			

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Endo / chemo	Outcome	Test or comparator [a]	Likelihood ratio χ^2	Increase in LR χ^2 over CP factors	C-index (AUC)	Increase in C-index (AUC) over CP factors[a]	Multivariable model (adj. for CP factors[a]): HR (95% CI)
Reanalyses of RCTs: LN+											
100% ET monotherapy											
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT; UK	ER+ HER2- [redacted]	LN1-3	All ET No CT	**DRFI 5 yr**	**EPClin**	[redacted]	[redacted]			
					DRFI 10yr		[redacted]	[redacted]			
100% CT&ET											
Martin 2016,[83] 2014[92]	**GEICAM 9906** R-RCT; Spain	ER+ HER2- N=536	LN1-3, 64% LN>3, 36%	All ET **All CT**	**DMFS 10yr**	**EPClin**			0.693	NR	
						EP vs CP[a]			0.672	p=0.0018	
						EP			0.657		1.1 (1.0, 1.2), p=0.003
						CP factors[a]			0.654		

ABCSG, Austrian Breast and Colorectal Cancer Study Group; AOL, Adjuvant! Online; CI, confidence interval; CP, clinical/pathological; CT, chemotherapy; CTS, Clinical Treatment Score; DMFS, distant metastasis-free survival; DRFS, distant recurrence free survival; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; LN, number of positive nodes; LR, likelihood ratio; R-RCT, reanalysis of RCT.

[a]CP factors (ABSCG) = age, grade, nodal status, tumour size, Ki67; CP factors (GEICAM) = age, grade, nodal status, tumour size, treatment, ER, PR, Ki67

[redacted]

4.7 Results: IHC4

4.7.1 Development and analytic validity: IHC4

The IHC4 score was derived in a sample of 1,125 patients from the TransATAC trial.[24] Tumour blocks were obtained from patients who had already undergone Oncotype DX testing (patients first reported in Dowsett *et al*. 2010)[35] and for whom sufficient tissue was available for IHC4 testing. Patients were HR+, 90% were HER2-, 26% were LN+ (but the percentage with >3 positive nodes was not reported) and 100% were post-menopausal. As such, the test was developed for a patient spectrum that is wider than the patients defined in the decision problem (which is HR+, HER2-, LN0-3 patients).

A summary of the technical methodology used to conduct the test is given in Appendix 3. In brief, the process involved constructing tissue microarrays with slides of three representative areas containing tumour cells, which were reviewed by a pathologist and/or experienced lab technician. Three cores were assembled for each patient. The immunohistochemistry and scoring of the slides was conducted as described elsewhere.[149, 150] ER was quantified using the H-score, and ER_{10} obtained by dividing the H-score by 30 (to give a value between 0 and 10). PGR10 was obtained by dividing the percent of cells stained positive for PgR by 10 (to give a value between 0 and 10). HER2 was scored according to manufacturer's recommendations (3+ was positive), with fluorescent *in situ* hybridisation to confirm equivocal (2+) samples. Ki-67 was scored as the percent positively stained cells.

The algorithm was developed in two parts, one using the four IHC components, the other using clinicopathological characteristics of nodal status, tumour size, grade, age and treatment (to account for survival advantages in patients whose endocrine therapy was anastrazole instead of tamoxifen). The most informative combination of the four IHC variables to predict time to distant recurrence (equivalent to DRFI, 100 months median follow-up) was derived using multivariable proportional hazard models and change in likelihood ratio X^2. The model derived was:

IHC4 = $94.7 \times (0.100\ ER_{10}\ 0.079\ PgR_{10} + 0.586\ HER2 + 0.240 \ln (1 + 10 \times Ki67))$.

with likelihood ratio X^2 4 *df*= 39.1; p<0.0001

A further model was developed that incorporated the clinicopathological variables, and the **IHC4+C** score was obtained by summing the scores provided from the two algorithms and multiplying by 100.

Clinical score = $100 \times (0.417N_{1\text{-}3} + 1.566N_4 + 0.930(0.497T_{1\text{-}2} + 0.882T_{2\text{-}3} + 1.838T_{>3} + 0.559Gr_2 + 0.970Gr_3 + 0.130Age_{\geq 65} - 0.149Ana))$

where N_j, T_j, Gr_j, and Age_j denote categories of nodal status, tumor size, grade, and age, respectively, and Ana denotes treatment with anastrozole as opposed to tamoxifen. A shrinkage factor was applied to account for overfitting. The likelihood ratio χ^2 for the clinical variables (9 *df*) was 147, p not reported.

Whilst the score was derived using DRFI, and in a cohort containing some LN+ and some HER2+ patients, the authors state that similar IHC4 scores and models were obtained using the endpoint "all recurrences" and LN0 only patients. In the LN0 group, the likelihood ratio χ^2 was 35.4 for the IHC4 component, but the clinical variables were less informative, with χ^2=40.7 (Table 64) compared to the models in the full cohort.

Table 64: Data relating to the derivation of IHC4 score and IHC3. DRFI (100 months median follow-up). All data from TransATAC

Reference; N	**Cohorts**	**Population**	**Nodal status**	**Endo / chemo**	**Likelihood ratio χ^2**	**DRFI: HR (95% CI) Unadjusted, 0-25th vs 75-100th percentile:**	**DRFI: HR (95% CI) Multivariable[a]**
Cuzick 2011[24] N=1,125	TransATAC	100% HR+ 90% HER2- Postmeno	LN+/-	100% ET monotherapy	**IHC4:** 39.1, p<0.0001 **Clin:** 147, p NR	**IHC4:** 5.7 (3.4 9.7)	**IHC4:** 3.9 (2.4, 6.7)
N=793			LN0		**IHC4:** 35.4, p NR **Clin:** 40.7, p NR		
N=1,066		100% HER2-	LN+/-		**IHC3:** 22.4, p<0.0001		

IHC4, IHC4 component alone; Clinical, clinical component alone
a multivariable model assumed to include IHC4 score and Clinical score as separate components

IHC3 Derivation

A further analysis was conducted in a group of patients who were HER2-, which negated the need for the HER2 component of the IHC4 score. A revised algorithm was developed:

$$IHC3 = 93.1 \times (0.086\ ER_{10} - 0.081\ PgR_{10} + 0.281 \ln (1 + 10 \times Ki67))$$

which was virtually identical to IHC4 when HER2 was negative and was also highly prognostic with χ^2 22.4, p<0.0001 (Table 64).

AIC analysis of HR+, HER2-, LN0-3 patient in TransATAC

The TransATAC team conducted analyses for the EAG in a subgroup of the TransATAC data set, specified by the decision problem. Patients who had been tested for any of IHC4, Oncotype DX, Prosigna or EndoPredict were included. [redacted] These data are presented alongside the other prognostic data for IHC4 (see Section 4.2.5.2), for ease of comparison, but it should be noted that these patients constitute the derivation cohort, and the prognostic value of IHC4 is likely to be overestimated in TransATAC as a consequence, and that the data reported in Table 64 is from the same patients. The new analyses used a cut off of <10% risk, 10-20% and >20% risk to define low, intermediate and high-risk groups.

Analytic Validity

A rapid review of the analytic validity of IHC4 will follow as an addendum to this report.

4.7.2 Prognostic performance: IHC4 and IHC4+C

In addition to the TransATAC derivation cohort[24, 43] (see Section 4.2.5.1 above), IHC4 has been reported in eleven separate cohorts, reported across fourteen publications (Table 65).[23, 24, 62, 85, 87, 88, 93, 108, 109, 111, 151-154] The size of the studies ranged from N=105[153] to 4,598.[151] Data relating to the subgroup of patients relevant to the decision problem (HR+, HER2-, LN0-3) from the derivation cohort (TransATAC) were provided in a personal communication from the transATAC team[43]. One cohort (Tamoxifen vs Exemestane Adjuvant Multinational(TEAM) trial) was reported in two separate analyses,[23, 62, 151] with different aims (validation of IHC4;[23, 151] prognosis of early or late recurrence)[62] and different numbers of patients (n=4598;[23, 151] n=2513)[62] as Stephen *et al.* 2014[62] recruited only those who had received endocrine monotherapy. Laboratory methodologies for conducting IHC4 varied across studies, and is discussed in more detail below (section "IHC4 methodology and cut-offs: IHC4 and IHC4+C prognostic performance").

Study designs: IHC4 and IHC4+C prognostic performance

Five of the validation cohorts[23, 62, 88, 93, 108, 109, 111, 151, 154] and the derivation cohort[43] were a reanalysis of prospectively collected RCT data, using archived tissue samples. The remaining six studies[24, 62, 85, 87, 152, 153] were analyses of cohorts of routinely collected patient data; one of these was a case-control study (Table 65).[87]

The derivation RCT was from the UK:

- ATAC[35, 43] – was an international trial, with a translational research continuation (TransATAC) that investigated prognosis of breast cancer recurrence. Only UK samples were included in this analysis. The trial evaluated anastrozole, tamoxifen, or the combination of

both treatments. Recruitment ended in 2006. There are numerous TransATAC publications that met the criteria for the review,[24, 34-42] but here we present data provided by the TransATAC team as a personal communication to the EAG, which restricts to HR+, HER2-, LN0-3 patients.[43]

Two RCTs were conducted in the UK and other countries:

- The TEAM trial[155] recruited patients between 2001 and 2006 and randomised them to exemestane alone or following tamoxifen.
- The IES (Intergroup Exemestane Study) trial[156] recruited patients between 1998 and 2003 and randomised them to one of two endocrine therapies: exemestane or tamoxifen.

The remaining three RCTs were conducted in Europe (Spain and Germany):

- WSG (West German Study Group) Plan B trial[157] recruited patients between 2009 to 2011, and randomised them to anthracycline-free or anthracycline-taxane based chemotherapy. In an early protocol amendment, patients with Oncotype DX RS <12 were not given chemotherapy.
- GEICAM 9906 (Grupo Espanol de Investigation en Cancer de Mama)[94, 154] randomised patients with node-positive disease to adjuvant fluorouracil, epirubicin, and cyclophosphamide versus fluorouracil, epirubicin, and cyclophosphamide followed by weekly paclitaxel, and patients with HR-positive disease subsequently received adjuvant endocrine therapy.
- WSG-AGO-Doc (West German Study Group epirubicine and cyclophosphamide-Doc)[158] recruited patients between 2000 and 2005 and randomised them to taxane or non-taxane-based chemotherapy regimens.

There were a total of six retrospective studies. Three studies were from the UK or Europe:

- A cohort from Nottingham, UK[24]
- A cohort from Edinburgh, UK[62]
- A cohort from France (Institut Curie).[153]

One study was from the USA, where clinical advice to the EAG suggests chemotherapy rates are generally higher:

- Patients in the Kaiser Permanente Northwest[87] database

A further two studies were from East Asia:

- A cohort from China[85] from the Sun Yat-sen Memorial Hospital and the Third Hospital of Nanchang City

- A cohort from Taiwan[152] from the National Taiwanese University Hospital.

Clinical advice received by the EAG suggests that these two East Asian studies may be less generalisable to the English context because: (a) patients were treated according to usual clinical practice and this may differ in these countries compared with the UK enough to affect prognostic outcomes, and (b) it is possible that people of different ethnicities have different underlying risk profiles and disease natural history. For this reason, data from these studies should be interpreted with caution and with reference to data from studies where the ethnic profile and clinical practice is similar to the UK.

Patients and treatments: IHC4 and IHC4+C prognostic performance

The studies were highly heterogeneous in terms of the patients recruited and the treatments given. Overall, only the derivation cohort (TransATAC)[43] reported an analysis of 100% ER+, HER2-, LN0-3 patients who had not undergone chemotherapy but had received 5 years of endocrine therapy. Data from this cohort were provided to the EAG as Academic in Confidence, and has limitations in that: (a) it is also the derivation cohort for the IHC4 score, so some overfitting (leading to overestimation of prognostic performance) can be expected, (b) it only recruited post-menopausal women, and (c) it did not recruit PR+ patients.

As such, most of the evidence base has low generalisability to the decision problem, and even the most relevant available evidence has limitations in that TransATAC is the derivation cohort for IHC4 and only recruited ER+ post-menopausal patients. These limitations along with the problems with patient cohorts and treatments given should be borne in mind when interpreting the evidence base.

What follows is a more detailed look at the evidence base from the perspective of each factor of importance to the decision problem:

Lymph node status: The IHC4 test was developed for use amongst LN+ or LN0 patients, though this assessment focusses on those with LN0-3. Amongst the RCT reanalysis studies (Table 65), TransATAC[24, 43] and WSG Plan B[108, 109, 111] recruited or reported a subgroup of patients with LN0-3, whilst TEAM[23, 62, 151] and IES[88] recruited patients with any lymph node status, and did not report the percentage with more than three positive nodes. GEICAM 9906[154] and WSG-AGO-Doc[93] recruited LN+ patients, with 38% patients having LN>3 in GEICAM 9906 but all patients being LN1-3 in WSG-AGO-Doc.

Amongst the retrospective cohort and case control studies, the Nottingham,[24] the Kaiser Permanente,[87] the Edinburgh (BCS),[62] the Chinese[85] and the Taiwanese[152] data sets all recruited both LN positive

and negative patients, but did not report the proportion who were LN>3. The cohort from the Institut Curie[153] were all LN0.

Hormone receptor status: IHC4 was intended for use in HR+ patients. All studies recruited HR+ or ER+ patients except the IES RCT[88] and the study from Taiwan,[152] both of which did not report the percentage of patients who were HR+ (Table 65)

HER2 status: The IHC4 test was developed for both HER2+ and HER2- patients, though this assessment focusses on HER2- patients. Amongst the RCT reanalysis studies (Table 65), TransATAC,[43] WSG Plan B,[108, 109, 111] GEICAM 9906[154] and WSG-AGO-Doc[93] recruited or reported a subgroup of HER2- patients, whilst TEAM[23, 62, 151] and IES[88] did not report the HER2 status of patients. Amongst the retrospective studies (Table 65), the Kaiser Permanente cohort,[87] Institut Curie[153] cohort and the Chinese[85] cohort all recruited 100% HER2- patients whist the Nottingham cohort,[24] Edinburgh (BCS)[62] cohort and the Taiwanese[152] cohort recruited a proportion who were HER2+, or did not report this.

Treatments: IHC4 was intended for use in predicting distant disease recurrence assuming 5 years of endocrine therapy in HER2- patients, and no chemotherapy. As such, failure to treat all HER2- patients with endocrine therapy or treatment of any patients with chemotherapy will affect the survival of patients, and the estimates of prognostic performance may also be affected, especially if the proportion of patients given or not given treatment differs in each risk group; in theory, assuming patients in the higher risk categories get chemotherapy more often (if there is some concordance between clinically-defined risk and tumour profiling test risk), this is likely to reduce the separation in observed risk between IHC4 risk categories reported in these studies. This type of problem is theoretically possible in the retrospective studies of routine practice, where the IHC4 markers alone are likely to have affected treatment decisions, but also in the RCT study WSG Plan B, where patients with Oncotype DX RS<12 were given endocrine monotherapy and those with RS≥12 were given chemotherapy and endocrine therapy, if there is some concordance between Onctoype-DX and IHC4 categorisations.

Only two data sets treated all HER2- patients with endocrine therapy and did not treat any patients with chemotherapy (TransATAC[24, 43] and the analysis of TEAM conducted by Stephen *et al*. 2014, (Table 65).[62] The analysis by Stephen *et al*. is likely to suffer from spectrum bias as patients were excluded if they received chemotherapy, and these patients are likely to be systematically different to those who did not as chemotherapy decisions were based on clinical practice in this trial (only exemestane/tamoxifen treatment was randomised). Five studies treated all HER2- patients with endocrine therapy but also treated some patients with chemotherapy, or were assumed to have treated

some patients with chemotherapy as they were treated according to routine practice (WSG Plan B,[108, 109, 158] IES,[88] GEICAM 9906, [154] China[85] cohort and the Bartlett *et al*. 2016[23, 151] analysis of TEAM, (Table 65). The Nottingham IHC4 validation cohort[24] included some HER2- patients who were not treated with endocrine therapy, but applied a correction in the analysis to account for this; however, as the cohort were patients undergoing routine therapy, it is likely that some received chemotherapy and no adjustment for this is reported (Table 65). Three studies (Kaiser Permanente,[87] WSG-AGO-Doc,[93] Taiwan[152] (Table 65) did not treat all patients with endocrine therapy or did not report the proportion who were treated, and one study (Institut Curie[153]) treated some patient with endocrine therapy, but none with chemotherapy.

IHC4 methodology and cut-offs: IHC4 and IHC4+C prognostic performance

The methodology for conducting IHC4 is well known to be problematic. Concerns centre on the performance of Ki-67, and specifically the lack of standardisation of laboratory and analytic methods.[23 159] We have documented the methods reported in the included studies in Appendix 3 for reference, but as it was beyond the expertise of the EAG to identify which methods are in accordance with UK practice, and the methods used by the derivation group, [24] we sought advice from the IHC4 team. Their judgement regarding the compatibility of the methods used in the studies to their own methodology (used in their laboratory) is given in Appendix 3, and in Table 65. Seven datasets were analysed using IHC4 methodologies that were the same or very similar to the IHC4 team's own methodology (referred to from here on in as the standard IHC4 methodology) (TransATAC AIC[43], TEAM,[23, 62, 151] the Nottingham cohort,[24] the BCS cohort,[62] the Institut Curie[153] cohort, GEICAM 9906[154] and WSG-AGO-Doc)[93] whilst the remaining five datasets were analysed with methodologies that were unclear or dissimilar to the IHC4 team's methods (WSG-Plan B,[108, 109, 111] the Kaiser Permanente cohort,[87] IES,[88] the Chinese cohort [85] and the Taiwanese cohort[152]). Results have not been excluded by IHC4 methodology, as methodologies are not currently standardised and as such all data is of some relevance.

A brief description of methods is given for each study in Table 65. Three studies were unclear whether it was the IHC4 score or the IHC4+C score, as they referenced Cuzick *et al*. 2011,[24] but not which score; attempts were made to clarify this point with the authors where contact details were available (IES;[88] Institut Curie cohort;[153] WSG-AGO-Doc).[93] Most other studies used only the IHC4 component of the IHC4 score, without using the clinical component (see section 4.7.1) (TEAM analyses by Barlett *et al* 2016[23] and Stephen *et al*. 2014;[62] Edinburgh cohort;[62] WSG Plan B;[108, 109, 111] GEICAM 9906; [154] Kaiser Permanente cohort;[87] China cohort[85]; Taiwan cohort).[152] Data definitely stated to relate to IHC4+C was only available for the Nottingham cohort[24] [redacted]

The original IHC4[24] analysis did not report numerical cut-offs for the definition of high, intermediate and low-risk patients, but used quartiles and tertiles, whilst the AIC analysis of TransATAC uses 10%, 10-20% and >20% risk or recurrence as cut offs. Other studies used quartiles and/or tertiles to define the cut-offs, or used the score as a continuous variable in cox proportional hazard models, except the Stephen *et al*. analysis of BCS and TEAM, [62] which stated that the same cut-offs as Cuzick *et al*.[24] were used.

The Insitut Curie trial,[153] which recruited all HER2- patients, stated that they used the IHC3 version of the IHC4 algorithm, where HER2 status is not incorporated. It is unclear whether other studies that recruited only HER2- patients and referenced Cuzick *et al*. 2011[24] as the source of the algorithm also used the IHC3 score, as reported by Cuzick *et al* 2011.[24]

Comparators: IHC4 and IHC4+C prognostic performance

No studies of IHC4 compared the score to a comparator. The TransATAC AIC study reported data with NPI and CTS as comparators. The Nottingham cohort analysis also reported a comparison to the clinical score component of the IHC4+C score.

Quality assessment: IHC4 and IHC4+C prognostic performance

The evidence base was of generally poor quality; no study scored well on all items (Table 66). Of particular concern was the high number of studies that included patients who had received chemotherapy treatment (see section entitled "*Treatments*" above), and the high number of studies that were not able to include all relevant patients due to missing samples or insufficient tissue. This is likely to introduce spectrum bias, as patients with smaller tumours are more likely to have been excluded due to insufficient tissue being available. Very few studies reported that they blinded test assessors, leaving the evidence base at high risk of ascertainment bias. The applicability of the IHC4 tests conducted to the decision problem is acceptable in seven studies (TransATAC AIC[43], TEAM,[23, 62, 151] the Nottingham cohort,[24] the BCS cohort,[62] the Institut Curie[153] cohort, GEICAM 9906[154] and WSG-AGO-Doc)[93], but unknown or not compatible in five (WSG-Plan B,[108, 109, 111] the Kaiser Permanente cohort,[87] IES,[88] the Chinese cohort [85] and the Taiwanese cohort[152]).

Table 65: Characteristics of prognostic studies: IHC4 and IHC4+C

Reference; N	Cohorts (Country)	Study design	Details of test[a]	Compatibility [b] &Algorithm	Population	Nodal status	Endo / chemo
Subgroup, relevant to the decision problem, of derivation cohort: LN0/+							
TransATAC [redacted][c]	TransATAC (UK)	R-RCT	FFPE. Biomarker expression was measured by IHC. HER2 was confirmed by FISH if ≥IHC2+. ER used 6F11 antibody (Vector Laboratories, Burlingame, CA), PgR used diluted 1:40, clone 16 (Vector Laboratories) and Ki-67 used the diluted 1:100, or SP6 antibody (Abcam, Cambridge, MA) diluted 1:100. ER positive if H>1; PR scored as % positive cells; HER2 by manufacturer's insturctions; Ki-67 using Ariol image system (Genetix, San Jose, CA). Similar methods and scoring algorithms were used for the Nottingham cohort, except that the MiB1 antibody was used on whole sections for Ki-67, and TMAs were used for ER, PgR, and HER2.	Compatible IHC4, IHC4+C Cuzick *et al*. 2011[24]	[redacted]	[redacted]	100% ET monotherapy
Validation cohorts: LN0/+							
Bartlett 2016[23] Christiansen 2012[151] N=2919[23] N=4598[151]	TEAM (UK/Eire, NL, Belgium, Germany, Greece)	R-RCT	FFPE samples Ariol SL50 image platform Staining as per Bartlett *et al*. 2011[160] Scoring as per Faratian *et al*. 2007[161] Scores normalised.	Compatible IHC4 Cuzick *et al*. 2011[24]	100% HR+ % HER2- NR 100% Postmeno % female NR	LN0/+, % NR	100% ET Some CT, % NR [160]
Cuzick 2011[24] N=786	Nottingham[redacted](UK)	R-RD	As TransATAC[43]	Compatible As TransATAC[43]	100% HR+ 95% HER2- Pre/postmeno	LN0 62% LN+ 38% (% LN>3 NR)	52% ET % CT NR

Reference; N	Cohorts (Country)	Study design	Details of test[a]	Compatibility [b] &Algorithm	Population	Nodal status	Endo / chemo
Nitz 2017 Gluz 2016a Gluz 2016b[108, 109, 111] N=2642 55 months follow-up	WSG Plan B (Germany)	R-RCT	Tissue microarrays (1.4mm diameter): ER (Rabbit [SP1]), PR (mouse monoclonal PgR636) and Ki-67 (clone 30-9 rabbit monoclonal). ER & PR positive if ≥1% stained. Ki-67 scored by one expert, >100 cells, semi- and quantitatively. FISH for HER2 (unclear if confirmatory). Instead of H-score a modified score was used as described in Prat *et al*. 2013[154]	Incompatible IHC4 Prat *et al*. 2013[154] Cuzick *et al*. 2011[24]	100% HR+ 100% HER2- Pre/post meno 100% female High clinical risk [d]	LN0-3 LN0 58.8% LN1-3 41.2%	RS<12 endo only; RS≥12, chemo + endo[e]
Rohan, 2014[87] N=295 (147 cases; 148 controls)[f]	Kaiser Permanente Northwest (USA)	CC, R-RD	FFPE samples ER, PR & HER2 according to ASCO-CAP[162, 163]. HER2 defined as ≥3.	Unclear/Unlikely IHC4 - UC if +C Cuzick *et al*. 2011[24]	100% ER+ 100% HER2- Meno NR 100% female	Any LN, % NR (for ER+/HER2- SG)	Some ET&CT, % NR (for ER+/HER2- SG)
Stephen, 2014[62] a) BCS N=831 b) TEAM N=2513	a) BCS b) TEAM (UK/Eire, NL, Belgium, Germany, Greece)	a) Cohort b) R-RCT	FFPE a) 0.6mm^2 TMA cores. Dual scoring by experts[164] b) as Bartlett 2016[23] Scores normalised. FISH for HER2- (unclear if confirmatory).	Similar IHC4 (personal communication) Cuzick *et al*. 2011[24]	100% ER+ % HER2- NR a) % Meno NR b) 100% Postmeno % female NR	LN0/+, % NR SG's: LN0 LN+	100% ET monotherapy
Viale 2013[88] N=1256	IES[156] (37 countries)	R-RCT	FFPE samples. Biomarker expression was measured by IHC. HER2 was confirmed by FISH if ≥IHC2+. Tumours were deemed positive for ER/PR if IHC ≥1% or Allred ≥3 & for HER2 if IHC 3+ or if FISH amplified. Ki67 was high if > 11% LI (median).	Unclear NR	% ER+ NR % HER2- NR 100% postmeno 100% female	LN NR (Source study recruited any LN status)[156]	100% ET 19% CT
Validation cohorts: LN0							
Vincent-Salomon, 2013[153] N=105	Institut Curie (France)	R-RD	FFPE. For each antibody, internal and external controls were included.	Compatible IHC3 - UC if +C Cuzick *et al*. 2011[24] Used IHC3 algorithm as patients HER2-	100% ER+ 100% HER2- <3cm Luminal BC	LN0 100%	9.5% ET 0% CT

Reference; N	Cohorts (Country)	Study design	Details of test[a]	Compatibility [b] &Algorithm	Population	Nodal status	Endo / chemo
Validation cohorts: LN+							
Prat, 2013[154] N=1,246	GEICAM 9906[94] (Spain)	R-RCT	Sections air-dried overnight. General intensity score instead of H-score for ER expression. [154]}	Compatible IHC4 – UC if +C Cuzick *et al*. 2011[24]	100% ER+ 100% HER2- 45% postmeno	100% LN+ %LN>3N R (37.5% LN>3 for unselecte d cohort, N=1,246)	ET if HER2- 100% CT
Gluz, 2016c[93] N=459	WSG-AGO-Doc[158] (Germany)	R-RCT	Paraffin-embedded tumour blocks, no further details.	Similar, lacks granularity IHC4 – UC if +C Prat *et al*. 2013[154] Cuzick *et al*. 2011[24]	100% HR+ 100% HER2- Menopause NR % Female NR	LN1-3	ET according to clinical guidelines[165] 100% CT
Retrospective studies: Uncertain generalisability to UK context							
Gong 2016[85] N=611	SYSMH; CCSYU; 3rdHNC (China)	R-RD	FFPE Scores normalised. Other details as per Cuzick *et al*. 2011.[160] FISH to confirm HER2 if ≥IHC2+.	Unclear IHC4 Cuzick *et al*. 2011[24]	100% HR+ 100% HER2- 61% postmeno % female NR non-metastatic	LN0 46.6% LN+ 53.4% (% LN>3 NR)	100% ET 76.8% CT
Lin, 2015[152] N=605	National Taiwan University Hospital (Taiwan)	R-RD	FFPE samples. Different IHC methodologies used, used percentiles to account for differences to Cuzick *et al*. 2011[24] FISH to confirm HER2 if ≥IHC2+.	Unclear/unlikely IHC4 Cuzick *et al*. 2011[24]	HR+ NR 76.2% HER2- Meno NR Female NR	Any LN, % NR	ET NR 74.6% CT

CA, California; MA, Massachusetts; R-RCT, retrospective analysis of RCT; R-RD, retrospective analysis of routine data; CC, case control study; FFPE, formalin fixed, paraffin embedded; TEAM, Tamoxifen versus Exemestane Adjuvant Multicentre trial; NL, the Netherlands; BCS, Edinburgh Breast Conservation Series; SG, subgroup; IES, Intergroup Exemestane Study; FISH, Fluorescent in situ hybridisation; GEICAM, Grupo Espanol de Investigacion en Cancer de Mama; SYSMH, Sun Yat-sen Memorial Hospital; CCSYSU, Cancer Centre of Sun Yat-sen University; 3rdHNC, Third Hospital of Nanchang City; HR+, hormone receptor positive; ER+, oestrogen receptor positive; HER2-, human epidermal growth factor receptor negative; RS, recurrence score; LN, lymph node; ET, endocrine therapy; CT, chemotherapy; NR not reported

Reference; N	Cohorts (Country)	Study design	Details of test[a]	Compatibility [b] &Algorithm	Population	Nodal status	Endo / chemo

[a] Full details provided in Appendix 5; [b] compatibility of test methodology to developer's methodology- further details in Appendix 3; [c]Data relating to the TransATAC study is also available in multiple publications, namely Sestak *et al*. 2016,[37] Sestak *et al*. 2013,[39] Sgroi *et al*. 2013,[41] and Dowsett *et al*. 2013,[36] all reporting slightly different analyses. The total analysed cohort N= 1,125 patients.

[redacted]; [d] HER2-negativity; pT1-T4c; LN+ [or LN0 with a risk factor (CpT2, grade 2/3, high uPA/PAI-1, <35 years, or HR-negative)][111] [e] patients were treated according to Oncotype DX score, with those with RS<12 receiving ET only, and those with RS≥12 receiving CT+ET; [f] controls could be matched to more than one case.

Table 66: Quality assessment of prognostic studies: IHC4 and IHC4+C

Reference(s); N	Cohort(s)	Derivation or validation?	Study design appropriate?	All eligible patients included?	Blinding (of test assessors to outcomes)?	Outcome definition standardised or *a priori*?	Applicability: Patient Spectrum	Applicability: Test as per decision problem?
Bartlett 2016[23] Christiansen 2012[151] N=2919[23] N=4598[151]	TEAM	V	N, Some CT	UC	UC	Y	UC (ER2- NR; LN>3 NR)	Y
Cuzick 2011[24] N=786	Nottingham	V	UC, % CT NR	UC	UC	Y	UC, %LN>3 NR, CT NR	Y
Gluz, 2016c[93] N=459	WSG-AGO-Doc	V	N, some CT	N, InsT,	UC	Y	Y	Y
Gong 2016[85] N=611	SYSMH; CCSYU; 3rdHNC	V	N, some CT	N InsT; MD	UC	Y	N, InT, MD, CT,	UC, assay methods unclear
Lin, 2015[152] N=605	National Taiwan University Hospital	V	N, some CT	N, InsT	UC	UC, unclear if DRFS includes deaths	N, InsT, CT, LN>3 NR	UC, assay methods unclear
Nitz 2017[108, 109, 111] N=2642	WSG-Plan B	V	N, some CT	N, MS	y	Y	Y, but high-risk	N, assay methods incompatible
Prat 2013[154]	GEICAM 9906	V	N, all CT	UC	UC	Y	N,	Y
Rohan, 2014[87] N=295 (147 cases; 148 controls)	Kaiser Permanente Northwest	V	N, Case control with some CT	N, InsT, MS, MC	Y	UC, unclear if deaths censored or an event	N, InsT, CT, LN>3 NR	N, some assay methods different
Stephen, 2014[62] a) BCS N=831 b) TEAM N=2513	a) BCS b) TEAM	V	Y, consecutive cohort; reanalysis of RCT	N, MS, InsT, MD	UC	Y	UC, (HER2 NR; LN>3 NR)	Y
TransATAC AIC N=1048	TransATAC	D	Y, reanalysis of RCT	N, InsT, MS	UC	Y	N, InsT, MS	Y
Viale 2013[88]	IES	V	N, some with CT	UC	UC	UC, unclear if deaths censored or an event	N, CT, % LN>3 NR, % HER2- NR	UC, assay methods unclear
Vincent-Salomon, 2013[153] N=105	Institut Curie	V	Y, Cohort	N, InsT, MS	UC	Y	N, InsT, MS	Y

V, validation; N, no, high risk of biase; UC unclear risk of bias; Y, yes, low risk of bias; NR, not reported; MS, missing samples; InsT, insufficient tissue; MS, missing sample; MD, missing data; CT, chemotherapy; MC, no eligible control; BCS, Edinburgh Breast Conservation Series ; SYSMH, Sun Yat-sen Memorial Hospital; CCSYSU, Cancer Centre of Sun Yat-sen University; 3rdHNC, Third Hospital of Nanchang City

Results: IHC4 prognostic performance: Unadjusted analyses

This section reports unadjusted analyses. Adjusted analyses, which show whether the test has prognostic value over clinicopathological variables, are reported in the section "Additional prognostic value"

DRFS: Three studies[85, 87, 152] reported unadjusted analyses for this outcome and results are reported in Table 67. None used methods compatible with the standard IHC4 methodology. Kasier Permanente[87] reported 5-year DRFS for LN0 patients, using tertiles with cut-offs defined as low-risk: ≤-7.81; intermediate-risk: >-7.81 to 88.32; high-risk: >88.32. Not all patients had endocrine therapy and some patients had chemotherapy. An odds ratio analysis of 5-year DRFS for intermediate vs low-risk (1.76 (95% CI 1.10 to 2.84)) and high vs low-risk patients (2.54 (95% CI 0.97 to 6.62)) gave a p value of 0.01. The C-index (AUC) was 0.62 (95% CI NR); values above 0.5 indicate the test is better than chance in placing patients into appropriate risk categories.

The two East Asian studies[85, 152] with uncertain generalisability to the UK context (recruited any lymph node status; variable endocrine and chemotherapy treatments; used methods not compatible with the standard IHC4 methodology) were in general agreement with Kaiser Permanente.[87] They reported statistically significant HRs for high-risk patients (above the 75th percentile) versus low-risk patients (below the 25th percentile) (1.454, (95% CI: 1.133, 1.866, p=0.003) and 2.33 (95% CI: 1.41: 3.85, p NR) respectively). Results for intermediate (between 25th to 75th percentile) vs low were not statistically significant[85] in one study and statistically significant in the other.[152]

DRFI: The Nottingham cohort and the IES study both[24, 88] reported unadjusted analyses for 5 year DRFI, and results are presented in Table 68. Only the Nottingham cohort[24] used the standard IHC4 methodology. Both studies reported statistically significant 5 year DRFI HRs for high versus low-risk groups, defined as quartiles (patients above the 75th quartile high-risk; patients below the 25th quartile low-risk) [24] or tertiles (not defined further)[88] but with different 5-year DRFI HRs (4.1 (95% CI: 2.5, 6.8) versus 2.3 (95% CI: 1.1, 4.7) respectively). This may be due to the different categorisation of patients (quartiles versus tertiles) or differences in patients recruited (LN0/+ versus LN0 respectively), or treatments given (not all patients received endocrine therapy in the Nottingham cohort; some patients received chemotherapy in the IES cohort). A comparison of patients between the second and first tertile to those below the first tertile in the IES study[88] was not statistically significant (5-year DRFI HR 1.4 (95% CI: 0.7 2.9)).

RFS: Both Bartlett *et al.*'s analysis of the TEAM trial[23, 151] and the Taiwanese cohort[152] reported 5-year RFS and results are presented in Table 69. Only the TEAM trial[23, 151] analysis used the standard IHC4 methodology. Both studies recruited LN0/+ patients, and both treated some patients with

chemotherapy. Both reported statistically significant differences for IHC4 risk categories (HR not reported, p<0.001 in TEAM;[23, 151] HR 2.33 (1.41, 3.85) in the Taiwan cohort)[152], except for an analysis of those below the 25th quartile to those between the 25th and 50th quartile in the TEAM[23, 151] trial (p=0.11).

IDFS: see Table 70. The WSG-Plan B[108, 109, 111] trial (LN0/+), where clinically high-risk patients were recruited, and patients with Oncotype DX <12 received endocrine monotherapy and those with RS ≥12 received endocrine and chemotherapy reported a statistically significant 5 year IDFS HR for those above the 75th versus those below the 25th quartile of 2.04 (95% CI: 1.47, 2.83, p<0.001). Similarly, 5 year IDFS results from the LN+ WSG-AGO-Doc trial,[93] where patients all received chemotherapy and the % receiving endocrine therapy was not reported, were statistically significant for the same analysis (HR 2.12 (95% CI: 1.32, 3.42, p 0.002)). Only the WSG-AGO-Doc trial[93] used the standard IHC4 methodology.

IDFI: See Table 71. The lymph node negative Insitut Curie[153] cohort, where some patients received endocrine therapy and none received chemotherapy, reported a non-statistically significant effect for an analysis of IHC3 as a continuous variable (HR 1.01 (95% CI: 1.00, 1.01, p=0.204)). This study was compatible with the standard IHC4 methodology.

Additional prognostic value: IHC4

This section reports adjusted analyses, which indicate the additional prognostic value of IHC4 over clinicopathological factors. The clinicopathological factors adjusted for vary from study to study, and are detailed in the footnotes to the tables.

None of the seven cohorts that reported data relating to the additional prognostic value of IHC4 over other clinicopathological risk scores or versus clinicopathological factors in multivariable analyses recruited HR+, HER2- LN0-3 patients and treated them with 100% endocrine therapy and 0% chemotherapy (Table 72). The closest study to the decision problem was the analysis of TEAM and the Edinburgh cohorts by Stephen *et al*. 2014,[62] though selection of chemotherapy-untreated patients in the Edinburgh cohort and from the TEAM trial may have led to spectrum bias, as patients not treated with chemotherapy in routine practice are likely to be systematically different to those who are treated with chemotherapy. As such, all estimates should be interpreted with caution. Three studies (WSG-Plan B, Kaiser Permanent cohort and the Taiwan cohort)[87, 108, 109, 111, 152] did not use methods compatible with standard IHC4 methodology.

Outcomes included DRFS, DRFI, DFS, IDFS and RFS. Across these outcomes, across the seven cohorts reporting relevant data (Edinburgh cohort, TEAM, WSG Plan B, Kaiser Permanente cohort,

WSG-AGO-Doc, GEICAM 9906, Taiwan cohort),[23, 62, 87, 93, 108, 109, 111, 151, 152, 154] the picture on additional prognostic value was mixed. The analysis conducted by Stephen *et al*.[62] analysed the Edinburgh cohort (median follow-up 12.9 years) and the TEAM cohort (median follow-up 6.2 years) separately, and reported HRs and D-statistics for IHC4 and clinical factors separately, where a difference in D statistics of 0.1 or more indicated improved prognostic separation. HRs (unclear which risk groups compared) were not statistically significant at 0-5 and 5-10 years for DRFI, but the separation in D-statistics between IHC4 and clinicopathological factors were greater at 0-5 year follow-up rather than at full follow-up in both cohorts, and the difference was 0.1 or more in all but the full follow-up analysis of the Edinburgh cohort. The authors interpreted these data as indicating that the additional prognostic value of IHC4 was restricted to the first five years of follow-up. Further to this, multivariable analyses of subgroups of LN0 and LN+ patients showed a statistically significant 0-5 year DRFI HR only for the LN0 subgroup of the Edinburgh cohort (HR 3.16 (95% CI: 1.03, 9.64).

The analysis by Bartlett *et al*.[23] of the TEAM trial (LN0/+, which did not select for endocrine monotherapy and therefore included some patients treated with chemotherapy) also reported a statistically significant HR of 1.006 (95% CI: 1.004, 1.008) when IHC4 was analysed as a continuous variable in a multivariable model including clinicopathological factors, with an increase in likelihood ratio χ^2 over clinicopathological factors of 38.5 (29%). WSG-Plan B,[108, 109, 111] in a mixed cohort of LN0/+, also reported a statistically significant HR of 1.59 (95% CI: 1.15, 2.2), p=0.005) when IHC4 was fractionally ranked by 75th to 25th percentiles in a multivariable model including clinicopathological factors. The Kaiser Permanente[87] LN0/+ cohort reported a statistically significant 5-year DRFS odds ratio of 1.06 (95% CI: 1.00, 1.13) when the score was analysed as a continuous variable in 10 unit increments in a multivariable model including clinicopathological factors, but not when an odds ratio was calculated (1.61 (95% CI: 0.48 5.47) for those above the highest tertile versus those below the lowest tertile). The Taiwanese study also reported a statistically significant HR for those above the 25th percentile versus those below the 25th percentile (1.90 (95% CI: 1.32, 2.73, p<0.001) in a multivariable model including clinicopathological factors.

No studies apart from Stephen *et al*.[62] reported on LN0 patients (see above). Stephen *et al*. [62] reported multivariable DRFI HRs corrected for clinicopathological variables at both 0-5 and 5-10 years in the TEAM and Edinburgh analyses. These were not statistically significant (which was also true for the HRs for the full LN0/+ analysis, where the D-statistic did show an effect), except for 0-5 years in the Edinburgh cohort (HR 3.16 (95% CI: 1.03, 9.64)), but no D-statistics were reported.

WSG-AGO-Doc[93] and GEICAM 9906[154] and the Stephen *et al*.[62] analysis of TEAM and Edinburgh cohorts (see above) reported LN+ cohorts. WSG-AGO-Doc[93] reported a non statistically significant

HR in a multivariable analysis corrected for clinicopathological variables, whilst GEICAM 9906[88] reported a statistically significant increase in likelihood ratio χ^2 over clinicopathological variables (13.5, $p<0.05$). As already stated, the analysis in TEAM and Edinburgh were not statistically significant in multivariable analyses at both 0-5 and 5-10 years for HRs, but no D-statistics were reported..[62]

Broadly speaking, results did not appear to be influenced by the compatibility of the IHC4 methodology with the standard methodology, with both statistically significant and non-significant results being reported in both compatible and non-compatible studies.

Table 67: Prognostic performance of IHC4: DRFS

Reference; N	Cohorts	Population	Nodal status	ET/CT	% pts per group			DRFS: HR (95% CI) unless stated otherwise
					Low	Inter	High	0-5 yr
LN0, some ET&CT								
Rohan, 2014[87] N=295 (147 cases; 148 controls)	Kaiser Permanente Northwest	100% ER+ 100% HER2-	Any LN, % NR (for ER+/HER2- SG)	Some ET&CT, % NR (for ER+/HER2- SG)	40.7[a]	51.9 [a]	7.5 [a]	**Odds Ratio** **Inter vs. Low:** [a] 1.76 (1.10, 2.84) **High vs Low:** [a] 2.54 (0.97, 6.62) p=0.01
					Continuous			**Odds Ratio Per 10 units:** 1.09 (1.03, 1.15) AUC: 0.62
Retrospective studies: Uncertain generalisability to UK context LN0/LN+, some/all ET&CT								
Gong 2016[85] N=611	SYSMH; CCSYU; 3rdHNC	100% HR+ 100% HER2-	LN0 46.6% LN+ 53.4% (% LN>3 NR)	100% ET 76.8% CT	25.7	48.4	25.9	**High vs. low:**[b] 1.454, (1.133, 1.866, p=0.003) **High vs. Inter:** [b] 1.370, (0.931,2.061, p=0.11) **Inter vs. low:** [b] 1.508 (0.941, 2.418, p=0.088) **AUC**: 0.692 (0.617, 0.767)
Lin, 2015 N=605[152]	National Taiwan University Hospital	HR+ NR 76.2% HER2-	Any LN, % NR	ET NR 74.6% CT	Used quartiles			**High vs. low:** [b] 2.33 (1.41, 3.85) **Inter vs. low:** [b] 1.88 (1.18, 2.99)

Pts per grp; patient per group; ER+, oestrogen receptor positive; HER2-, human epidermal growth factor receptor negative; yr, year; Endo, endocrine therapy; chemo, chemotherapy; ET, endocrine therapy; CT, chemotherapy; LN, lymph node; DRFS, distant recurrence free survival; Inter, intermediate-risk group; AUC, area under the curve; SYSMH, Sun Yat-sen Memorial Hospital; CCSYSU, Cancer Centre of Sun Yat-sen University; 3rdHNC, Third Hospital of Nanchang City

[a] Defined via tertiles: Low: ≤-7.81; Intermediate: >-7.81 to 88.32; High: >88.32. DRFS definition unclear regarding whether non-cancer deaths were events or censored; [b] High defined as patient above the 75th percentile; Low defined as patients below the 25th percentile; Intermediate patients from 25th to 75th percentile

Table 68: Prognostic performance of IHC4: DRFI

Reference; N	Cohorts	Population	Nodal status	ET/CT	% pts per group			DRFI: HR (95% CI)
					Low	Inter	High	0-5 yr
LN0/+, some ET&CT								
Cuzick 2011[24] N=786	Nottingham	100% HR+ 95% HER2- Pre/postmeno	LN0 62% LN+ 38% (% LN>3 NR)	52% ET % CT NR	0-25th, 26th-75th, 76th-100			**Below 25th vs above 75th quartile:** 4.1 (2.5, 6.8)
LN0, some/all ET&CT								
Viale 2013[88]	IES	% ER+ NR % HER2- NR	LN0	100% ET 19% CT	Used Tertiles (not further defined)			**2nd T vs. 1st T:** 1.4 (0.7 2.9) **3rd T vs 1st T:** 2.3 (1.1, 4.7) p=0.04

DRFI, distant recurrence free interval; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; yr, year; ET, endocrine therapy; CT, chemotherapy; LN, lymph node; T, tertile; IES, Intergroup Exemestane Study

Table 69: Prognostic performance of IHC4: RFS

Reference; N	Cohorts	Population	Nodal status	ET/CT	% pts per group			RFS: HR (95% CI) unless stated otherwise
					Low	Inter	High	0-5 yr
LN0/+, 100% ET, some CT								
Bartlett 2016[23] Christiansen 2012[151] N=2919[23] N=4598[151]	TEAM	100% HR+ % HER2- NR	LN0/+, % NR	100% ET Some CT, % NR [160]	Used Quartiles			**8 year (n=2919):** continuous: 1.008 (1.006, 1.009, p<0.001) [23] Quartiles: p<0.001[23] Q1 vs Q2: p=0.11[23] **Yr NR (n=4598):** continuous: 1.008 (1.007, 1.010)[151]
Retrospective studies: Uncertain generalisability to UK context LN0/LN+, some ET&CT								
Lin, 2015[152] N=605	National Taiwan University Hospital	HR+ NR 76.2% HER2-	Any LN, % NR	ET NR 74.6% CT	Used Quartiles			**High vs. low:** [a] 2.33 (1.41, 3.85) **Intermediate vs. low:** [a] 1.88 (1.18, 2.99)

Pts per grp; patient per group; HR+, hormone receptor positive; HER2-, human epidermal growth factor receptor negative; NR, not reported; Q1, first quartile (0.-25%); Q2, second quartile (26-50%); RFS, relapse free survival; NR, not reported; LN, lymph node; ET, endocrine therapy; CT, chemotherapy; HR, hazard ratio; CI, confidence; yr, year

[a] High defined as above 75th percentile; low defined as below 25th percentile; intermediate 25th to 75th percentile.

Table 70: Prognostic performance of IHC4: IDFS

Reference; N	Cohorts	Population	Nodal status	ET/CT	Test or comp.	% pts per grp	Other analyses
LN0/+, 100% ET, some CT							
Nitz 2017[108, 109, 111] N=2642	WSG-Plan B	100% HR+ 100% HER2- High clinical risk 100% female	LN0-3 LN0 58.8% LN1-3 41.2%	RS<12 endo only; RS≥12, chemo + endo	**IHC4**	Used quartiles	**0-5 yr: HR 100th-75th to 0-25th percentil**e: 2.04 (95% CI: 1.47, 2.83, p<0.001)
LN+, ET NR, 100% CT							
Gluz, 2016c[93] N=459	WSG-AGO-Doc[158]	100% HR+ 100% HER2-	LN1-3	% ET NR 100% CT	**IHC4**	Used quartiles	**0-5 yr: HR 100th-75th to 0-25th percentil**e: 2.12 (95% CI: 1.32, 3.42, p 0.002)
Pts per grp; patient per group; RS, recurrence score; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; LN, lymph node; ET, endocrine therapy; CT, chemotherapy; HR, hazard ratio; CI, confidence interval; yr, year							

Table 71: Prognostic performance of IHC4: IDFI

Reference; N	Cohorts	Population	Nodal status	ET/CT	Test or comp.	% pts per grp	IDFI: HR (95% CI, p)
LN0, some ET, 0% CT							
Vincent-Salomon, 2013[153] N=105	Institut Curie	100% ER+ 100% HER2- <3cm	LN0 100%	9.5% ET 0% CT	**IHC3**	NR	**HR continuous:** 1.01 (1.00, 1.01, p=0.204)
Pts per grp; patient per group; IDFI, invasive disease free survival; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; LN, lymph node; ET, endocrine therapy; CT, chemotherapy; HR, hazard ratio; CI, confidence interval; NR not reported							

Table 72: Additional prognostic value, all outcomes: IHC4

Reference; N	Cohorts	Population	Nodal status	ET/CT	Outcome	Test or comparator[a]	Likelihood ratio χ^2	Increase in LR χ^2 over CP factors[a]	Other analyses
LN0/+, 100% ET, 0% CT									
Stephen, 2014[62] a) BCS N=831 b) TEAM N=2513	a) BCS	100% ER+ % HER2- NR	LN0/+, % NR	100% ET monotherapy	**DRFI**	**IHC4 vs CP factors**			**MV model (adj. for CP factors[a]): HR (95% CI):[g]** **0-5 years:** 1.79 (0.87, 3.71) **5-10 years:** 1.20 (0.59, 2.44) **Full follow-up [g]** **%R^2 (95% CI): IHC4:** 26.3 (17.4, 35.1); **CP factors:** 25.7 (16.7, 34.6) **D-statistic (95%CI):[b] IHC4:** 1.22 (0.94, 1.50); **CP factors:** 1.20 (0.92, 1.48) **5 years [g]** **%R^2 (95% CI): IHC4:** 39.0 (27.2, 50.7); **CP factors:** 35.3 (23.3, 47.4) **D-statistic (95%CI):[b] IHC4:** 1.63 (1.23, 2.04); **CP factors:** 1.51 (1.12, 1.91) **CP+ IHC4 vs CP:** Wald test: 6.4 (0.01) ; Change R^2 (%): 3.7; Change D stat: 0.12

Reference; N	Cohorts	Population	Nodal status	ET/CT	Outcome	Test or comparator[a]	Likelihood ratio χ^2	Increase in LR χ^2 over CP factors[a]	Other analyses
	b) TEAM		LN0/+, % NR						**MV model (adj. for CP factors[a]): HR (95% CI):** [g] **HR (95%CI)** **0-5 years:** 1.34 (0.85, 2.10) **5-10 years:** 0.89 (0.44, 1.78) **Full follow-up** [g] **%R^2 (95% CI): IHC4:** 32.8 (27.0, 38.4); **CP factors:** 29.5 (23.6, 35.3) **D-statistic (95%CI):**[b] **IHC4:** 1.43 (1.24, 1.62); **CP factors:** 1.33 (1.14, 1.51) **5 years** [g] **%R^2 (95% CI): IHC4:** 34.9 (28.3, 41.2); **CP factors:** 30.5 (23.7, 37.0) **D-statistic (95%CI):**[b] **IHC4:** 1.50 (0.29, 1.71); **CP factors:** 1.36 (1.14, 1.57) **CP+ IHC4 vs CP:** Wald test: 34.5 (<0.001); Change R2 (%): 4.4; Change D stat: 0.14
LN0/+, 100% ET, some CT (or Ct NR)									
Bartlett 2016[23] Christiansen 2012[151] N=4598[151]	TEAM	100% HR+ % HER2- NR	LN0/+, % NR	100% ET Some CT, % NR [160]	**IDFS**	**IHC4 vs CP factors**	170.0[23]	38.5 (29%)[23]	**8 years. MV model (adj. for CP factors[a]): HR (95% CI):**[c] 1.007 (1.005, 1.009)[151]
Nitz 2017[108, 109, 111] N=2642	WSG-Plan B	100% HR+ 100% HER2- High clinical risk 100% female	LN0-3 LN0 58.8% LN1-3 41.2%	100% ET, RS<12 no chemo; RS≥12, chemo	**IDFS**	**IHC4 vs CP factors**			**MV model (adj. for CP factors[a]): HR (95% CI):** 1.59 (95 CI 1.15, 2.2), p=.005)[d]
LN0/+, some ET&CT									

Reference; N	Cohorts	Population	Nodal status	ET/CT	Outcome	Test or comparator[a]	Likelihood ratio χ^2	Increase in LR χ^2 over CP factors[a]	Other analyses
Rohan, 2014[87] N=295 (147 cases; 148 controls)	Kaiser Permanente Northwest	100% ER+ 100% HER2-	Any LN, % NR (for ER+/HER2-SG)	Some ET&CT, % NR (for ER+/HER2-SG)	**DRFS**[e]	**IHC4 vs CP factors**			**Follow-up year NR** **Odds Ratio (95% CI)** **Inter vs. Low:** [e] 1.62 (0.94, 2.81) **High vs Low:** [e] 1.61 (0.48, 5.47) p=0.12 **Continuous per 10 units:** 1.06 (1.00, 1.13)
LN0, 100% ET, 0% CT									
Stephen, 2014[62]	a) BCS	100% ER+ % HER2- NR	LN0	100% ET 0% CT	**DRFI**	**IHC4 vs CP factors**			**MV model (adj. for CP factors[a]): HR (95% CI):** **0-5 years:** 3.16 (1.03, 9.64) **5-10 years:** 2.61 (0.88, 7.75)
a) BCS N=657 b) TEAM N=1,208	b) TEAM		LN0						**MV model (adj. for CP factors[a]): HR (95% CI):** **0-5 years:** 1.29 (0.58, 2.90) **5-10 years:** 0.73 (0.23, 2.31)
LN+, 100% ET, 0% CT									
Stephen, 2014[62]	a) BCS	100% ER+ % HER2- NR	LN+	100% ET monotherapy	**DRFI**	**IHC4 vs CP factors**			**MV model (adj. for CP factors[a]): HR (95% CI):** **0-5 years:** 1.02 (0.33, 3.15) **5-10 years:** 0.53 (0.17, 1.68)
a) BCS N=174 b) TEAM N=1,296	b) TEAM		LN+						**MV model (adj. for CP factors[a]): HR (95% CI):** **0-5 years:** 1.39 (0.81, 2.40) **5-10 years:** 0.98 (0.40, 2.36)
LN+, % ET NR, 100% CT									
Gluz, 2016c[93] N=NR[d]	WSG-AGO-Doc[158]	100% HR+ 100% HER2-	LN1-3	% ET NR 100% CT	**IDFS**	**IHC4 vs CP factors**			**5 year MV model (adj. for CP factors[a]): HR (95% CI):** IHC4 (dichotomous) not significant in multivariable analysis (HR NR)[f]
LN+, 100% ET, 100% CT									
Prat, 2013[154] N=580	GEICAM 9906[94]	100% ER+ 100% HER2- 45% postmeno	100% LN+ %LN>3NR	100% ET 100% CT	**IDFS**	**IHC4 score vs CP**		Follow-up year NR 13.5, p<0.05 (estimated off graph)	

Retrospective studies: Uncertain generalisability to UK context LN0/LN+, variable ET&CT									
Lin, 2015[152] N=605	National Taiwan University Hospital	HR+ NR 76.2% HER2-	Any LN, % NR	ET NR 74.6% CT	**RFS**				**MV model (adj. for CP factors[a]): HR (95% CI): High/Inter vs Low:*** 1.90 (1.32, 2.73, p<0.001)
GGI, genomic grade index, DRFS, distant recurrence free survival; DRFI, distant recurrence free interval; IDFS, invasive disease free survival; RFS, relapse free survival; LN, lymph node; ET, endocrine therapy; CT, chemotherapy; HR, hazard ratio; CI, confidence interval; yr, year; CP, clinicopathological; MV, multivariate; LR, likelihood ratio [a] adjusted for: **Bartlett 2016**, grade, tumour size, age, nodal status, type of endocrine treatment (exemestane vs exemestane+tamoxifen), chemotherapy, radiation therapy; **Stephen 2014,** age, grade, tumour size, nodal status, treatment; **Gluz 2016c**, central grade, genomic grade, Ki-67, Molecular subtype, IHC4; **Nitz 2017,** Nodal status, Tumour stage, local grade, central grade; **Rohan 2014,** nodal status, tumour size, tumour grade, hormone therapy, age at diagnosis, duration of follow-up; **Prat 2012,** treatment arm, histological grade, tumour stage, nodal status, age; [b] difference in D of at least 0.1 indicates improved prognostic separation; [c] C-index reported in Christiansen 2012[151] poster presentation, but text was illegible. [d]personal communication with Professor Gluz, 27th August 2017; [e] High/Intermendiate/Low defined via tertiles: Low: ≤-7.81; Intermediate: >-7.81 to 88.32; High: >88.32. DRFS definition unclear regarding whether non-cancer deaths were events or censored; [f] subgroup with GGI available; [g] High risk >20% risk in orignial (TransATAC) cohort; low risk <10% risk in original (TransATAC) cohort								

Results: IHC4+C prognostic performance: Unadjusted analyses

This section reports unadjusted analyses. Adjusted analyses are reported in the section "Additional prognostic value".

DRFI: Both the Nottingham cohort[24] and the TransATAC[43] derivation cohort re-analysis reported DRFI for IHC4+C, and results are presented in Table 73. The TransATAC analysis used the cut-offs of <10% risk, 10-20% and >20% risk to define low, intermediate and high-risk groups and reported data for LN0-3, LN0 and LN1-3. TransATAC analysis reports [redacted]

The IES study in LN0 patients (100% endocrine therapy, 19% chemotherapy) reported that "*addition of clinical variable to IHC made the effect more profound*" which is ambiguous but could indicate that the addition of the clinical score to the IHC4 score increased the 5 year DRFI HR (those below the 1st tertile versus those above the 3rd tertile), which was 2.3 (95% CI: 1.1, 4.7).

Broadly speaking, results did not appear to be influenced by the compatibility of the IHC4 methodology with the standard methodology.

OS: [redacted]

Additional prognostic value: IHC4+C

This section report adjusted analyses, which indicate the additional prognostic value of IHC4+C over clinicopathological factors. The clinicopathological factors adjusted for vary from study to study, and are detailed in the footnotes to the tables.

The additional prognostic value of IHC4+C was analysed in the TransATAC (derivation) cohort[43] and the Nottingham cohort[24] (Table 75). Both studies used methodologies compatible with the standard IHC4 methodologies. In the TransATAC analysis, additional prognostic value was assessed via increases in likelihood ratio χ^2 for 5-year and 10-year DRFI, for IHC4+C plus NPI or CTS, over NPI or CTS alone (Table 75). Increases in likelihood ratio χ^2 at 5 and 10 years were [redacted] Table 75). Similarly, the Nottingham cohort reported an increase in likelihood ratio χ^2 over the clinical score component of the IHC4 total score of 25.89 ($p<0.0001$) and an HR of 3.9 (95% CI: 2.3, 6.5) in a multivariable analysis adjusted for clinicopathological variables. If the CTS is the same as the clinical component of IHC4+C, then likelihood ratio χ^2 provides the additional prognostic value of IHC4, over CTS.

Discussion: IHC4 & IHC4+C

IHC4: Eleven separate validation cohorts[23, 24, 62, 85, 87, 88, 93, 108, 109, 111, 151-154] have reported prognostic performance data for IHC4, with a total of 13,434 patients. Five cohorts (TEAM, the WSG Plan B, IES, GEICAM 9906 and WSG-AGO-Doc) [23, 62, 88, 93, 108, 109, 111, 151, 154] were re-analyses of RCT data (three LN+/- studies, patients N=8496; no LN0 studies; two LN+ studies, patients N=1705) and six[24, 62, 85, 87, 152, 153] were reanalyses of routinely collected data where patients were treated according to usual practice without use of IHC4 or IHC4+C (five LN+/- studies, patients N=3128; one LN0 study, patients N=105; no LN+ studies). Only one validation cohort treated 100% patients with endocrine monotherapy, whilst the remainder treated varying proportions of patients with endocrine therapy and chemotherapy. Many studies excluded patients on the basis of insufficient tissue sample being available, meaning patients with smaller tumours may have been excluded.

Most analyses used IHC4 as a continuous score, or used quartiles and tertiles as cut-offs, which differs from the cut-points as defined in the NICE scope;[166] as such there was no data relating to how many patients were assigned to each risk group using the pre-defined cut-points except from the derivation cohort (TransATAC). Across the studies reporting prognostic performance data from unadjusted analyses, none reported survival or recurrence outcomes per risk group. HR analyses showed statistically significant performance when high-risk groups (defined by quartiles or tertiles) were compared to low-risk groups, whether in LN0/+, LN0 alone or LN+ alone, and regardless of patient spectrums and treatments received. The use of continuous scores, quartiles and tertiles allows for broad conclusions to be drawn about the potential for IHC4 to be clinically useful in the prognosis of recurrence, and allows for consistency in terms of comparisons between cohorts where assay

methods may have affected absolute values, but it does not allow conclusions to be drawn about which cut offs should be used in clinical practice, and how these would perform. The only validation study which used the same IHC4 cut-offs as in the derivation study[24] was the Stephen *et al*.[62] analysis of TEAM and BCS, which reported multivariable analyses rather than unadjusted analyses (see next paragraph). In addition, very little data relating to the calibration of the test were evident (only Stephen *et al*.[62] reported calibration slopes for the BCS cohort, with a value at 5 years of 1.0 (95% CI 0.8 to 1.1)), and analyses relating to discrimination were generally HRs with 95% CIs rather than more formal tests.

Data from Stephen *et al*. [62] in the separate cohorts (BCS and TEAM)[62] indicated that IHC4 provided more prognostic information than clinicopathological variables in the LN0/+ mixed group, based on D-statistics but not when considering HRs, and was more informative for years 0-5 than 5-10. The same study reported HRs only for LN0 and LN+ subgroups adjusted for CP factors and these were not statistically significant. No other studies reported data for LN0 subgroups, whilst three further studies[93, 152, 154] reported on LN+ subgroups, two of which[152, 154] reported statistically significant additional prognostic value of IHC4 over CP factors.

Interestingly, the methodologies used to conduct IHC4 did not appear to impact on the statistical significance or otherwise of either unadjusted or adjusted analyses. However, without a more thorough consideration of the evidence and the size of the effects, which due to time constraints has not been conducted, it is not possible to conclude how methodologies may impact on the prognostic performance of the test.

IHC4+C: Most information relating to IHC4+C comes from the TransATAC trial, which was the derivation cohort, where [redacted]. Additional data from the Nottingham cohort and IES (though the description is ambiguous) are limited in nature, but support the observations in the TransATAC derivation trial. The TransATAC results suggest IHC4+C is prognostic for DFRI, with HRs for high versus low-risk groups (for different subgroups and timepoints) ranging from [redacted].

IHC4+C appeared to have additional prognostic value over NPI and CTS, but this was based on the derivation cohort (TransATAC), [redacted] In the validation cohort (Nottingham), the HR adjusted for CP variables remained statistically significant (HR 3.9 (95% CI2.3, 6.5)).

Conclusions: IHC4 and IHC4+C

The IHC4 score has been validated in five re-analyses of RCTs and six retrospective cohort studies, and provides statistically significant prognostic information consistently in unadjusted analyses in LN+/-, LN0 and LN+ groups. However, most studies used quartiles or tertiles to define risk groups, and these will have been specific to each cohort. Also, many used laboratory methods that differed from the derivation study methodology. Only one validation study, Stephen *et al*,[62] reports using the cut-offs from the original analysis,[24] which provides a second and third validation cohort (BCS and TEAM), but only for the IHC4 component of the test, not including the clinical factors component (i.e. IHC4+C). IHC4 was shown to have additional prognostic value over clinicopathological factors in some studies. Test methodologies did not appear to impact on the statistical significance of test results, but this does not mean their performance is necessarily the same, and concerns remain about the conduct of the test in laboratories other than that used to derive the score. IHC4+C had prognostic value in unadjusted analyses in the derivation and one validation cohort. Additional prognostic value has been reported in the derivation cohort where IHC4+C provided statistically significantly more information than NPI in LN0 but not LN+ patients, and in one validation cohort (Nottingham) where statistical significance was maintained after adjustments for CP factors.

Table 73: Prognostic performance of IHC4+C: DRFI

Reference; N	Cohorts	Population	Nodal status	ET/CT	% pts per group			% DRFI risk: 0-5 yr			% DRFI risk: 0-10 yr			DRFI: HR (95% CI)	
					Low	Inter	High	Low	Inter	High	Low	Inter	High	0-5 yr	0-10 yr
LN0/+, [redacted]															
TransATAC [redacted] N=1005	TransATAC	[redacted]	[redacted]	100% ET 0% CT	[redacted]										[redacted]
LN0/+, 52% ET, NR CT															
Cuzick 2011[24] N=786	Nottingham	100% HR+ 95% HER2- Pre/postmeno	LN0 62% LN+ 38% (% LN>3 NR)	52% ET % CT NR	Tertiles			Visual inspection of predicted versus observed DRFI plot (Kaplan Meier Curves plus 95% CIs, Figure 5 in Cuzick *et al*. 2011[24]) showed good agreement between predicted and observed scores, though agreement appeared to decrease over time, with lines diverging after 6 years in the high-risk group (67-100%), and 8 years in the low (0-33%) and intermediate-risk (33-66%) groups.							
LN0, 100% ET, 0% CT															
TransATAC [redacted] N=1005	TransATAC	[redacted]	[redacted]	100% ET 0% CT	[redacted]			[redacted]			[redacted]			[redacted]	[redacted]
LN0, 100% ET, some CT															
Viale 2013[88]	IES[156]	% ER+ NR % HER2- NR	LN0	100% ET 19% CT										States "addition of clinical variable to IHC made the effect more profound."	
LN+, 100% ET, 0% CT															
TransATAC [redacted] N=219	TransATAC	[redacted]	[redacted]	100% ET 0% CT	[redacted]			[redacted]			[redacted]			[redacted]	[redacted]

Pts per grp; patient per group; DRFI, distant recurrence free interval; Yr, year; Endo, endocrine therapy; chemo, chemotherapy ET, endocrine therapy; CT, chemotherapy; LN, lymph node; HR, hazard ratio; CI, confidence interval

[a] These analyses used a cut off of <10% risk, 10-20% and >20% risk to define low, intermediate and high-risk groups; [b] this data from the reduced data set

Table 74: Prognostic performance of IHC4+C: OS

Reference; N	Cohorts	Population	Nodal status	ET/CT	% pts per group			% OS risk: 0-5 yr			% OS risk: 0-10 yr			OS: HR (95% CI)	
					Low	Inter	High	Low	Inter	High	Low	Inter	High	0-5 yr	0-10 yr
LN0; LN+ subgroups, 100%ET, 0% CT															
TransATAC [redacted]	TransATAC	[redacted]	[redacted]	100% ET 0% CT	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
			[redacted]		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Pts per grp; patient per group; OS, overall survival; Yr, year; Endo, endocrine therapy; chemo, chemotherapy ET, endocrine therapy; CT, chemotherapy; LN, lymph node; HR, hazard ratio; CI, confidence interval

[a] These analyses used a cut off of <10% risk, 10-20% and >20% risk to define low, intermediate and high-risk groups; [b] this data from the reduced data set

Table 75: Additional prognostic value, all outcomes: IHC4+C

Reference; N	Cohorts	Population	Nodal status	ET/CT	Outcome	Test or comparator[a]	Likelihood ratio χ^2	Increase in LR χ^2 over CP factors[a]	Other analyses
LN0/+, 100% ET, 0% CT									
TransATAC [redacted]	TransATAC	[redacted]	[redacted]	100% ET 0% CT	**DRFI**	[redacted]	[redacted]	[redacted]	
Cuzick 2011[24] N=786	Nottingham	100% HR+ 95% HER2- Pre/postmeno	LN0 62% LN+ 38% (% LN>3 NR)	52% ET % CT NR	**DRFI**	**IHC4+C vs Clinical score**		25.89, p<.0001	**MV model (adj. for CP factors[a]): HR (95% CI):** 3.9 (2.3, 6.5)
LN0, 100% ET, 0% CT									
TransATAC [redacted]	TransATAC	[redacted]	[redacted]	100% ET 0% CT	**DRFI**	[redacted]	[redacted]	[redacted]	[redacted]
LN+, 100% ET, 0% CT									
TransATAC [redacted]	TransATAC	[redacted]	[redacted]	100% ET 0% CT	**DRFI**	[redacted]	[redacted]	[redacted]	[redacted]

DRFI, distant recurrence free interval; LN, lymph node; ET, endocrine therapy; CT, chemotherapy; LR, likelihood ratio; CI, confidence interval; yr, year; MV, multivariate; CP, clinicopathological;

4.8 Results: All tests compared to each other

This section provides an overview of two types of studies that allow some form of comparison between tests:

- 4.8.1 Studies reporting more than one test – these are studies where two or more of the tests were conducted and patient outcomes were reported, such that the prognostic performance of two or more tests in the same cohort can be compared. Very few studies conduct formal comparisons between tests.
- 4.8.2 Microarray studies – these are studies where the commercial version of the tests was not conducted, rather test algorithms were applied to genetic profiles obtained using microarray techniques. Mostly these are publically available *in silico* (electronic database) genetic profiles, complete with patient outcome data. As with the studies that report more than one test, the comparions provided are not always formal.
- 4.8.3 Concordance in risk categorisation between tests – focussing on the OPTIMA Prelim study.

4.8.1 Studies reporting more than one test

Prognostic performance: Studies assessing multiple tests

Few studies assessed multiple tests in the same cohort. This section of the report focuses on how the tests compare to each other in terms of prognostic performance. Evidence is often limited and formal statistical comparisons lacking. Further data relating to studies assessing multiple tests can be found in Section 4.8.2, which focuses on studies which used microarray data (*in silico* data, i.e. held on electronic databases) rather than the commercial versions of the test.

Study designs: Studies assessing multiple tests

Data were reported for six cohorts (Table 76). Four studies were reanalyses of RCTs (TransATAC;[43] ABCSG6+8[57-59] and ABCSG-8 alone;[54, 55] GEICAM 9906;[83, 92] and WSG Plan B[108, 109, 111]). The most comprehensive analysis in terms of the number of tests compared was the translational research analysis of UK-based patients from the ATAC[167] trial (TransATAC), which assessed four tests: EndoPredict, Prosigna, Oncotype DX and IHC4+C. Analyses were reported across ten publications,[24, 34-42] but none reported only ER+, HER2-, LN0-3 patients. The EAG were provided with an analysis[43] from the TransATAC team, largely based on Sestak 2016a,[37] in ER+, HER2-, LN0 patients, which is used as the primary source of data in this review (though the published articles were referred to for data on methods). The TransATAC analysis[43] reported two analysis sets: a full set of patients (N=[redacted] Oncotype DX; N=[redacted] Oncotype RSPC; N=[redacted] IHC4+C; N=[redacted] ROR46; N=[redacted] EPClin), and a reduced set of patients who had received all four of the tests, N=[redacted] (Oncotype RSPC

in LN0 patients only, N=[redacted]). In this section of the report we use the reduced data set, whereas we have used the data from the full analysis set in the sections relating to each of the tests individually (see Sections 4.3 to 4.7). A pooled analysis of 1702 patients from the Austrian Breast and Colorectal Cancer Study Group (ABCSG)-6 and ABCSG-8 trials assessing EndoPredict (EP and EPClin) was reported by Dubsky *et al* (2013a and 2013b) plus subgroup analyses submitted to NICE by Myriad Genetics,[48, 57-59] while an analysis of 1397 patients from ABCSG-8 alone assessed Prosigna (Gnant *et al*., 2014 and Filipits *et al*., 2014).[54, 55] Since these two analyses have a large overlap (majority of patients are from ABCSG-8), they are used here to compare EndoPredict and Prosigna. Finally, 555 patients from the Spanish GEICAM 9906 trial were analysed for EndoPredict (EP and EPClin) and Prosigna by Martin *et al* (2014, 2016).[83, 92] WSG/Plan B[108, 109, 111] was a re-analysis of RCT data from Germany, but has limitations in its use for assessing prognostic performance (discussed below).

The three remaining studies (Russell *et al*. 2016; WSG/Plan B and Gong *et al*. 2016, Table 76)[85, 100, 108, 109, 111] all had limitations. Russell *et al*. 2016 was an observational study of Oncotype DX and MammaPrint where patients were treated according to MammaPrint results, and is therefore confounded as a prognostic study as chemotherapy treatment is likely to have differed across risk groups. However, as there were no other data that compared MammaPrint to other tests, except from microarray studies (see Section 4.8.2), it has been included as the next available level of evidence. Two studies (WSG Plan B and Gong *et al*. 2016)[85, 108, 109, 111] which both have limitations were included because they compared Oncotype DX to IHC4, and the only other data (apart from microarray studies) which compares Oncotype DX to IHC4 is the IHC4 derivation cohort (TransATAC). WSG/Plan B[108, 109, 111] was a reanalysis of RCT data from Germany and was included as a clinical utility study for Oncotype DX (Section 4.3.4) as patients were not treated with chemotherapy where RS<12, but as a prognostic study for IHC4 (Section 4.7.2). Gong *et al*. 2016[85] is an observational study where patients were treated according to usual practice and it was not clear if this included the test result, and the assay used was not the commercial version of Oncotype DX.

As the TransATAC analysis is key to this assessment and compares the most in-scope tests (n=4), to simplify the write up we have structured this section of the report around the TransATAC data and compared other data to these, or used other data to provide comparative data where TransATAC data are lacking. The subheadings are as follows:

- TransATAC[43] comparing Oncotype DX, EPClin, Prosigna and IHC4+C
- EndoPredict and EPClin (n=2 studies, ABCSG-6+8; GEICAM 9906)[57-59, 83, 92]
- EPClin and Prosigna (n=3 studies, TransATAC; GEICAM 9906; ABCSG-6+8 or ABCSG-8)[40, 54, 55, 57-59, 83, 92]

- Oncotype DX and MammaPrint (n=1 study, Russell *et al*. 2016).[100] The limitations of Russell *et al*. (2016)[100] are discussed below.
- Oncotype DX and IHC4 or IHC4+C (n=3 studies, TransATAC; WSG Plan B, Gong *et al*. 2016)[43, 85, 108, 109, 111] The limitations of WSG Plan B[108, 109, 111] and Gong *et al*. 2016[85] are discussed below.

Patients and treatments: Studies assessing multiple tests

Patient characteristics and details of the treatments received are presented in Table 76. Six of the seven data sets either consisted of, or had analyses available for, ER+, HER2- patients,[43, 54, 55, 57-59, 83, 85, 92, 108, 109, 111] whilst Russell *et al*. 2016[41] consisted of all ER+ patients, but did not report the proportion who were HER2-.[100] In terms of nodal status, one study was in LN0 patients only (Gong *et al*. 2016),[85] one study was in LN+ patients only (GEICAM 9906)[83, 92] and one did not report nodal status (Russell *et al*. 2016).[100] Three data sets included node negative and node positive patients (TransATAC, ABCSG-6+8, WSG Plan B, [54, 55, 57-59]). In GEICAM 9906[83, 92] 36% had >3 positive nodes and in ABCSG6+8[57-59] 5% had >3 positive nodes. In WSG Plan B patients were at clinically high-risk defined as LN+ or LN0 with a risk factor (CpT2, grade 2/3, high uPA/PAI-1, <35 years).[108, 109, 111] Patients in all analyses received 5 years of endocrine therapy, apart from Russell *et al* 2016[100] where this was not reported, and Gong *et al*. 2016[85] where 100% received endocrine therapy, but the duration was not reported. Patients in the GEICAM 9906 analysis[83, 92] also received adjuvant chemotherapy, Russell *et al*. 2016[100] did not report how many patients received chemotherapy, WSG Plan B[108, 109, 111] patients with RS≥12 received chemotherapy and 79% of patients in Gong *et al*. 2016[85] received chemotherapy.

Tests and comparators: Studies assessing multiple tests

Details of the tests conducted and the cut-offs applied are presented in Table 76. All data sets which included EndoPredict and Prosigna assessed EndoPredict as marketed, using qRT-PCR and standard cut-offs for risk groups (5 for EP and 3.3 for EPClin). In two analyses (TransATAC and ABCSG-6+8[54, 55]), Prosigna was assessed using the nCounter device and cut-offs of 40 and 60 (LN0) or 15 and 40 (LN1-3), while GEICAM 9906[83, 92] used a "research-based non-standardised" PAM50 ROR-PT assay, using qRT-PCR then microarray rather than nCounter, with cut-offs of 18 and 65 (LN+). Russell *et al*. 2016 did not report how Oncotype DX and MammaPrint were obtained. WSG Plan B[108, 109, 111] ordered Oncotype DX from Genomic Health, and conducted IHC4 tests according to Prat *et al*. 2013[168] and Cuzick *et al*. 2011, [24] and used 25th to 75th percentiles as cut points for Oncotype DX and IHC4. Gong *et al*. 2016[85] conducted Oncotype DX assays using Surexam (Guangzhou, China) and IHC4 according to Cuzick *et al*. 2011, [24] also using 25th to 75th percentiles as cut points.

Comparators in TransATAC[37, 43] included the CTS score and NPI. ABCSG-6+8[57-59] compared AOL to EndoPredict.

Quality assessment: Studies assessing multiple tests

A summary of the quality of the studies is presented in Table 77. Two data sets (TransATAC and ABCSG-8 or 6+8) [39, 43, 54, 55, 57-59] were re-analyses of RCTs where no patients received chemotherapy and all received adjuvant endocrine therapy. Two (GEICAM and WSG Plan B)[83, 92, 108, 109, 111] were re-analyses of RCTs where some patients received chemotherapy, and two[85, 100] were observational studies where patients were either treated according to routine practice but it was not clear if the test results was known,[85] or were treated according to routine practice including a test result (MammaPrint).[100] None of the studies reported including all relevant patients, meaning there is a risk of bias and the generalisability of the cohort to the decision problem is uncertain. Test assessors were blind to patient outcomes in four studies.[43, 54, 55, 83, 92, 108, 109, 111] All used standardised outcomes. Two studies[83, 85, 92] used assays that were not the same as the commercially marketed version of the test (Prosigna not using nCounter in one study;[83, 92] Oncotype DX performed by Surexam (Guangzhou, China) and IHC4 process was not clear in one study.[85]

Results: Studies assessing multiple tests

Table 78 to Table 81 present the data for all patients (node-positive or node-negative) and separate data for node-positive and node-negative patients.

Prognostic Performance

Distribution of patients by risk group, event rates (DRFI/DMFS/DRFS) and HRs (unadjusted analyses)

This section reports unadjusted analyses. Adjusted analyses, which show whether the test has prognostic value over clinicopathological variables, are reported in the section "Additional prognostic value"

TransATAC data: In the TransATAC cohort (Table 78),[40, 43] the proportion of patients categorised as low-risk was similar for [redacted]

[redacted][43]

[redacted][43]

EndoPredict vs EPClin: Results are presented in Table 78. Similar proportions of low-risk patients as seen in the TransATAC cohort were reported for EPClin in ABCSG6+8 (63% LN0-3; [redacted] LN0; [redacted] LN+). When comparing EndoPredict to EPClin in the ABCSG6+8 cohort, EndoPredict placed fewer patients in low-risk category (49% for LN0-3, NR for LN0 and LN+) than EPClin (63%).[57-59] Only DRFS rates for EPClin were reported, which for low-risk patients at 5 years were 95.3%. In contrast, in the LN+ study GEICAM 9906, EndoPredict placed more patients in the low-risk category (25%) than EPClin (13%), but event rates were lower in EPClin low-risk groups (100%) than the EP (93%). HRs for EPClin were higher than for EndoPredict, e.g 0-5 year HR for low vs high 4.82 (EP clin) and 2.80 (EndoPredict).[83, 92]

Prosigna vs EPClin: Results are presented in Table 78. For LN0-3 cohorts, data from ABCSG6+8[57-59] was consistent with TransATAC:[43] Prosigna/ROR-PT placed a [redacted] of patients in the low-risk group than EPClin in two cohorts (TransATAC and ABCSG-6+8/ABCSG-8),[54, 55, 57-59] with [redacted] (TransATAC) and 35% vs 63% (ABCSG trials) respectively. This was also true in LN0 subgroups [redacted] and 48% vs [redacted] (ABCSG) respectively), and LN+ subgroups [redacted] and 4% vs [redacted] (ABCSG) respectively), though in the GEICAM 9906 data set,[83, 92] the direction was [redacted] 10-year DRFS/DRFI rates in LN0 patients were better in Prosigna/ROR-PT [redacted] 96.5%

(ABCSG-8)[54, 55] than EPClin [redacted] and [redacted] respectively), and also in LN+ patients [redacted] and 100%, [redacted] and [redacted] respectively). In GEICAM 9906, event rates were 92% and 100% at 10 years respectively.

Oncotype DX vs MammaPrint: Results are presented in Table 78. Only one study reported data for both tests.[100] MammaPrint assigned a larger proportion of patients (63%) to the low-risk category than Oncotype DX (53%) in the observational study by Russell et al. 2016.[100] Event rates were not reported, and only p-values for log rank tests given, where both tests showed a statistically significant difference in DRFS at the $p<0.05$ level for high versus low-risk group comparisons.

Oncotype DX vs IHC4 and IHC4+C: Results are presented in Table 78. Two studies reported Oncotype DX and IHC4 analyses (WSG Plan B[108, 109, 111] (LN0-3 only) and Gong *et al*. 2012[85] (LN0 only)), and both used quartiles to define boundaries for risk categories, making the comparisons of proportions in risk categories and event rates in risk categories of little relevance to the decision problem. For IHC4 alone, Gong *et al*. 2016 reported C-indexes (AUC; which analyse IHC4 and Oncotype DX as continuous variables) in LN0 patients, which indicate that the two tests have similar prognostic performance (Oncotype DX 0.685 (95% CI: 0.540, 0.830) and IHC4 0.602 (95% CI: 0.436, 0.767).

TransATAC[43] (LN0-3, LN0, LN+, Table 78) reported Oncotype DX and IHC+C (rather than IHC4 alone), and reported [redacted] proportions of patients in low-risk in LN0-3 patients ([redacted] respectively) and LN0 patients ([redacted] respectively) but [redacted] in LN+ patients ([redacted] respectively). In LN0 patients, 10-year DRFI was [redacted] between the two tests in low-risk patients ([redacted] respectively)[43]. HRs for low vs high-risk patients were [redacted] _respectively.[43] In LN+ patients, 10-year DRFI was [redacted] for IHC4+C ([redacted]) than for Oncotype DX ([redacted]).[43]

Impact of menopausal status: Patients were subgrouped according to menopausal status (pre- or post-menopausal), in GEICAM 9906[83, 92]. For EndoPredict, event rates in the low-risk groups were similar in pre and post menopausal patients (93% and 92% respectively), though HRs were somewhat different at 6.7 ($p<0.0001$) and 3.3 ($p=0.069$) respectively. For EPClin, DRFS rates in the low-risk groups were identical (100%). HRs between groups were not reported, but between-group differences were statistically significant.

Overall survival: Data relating to OS are reported in Table 79. Only TransATAC[43] and GEICAM 9906[83, 92] report OS. For 0-10 years in LN0-3 groups,[43] HRs are [redacted] and for low versus high group comparisons range from [redacted] (EPClin) to [redacted] (Prosigna). In LN0

patients, HRs comparing low to high-risk groups range from [redacted] (EP Clin) to [redacted] (Prosigna). In LN+ groups, however, the low to high-risk groups show more variation, ranging from [redacted] (EPClin) to [redacted] (Prosigna) in TransATAC, and 19.38 for EPClin in GEICAM 9906.

Additional prognostic value

This section reports adjusted analyses, which indicate the additional prognostic value of IHC4 over clinicopathological factors. The clinicopathological factors adjusted for vary from study to study, and are detailed in the footnotes to the tables.

Likelihood ratios: In TransATAC,[43] additional prognostic value was assessed via increases in likelihood ratio χ^2 for 10-year DRFI, for each test plus NPI or CTS, over NPI or CTS alone (Table 80). In LN0 patients, increases in likelihood ratio χ^2 over CTS and NPI were [redacted]. However, for LN+ patients, increases in likelihood ratio χ^2 were [redacted].

In ABCSG-8,[54] likelihood ratios also showed a statistically significant increase for Prosigna over the Clinical Linear Predictor (same variables as CTS) in node-negative patients ($p<0.0001$) and node-positive patients ($p=0.0002$).

C-indexes (AUC): In node-positive patients in GEICAM 9906,[83, 92] the C-index was higher for EPClin (0.693) and EP (0.657) than for the research-based ROR-PT (0.644) (Table 81), though the lack of p-values and/or confidence intervals mean it is unclear whether the difference in C-indexes were statistically significant. Adding EPClin to ROR-PT plus clinical variables increased the statistical significance of the test of the C-index (C-indexes not reported; $p<0.001$). Conversely, adding ROR-PT to EPClin plus clinical variables did not increase the statistical significance of the test of the C index ($p=0.567$) (Table 81), though this finding should be interpreted with caution due to the non-standard ROR-PT assay.

In ABCSG-6+8, a C-index for EPClin was only reported for a mixed node-negative and node-positive population (including 5% with >3 positive nodes) and for years 5-10 (no data for years 0-5).[58] In this period, the C-index statistically significantly increased when adding EP to a combination of clinical variables or to AOL (both $p<0.001$; Table 81). In the ABCSG-8 analysis of Prosigna,[54] C-indexes were numerically higher for Prosigna (0.720) than for the Clinical Linear Predictor (0.688), but any statistical significance of the difference was not reported.

Multivariable Cox models: Both ABCSG-6+8[57-59] and GEICAM 9906[83, 92] used multivariable analyses and showed that EP was an independent prognostic parameter for 10-year DMFS/DRFS after adjustment for clinical variables (Table 81), while ABCSG-8[54] showed a similar finding for Prosigna.

Discussion: Studies assessing multiple tests

Few studies reported data from multiple tests and no study reported all comparisons of interest to the decision problem. Of most relevance to the decision problem was the TransATAC analysis,[43] as this includes patients from the UK, analyses four of the five tests, reports ER+, HER2- LN0-3 patients only, and provides change in likelihood ratios which allows comparisons between tests to be made. However, the TransATAC data also has limitations: it is the derivation set for IHC4 and is therefore likely to be subject to some over-fitting and overestimation of prognostic performance; only menopausal patients were recruited; and MammaPrint was not tested. It is also only a single cohort and ideally all comparisons would be available in multiple independent cohorts. Data from other cohorts also have limitations: ABCSG6+8[57-59] only recruited LN0 patients and only evaluated Prosigna for a proportion of patients (ABCSG-8);[54, 55] WSG Plan B recruited only high-risk patients, and patients were treated with chemotherapy according to Oncotype DX score;[108, 109, 111] Russell *et al*. 2016[100] was an observational study and reported only very limited study characteristics and analyses, Gong *et al*. 2016[85] used non-standard test methods for Onctoype-DX and IHC4, and was conducted in population of different ethnicity to the decision problem population; and GEICAM 9906[83, 92] included a high proportion of LN>4 patients (36%) and used a non-standard ROR-PT assay.

As the data comparing the tests to each other is limited so are the conclusions that can be drawn. Broad observations include that generally speaking, the more patients are placed in a low-risk category, the poorer the event-free survival for that group. For example, [redacted] Another broad observation is that the tests generally perform differently in LN+ and LN0 patients. In TransATAC, [redacted]. Data from other cohorts generally supported these broad observations.

In terms of how much additional prognostic information the tests provide over clinicopathological variables or algorithms (e.g. NPI, AOL, CTS), most data came from TransATAC,[43] where increases in

likelihood ratio χ^2 over CTS or NPI were [redacted]. One analysis[83] suggested EPClin could provide additional information over ROR-PT (plus clinicopathological variables), whilst ROR-PT could not provide additional information over EPClin (plus clinicopathological variables), but this was limited by the use of a non-standard version of ROR-PT.[83, 92]

Table 76: Characteristics of prognostic studies: Multiple tests

Reference(s)	Cohort(s)	N pts	Country	Study design	Test	Details of test	Cut-offs	Population	Nodal status	Endo / chemo
Sestak 2017 (data request)[43]	**TransATAC**	[redacted]	UK	R-RCT	EPClin	FFPE qRT-PCR, Sividon	3.3	ER+ HER2- Postmeno 100% female	LN0, [redacted] LN1-3, [redacted]	ET 5yr No chemo
					O-DX RS	FFPE Gen Health	18-30			
					O-DX RSPC	FFPE Gen Health	10yr DR risk <10%, 10-20%, >20%			
					Prosigna	FFPE NanoString nCounter	LN0: 41-60 LN+: 16-40			
					IHC4+C	FFPE Cuzick *et al*. 2011[24]	10-20			
Dubsky 2013a,[57] 2013b,[58] Myriad[59]	**ABCSG-6+8**	[redacted] (LN0-3) 1702 (all)	Austria	R-RCT	EP EPClin	FFPE qRT-PCR	5 3.3	ER+ HER2- Postmeno Stage I-II 100% female	LN0, 68% LN1-3, 27% LN>3, 5%	ET 5yr No CT
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8**	1397			Prosigna	FFPE nCounter	LN0: 40-60 LN1-3: 15-40 LN>3: all high	ER+ HER2- Postmeno 100% female	LN0, 71%[a] LN1-3, 26%[a] LN>3, 3%[a]	
Martin 2016,[83] 2014[92]	**GEICAM 9906**	555	Spain	R-RCT	EP EPClin	FFPE qRT-PCR	5 3.3	ER+ HER2- 46% postmeno Stage II-III 100% female	All LN+ LN1-3, 64% LN>3, 36%	Adj CT (FEC / FEC-P) ET 5yr
					Prosigna	qRT-PCR then microarray	18-65			
Russell 2016[100]	U South Florida; Morton Plan Hospital	135	USA	Obs, RPWT	O-DX	NR	NR	100% ER+ HER2- NR Meno NR Female NR	NR	NR – RPWT
					MMP	NR				
Nitz 2017,[111] Gluz 2016a,[108] Gluz	**WSG Plan B**	2642	Germany	R-RCT	O-DX	NR Genomic Health	25th-75th percentile	HR+ HER2- Pre/post meno	LN0-3 LN0 58.8%	RS<12 ET mono;

Reference(s)	Cohort(s)	N pts	Country	Study design	Test	Details of test	Cut-offs	Population	Nodal status	Endo / chemo
2016b[109]					IHC4	PE, IHC4 Prat *et al*. 2013[168] Cuzick *et al*. 2011[24]	25th-75th percentile	100% female High clinical risk[b]	LN1-3 41.2%	RS≥12, CT+ET[c]
Gong 2016[85]	SYSMH; CCSYU; 3rdHNC	153	China	Obs, RPWOT	O-DX	FFPE Multiplex branched-DNA liquid chip technology; Surexam, Guangzhou, China	NR	100% HR+ 100% HER2- 61% postmeno[d] % female NR non-metastatic	LN0	100% ET 79% CT
					IHC4	IHC4 Cuzick *et al*. 2011[24]	25th-75th percentile			

O-DX, Onoctype-DX; MMP, MammaPrint; EP, EndoPredict; ERclin, EndoPredict with Clinical score; ABCSG, Austrian Breast and Colorectal Cancer Study Group; adj, adjuvant; CT, chemotherapy; ET, endocrine therapy; ER, oestrogen receptor; FEC, 5-Fluorouracil, epirubicin, and cyclophosphamide; FEC-P, FEC + paclitaxel; FFPE, formalin-fixed paraffin-embedded; HER2, human epidermal growth factor receptor 2; N0, node-negative; N+, node-positive; qRT-PCR, quantitative reverse transcription polymerase chain reaction; R-RCT, reanalysis of RCT; Obs, observational trial; RPWT, routine practice with MMP test results; U, University; SYSMH, Sun Yat-sen Memorial Hospital; CCSYSU, Cancer Centre of Sun Yat-sen University; 3rdHNC, Third Hospital of Nanchang City

[a]Nodal status for all 1478 patients; NR for 1397 who were HER2-; [b]HER2-negativity; pT1-T4c; LN+ [or LN0 with a risk factor (CpT2, grade 2/3, high uPA/PAI-1, <35 years, or HR-negative)]; [c] patients were treated according to Oncotype DX score, with those with RS<12 receiving ET only, and those with RS≥12 receiving CT+ET

Table 77: Quality assessment of prognostic studies: Multiple tests

Reference(s)	Cohort(s)	Derivation or validation?	Study design appropriate?	All eligible patients included?	Blinding (of test assessors to outcomes)	Outcome definition standardised *or a priori*?	Applicability: Patient Spectrum	Applicability: Test as per decision problem?
Sestak 2017 (data request)[43]	**TransATAC**	V	Y, R-RCT, no chemo	N InT, FT	Y	Y	Y	Y
Dubsky 2013a,[57] 2013b,[58] Myriad[59]	**ABCSG-6+8**	V	Y, R-RCT, no chemo	UC	UC	Y	Y (for subgroup analysis)	Y
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8**	V	Y, R-RCT, no chemo	N InT, MS, TF	Y	Y	Y (for subgroup analysis)	Y
Martin 2016,[83] 2014[92]	**GEICAM 9906**	V	N, R-RCT, adj chemo	N (reason NR)	Y	Y	N (36% LN>3)	N, Prosigna via qRT-PCR then microarray
Nitz 2017,[111] Gluz 2016a,[108] Gluz 2016b[109]	**WSG-Plan B**	V	N, some CT	N, MS	y	Y	Y, but high-risk	Y
Russell 2016[100]	**University of South Florida; Morton Plan Hospital**	V	N, cohort study, usual practice (some CT)	N InT, SfT	UC	Y	N InT	Y
Gong 2016[85] N=611	**SYSMH; CCSYU; 3rdHNC**	V	N, some CT	N InsT; MD	UC	Y	N, InT, MD, CT,	N – Oncotype DX algorithm, but used Surexam, Guangzhou, China assay.

Y, yes; N, no; UC, unclear; ABCSG, Austrian Breast and Colorectal Cancer Study Group; D, Development; InT, insufficient tissue; MS, missing samples; FT, failed test; N, number of positive nodes; qRT-PCR, quantitative reverse transcription polymerase chain reaction; R-RCT, reanalysis of RCT; TF, test failure; V, validation; SYSMH, Sun Yat-sen Memorial Hospital; CCSYSU, Cancer Centre of Sun Yat-sen University; 3rdHNC, Third Hospital of Nanchang City

Table 78: Prognostic performance of multiple tests: DRFI/DMFS/DRFS[a]

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Endo / chemo	Test	% pts per group			% risk: 0-5 yr			% risk: 0-10 yr			DRFI/DMFS/DRFS[a]: HR (95% CI)
						Low	Int	High	Low	Int	High	Low	Int	High	
Node-negative and node-positive															
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT; UK	ER+ HER2- N=[redacted]	LN0, [redacted] LN1-3, [redacted]	All ET No CT	**EPClin**	[redacted]									[redacted]
					O-DX	[redacted]									[redacted]
					Prosigna	[redacted]									[redacted]
					IHC4+C	[redacted]									[redacted]
Dubsky 2013a,[57] 2013b,[58] Myriad[59]	**ABCSG-6+8** R-RCT; Austria	ER+ HER2- N=1702	LN0, 68% LN1-3, 27% LN>3, 5%	All ET No CT	**EP**	49	-	51				NR	-	NR	**0-5 yr:** 2.80 (1.81, 4.34), p<0.001 **5-10yr:** 3.28 (1.48, 7.24), p=0.002
					EPClin	63	-	37				95.3	-	NR	**0-5 yr:** 4.82 (3.12, 7.44), p<0.001 **0-10 yr:** 5.11 (3.48, 7.51), p<0.001 **5-10yr:** 6.25 (2.72, 14.36), p<0.001
		[redacted]	LN0, [redacted] LN1-3, [redacted]	All ET No CT	**EPClin**	[redacted]						[redacted]			[redacted]
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8** R-RCT; Spain	ER+ HER2- N=1397	LN0, 71%[b] LN1-3, 26%[b] LN>3, 3%[b]	All ET No CT	**Prosigna**	35	32	33				96.6	91.1	79.9	**5-15 yr:** I vs L: 3.74 (NR), p=0.002[c] H vs L: 6.90 (3.08, 15.45), p<0.001[c]
Nitz 2017,[111] Gluz 2016a,[108] Gluz 2016b[109]	**WSG Plan B**	HR+ HER2- N=2642	LN0-3 LN0 58.8% LN1-3 41.2%	RS<12 ET; RS≥12, CT&ET[e]	**O-DX**	17[d]	58[d]	21[d]	93.6[e]	94.3[e]	84.2[e]				**0-5 yr:** 2.33 (1.73, 3.14), p<0.001
					IHC4	NR	NR	NR	NR	NR	NR				**0-5 yr:** 2.04 (1.47, 2.83), p<0.001
Russell 2016[100]	U South Florida; Morton Plan Hospital	ER+, NR HER2- N=135	NR	NR	**O-DX**	53	26	21							**Log Rank 0-5 yr** **I vs L:** p=0.760 **H vs L:** p=0.036
					MMP	63		72							**Log Rank, 0-5 yr** p=0.032
Node-negative															
Sestak 2017 (data	**TransATAC** R-RCT; UK	ER+ HER2- N=[redacted]	LN0	All ET No CT	**EPClin**	[redacted]			[redacted]		[redacted]	[redacted]		[redacted]	[redacted]

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Endo / chemo	Test	% pts per group			% risk: 0-5 yr			% risk: 0-10 yr			DRFI/DMFS/DRFS[a]: HR (95% CI)
						Low	Int	High	Low	Int	High	Low	Int	High	
request)[43] [redacted]					**O-DX**	[redacted]									[redacted]
					O-DX RSPC	[redacted]									[redacted]
					Prosigna	[redacted]									[redacted]
					IHC4+C	[redacted]									[redacted]
Dubsky 2013a,[57] 2013b,[58] Myriad[59]	**ABCSG-6+8** R-RCT; Austria	ER+ HER2- [redacted]	LN0	All ET No CT	**EP**	NR	-	NR				NR	-	NR	-
					EPClin	[redacted]									[redacted]
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8** R-RCT; Austria	ER+ HER2- N=984	LN0	All ET No CT	**Prosigna**	48	32	20				96.5	90.0	84.7	**5-15 yr:** I vs L: 4.03 (NR), p=0.002[c] H vs L: 4.74 (1.89, 11.87), p<0.001[c]
Gong 2016[85]	SYSMH; CCSYU; 3rdHNC	ER+ HER2- N=153	LN0	100% ET 79% CT	**O-DX**	49	26	25							**0-10 yr C-index (AUC):** 0.685 (95% CI: 0.540, 0.830)
					IHC4	29	48	23							**0-10 yr C-index (AUC):** 0.602 (95% CI: 0.436, 0.767)
Node-positive															
Sestak 2017 (data	**TransATAC** R-RCT; UK	ER+ HER2- N=[redacted]	LN1-3	All ET No CT	**EPClin**	[redacted]									[redacted]

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Endo / chemo	Test	% pts per group			% risk: 0-5 yr			% risk: 0-10 yr			DRFI/DMFS/DRFS[a]: HR (95% CI)
						Low	Int	High	Low	Int	High	Low	Int	High	
request)[43] [redacted]					**O-DX**	[redacted]									[redacted]
					Prosigna	[redacted]									[redacted]
					IHC4+C	[redacted]									[redacted]
Dubsky 2013a,[57] 2013b,[58] Myriad[59]	**ABCSG-6+8** R-RCT; Austria	ER+ HER2- [redacted]	LN1-3	All ET No CT	**EP**	NR	-	NR				NR	-	NR	-
					EPClin	[redacted]									[redacted]
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8** R-RCT; Austria	ER+ HER2- N=413	LN1-3, 89%[b] LN>3, 11%[b]	All ET No CT	**Prosigna**	4	34	62				100	93.6	76.2	**5-15 yr:** I vs L: no events; H vs L: no events H vs I: 3.15 (1.20, 8.24), p=0.020[c]
Martin 2016,[83] 2014[92]	**GEICAM 9906** R-RCT; Spain	ER+ HER2- N=536	LN1-3, 64% LN>3, 36%	All ET **All CT**	**EP**	25	-	75				93	-	69	**0-10 yr:** 4.7 (CI NR), p<0.0001
					EPClin	13	-	87				100	-	71	**0-10 yr:** Not estimable, p<0.0001
					ROR-PT (research)	19	56%	26				92	74	66	**0-10 yr:** 4.4 (L vs I) 5.8 (L vs H) (CI NR), p<0.0001

ABCSG, Austrian Breast and Colorectal Cancer Study Group; chemo, chemotherapy; CI, confidence interval; DMFS, distant metastasis-free survival; DRFS, distant recurrence free survival; endo, endocrine therapy; ER, oestrogen receptor; H, high; HR, hazard ratio; HER2, human epidermal growth factor receptor 2; int/I, intermediate; L, low; N0, node-negative; N+, node-positive; OS, overall survival.

[a]DMFS (GEICAM, ABSCG); DRFI (TransATAC); DFS (Nitz 2017); [b]Nodal status for all patients; NR for HER2- subgroup; [c]5-15 yr in ABCSG-8 analysis of Prosigna; [d]For cut offs <12, 12-25, >25; [e] Patients treated according to RS score: RS<12 no CT, RS≥12 CT;

[redacted]

Table 79: Prognostic performance of multiple tests: overall survival

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Endo / chemo	Test	% per risk group			OS 5 yr			OS 10 yr			HR, low vs high (95% CI)
						Low	Int	High	Low	Int	High	Low	Int	High	
Node-negative and node-positive															
Sestak 2017[43] [redacted][a]	**TransATAC** R-RCT; UK	ER+ HER2- N=[redacted]	LN0, [redacted] LN1-3, [redacted]	All ET No CT	**EPClin**	[redacted]									
					O-DX	[redacted]									
					Prosigna	[redacted]									
					IHC4+C	[redacted]									
Node-negative															
Sestak 2017[43] [redacted][a]	**TransATAC** R-RCT; UK	ER+ HER2- N=[redacted]	LN0	All ET No CT	**EPClin**	[redacted]									
					O-DX	[redacted]									
					O-DX RSPC	[redacted]									
					Prosigna	[redacted]									
					IHC4+C	[redacted]									
Node-positive															
Sestak 2017[43] [redacted][a]	**TransATAC** R-RCT; UK	ER+ HER2- N=[redacted]	LN1-3	All ET No CT	**EPClin**	[redacted]									

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Endo / chemo	Test	% per risk group			OS 5 yr			OS 10 yr			HR, low vs high (95% CI)
						Low	Int	High	Low	Int	High	Low	Int	High	
					O-DX	[redacted]									[redacted]
					Prosigna	[redacted]									[redacted]
					IHC4+C	[redacted]									[redacted]
Martin 2016,[83] 2014[92]	**GEICAM 9906** R-RCT; Spain	ER+ HER2- N=536	LN1-3, 64% LN>3, 36%	All ET All CT	**EP**	25		75				92		6	**0-10 yr:** 3.9 (2.0 to 7.5), p<0.0001
					EPClin	13		87				99		69	**0-10 yr:** 19.4 (2.7 to 138.7), p<0.0001

ABCSG, Austrian Breast and Colorectal Cancer Study Group; chemo, chemotherapy; CI, confidence interval; DMFS, distant metastasis-free survival; DRFS, distant recurrence free survival; endo, endocrine therapy; ER, oestrogen receptor; HR, hazard ratio; HER2, human epidermal growth factor receptor 2; int, intermediate; N0, node-negative; N+, node-positive; OS, overall survival; ET, endocrine therapy; CT, chemotherapy; LN, lymph node

[redacted]

Table 80: Additional prognostic value (likelihood ratio χ^2 values): Multiple tests

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Outcome	Test or comparator[a]	Likelihood ratio χ^2	Increase in likelihood ratio χ^2 over CTS/CLP[a]
Node-negative and node-positive							
Sestak 2017 (data request)[43] [redacted]	**TransATAC** R-RCT; UK	ER+ HER2- ET, no CT N=[redacted]	LN0, [redacted] LN1-3, [redacted]	**DRFI 10yr**	**EPClin**	[redacted]	[redacted]
					O-DX	[redacted]	[redacted]
					Prosigna	[redacted]	[redacted]
					IHC4+C	[redacted]	[redacted]
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8** R-RCT; Austria	ER+ HER2- ET, no CT N=1397	LN0, 71% LN1-3, 26% LN>3, 3%	**DRFS 10yr**	**Prosigna**		Over CLP: 29.94 (p<0.0001)
Node-negative							
Sestak 2017 (data request)[43] [redacted][d]	**TransATAC** R-RCT; UK	ER+ HER2- ET, no CT N=[redacted]	LN0	**DRFI 10yr**	**EPClin**	[redacted]	[redacted]
					O-DX	[redacted]	[redacted]
					O-DX RSPC	[redacted]	[redacted]
					Prosigna	[redacted]	[redacted]
					IHC4+C	[redacted]	[redacted]
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8** R-RCT; Austria	ER+ HER2- ET, no CT N=984	LN0	**DRFS 10yr**	**Prosigna**		Over CLP: 20.32 (p<0.0001)
Node-positive							
Sestak 2017 (data request)[43] [redacted][d]	**TransATAC** R-RCT; UK	ER+ HER2- ET, no CT N=[redacted]	LN1-3	**DRFI 10yr**	**EPClin**	[redacted]	[redacted]
					O-DX	[redacted]	[redacted]
					Prosigna	[redacted]	[redacted]
					IHC4+C	[redacted]	[redacted]

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Outcome	Test or comparator[a]	Likelihood ratio χ^2	Increase in likelihood ratio χ^2 over CTS/CLP[a]
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8** R-RCT; Austria	ER+ HER2- ET, no CT N=413	LN1-3, 89%[b] LN>3, 11%[b]	**DRFS 10yr**	**Prosigna**		Over CLP: 17.45 (p=0.0002)

ABCSG, Austrian Breast and Colorectal Cancer Study Group; AOL, Adjuvant! Online; chemo, chemotherapy; CI, confidence interval; CLP, Clinical Linear Predictor; CP, clinical/pathological; CTS, Clinical Treatment Score; DMFS, distant metastasis-free survival; DRFS, distant recurrence free survival; endo, endocrine therapy; ER, oestrogen receptor; HR, hazard ratio; HER2, human epidermal growth factor receptor 2; N0, node-negative; N+, node-positive; NPI, Nottingham Prognostic Index

[a]CP factors **(ABSCG)** = age, grade, nodal status, tumour size, Ki67. CP factors **(GEICAM)** = age, grade, nodal status, tumour size, treatment, ER, PR, Ki67. CTS **(TransATAC)** and CLP **(ABCSG-8)** = age, grade, nodal status, tumour size, treatment; CP factors **(WSG-Plan B)** = Nodal status, Tumour stage, local grade, central grade, Ki-67, ER, PR, IHC4, O-DX RS; [b] Nodal status for all patients; NR for HER2- subgroup; [c] Patients treated according to RS score: RS<12 no CT, RS≥12 CT;

Table 81: Additional prognostic value (C-indexes and multivariable analyses): Multiple tests

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Outcome	Test or comparator[a]	C-index (AUC)	Increase in C-index (AUC) over CP factors[a]	Multivariable Cox model (adjusted for CP factors[a]): HR (95% CI)
Node-negative and node-positive								
Dubsky 2013a,[57] 2013b[58]	**ABCSG-6+8** R-RCT; Austria	ER+ HER2- Endo, no chemo N=1702	LN0, 68% LN+, 32%	**DMFS 0-5yr**	**EP**			1.20 (1.10, 1.31), p<0.001
				DMFS 5-10yr	**EP**			1.28 (1.10, 1.48), p=0.001
					EPClin	0.786		
					EP + AOL	0.765	EP+AOL vs. AOL: p<0.001	
					EP + CP factors	0.716	EP+CP factors vs. CP factors: p<0.001	
					AOL	0.674		
					CP factors	0.644		
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8** R-RCT; Austria	ER+ HER2- Endo, no chemo N=1397	LN0, 71% LN1-3, 26% LN>3, 3%	**DRFS 10yr**	**Prosigna**	0.720	NR	HR (int vs low) 2.15 (1.21, 3.81), p=0.009 HR (high vs low) 4.26 (2.44, 7.43), p<0.0001
					CLP	0.688		
Nitz 2017[108, 109, 111] N=2642	**WSG-Plan B**	100% HR+ 100% HER2- High clinical risk N=2642	LN0, 58.8% LN1-3, 41.2%	RS<12 ET; RS≥12, CT&ET[c]	**O-DX**			HR (25th-75th percentile) 1.73 (1.21, 2.47, P0.001)
					IHC4			HR (25th-75th percentile) NS
Node-negative								
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8** R-RCT; Austria	ER+ HER2- Endo, no chemo N=984	LN0	**DRFS 10yr**	**Prosigna**	0.692	NR	
					CLP	0.639		
Node-positive								
Gnant 2014,[54] Filipits 2014[55]	**ABCSG-8** R-RCT; Austria	ER+ HER2- Endo, no chemo N=413	LN1-3, 89%[b] LN>3, 11%[b]	**DRFS 10yr**	**Prosigna**	0.743	NR	
					CLP	0.667		
Martin 2016,[83] 2014[92]	**GEICAM 9906** R-RCT; Spain	ER+ HER2- Chemo-treated N=536	LN1-3, 64% LN>3, 36%	**DMFS 10yr**	**EPClin**	0.693	Adding EP-clin to ROR-PT + CP factors: p<0.001	
					EP + CP factors*	0.672	EP+CP factors vs. CP factors: p=0.0018	

Reference(s)	Cohort(s) Design; Country	Population	Nodal status	Outcome	Test or comparator[a]	C-index (AUC)	Increase in C-index (AUC) over CP factors[a]	Multivariable Cox model (adjusted for CP factors[a]): HR (95% CI)
					EP	0.657		1.1 (1.0, 1.2), p=0.003
					CP factors*	0.654		
					ROR-PT (research-based)	0.644	Adding ROR-PT to EP-clin + CP factors: p=0.567	

ABCSG, Austrian Breast and Colorectal Cancer Study Group; AOL, Adjuvant! Online; chemo, chemotherapy; CI, confidence interval; CLP, Clinical Linear Predictor; CP, clinical/pathological; CTS, Clinical Treatment Score; DMFS, distant metastasis-free survival; DRFS, distant recurrence free survival; endo, endocrine therapy; ER, oestrogen receptor; HR, hazard ratio; HER2, human epidermal growth factor receptor 2; N0, node-negative; N+, node-positive.

[a]CP factors **(ABSCG)** = age, grade, nodal status, tumour size, Ki67. CP factors **(GEICAM)** = age, grade, nodal status, tumour size, treatment, ER, PR, Ki67. CTS **(TransATAC)** and CLP **(ABCSG-8)** = age, grade, nodal status, tumour size, treatment; CP factors **(WSG-Plan B)** = Nodal status, Tumour stage, local grade, central grade, Ki-67, ER, PR, IHC4, O-DX RS; [b]Nodal status for all patients; NR for HER2- subgroup; [c]Patients treated according to RS score: RS<12 no CT, RS≥12 CT;

4.8.2 Microarray studies

Microarray studies are defined, for the purposes of this review, as any study that applied a test algorithm (e.g. Oncotype DX, MammaPrint) to either *in silico* data (microarray gene expression data held electronically, usually accessed from the National Centre for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO))[169] or to a *de novo* microarray assessment conducted for the purpose of the study. These studies differ from studies that used the commercially offered assays in that the agreement between microarray and commercial assays is unknown, and as such the generalisability of the findings to the decision problem is also unknown.

It should be noted that some of the early MammaPrint studies were conducted using a 25,000 gene microarray platform, until the mini-array specific to the 70 MammaPrint genes was developed (see Section 4.4.1). To minimise heterogeneity between studies, MammaPrint studies conducted after the development of the mini-array that used wider microarray data are included here as "microarray studies" rather than alongside studies using the mini-array (see Section 4.2.2).

Given the limitations of these studies in terms of analytic validity and due to time constraints, we have conducted a rapid review rather than a full systematic review. This section of the report differs from other sections in that:

- No quality assessment of studies has been conducted
- Data were not checked by a second reviewer

It should also be noted that due to time and expertise constraints the EAG were not able to fully consider the following factors:

- The degree to which the same cohorts of patients are included in multiple studies. There is likely to be considerable overlap
- The quality of the methodology used to conduct the microarray analyses
- The cut-points used across the studies
- The proportion of ER+ and HER2- patients in each cohort
- The proportion receiving endocrine or chemotherapy in each cohort
- The ethnic composition of the cohorts used.

Further general limitations of the studies as a whole include:

- A lack of clarity as to the characteristics of the patients
- A lack of clarity as to whether patients were treated with endocrine or chemotherapy
- A lack of clarity as to whether patients were treated according to a protocol or according to routine practice, and whether the exclusion of patients who were treated would therefore lead to spectrum bias.

Some of this information may have been obtainable by reference to the GEO, or to the primary publications relating to each cohort, but due to time constraints these data were not sought.

Whilst acknowledging the considerable limitations of these studies and the review methodology, microarray studies hold some value as they report data on more than one test. This is important as there are very few studies using the commercial versions of the assays that report data for more than one test (see Section 4.8.1). Specifically, there are few studies which report data for MammaPrint compared to any other test, meaning it is difficult to assess the relative merits of this test compared to others.

As such, the review of these studies will focus on the information provided relating to the prognostic performance and additional prognostic value of the tests in comparison to each other, rather than on absolute values provided for individual tests, which may not be generalisable. It is of course entirely possible that such comparison between tests are not generalisable either, but given the lack of data comparing the commercial tests, the information provided has some value to the decision problem.

Microarray studies

A total of eighteen studies[170-187] reported data from microarray analyses (Table 82). Of these, five reported only data for one test (three reported Oncotype DX[172, 178, 179] and two reported MammaPrint[171, 186]); the results of these studies are presented in Appendix 4 but are not considered further. Of the remaining 13 studies, six[170, 173, 175-177, 181] reported 7 cohorts of data from single institutions, five[174, 175, 180, 183, 187] reported pooled *in silico* data from multiple cohorts, three[182, 184, 187] reported data from METABRIC (a UK-Canada dataset), one[185] analysed TRANSBIG data (an international collaboration of 22 countries) and one analysed four previously reported cohorts[184] in addition to METABRIC.

All studies reported data on Oncotype DX and MammaPrint, whilst two[174, 187] also reported data on EndoPredict. For the most part, only HRs for recurrence/survival rates between test risk groups were reported, which give an indication of the test's association with an outcome, but do not allow conclusions to be drawn about the prognostic ability of one test versus another. These data are presented in Table 83, and C-index (AUC) data in Table 84, whilst data that provide direct comparisons of the prognostic performance of one test compared to another are presented in Table 85.

Prognostic performance in microarray studies.

Categorisation

Only four studies[173, 176, 180, 181] reported the number of patients in each risk category, and these only included Oncotype DX and MammaPrint (Table 83). In LN+/- cohorts for Oncotype DX there were

24%, 31% and 37% low-risk, 11%, 16% and 19% intermediate-risk and 44%, 53% and 65% high-risk. In LN0 groups there were 14% and 19% low-risk, 19% and 45% intermediate-risk and 67% and 36% high-risk. For MammaPrint there were 39%, 48% and 51% low-risk, 61%, 52% and 49% high-risk patients in LN+/- patients, and similar proportions in LN0 patients (40% and 48% low-risk and 60% and 52% high-risk).

Hazard ratios

Nine studies[173-176, 180-183, 187] reported HR data (Table 83). Data for Oncotype DX and MammaPrint were reported in four studies[173, 175, 181, 182] with a mix of LN+/- patients, four studies[175, 176, 180, 183] with LN0 only patients and one in LN+ patients.[180] Two studies [174, 187] reported HRs for Oncotype DX, MammaPrint and EndoPredict.

Oncotype DX vs MammaPrint, LN+/-: Six studies[173-175, 181, 182, 187] reported data for both Oncotype DX and MammaPrint in a mixed LN+/- cohort (Table 83, seven cohorts/pooled cohorts analysed, including the two that also report EndoPredict HRs). Across various outcome measures including DRFS, RFS, OS and BCSS, all reported statistically significant HRs between test risk groups for both tests, apart from Vollan *et al*. 2015[182] where the HR for BCSS for MammaPrint was not significant (HR 1.25 (95% CI: 0.95, 1.64, p=0.11)), and[182] Zhao *et al*. 2014 which reported HRs at 5 and 10 years, and the 10 year HRs were not statistically significant.[187] As both Zhao *et al*. 2014 and Vollan *et al*. 2015 used the METABRIC cohort, and Vollan *et al*. 2015 did not report the length of follow-up it is possible the statistically non-significant result was for 10 or more years of follow-up. Oncotype DX had higher HRs in three studies (HR 2.65 vs 1.91; 2.57 vs 1.96; 2.05 vs 1.5 for Oncotype DX vs MammaPrint respectively)[174, 175, 181] whilst MammaPrint HRs were higher in two studies (3.40 vs 2.82; 4.61 vs 2.87 for MammaPrint versus Oncotype DX respectively).[175, 181] Whether the HR was higher in Oncotype DX or MammaPrint did not appear to depend on whether the tests were analysed categorically or as continuous variables.

Oncotype DX vs MammaPrint, LN0: Three studies[176, 180, 183] reported data for Oncotype DX and MammaPrint in LN0 patients (Table 83). Neither test was statistically significant in Jonsdottir *et al*. 2014,[176] (Oncotype DX p=0.522; MammaPrint p=0.287) where DMFS was measured at 14 years. HRs were statistically significant (HR 2.7 (95% CI: NR, p<0.001); 2.5 (95% CI: NR, p<0.001) respectively) in Xu *et al*. 2017[183] where RFS was measured at 15 years and in Prat *et al*. 2012[180] (HR 1.97, p<0.0001 and 1.42, p<0.005 respectively, 95% CIs not reported), where outcomes were censored at 8.5 years. NPI was also measured in Xu *et al*. 2017,[183] with a HR a little higher than MammaPrint and a little lower than Oncotype DX at 2.6 (p<0.001).

Oncotype DX vs MammaPrint, LN+: Only one[180] study reported results in a subgroup of LN+ patients (Table 83). Both Oncotype DX and MammaPrint had statistically significant HRs (4.67 (95% CI: NR, p=0.01)) and 2.12 (95% CI: NR, p=0.03, respectively).

Oncotype DX vs MammaPrint vs EndoPredict, LN+/-: Two studies[174, 187] reported two pooled analyses of 33 cohorts[174]and 6 cohorts,[187] and one analysis using METABRIC data (Table 83).[187] These cohorts are likely to contain some of the same patients. All three tests reported statistically significant HRs for DRFS at time points <10 years, but an analysis of 0-5, 5-10 and 0-10 year HRs in Zhao *et al*. 2014[187] only reported statistically significant HRs in the period 0-5 years, for all three tests. Oncotype DX high vs low HR was the highest in the Finetti *et al*. 2014 analysis (HR 2.05 (95% CI: 1.59, 2.63, p<0.001) compared with an HR for MammaPrint of 1.5 (95% CI: 1.21, 1.85, p=0.0002) and an HR for EndoPredict of 1.88 (95% CI: 1.52, 2.32, p<0.001)), though when all tests were analysed as continuous variables in Zhao *et al*. 2014, the HR was highest for EndoPredict (1.97 (95% CI: 1.66, 2.33, p<0.0001) compared with MammaPrint (1.70 (95% CI: 1.43, 2.03, p<0.0001)) and Oncotype DX (1.79, (95% CI: 1.55, 2.07, p<0.0001).

C-index (AUC) and other comparative data

Data relating to C-indexes and other outcomes are presented in Table 84.

Oncotype DX and MammaPrint, LN+/-: Pairs of C-indexes (AUC) for Oncotype DX and MammaPrint were reported in three studies [177, 181, 184] in LN+/- patients (for 8 cohorts). Outcomes included DRFS, DFS, OS and BCSS. The C-index for ranged from 0.372[184] to 0.84,[170, 181] indicating a wide range of fit. Notably, the worst fit was reported for an analysis in Yang *et al*. 2014[184] of cohort GSE19615, where Oncotype DX had a C-index of 0.435 (p<0.05) and MammaPrint had a C-index of 0.372 (p<0.05), both indicating that the test was worse than chance alone at categorising patients into risk groups. Apart from these data, C-indexes for Oncotype DX ranged from 0.59[177] to 0.73[181] and for MammaPrint from 0.606 to 0.84. Oncotype DX had a higher C-index in four cohorts (METABRIC; GSE6532; GSE22219; GSE19615)[184], whilst MammaPrint had a higher C-index in three (Fundan University; Uppsala cohort; Stockholm cohort). [177, 181] P-values were only reported in one study[184] (four out of five cohorts) and were all statistically significant. 95% CIs were not reported in any analyses, meaning it was not possible to determine if the C-indexes were substantially different to each other.

One further study[170] reported data (see Table 84) which explored the prognostic value of MammaPrint in a group of patients with intermediate Oncotype DX. MammaPrint still had prognostic value in this group, with a statistically significant difference between risk groups (HR not reported, p=0.013) and a

C-index of 0.844, indicating MammaPrint was able to further discriminate between patients with and without OS events.

A further study[180] reported increases in likelihood ratio χ^2 for Oncotype DX over MammaPrint and vice versa (see Table 84). This showed that the likelihood ratio χ^2 increased by 14.4 units ($p<0.001$) when Oncotype was added to MammaPrint, and of 9.2 ($p=0.002$) when MammaPrint was added to Oncotype DX, indicating both tests had added prognostic value over the other, but Oncotype DX added a little more.

Oncotype DX and MammaPrint, LN0: Pairs of C-indexes (AUC) for Oncotype DX and MammaPrint were reported in four studies[180, 181, 183, 184] (for 8 cohorts, two of which were pooled analyses). C-indexes for Oncotype DX ranged from 0.608 to 0.71 and for MammaPrint from 0.604 to 0.81. P-values were only reported in one study[184] (5 cohorts) and were not always statistically significant, possibly due to smaller sample sizes in these subgroup analyses compared to the full LN+/- cohorts. Oncotype DX had a higher C-index in five cohorts (Prat *et al*. 2014 and four of the cohorts reported in Yang *et al*. 2014),[180, 184] and MammaPrint had a higher C-index in three (Tobin *et al*. 2014; Xu 2017; GSE19615 from Yang *et al*. 2014).[181, 183, 184]

Oncotype DX and MammaPrint, LN+: One study[180] reported the C-index for LN+ patients. This was 0.64 for Oncotype DX and 0.61 for MammaPrint.

Additional prognostic value in microarray studies

Oncotype DX, MammaPrint and EndoPredict in LN+/-: One study[173] reported a multivariable analysis including Oncotpye-DX and MammaPrint separately alongside ER status, tumour grade, nodal status, age, tumour size and treatment (endocrine therapy, chemotherapy or both) in patients with mixed nodal status (Table 85). The cohort used was the derivation cohort for MammaPrint (and there may therefore be some overfitting of the model, resulting in overestimation of the prognostic performance for MammaPrint) and a subgroup of ER+ only patients. Tests were analysed as categorical rather than continuous variables. All high vs. low HRs were statistically significant though the intermediate vs. low analyses (Oncotype DX only) were not. High vs. low HRs were higher for Oncotype DX than for MammaPrint, though this is perhaps to be expected as Oncotype DX high vs. low comparisons do not account for the intermediate patients while MammaPrint has only two categories and the analyses are therefore not comparable.

One study reported a multivariable analysis in Oncotype DX intermediate patients (Table 85), and MammaPrint was shown to have additional prognostic value in this subgroup of patients (adjusted for

tumour size, nodal status, PR and chemotherapy treatment) with an HR of 10.19 (95% CI: 1.05, 99.01, p=0.045).[170]

One study[187] reported likelihood ratio χ^2 and differences in likelihood ratio χ^2 for Oncotype DX, MammaPrint and EndoPredict (Table 85). EndoPredict had the highest Likelihood ratio χ^2 at 53.6 ($p<0.0001$) compared to 43.6 (Oncotype DX) and 36.0 (MammaPrint), both $p<0.0001$. In an analysis which adjusted for nodal status, grade and tumour size the difference in likelihood ratio χ^2 over these clinicopathological variables was also highest for EndoPredict (31.4 verusus 23.1 and 21.5 respectively, all $p<0.0001$), indicating that all these tests have prognostic value over these clinical factors, and EndoPredict appears to perform best.

Oncotype DX and MammaPrint versus NPI and Adjuvant! Online in LN0 patients: One study reported data LN0 patients (Table 85). The increase in likelihood ratio χ^2 over clinicopathological variables was reported for Oncotype DX, MammaPrint, NPI and AOL. For DMFS, Oncotype DX had the highest increase at 13.734 (p=0.004) compared to MammaPrint (3.038, p=0.986), AOL (3.325, p=0.601) and NPI (6.823, p=0.131) and was the only test to report a statistically significant change. Results were similar for OS.

Discussion: Microarray studies

Data from microarray studies have been included in this report to provide additional information relating to the comparative prognostic value of the tests, as comparative data from studies using the commercial versions of the tests are limited in number (see Section 4.8.1). In particular, comparisons between MammaPrint and other tests (specifically Oncotype DX and EndoPredict) were made in microarray studies but rarely in the studies using the commercial tests. However, these data should be interpreted with caution because of the unknown comparability of microarray studies and the commercial versions.

Data relating to HRs for outcomes between test risk groups support the data from studies using the commercial assays that show a statistically significant difference between test risk categories for outcomes such as DRFS, DFS, OS and BCSS for Oncotype DX, MammaPrint and EndoPredict (no microarray studies were identified assessing Prosigna or IHC4). One study did not report statistically significant HRs at ≥10 years. However, conversely, three studies reported statistically significant HRs at ≥10 years,[173, 181, 183] suggesting that the assumption of proportional hazards may not hold in all cohorts, and the tests are likely to be more often accurate at 0-5 years than at time points beyond. HRs were generally statistically significant in LN+/- cohorts, LN0 cohorts and in LN+ cohorts, though the evidence base for the latter two was limited and one study did not report a statistically significant HR in a LN0 cohort, which may have been due to small sample size (n=94) or follow-up duration (14

years).[176] No study reported HRs in LN+/-, LN0 and LN+ patients separately, so it is difficult to draw any conclusions about whether HRs differ according to LN status.

C-indexes (AUC) were generally good for all tests, and did not appear to differ according to LN status. Conclusions that can be drawn from the data reporting C-indexes were limited by the non-reporting of 95% CIs, meaning it was not possible to tell whether the tests were substantially better or worse than each other. One further problem with determining the superiority of tests was that Oncotype DX has three risk categories (high, intermediate and low) whilst MammaPrint and EndoPredict have only two (high and low); C-index analyses represent the prognostic potential of the test, but do not indicate which cut-offs should be used, what clinical decisions should be made for intermediate-risk patients, or what the long-term clinical outcomes would be for patients treated according to the test as commercially marketed. One study showed that MammaPrint could further categorise Oncotype DX intermediate-risk patients into high and low-risk patients, with an excellent C-index of 0.844. However, without seeing the overall performance of MammaPrint in this cohort, it is not possible to conclude that MammaPrint outperforms Oncotype DX. As such, it is difficult to draw any conclusions about superiority given these differences in categories and the clinical significance in terms of treatment options.

As in previous sections of this report, it can be argued that the true value of the test lies in how much additional prognostic information is provided over and above clincial factors. The one study[187] to report such data across three tests (Oncotype DX, MammaPrint, EndoPredict), reported likelihood ratio χ^2 and change in likelihood ratio χ^2 in analyses adjusted for clinicopathological variables suggesting that EndoPredict had the greatest additional value, followed by Oncotype DX, then MammaPrint, [redacted]. One study[185] (not adjusted for clinicopathologial variables) in LN0 patients reported that only Oncotype DX had a statistically significant change in likelihood ratio χ^2 whereas AOL, NPI and MammPrint did not, which supports the order of prognostic performance reported in TransATAC.

Conclusions: Microarray studies

Microarray studies support conclusions from studies using the commercial versions of the assays in suggesting that Oncotype DX, MammaPrint and EndoPredict can discriminate between high and low-risk patients regardless of LN status (data limited to mixed LN+/- patients for EndoPredict); the utility of the intermediate-risk group in Oncotype DX is uncertain; the additional prognostic performance of the tests over clinicopathological variable is less certain for MammaPrint, though the order of superiority [redacted] namely EndoPredict, then Oncotype DX, then MammaPrint, though the evidence base is limited.

Table 82: Characteristics of Microarray studies

Author, year, Number patients	Cohorts	Country	O-DX	EP	MMP	PRO	Other tests	Population	Nodal Status	Endo/Chemo
O-DX vs MMP vs EP										
Finetti 2014[174] N=1,229	33 publicly available gene expression datasets from NCBI GEO database	NR	O-DX	EP	MMP			ER+, HER2- NR (N=1,299) 95% ER+, 92% HER2- (N=3,074) All luminal (A or B)	NR (N=1,299) 58% LN0 (% LN>3 NR) (N=3,074)	NR
Zhao 2014[187] a) N=912 a-i) N=692 b) N=996	a) GSE6532, GSE3494, GSE1456, GSE7390, GSE2603, E-TABM-158 b) METABRIC cohort		O-DX	EP	MMP	Excluded[a]		a)ER+ 76% HER2- 85% **SG** a-i) ER+ 100%, HER2- NR b) ER+ NR, HER2- NR	a) LN0 67% (LN>3 NR) a-i) NR B) NR	NR
O-DX vs MMP studies										
Ahn 2013[170] a)N=186 b)N=82	Gananam Severance Hospital (1997-2007)	Korea	O-DX		MMP			100% ER+ 12% HER2+ a) all patients b) subset with RS 19-30	a)47.8% LN+ (% LN>3 NR) b)43.9% LN+ (LN>3 NR)	a)84% ET 13% CT b) 94% ET 82% CT
Fan 2006[173] Microarray a) N=295 b) SG N=225	NKI (Derivation cohort for MMP)	Neths	O-DX		MMP			a) 77% ER+ HER2 NR Age ≤52 100% female b)100% ER+	a) LN0, 51% LN1-3, 36% LN>3, 13% b) NR	a) 14% ET 37% CT b) NR
Jonsdottir, 2014[176] N=94	NR	Norway	O-DX		MMP			a) ER+ NR 85% HER2- a-i) 100% ER+, HER2- NR	LN0 100% (% LN>3 NR)	a) 14% ET 11% CT a-i) NR

Author, year, Number patients	Cohorts	Country	O-DX	EP	MMP	PRO	Other tests	Population	Nodal Status	Endo/Chemo
Li 2009[177] N=27	Fudan University Cancer Hospital	China	O-DX		MMP			HR+ NR 70% HER2-	LN0 56% (% LN>3 NR)	ET NR 100% CT
Gyorffy 2015[175] a) N=3,534 b) N=325	a) 25 data sets from GEO[b] b) University Hospitals (Frankfurt & Hamburg)	a) NR b) Germany	O-DX		MMP			a) 83.1% ER+ 84.4% HER2+ NR **SG:** 100% ER+, HER2- b) 81.1% ER+, HER2- NR **SG:** i) 100% ER+; HER2- NR	a) LN+ 30.8% b)LN+ 39.4% (LN>3 NR) **SG:** ER+, LN0	a) ET NR 19% CT **SGs:** i) ER+, HER2-, untreated; ii) ER+, HER2- treated b) ET& CT NR **SG:** i) NR
Prat, 2012[180] N=594/1380 a) N=339 b) N=171	GSE17705, GSE6532, GSE12093, GSE1456, MDACC133	NR	O-DX		MMP	Excluded*		ER+, HER2- NR (N=549) 100% ER+ HER2- NR (n=1380)	NR (N=549) LN0 47% (% LN>3 NR) (N=1380) a) LN0 100% b) LN+ 100%	ET 100% CT 0%
Tobin, 2014[181] a)N=253 b) N=159	a) Uppsala cohort b) Stockholm cohort (Karolinska Hospital)	Sweden	O-DX		MMP			HR+ NR HER2- NR **SG:** a-i) ER+ 100%	a) LN0 63% b) LN0 59%	a) ET 58%; CT 11% b)ET 72%; CT 19%
Vollan, 2015[182] N=1412	METABRIC	International	O-DX		MMP			ER+ 100% HER2- NR	NR	NR

Author, year, Number patients	Cohorts	Country	O-DX	EP	MMP	PRO	Other tests	Population	Nodal Status	Endo/Chemo
Xu 2017[183] a) N=917	a) METABRIC / Bioconductor datasets: GSE11121, GSE7390, GSE3494, GSE2990, Breast Cancer NKI	Internati onal	O-DX		MMP	Exclu ded[a]	NPI	ER+ 100% HER2- NR	LN0 100%	NR
Yang 2014[184] N, (i)LN0 subgroup; ii) ER+ subgroup) a) N=1981 (1037;1526) b) N=216 (125; 134) c) N=393 (250; 348) d) N=115 (64; 66) e) N=236 (158; 201)	a) METABRIC b) Loi (GSE6532) c) Buffa (GSE22219 d) Wang (GSE19615) e) Miller (GSE3494)	Internati onal; NR	O-DX		MMP	Exclu ded[a]		**ER+** a) 77% b) 62% c) 89% d) 57% e) 85% **SG** i) ER+ NR ii) ER+ 100%. HER2- NR	**LN0** a) 52% b) 58% c) 64% d) 56% e) 67% **SG** i) LN0 100%, ii)LN0 NR	NR
Yin, 2014[185] N=198	TRANSBIG GSE7390	France, Sweden, UK	O-DX		MMP		AOL NPI	ER+ NR HER2- NR	LN0 100%	ET 0% CT 0%s

O-DX, Oncotype DX; EP, EndoPredict; MMP, MammaPrint; PRO, Prosigna, NR, not reported; ER+, Oestrogen receptor positive; HER2, human epidermal growth factor receptor 2; LN, lymph node; NKI, Netherlands Cancer Institute; Neths, Netherlands; ET, endocrine therapy; CT, chemotherapy; SG, subgroup; NR, not reported; AOL, Adjuvant! Online; NPI, Nottigham Prognostic Index

[a] **Cockburn 2016:** Data was reported in this study for a simulation of Prosigna. However, only 45 of the 50 Prosigna genes were available for analysis and the data is excluded as it does not conform to algorithm used in the commercially offered test; **Prat 2012:** ROR-P, not ROR-PT; **Xu 2017:** ROR-S not ROR-PT; Yang 2014 used ROR-T and ROR-S not ROR-PT; **Zhao 2014:** ROR-S only, not ROR-PT; [b]GSE1456, GSE4922, GSE5327, GSE6532, GSE7390, GSE9195, GSE11121, GSE12093, GSE12276, GSE2034, GSE16391, GSE16446, GSE17705, GSE17907, GSE19615, GSE2603, GSE20685, GSE20711, GSE21653, GSE25066, GSE2990, GSE31519 and GSE3494;

Table 83: Microarray results: Hazard ratios

Author, year, Number patients	Cohorts	Population	Nodal status	ET/CT	% pts per group			Outcome	Test	Outcomes HR (95% CI) unless stated otherwise		
					Low	Inter	High			0-5 yr	0-10 yr	5-10yr
O-DX & MMP												
LN0/+												
Fan 2006[173] a) N=295	NKI (Derivation cohort for MMP)	a) 77% ER+ HER2 NR	a) LN0, 51% LN1-3, 36% LN>3, 13%	a) 14% ET 37% CT	24	11	65	**RFS**	**O-DX**	-	NR, p<0.001	-
					39	-	61	**RFS**	**MMP**	-	NR, p<0.001	-
					24	11	65	**OS**	**O-DX**	-	NR, p<0.001	-
					39	-	61	**OS**	**MMP**	-	NR, p<0.001	-
Gyorffy 2015[175] a) N=3,534	a) 25 data sets from GEO	a) 83.1% ER+ 84.4% HER2+ NR	a) LN+ /-, LN+ 30.8%	a) ET NR 19% CT	-	-	-	**RFS**	**O-DX**	2.55 (2.21, 2.94, p<0.001)		
					-	-	-		**MMP**	3.40 (2.47, 4.68, p<0.001)		
Gyorffy 2015[175] a-i) N=672		**SGs a-i&ii):** 100% ER+, HER2-		a-i) untreated	-	-	-		**O-DX**	2.82 (2.04, 3.90, p<0.001)		
					-	-	-		**MMP**	3.07 (1.87, 5.04, p<0.001)		
Gyorffy 2015[175] a-ii) N=1,316				a-ii) treated	-	-	-		**O-DX**	2.47 (2.14, 3.49, p<0.001)		
					-	-	-		**MMP**	3.01 (1.85, 4.90, p<0.001		
Gyorffy 2015[175] b) N=325	b) University Hospitals (Frankfurt & Hamburg)	b) ER+, HER2- NR	b)LN+/- (LN>3 NR)	b) ET& CT NR	-	-	-		**O-DX**	2.65 (1.73, 4.07, p<0.001)		
					-	-	-		**MMP**	1.91 (1.05, 3.50, p=0.0322)		
Tobin, 2014[181] a)N=253	a) Uppsala cohort	HR+ NR HER2- NR	a) LN0 63%	a) ET 58%; CT 11%	37	19	44	**BCSS**	**O-DX**	**21 year follow-up:** HR continuous NR, p=0.004 **Inter vs Low:** HR NR, p=0.018 **High vs Low:** HR NR p=0.001 **High/inter vs Low:** 2.57 (1.43, 4.62)		
					51	-	49		**MMP**	**21 year follow-up:** HR continuous NR, p=0.005 **High vs low:** 1.96 (1.21, 3.17)		

Author, year, Number patients	Cohorts	Population	Nodal status	ET/CT	% pts per group			Outcome	Test	Outcomes HR (95% CI) unless stated otherwise		
					Low	Inter	High			0-5 yr	0-10 yr	5-10yr
Tobin, 2014[181] b) N=159	b) Stockholm cohort (Karolinska Hospital)	HR+ NR HER2- NR	b) LN0 59%	b) ET 72%; CT 19%	31	16	53		**O-DX**	**Follow-up NR:** HR NR p=0.006 **High/Inter vs Low:** 2.87 (1.43, 5.75)		
					48	-	52		**MMP**	**Follow-up NR:** HR NR p<0.001 **High vs Low:** 4.61 (2.12, 10.03)		
Vollan, 2015[182] N=1412	METABRIC	ER+ 100% HER2- NR	NR	NR	-	-	-	**BCSS**	**O-DX**	**Follow-up NR** **Inter vs Low:** 1.23 (0.91, 1.68, p=0.179) **High vs Low:** 2.35 (1.64 3.36, p<0.001)		
					-	-	-		**MMP**	**Follow-up NR** **High vs Low:** 1.25 (0.95, 1.64, p=0.11)		
LN0												
Jonsdottir, 2014[176] N=94	NR - Norway	a) ER+ NR 85% HER2-	LN0 100%	a) 14% ET 11% CT	19	45	36	**DRFS**	**O-DX**	**14 year HR** Inter vs low: 1.2 (0.3, 4.4) High vs low: 1.8 (0.5, 6.5) p=0.522 **Rates:** low: 83%; Inter: 79%; High: 68%		
					48	-	52		**MMP**	**14 year HR** 1.6 (0.7, 3.6, p=0.287) **Rates:** Low: 80%; High: 71%.		
Gyorffy 2015[175] b-i) N=113	b) University Hospitals (Frankfurt & Hamburg)	SG b-i): 100% ER+; HER2- NR	SG b-i): ER+, LN0	NR	-	-	-	**RFS**	**O-DX**	O-DX Sens 0.80 (0.76, 0.82) Spec 0.55 (0.53, 0.58) Accuracy: 0.64 (0.62, 0.65) MMP Sens 0.98 (0.96, 0.98) Spec 0.14 (0.12, 0.16) Accuracy: 0.47 (0.46, 0.47)		

Author, year, Number patients	Cohorts	Population	Nodal status	ET/CT	% pts per group			Outcome	Test	Outcomes HR (95% CI) unless stated otherwise		
					Low	Inter	High			0-5 yr	0-10 yr	5-10yr
Prat, 2012[180] a) N=339	GSE17705, GSE6532, GSE12093, GSE1456, MDACC133	100% ER+ HER2- NR	a) LN0 100%	ET 100% CT 0%	14	19	67	**DRFS**	**O-DX**	**Rates:** Low:98% Inter:95% High:86% p=0.004	**DRFS censored at 8.5 years** **Continuous:**1.97, p<0.0001 **High vs Low:**3.79, p<0.0023	
					40	-	60		**MMP**	**Rates:** Low:95% High:84% p=0.004	**DRFS censored at 8.5 years** **Continuous:**1.42, p<0.005 **High vs Low:** 2.6, p=0.0054	
Xu 2017[183] a) N=917	METABRIC / Bioconductor datasets: GSE11121, GSE7390, GSE3494, GSE2990, breastCancer NKI	ER+ 100% HER2- NR	LN0 100%	NR	-	-	-	**RFS**	**O-DX**	**15 years** 2.7 (95% CI NR, p<0.001)		
					-	-	-		**MMP**	**15 years** 2.5 (95% CI NR, p<0.001)		
					-	-	-		**NPI**	**15 years** 2.6 (95% CI NR, p<0.001)		
LN+												
Prat, 2012[180] b) N=171	GSE17705, GSE6532, GSE12093, GSE1456, MDACC133	100% ER+ HER2- NR	b) LN+ 100% (% LN>3 NR)	ET 100% CT 0%	8	12	80	**DRFS**	**O-DX**	**Rates:** Low:91% Inter:95% High:72% p=0.015	**DRFS censored at 8.5 years** **Continuous:** 1.51, p=0.01 **High vs Low:** 4.67, p=0.01	
					31	-	69		**MMP**	**Rates:** Low:85% High:72% p=0.03	**DRFS censored at 8.5 years** **Continuous:** 1.26, p=0.06 **High vs Low:** 2.12, p=0.03	
O-DX & MMP & EP												

Author, year, Number patients	Cohorts	Population	Nodal status	ET/CT	% pts per group			Outcome	Test	Outcomes HR (95% CI) unless stated otherwise		
					Low	Inter	High			0-5 yr	0-10 yr	5-10yr
Finetti 2014[174] N=1,229	33 publicly available gene expression datasets from NCBI GEO database	NR	LN+/-	NR	-	-	-	**DRFS**	**O-DX**	**Median follow-up 7.8 years** **Inter vs Low:** 1.82 (1.44, 2.3, p<0.001) **High vs Low:** 2.05 (1.59, 2.63, p<0.001)		
					-	-	-		**MMP**	**Median follow-up 7.8 years** 1.5 (1.21, 1.85, p=0.0002)		
					-	-	-		**EP**	**Median follow-up 7.8 years** 1.88 (1.52, 2.32, p<0.001)		
Zhao 2014[187] a-i) N=692	a) GSE6532, GSE3494, GSE1456, GSE7390, GSE2603, E-TABM-158	a-i) ER+ 100%, HER2- NR	a) LN0 67% (LN>3 NR) a-i) NR		-	-	-		**O-DX**	1.79 (1.55, 2.07, p<0.0001)	0.65 (0.26, 1.61, p=0.3535)	1.06 (0.78, 1.43, p=0.7311)
					-	-	-		**MMP**	1.70 (1.43, 2.03, p<0.0001)	1.06 (0.57, 1.96, p=0.8468)	1.16 (0.87, 1.55, p=0.3054)
					-	-	-		**EP**	1.97 (1.66, 2.33, p<0.0001)	1.02 (0.55, 1.91, p=0.9462)	1.13 (0.83, 1.53, p=0.4393)
Zhao 2014[187] b) N=996	b) METABRIC cohort	b) ER+ NR, HER2- NR	b) NR		-	-	-		**O-DX**	1.94 (1.69, 2.24, p<0.0001)	1.19 (0.86 1.65, p=0.2856)	1.11 (0.89, 1.38, p=0.3481)
					-	-	-		**MMP**	1.99 (1.63, 2.41, p<0.0001)	1.21 (0.87, 1.68, p=0.2545)	1.11 (0.89, 1.38, p=0.3514)
					-	-	-		**EP**	1.96 (1.64, 2.33, p<0.0001)	1.13 (0.82, 1.55, p=0.4593)	1.29 (1.04, 1.59, p=0.0183

N, number of patient; CT, chemotherapy; ET, endocrine therapy; yr, year; Inter, intermediate; HR Hazard ratio; CI, confidence interval; HR+, hormone receptor positive; HER2, human epiderma1 growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; RFS, relapse free survival; OS, overall survival; O-DX, Oncotype DX; MMP, MammaPrint; NR,not reported; SG, subgroup; BCSS, breast cancer specific survival; DRFS, distant metastases free survival; Sens, sensitivity; Spec, specificity; EP, EndoPredict

Table 84: Microarray results: C-index (AUC) data

Author, year	Cohorts	Population	Nodal status	Endo / chemo	% pts per group			Outcome	Test	Outcomes
					Low	Inter	High			
Unique cohorts										
O-DX vs MMP										
LN0/+										
Li 2009[177] N=27	Fudan University Cancer Hospital	HR+ NR 70% HER2-	LN0 56%	ET NR 100% CT	-	-	-	**DFS**	**O-DX**	**5 year C-index (AUC):** 0.59; **Sens** 68%; **Spec** 50.0%
					-	-	-		**MMP**	**5 year C-index(AUC):** 0.691; **Sens** 72%; **Spec** 66.2%
Studies drawing from more than one data source, with multiple overlaps between studies										
O-DX vs MMP studies										
LN0/+										
Ahn 2013[170] b)N=82	Gananam Severance Hospital	100% ER+ 12% HER2+ b) subset with RS 19-30	b)43.9% LN+ (LN>3 NR)	b) 94% ET 82% CT				**OS**	**MMP vs O-DX**	**10 year** **O-DX intermediate (RS 19-30) risk group** **KM curve:** MMP low vs high: HR NR p=0.013 **C-index (AUC) MMP:** 0.844
Prat, 2012[180] N=1380	GSE17705, GSE6532, GSE12093, GSE1456, MDACC133	100% ER+ HER2- NR	LN0 47% (% LN>3 NR)	ET 100% CT 0%				**DRFS**	**O-Dx vs MMP**	**8.5 years** **Increase in LR χ^2 of O-DX over MMP:** 14.4, p<0.001 **Increase in LR χ^2 of MMP over O-DX:** 9.2, p=0.002
Tobin, 2014[181] a)N=253	a) Uppsala cohort	a-i) ER+ 100%	NR	NR				**BCSS**	**O-DX**	**13 years** **C-index(AUC):** 0.68
									MMP	**13 years** **C-index(AUC):** 0.81
Tobin, 2014[181] b) N=159	b) Stockholm cohort (Karolinska Hospital)	HR+ NR HER2- NR	b) LN0 59%	b) ET 72%; CT 19%	31	16	53	**BCSS**	**O-DX**	**14.5 years C-index(AUC):** 0.72, p=NR
					48	-	52		**MMP**	**14.5 years C-index(AUC):** 0.76, p=NR

Author, year	Cohorts	Population	Nodal status	Endo / chemo	% pts per group			Outcome	Test	Outcomes
					Low	Inter	High			
Yang 2014[184] N=1981 a-ii) N= 1526	a)METABRIC	ER+ 100% HER2- NR	NR		-	-	-	**BCSS**	**O-DX**	**0-10 year C-index(AUC):** 0.657, p=NR
					-	-	-		**MMP**	**0-10 year C-index(AUC):** 0.612, p=NR
Yang 2014[184] b-ii) N= 134	b) Loi (GSE6532)	ER+ 100% HER2- NR	NR		-	-	-		**O-DX**	**0-10 year C-index(AUC):** 0.640, p<0.05
					-	-	-		**MMP**	**0-10 year C-index(AUC):** 0.606, p<0.05
Yang 2014[184] c-ii)348	c) Buffa (GSE22219)	ER+ 100% HER2- NR	NR		-	-	-		**O-DX**	**0-10 year C-index(AUC):** 0.727, p<0.05
					-	-	-		**MMP**	**0-10 year C-index(AUC):** 0.647, p<0.05
Yang 2014[184] d-ii) 66	d) Wang (GSE19615)	ER+ 100% HER2- NR	NR		-	-	-		**O-DX**	**0-10 year C-index(AUC):** 0.435, p<0.05
					-	-	-		**MMP**	**0-10 year C-index(AUC):** 0.372, p<0.05
Yang 2014[184] e-ii) 201	e) Miller (GSE3494)	ER+ 100% HER2- NR	NR		-	-	-		**O-DX**	**0-10 year C-index(AUC):** 0.645, p<0.05
					-	-	-		**MMP**	**0-10 year C-index(AUC):** 0.650, p<0.05
LN0										
Tobin, 2014[181] a)N=253	a) Uppsala cohort	HR+ NR HER2- NR	a) LN0 63%	a) ET 58%; CT 11%	37	19	44	**BCSS**	**O-DX**	**13 years C-index (AUC):** 0.73
					51	-	49		**MMP**	**13 years C-index (AUC):** 0.84
Prat, 2012[180] a) N=610	GSE17705, GSE6532, GSE12093, GSE1456, MDACC133	100% ER+ HER2- NR	a) LN0 100%	ET 100% CT 0%				**DRFS**	**O-Dx vs MMP**	**8.5 years C-index(AUC):** O-DX: 0.71 MMP: 0.64
Xu 2017[183] a) N=917	METABRIC / Bioconductor	ER+ 100% HER2- NR	LN0 100%	NR	-	-	-	**RFS**	**O-DX**	**15 years C-index(AUC):** 0.68 (estimate off graph)
					-	-	-		**MMP**	**15 years C-index(AUC):** 0.71 (estimate off graph)

Author, year	Cohorts	Population	Nodal status	Endo / chemo	% pts per group			Outcome	Test	Outcomes
					Low	Inter	High			
	datasets: GSE11121, GSE7390, GSE3494, GSE2990, breastCancer NKI				-	-	-		NPI	**15 years C-index(AUC):** 0.68 (estimate off graph)
Yang 2014[184] N=1981a-i) N= 1037	a)METABRIC	NR	LN0 100%	NR	-	-	-	**BCSS**	**O-DX**	**0-10 year C-index(AUC):** 0.650, p=NR
					-	-	-		**MMP**	**0-10 year C-index(AUC):** 0.641, p=NR
Yang 2014[184] b-i) N=125	b) Loi (GSE6532)	NR	LN0 100%	NR	-	-	-		**O-DX**	**0-10 year C-index(AUC):** 0.635, p<0.05
					-	-	-		**MMP**	**0-10 year C-index(AUC):** 0.604, p<0.05
Yang 2014[184] c-i) 250	c) Buffa (GSE22219	NR	LN0 100%	NR	-	-	-		**O-DX**	**0-10 year C-index(AUC):** 0.681, p=NS
					-	-	-		**MMP**	**0-10 year C-index(AUC):** 0.628, p<0.05
Yang 2014[184] d-i) 64	d) Wang (GSE19615)	NR	LN0 100%	NR	-	-	-		**O-DX**	**0-10 year C-index(AUC):** 0.665, p<0.05
					-	-	-		**MMP**	**0-10 year C-index(AUC):** 0.674, p=NS
Yang 2014[184] e-i) 158	e) Miller (GSE3494)	NR	LN0 100%	NR	-	-	-		**O-DX**	**0-10 year C-index(AUC):** 0.608, p=NS
					-	-	-		**MMP**	**0-10 year C-index:** 0.604, p=NS
LN+										
Prat, 2012[180] b) N=699	GSE17705, GSE6532, GSE12093, GSE1456, MDACC133	100% ER+ HER2- NR	b) LN+ 100% (% LN>3 NR)	ET 100% CT 0%				**DRFS**	**O-Dx vs MMP**	**0-10 year C-index(AUC):** O-DX: 0.64 MMP: 0.61
O-DX vs MMP vs EP										
Zhao 2014[187] a) N=912	a) GSE6532, GSE3494, GSE1456,	a)ER+ 76% HER2- 85%		NR	-	-		**DRFS**	**O-DX**	**Follow-up year NR for C-index analysis** **C-index(AUC):** 0.648 (95% CI: 0.63, 0.67) **PVE:** 4.05

Author, year	Cohorts	Population	Nodal status	Endo / chemo	% pts per group			Outcome	Test	Outcomes
					Low	Inter	High			
	GSE7390, GSE2603, E-TABM-158				-	-	-		MMP	**C-index(AUC):** 0.612 (95% CI: 0.60, 0.63) **PVE:** 4.76
					-	-	-		EP	**C-index(AUC):** 0.648 (95% CI: 0.63, 0.67) **PVE:** 4.78

N, number of patient; CT, chemotherapy; ET, endocrine therapy; yr, year; Inter, intermediate; CI, confidence interval; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; DFS, disease free survival; DRFS, distant recurrence free survival; O-DX, Oncotype DX; MMP, MammaPrint; EP, EndoPredict; HR, hazard ratio; KM, Kaplan-Meier; LR, likelihood ratio; BCSS, breast cancer specific survival

Table 85: Microarray results: Additional prognostic value

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Outcome	Test or compar-ison	Likelihood ratio χ^2	Increase in LR χ^2 over CP factors[a]	Other analyses
LN+/- or NR									
Ahn 2013[170] b)N=82	Gananam Severance Hospital	100% ER+ 12% HER2+ b) subset with RS 19-30	b)43.9% LN+ (LN>3 NR)	b) 94% ET 82% CT	**OS**	**MMP vs O-DX**			**O-DX intermediate (RS 19-30) risk group Adjusted HR[a] of MMP :** 10.19 (95% CI: 1.05, 99.01, p=0.045)
Fan 2006[173] a) N=295	NKI (Derivation cohort for MMP)	a) 77% ER+ HER2 NR	a) LN0, 51% LN1-3, 36% LN>3, 13%	a) 14% ET 37% CT	**RFS**	**OD-X**			**Adjusted HR[a]** **Inter Vs Low:** 1.81 (95% CI: 0.70, 4.68, p=0.22) **High Vs Low:** 4.27 (95% CI: 2.05, 8.92, p=0.001)
						MMP			**Adjusted HR[a]:** 3.44 (95% CI: 1.98, 5.99, p<0.001)
					OS	**OD-X**			**Adjusted HR[a]** **Inter Vs Low:** 1.81 (95% CI: 0.39, 8.27, p=0.45) **High Vs Low:** 6.14 (95% CI: 1.84, 20.4, p=0.003)
						MMP			**Adjusted HR[a]:** 4.71 95% CI: (2.02, 11.00, p<0.001)
Fan 2006[173] b) SG N=225		b) 100% ER+	b) NR	b) NR	**RFS**	**OD-X**			**Adjusted HR[a]** **Inter Vs Low:** 0.82 (95% CI: 0.27, 2.46, p=0.72) **High Vs Low:** 2.59 (95% CI: 1.44, 4.65, p=0.001)
						MMP			**Adjusted HR[a]:** 3.88 (95% CI: 2.15, 7.02, p<0.001)
					OS	**OD-X**			**Adjusted HR[a]** **Inter Vs Low:** 1.42 (95% CI: 0.27, 7.50, p=0.68) **High Vs Low:** 4.95 (95% CI: 1.82, 13.4, p=0.002)

Reference; N	Cohorts	Population	Nodal status	Endo / chemo	Outcome	Test or compar-ison	Likelihood ratio χ^2	Increase in LR χ^2 over CP factors[a]	Other analyses
						MMP			**Adjusted HR[a]:** 5.47 (95% CI: 2.13, 14.1, p<0.001)
Zhao 2014[187] a-i) N=692	a) GSE6532, GSE3494, GSE1456, GSE7390, GSE2603, E-TABM-158	a-i) ER+ 100%, HER2- NR	a) LN0 67% (LN>3 NR) a-i) NR	NR	**DRFS**	**O-DX**	43.6, p<0.0001	23.1, p<0.0001 [a]	
						MMP	36.0, p<0.0001	21.5, p<0.0001 [a]	
						EP	53.6, p<0.0001	31.4, p<0.0001 [a]	
LN0									
Yin, 2014[185] N=198	TRANSBIG GSE7390	ER+ NR HER2- NR	LN0 100%	"Systemically untreated patients"[b]	**DRFS**	**O-DX**		13.734, p=0.004	
						MMP		3.038, p=0.986	
						AOL		3.325, p=0.601	
						NPI		6.823, p=0.131	
					OS	**O-DX**		13.286, p=0.002	
						MMP		0.221, p=0.647	
						AOL		0.377, p0.551	
						NPI		3.658, p=0.16	

N, number of patient; CT, chemotherapy; ET, endocrine therapy; Inter, intermediate; CI, confidence interval; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; RFS, relapse free survival; DRFS, distant recurrence free survival; O-DX, Oncotype DX; MMP, MammaPrint; EP, EndoPredict; HR, hazard ratio; KM, Kaplan-Meier; LR, likelihood ratio; BCSS, breast cancer specific survival; NKI, Netherlands Cancer Institute;

[a] Multivariable analysis covariates: **Ahn 2013:** tumour size; nodal status; PR; CT treatment; **Fan 2006 data set a):** ER status, tumour grade, nodal status, age, tumour size, treatment (ET, CT or both); **Fan 2006 data set b)**: as a) but omitting ER status; **Zhao 2014:** nodal status, grade, tumour size**; Yin 2014,** not adjusted, but gives values for AOL and NPI on same cohort for comparison.

[b] from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7390

4.8.3 OPTIMA Prelim: a study of concordance between tests

Concordance between tests

Concordance is defined in this review as the degree to which tests assign the same patients to the same risk groups. They do not report long-term outcomes. They are distinct from decision impact studies, where patients are actually assigned to treatment or not based on the test result and clinician and patient preference.

In accordance with the scope[22] and the protocol[188] we did not conduct a systematic review of concordance. Instead, we present a summary of one high quality, highly relevant study (the Optimal Personalised Treatment of early breast cancer using Multiparameter Analysis (OPTIMA) prelim study)[75] conducted in the UK.

OPTIMA prelim: Methods

OPTIMA prelim (ISRCTN42400492) was a feasibility phase of OPTIMA.[189] OPTIMA is an ongoing trial which aims to test the effectiveness of multiparameter testing in identifying a subgroup of patients (amongst those who would ordinarily be offered chemotherapy) who will not respond to chemotherapy and can therefore avoid it and move more quickly to more appropriate treatments (endocrine therapy and radiotherapy). OPTIMA prelim was designed to help select which of six available tests (Oncotype DX, MammaPrint, Prosigna, IHC4, MammaTyper, and NexCourse Breast by Aqua (IHC4-AQUA)) to use in the trial. Where possible, here we only report data for the four in-scope tests (Oncotype DX, MammaPrint, Prosigna and IHC4). Three clinical prognostic scores were also used, namely AOL, NPI and PREDICT, but these were only compared to each other.

OPTIMA prelim selected women who would routinely be offered chemotherapy, specifically women aged 40 years or older with ER+, HER2- early breast cancer with either 1-9 positive lymph nodes or a tumour 30mm or greater if node negative. Women were randomised to test-directed therapy or standard treatment (chemotherapy followed by endocrine therapy). Patients in the test-directed arm received Oncotype DX testing and those with RS 25 or lower received endocrine monotherapy.

OPTIMA prelim: Results

Results are presented in Table 86. 313 patients from 35 UK hospitals were recruited and randomised. 302 patients received multiple tests. Eleven patients were excluded: four withdrew consent, one was ineligible and four had insufficient tissue for all tests to be performed.

NPI, PREDICT and Adjuvant! Online

By NPI, patients were at high (21%), intermediate (75%) and low (4%) risk. All patients with NPI≤3.4 had tumours 30mm or larger. PREDICT and AOL predict a risk for patients depending on

whether they either take only endocrine monotherapy or take chemotherapy and endocrine therapy; the difference between PREDICT and AOL median predicted 10 year overall survival within each treatment type ranged from 6.2% to 8.4%.

Oncotype DX, MammaPrint, Prosigna and IHC4, MammaTyper, NexCourse Breast by Aqua (IHC4-AQUA)

Results for all tests were available from 236 (78%) of patients. IHC4 could not be determined for 45 (15%) patients; one patient did not have enough tissue for Oncotype DX testing, whilst three Prosigna and seven BluePrint (MammaPrint) tests were unobtainable.

Table 86: Percentage in each risk category and Kappa statistics between tests

Test	% tested	% Low	% Inter	% High	Kappa statistic (95% CI)		
					MMP	Prosigna (L/I)	IHC4 (L/I)
Oncotype DX	99.7	54	28	18	0.40 (0.30 to 0.5)	0.44 (0.3 to 0.5)	0.53 (0.4 to 0.7)
MMP	98.9	61		39	-	0.53 (0.4 to 0.6)	0.33 (0.2 to 0.4)
Prosigna (L/I)	99.0	36	29	35	-	-	0.39 (0.3 to 0.5)
IHC4 (L/I)	85.1	24	48	28	-	-	-
MMP, MammaPrint; L, low; I, intermediate; Inter, intermediate							

Out of the four in-scope tests, MammaPrint assigned the most patients to the low-risk category (61%), though when low and intermediate categories were treated as one category for the three tests that have three risk groups (Oncotype DX, Prosigna and IHC4), Oncotype DX assigned the most to low/intermediate category (82%), and MammaPrint the least (61%) (Table 86).

Kappa statistics indicated modest agreement between tests, ranging from 0.33 (95% CI 0.3 to 0.5) between MammaPrint and IHC4 and 0.53 (95% CI 0.4 to 0.7) between MammaPrint and Prosigna, and 0.53 (95% CI 0.4 to 0.7) between Oncotype DX and IHC4 (Table 86). Data are not reported for the four in-scope tests alone, but across all five tests (that have risk groups rather than intrinsic subtypes, i.e. Oncotype DX, MammaPrint, Prosigna, IHC4 and IHC4-AQUA), 61% of tumours gave no consensus, and only 119 (39%) tumours were uniformly classified as either low-intermediate or high by all five tests. Of these, 93 (31%) were low-intermediate by all tests and 26 (8%) were high-risk by all tests. An exploratory analysis using high/intermediate versus low-risk patients also showed only moderate agreement.

The authors report a number of further analyses which demonstrate that no tests appeared to be more in agreement than others, and that there were no statistically significant differences in clinicopathological variables between concordant and discordant patients. Disagreement spanning two categories (i.e. between low and high-risk) was not infrequent. Agreement was not better at the extremes of the ranges of the tests (the very low- and very high-risk tumours).

Conclusions

The authors concluded that whilst tests assigned similar proportions of patients to low/intermediate and high-risk categories, test results for an individual patient could differ markedly depending on which test was used.

4.9 Results: Decision impact studies

Decision impact: Study and patient characteristics

Decision impact studies assess how decisions to use or not use chemotherapy changed pre- and post-availability of the test. Table 87 to Table 91 show the study characteristics of the included decision impact studies, including whether studies were prospective or retrospective and whether the data were for chemotherapy recommendations or actual treatment decisions. Also shown are the ER, HER2 and nodal status.

Six UK studies[113, 190-196] and twelve other European studies[197-210] assessed decision impact of Oncotype DX (Table 87). One UK study[76] and three other European studies[211-213] assessed decision impact of EndoPredict (EPClin) (Table 88). One UK study[214] and no other European studies assessed decision impact of IHC4+C (Table 89). No UK studies and three European studies[80, 82, 215] assessed decision impact of Prosigna (Table 90). No UK studies and eight European studies[123, 216-222] assessed decision impact of MammaPrint (Table 91). [redacted]

Decision impact: Results

Format of results: In most studies, patients were allocated pre-test to either chemotherapy or no chemotherapy. This could be a recommendation (by physician or multidisciplinary team (MDT)) or actual treatment decisions (what the patient actually received). They were then split into four post-test groups: those whose decision/recommendation remained as chemotherapy, remained as no chemotherapy, changed from no chemotherapy to chemotherapy, or changed from chemotherapy to no chemotherapy. Table 92 to Table 96 illustrate the above data. These data are also summarised in terms of: the proportion of patients undergoing any treatment change (either to or from chemotherapy); the total proportion allocated to chemotherapy both pre- and post-test; and the net change in chemotherapy use. Within each results table sub-heading, studies are broadly ordered as LN0, then mixed nodal status, then LN+.

Oncotype DX: Among four UK studies,[190-195] the percentage of patients with any change in treatment recommendation or decision (either to or from chemotherapy) ranged from 29% to 49% ([redacted]) (Table 92). Across eleven European (non-UK) studies,[197-209] the percentage with any change in treatment recommendation or decision ranged from 5% to 70%. There was little clear difference in results according to LN status.

Among UK studies, the net reduction in chemotherapy recommendations (pre-test to post-test) was 14% to 23% across two studies,[194, 195] while the net reduction in chemotherapy decisions was 8% to 14% across two studies[192, 193, 195] ([redacted]). Two further UK studies[190, 191, 196] reported changes from pre-test chemotherapy recommendation to post-test decision, which may overestimate the net change; one reported a reduction of 23% in chemotherapy use;[190, 191] the other only assessed patients with an initial recommendation for chemotherapy so it is misleading to calculate the absolute change.[196] Across eleven European (non-UK) studies,[197-200, 202-210] the net reduction in chemotherapy recommendations or decisions ranged from 0% to 64%. Again there was little clear difference in results according to LN status.

EndoPredict: In the one UK study of EndoPredict,[76] 37% had a change in treatment decision (either to or from chemotherapy; Table 93). Across three European (non-UK) studies,[211-213] the percentage of patients with any change in treatment recommendation ranged from 38% to 41%. In the UK study, the net change in chemotherapy use (pre-test to post-test) was +1% (since treatment changes occurred in both directions).[76] However, across three European (non-UK) studies,[211-213] there was a net reduction in chemotherapy recommendations ranging from 13% to 26%. There was insufficient data to assess results by LN status.

IHC4+C: In the one UK study of IHC4+C (mix of LN+/-),[214] 27% had a change in treatment recommendation (either to or from chemotherapy; Table 94). Pre-test decisions included either "recommend chemotherapy" or "discuss chemotherapy". The net reduction in patients definitively recommended chemotherapy was 2%. However, if pre-test chemotherapy recommendations were assumed to include both "recommend chemotherapy" and "discuss chemotherapy", the net reduction could be up to 26%. There were no other European studies of IHC4.

Prosigna: There were no UK studies of Prosigna. Across three European (non-UK) studies (either LN0 or not reported),[80, 82, 215] the percentage with any change in treatment recommendation ranged from 14% to 41% (Table 95). The net change in chemotherapy recommendations (pre-test to post-test) was a reduction of 2% in one study[80] and an increase of 2% to 9% in two studies.[82, 215]

MammaPrint: There were no UK studies of MammaPrint. Across seven European (non-UK) studies,[123, 216-220, 222] the percentage with any change in treatment recommendation or decision ranged from 13% to 51% (Table 96). The net change in chemotherapy recommendations (pre-test to post-test) ranged from a reduction of 31% to an increase of 8% across six studies.[123, 216-218, 220, 222] Again there were insufficient data to assess results by LN status.

Summary and discussion of decision impact studies

The percentage of patients with any change in treatment recommendation or decision (either to or from chemotherapy) among UK studies was 29% to 49% across four Oncotype studies,[113, 190-193, 195] 37% in one EndoPredict study,[76] and 27% in one IHC4+C study.[214] Ranges across European (non-UK) studies were 5% to 70% for Oncotype,[197-209] 38% to 41% for EndoPredict,[211-213] 14% to 41% for Prosigna[80, 82, 215] and 13% to 51% for MammaPrint.[123, 216-220, 222]

The net change in the percentage of patients with a chemotherapy recommendation or decision (pre-test to post-test) among UK studies was a reduction of 8% to 23% across four Oncotype studies,[192-195] an increase of 1% in one EndoPredict study,[76] and a reduction of between 2-26% in one IHC4+C study (unclear due to category definitions).[214] Net changes across European (non-UK) studies were a reduction of 0% to 64% for Oncotype,[197-200, 202-210] reduction of 13% to 26% for EndoPredict,[211-213] and reduction of 2% to increase of 9% for Prosigna,[80, 82, 215] and reduction of 31% to increase of 8% for MammaPrint.[123, 216-218, 220, 222]

Table 87: Study characteristics: Oncotype DX

Study	Country (area)	N patients	Population	Nodal status	Prosp / retro	N centres	Pre-test	Pre-test by (based on)	Post-test	Post-test by (based on)	Risk group (%)		
											Low	Inter	High
UK studies													
Hassan 2015[190]; Hassan 2015[191]	UK (Bolton)	26	ER+ HER2- (assumed)	LN0 (assumed)	Prosp	1	Recomm	MDT (NR)	Decision	MDT & patient (NR)	81%		19%
Holt 2013[192] Albanell 2016[193] (subgroup)	UK (Wales)	All: 142 Sub: 94	All: ER+ HER2+/- Sub: ER+ HER2-	All: LN0/N1mi Sub: LN0	Prosp	1	Decision	Physician & patient (CP factors+AOL)	Decision	Physician & patient (NR)	All: 56% Sub: NR	All: 28% Sub: NR	All: 17% Sub: NR
Kiernan 2016[194]	UK	50	ER+ HER2- (assumed)	LN0 (assumed)	Retro	2	Recomm	Physician (NR)	Recommend ation	Physician	NR	NR	NR
Kuchel 2016[195]	UK	135	ER+ HER2-	LN0-3	Prosp	Multi	Recomm and Decision	NR	Recomm and Decision	NR	52%	42%	6%
Loncaster 2017[196]	UK (Manches ter)	All: 201 LN0: 136 LN+: 65	ER+ HER2-	LN0 68% LN+ 32%	Pilot + retro.	NR	Recomm	MDT (CP factors + PREDICT)	Decision	NR (test for low/high RS; test + patient discussion for inter RS)	All: 43% LN0: 34% LN+: 62%	All: 44% LN0: 51% LN+: 29%	All: 13% LN0: 15% LN+: 9%
[redacted]													
European studies													
Albanell 2012[197] (trans-GEICAM)	Spain	107	ER+ HER2-	LN0	Prosp	6	Recomm	Physician (CP factors)	Recomm	Physician	58%	33%	9%
Bodmer 2015[198]	Switzerla nd	60	ER+ HER2- Pre/postmeno Inter clin risk	LN0 or LN+	Prosp	1 area	Recomm	Physician (CP factors)	Recomm	MDT	52%	40%	8%
De San Vicente 2015[199]	Spain	37	HR+ HER2- Inter O-DX	LN0, 73% LN+, 27%	Retro	1	Recomm	Physician (CP factors)	Decision	Physician & patient	0%	100%	0%

Study	Country (area)	N patients	Population	Nodal status	Prosp / retro	N centres	Pre-test	Pre-test by (based on)	Post-test	Post-test by (based on)	Risk group (%)		
											Low	Inter	High
Dieci 2016[200]	Italy	123	ER+ HER2- T1-3 Inter clin risk	LN0	Prosp	9	Recomm	Physician (NR)	Recomm & Decision	Physician Physician & patient	61%	33%	6%
Dreyfus 2015[201]	France	39	HR+ HER2- Indicated for CT	LN0, 39% LN1-3, 51%	Prosp	2	Recomm	MDT	Recomm	MDT	49%	46%	5%
Eiermann 2013[202]	Germany	244 LN0 122 LN+	ER+ HER2-	LN0, 67% LN1-3, 33%	Prosp	15	Recomm	MDT (CP factors & local protocol)	Recomm	MDT	54%	38%	8%
Gligorov 2015[203] (SWITCH)	France	95	ER+ HER2-	LN0-mic	Prosp	7	Recomm	MDT (CP factors + French guidelines)	Recomm	MDT	55%	40%	5%
Hejduk 2016[204] Petrakova 2016a, b[205, 206]	Czech Republic	196	ER+ HER2- grade 2 + other risk factor	LN0	Prosp	13	Recomm	NR	Recomm	NR	56%	38%	6%
Mouysset 2016[207]	France	603	ER+ HER2-	LN0, 61% LN+, 39%	Prosp	Multi	Recomm	MDT (CP factors)	Recomm	MDT	60%	34%	6%
Novas 2016[208]	Spain	35	NR	N1mic	Retro	NR	Recomm	Physician (NR)	Recomm	Physician	54%	43%	3%
Pestalozzi 2015[209]	Switzerland	221	ER+ HER2-	pN0 or pN1a	Prosp	Multi	Recomm	MDT (NR)	Recomm	MDT	NR	NR	NR
Wassermann 2015[210]	France	72	HR+ HER2- Pre/postmeno	LN0, 86% LNmic, 6% LN1-3, 9%	Prosp	4	Recomm	MDT (NR)	Recomm	MDT	NR	NR	NR

N, number of patient;AOL, Adjuvant! Online; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; RS, Oncotype DX recurrence score; Inter, intermediate; sub, subgroup; NR, not reported; Prosp, prospective; Retro, retrospective; Multi, multinational; CP, clinicopathological; Recomm, recommendation; MDT, multidisciplinary team; pre/postmeno, pre and post menopausal women

Table 88: Study characteristics: EndoPredict (EPClin)

Study	Country (area)	N patients	Population	Nodal status	Prosp / retro	N centres	Pre-test	Pre-test by (based on)	Post-test	Post-test by (based on)	Risk group (%)		
											Low		
UK studies													
Bloomfield 2017[76] (abstract)	UK	149	ER+ HER2-	NR	Prosp	8	Decision	Physician & patient (CP factors)	Decision	Physician & patient	50%	-	50%
European studies													
Ettl 2015[211]	Germany	217	ER+ HER2-	LN0, 73% LN+, 27%	Prosp	1	Recomm	MDT (CP factors + uPA/PAI-1)	Recomm	MDT	61%	-	39%
Muller 2013[212]	Germany	130	ER+ HER2-	LN0, 62% LN1-3, 35.5% LN4+, 2.5%	Retro	1	Recomm	Physician (CP factors)	Recomm	Physician	48%	-	52%
Penault-Llorca 2016[213] (ADENDOM)	France	200	ER+ HER2- Clinically inter. risk	LN0-mic	Prosp		Recomm	MDT (CP factors)	Recomm & Decision	MDT	67%	-	33%

N, number of patient; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; Inter, intermediate; -, not reported; Prosp, prospective; Retro, retrospective; CP, clinicopathological; Recomm, recommendation; MDT, multidisciplinary team;

Table 89: Study characteristics: IHC4+C

Study	Country (area)	N patients	Population	Nodal status	Prosp / retro	N centres	Pre-test	Pre-test by (based on)	Post-test	Post-test by (based on)	Risk group (%)		
											Low		
UK studies													
Yeo 2015[214]	UK (London)	124	ER+ HER2-	LN0 74% LN1-3 26%	Prosp	1 (Royal Marsden)	Recomm	MDT (NR)	Recomm	MDT	NR	NR	NR
European studies													
None													

N, number of patient; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; Inter, intermediate; -, not reported; Prosp, prospective; Recomm, recommendation; MDT, multidisciplinary team

Table 90: Study characteristics: Prosigna

Study	Country (area)	N patients	Population	Nodal status	Prosp / retro	N centres	Pre-test	Pre-test by (based on)	Post-test	Post-test by (based on)	Risk group (%)		
											Low		
UK studies													
None													
European studies													
Martin 2015[80] (GEICAM)	Spain	200	ER+, HER2- Stage 1-2 T<5cm postmeno	LN0	Prosp	15	Recomm	Physician (CP variables or AOL & immunohistochemistry)	Recomm	Physician	51%	33%	17%
Van Asten 2016[215]	Belgium	51	ER+, HER2- Unclear if CT needed	NR	Prosp	1	Recomm	MDT (CP factors)	Recomm	MDT	NR	NR	NR
Wuerstlein 2016[82]	Germany	198	ER+, HER2- postmeno	LN0	Prosp	11	Recomm	Physician (CP factors)	Recomm	Physician	43%	35%	22%

N, number of patient; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; Inter, intermediate; NR, not reported; Prosp, prospective; CP, clinicopathological; Recomm, recommendation; MDT, multidisciplinary team; postmeno, postmenopausal; AOL, Adjuvant! Online;

Table 91: Study characteristics: MammaPrint

Study	Country (area)	N patients	Population	Nodal status	Prosp / retro	N centres	Pre-test	Pre-test by (based on)	Post-test	Post-test by (based on)	Risk group (%)		
											Low		
UK studies													
None													
European studies													
Bueno-de-Mesquita 2007[123] (RASTER)	Netherlands	427	80% ER+ 84% HER2- T1-4, M0 <61 yrs	LN0-micro	Prosp	16	Recomm	Physician (Dutch CBO guidelines)	Recomm & Decision	Physician; physician & patient	51%	-	49%
Cusumano 2014[216]	Netherlands, Belgium, Italy, Spain	151	ER+ HER2- T1-3, M0	LN0 LN1-3	Prosp	4	Recomm	MDT (NR)	Recomm	MDT	NR	-	NR
Drukker 2014[217] (subset of RASTER)	Netherlands, Germany, France, Italy, Portugal	37	ER+/- HER2+/- T1-3, M0	LN0	Selected cases	12 oncologists	Recomm	Physician (tools & CP factors)	Recomm	Physician	NR	-	NR
Exner 2014[218]	Austria	75	ER+ HER2- Grade 1-2 T 1-3cm	LN0	Prosp	1 hospital	Recomm	MDT (closely followed St Gallen 2009)	Recomm	MDT	76%	-	24%
Hartmann 2012[219]	Germany	60	HR+ HER2- ≥60 years Grade 2-3 T >1cm	LN0 LN1-3	Prosp	2 hospitals	Decision	MDT (national guidelines) + patient preference	Recomm	MDT	63%	-	37%
Kuijer 2016[220]	Netherlands	377	ER+ (HER2 NR)	NR	Prosp	33 hospitals	Recomm	Physician (CP factors)	Recomm	Physician	57%	-	43%
Rullan 2016[221]	Spain	129	HR+ HER2- 35-70 yr T 1-3cm, grade 2	94% LN0-mic	NR	3 hospitals	Recomm	Physician (CP factors & local protocol)	Decision	Physician + patient	NR	-	NR
Wuerstlein 2016[222] (WSG PRIMe)	Germany	430	HR+ HER2-	LN0 (72%) LN1-3 (28%)	Prosp	27 hospitals	Recomm	Physician (CP factors and/or IHC for ER/PR/Ki67)	Recomm (unclear)	Physician	NR	-	NR

N, number of patient; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; Inter, intermediate; NR or -, not reported; Prosp, prospective; Retro, retrospective; CP, clinicopathological; Recomm, recommendation; MDT, multidisciplinary team

Table 92: Decision impact results: Oncotype

Study	Country	Population	Nodal status	Pre-test	Post-test	N	No CT unchanged	No CT changed to CT	CT unchanged	CT changed to no CT	Treatment change (%)	Pre-test CT (%)	Post-test CT (%)	Net change in CT (%)
UK studies: Recommendation														
Kiernan 2016[194] [a]	UK	ER+ HER2- (assumed)	LN0 (assumed)	Recomm	Recomm	50	NR	NR	NR	NR	NR	21 (42%)[a]	14 (28%)[a]	-7 (-14%)
Kuchel 2016[195]	UK	ER+ HER2-	LN0-3	Recomm	Recomm	135	54	12	26	43	55 (41%)	69 (51%)	38 (28%)	-31 (-23%)
		ER+ HER2- NPI inter.	LN0-3	Recomm	Recomm	67	17	10	17	23	33 (49%)	40 (60%)	27 (40%)	-13 (-19%)
UK studies: Decision														
Holt 2013[192] Albanell 2016[193]	UK	ER+ HER2- (subgroup)	LN0	Decision	Decision	94	45	9	18	22	31 (33%)	40 (43%)	27 (29%)	-13 (-14%)
[redacted]														
Kuchel 2016[195]	UK	ER+ HER2-	LN0-3	Decision	Decision	131	66	13	24	28	41 (31%)	52 (40%)	37 (28%)	-15 (-11%)
		ER+ HER2- NPI inter.	LN0-3	Decision	Decision	65	31	7	15	12	19 (29%)	27 (42%)	22 (34%)	-5 (-8%)
UK studies: Recommendation to decision														
Hassan 2015[190]; Hassan 2015[191]	UK	ER+ HER2- (assumed)	LN0 (assumed)	Recomm	Decision	26	9	2	7	8	10 (38%)	15 (58%)	9 (35%)	-6 (-23%)
Loncaster 2017[196]	UK	ER+ HER2-	LN0	Recomm	Decision (largely on test)	136	NR	NR	NR	NR	NR	136 (100%)	54 (40%)	NA
			LN+			65	NR	NR	NR	NR	NR	65 (100%)	20 (31%)	NA
European studies: Recommendation														
Albanell 2012[197] (trans-GEICAM)	Spain	ER+ HER2-	LN0	Recomm	Recomm	107	56	12	17	22	34 (32%)	39 (36%)	29 (27%)	-10 (-9%)
Dieci 2016[200]	Italy	ER+ HER2-	LN0	Recomm	Recomm	123	71	5	37	10	15 (12%)	47 (38%)	42 (34%)	-5 (-4%)
Eiermann 2013[202]	Germany	ER+ HER2-	LN0	Recomm	Recomm	244	99	28	72	45	73 (30%)	117 (48%)	100 (41%)	-17 (-7%)

Study	Country	Population	Nodal status	Pre-test	Post-test	N	No CT unchanged	No CT changed to CT	CT unchanged	CT changed to no CT	Treatment change (%)	Pre-test CT (%)	Post-test CT (%)	Net change in CT (%)
Hejduk 2016[204] Petrakova 2016a, b[205, 206]	Czech Republic	ER+ HER2-	LN0	Recomm	Recomm	196	43	3	27	123	126 (64%)	150 (77%)	30 (15%)	-120 (-61%)
Gligorov 2015[203] (SWITCH)	France	ER+ HER2-	LN0-mic	Recomm	Recomm	95	41	5	19	30	35 (37%)	49 (52%)	24 (25%)	-25 (-26%)
Novas 2016[208]	Spain	NR	N1mic	Recomm	Recomm	35	21	1	5	8	9 (26%)	13 (37%)	6 (17%)	-7 (-20%)
Bodmer 2015[198]	Switzerland	ER+ HER2-	LN0 or LN+	Recomm	Recomm	60	19	3	13	25	28 (47%)	38 (63%)	16 (27%)	-22 (-37%)
Dreyfus 2015[201]	France	HR+ HER2- Indicated for CT	LN0, 39% LN1-3, 51%	Recomm	Recomm	39	0	0	13	26	26 (67%)	39 (100%)	13 (33%)	NA
Mouysset 2016[207]	France	ER+ HER2-	LN0, 61% LN+, 39%	Recomm	Recomm	603	NR	NR	NR	NR	425 (70%)	529 (88%)	145 (24%)	-384 (-64%)
Pestalozzi 2015[209]	Switzerland	ER+ HER2-	pN0 or pN1a	Recomm	Recomm	221	124	8	52	37	45 (20%)	89 (40%)	60 (27%)	-29 (-13%)
Wassermann 2015[210]	France	HR+ HER2-	LN0, 86% LNmic or 1-3, 14	Recomm	Recomm	72	NR	NR	NR	NR	NR	41 (57%)	14 (19%)	-27 (-38%)
Eiermann 2013[202]	Germany	ER+ HER2-	LN1-3	Recomm	Recomm	122	18	12	58	34	46 (38%)	92 (75%)	70 (57%)	-22 (-18%)
European studies: Recommendation to decision														
Dieci 2016[200]	Italy	ER+ HER2-	LN0	Recomm	Decision		73	3	31	16	19 (15%)	47 (38%)	34 (28%)	-13 (-11%)
Eiermann 2013[202]	Germany	ER+ HER2-	LN0	Recomm	Decision	244	NR	NR	NR	NR	NR	117 (48%)	83 (34%)	-34 (-14%)
De San Vicente 2015[199]	Spain	HR+ HER2- Intermediate O-DX	LN0, 73% LN+, 27%	Recomm	Decision	37	27	1	8	1	2 (5%)	9 (24%)	9 (24%)	0 (0%)
Eiermann 2013[202]	Germany	ER+ HER2-	LN1-3	Recomm	Decision	122	NR	NR	NR	NR	NR	92 (75%)	57 (47%)	-35 (-29%)

N, number of patient; HR+, hormone receptor positive; CT, chemotherapy;HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; Inter, intermediate; NR or -, not reported; Prosp, prospective; Retro, retrospective; CP, clinicopathological; Recomm, recommendation; MDT, multidisciplinary team; mic, micrometastases

[a] Pre/post-test CT includes "CT recommended" and "bias towards CT recommended", while pre/post test no CT includes "ET alone advised" and "bias towards ET alone"

[redacted]

Table 93: Decision impact results: EndoPredict (EPClin)

Study	Country	Population	Nodal status	Pre-test	Post-test	N	No CT unchanged	No CT changed to CT	CT unchanged	CT changed to no CT	Treatment change (%)	Pre-test CT (%)	Post-test CT (%)	Net change in CT (%)
UK studies: Decision														
Bloomfield 2017[76]	UK	ER+ HER2-	NR	Decision	Decision	149	60	28	34	27	55 (37%)	61 (41%)	62 (42%)	+1 (+1%)
European studies: Recommendation														
Penault-Llorca 2016[213] (ADENDOM)	France	ER+ HER2-	LN0-mic	Recomm	Recomm	200	85	20	40	55	75 (38%)	95 (48%)	60 (30%)	-35 (-18%)
Ettl 2015[211]	Germany	ER+ HER2-	LN0, 73% LN+, 27%	Recomm	Recomm	217	NR	16	NR	73	89 (41%)	NR	NR	-57 (-26%)
Muller 2013[212]	Germany	ER+ HER2-	LN0, 62% LN1-3, 35.5% LN4+, 2.5%	Recomm	Recomm	130	31	16	50	33	49 (38%)	83 (64%)	66 (51%)	-17 (-13%)
European studies: Recommendation to decision														
Penault-Llorca 2016[213] (ADENDOM)	France	ER+ HER2-	LN0-mic	Recomm	Decision	200	90	15	38	57	72 (36%)	95 (48%)	53 (27%)	-42 (-21%)

N, number of patient; CT, chemotherapy; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; Inter, intermediate; NR or -, not reported; Prosp, prospective; Retro, retrospective; CP, clinicopathological; Recomm, recommendation; MDT, multidisciplinary team; mic, micrometastases

Table 94: Decision impact results: IHC4+C

Study	Country	Population	Nodal status	Pre-test	Post-test	N	No CT unchanged	No CT changed to CT	CT unchanged	CT changed to no CT	Treatment change (%)	Pre-test CT (%)	Post-test CT (%)	Net change in CT (%)
UK studies: Recommendation														
Yeo 2015[214] [a]	UK	ER+ HER2-	LN0 74% LN1-3 26%	Recomm	Recomm	124	49	1	41	33	34 (27%)	45 (36%) to 74 (60%)[a]	42 (34%)	-3 (-2%) to -32 (-26%)
European studies														
None														

N, number of patient; CT, chemotherapy; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; Inter, intermediate; NR or -, not reported; Prosp, prospective; Retro, retrospective; CP, clinicopathological; Recomm, recommendation; MDT, multidisciplinary team; mic, micrometastases

[a]Pre-test CT: lower estimate includes only those classed as "recommend CT" while upper estimate includes both those classed as "recommend CT" and "discuss CT"

Table 95: Decision impact results: Prosigna

Study	Country	Population	Nodal status	Pre-test	Post-test	N	No CT unchanged	No CT changed to CT	CT unchanged	CT changed to no CT	Treatment change (%)	Pre-test CT (%)	Post-test CT (%)	Net change in CT (%)
UK studies														
None														
European studies: Recommendation														
Martin 2015[80] (GEICAM)	Spain	ER+, HER2-	LN0	Recomm	Recomm	200	122	18	38	22	40 (20%)	60 (30%)	56 (28%)	-4 (-2%)
Wuerstlein 2016[82]	Germany	ER+, HER2-	LN0	Recomm	Recomm	198	131	22	40	5	27 (14%)	45 (23%)	62 (31%)	+17 (+9%)
Van Asten 2016[215]	Belgium	ER+, HER2-	NR	Recomm	Recomm	51	15	11	15	10	21 (41%)	25 (49%)	26 (51%)	+1 (+2%)

N, number of patient; CT, chemotherapy; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; Inter, intermediate; NR or -, not reported; Prosp, prospective; Retro, retrospective; CP, clinicopathological; Recomm, recommendation; MDT, multidisciplinary team; mic, micrometastases

Table 96: Decision impact results: MammaPrint

Study	Country	Population	Nodal status	Pre-test	Post-test	N	No CT unchanged	No CT changed to CT	CT unchanged	CT changed to no CT	Treatment change (%)	Pre-test CT (%)	Post-test CT (%)	Net change in CT (%)
UK studies														
None														
European studies: Recommendation														
Drukker 2014[217b] (RASTER)	Netherlands, Germany, France, Italy, Portugal	ER+/- HER2+/-	LN0	Recomm	Recomm	37 (414)[b]	202	9	144	59	68 (16%)	203 (49%)	153 (37%)	-50 (-12%)
Exner 2014[218]	Austria	ER+ HER2-	LN0	Recomm	Recomm	75	40	4	21	10	14 (19%)	31 (41%)	25 (33%)	-6 (-8%)
Bueno-de-Mesquita 2007[123] (RASTER)	Netherlands	80% ER+ 84% HER2-	LN0-micro	Recomm	Recomm	427	NR	NR	NR	NR	NR	186 (44%)	219 (51%)	+33 (+8%)
Cusumano 2014[216a]	Netherlands, Belgium, Italy, Spain	ER+ HER2-	LN0 LN1-3	Recomm	Recomm	151 (453)[a]	149	68	161	75	143 (32%)	236 (52%)	229 (51%)	-7 (-2%)
Kuijer 2016[220c]	Netherlands	ER+ (HER2 NR)	NR	Recomm	Recomm	377[c]	69	38	114	156	194 (51%)	270 (72%)	152 (40%)	-118 (-31%)
Wuerstlein 2016[222] (WSG PRIMe)	Germany	HR+ HER2-	LN0 (72%) LN1-3 (28%)	Recomm	Recomm (unclear)	430	201	65	107	57	122 (28%)	164 (38%)	172 (40%)	+8 (+2%)
European studies: Recommendation to decision														
Bueno-de-Mesquita 2007[123] (RASTER)	Netherlands	80% ER+ 84% HER2-	LN0-micro	Recomm	Decision	427	206	35	167	19	54 (13%)	186 (44%)	202 (47%)	+16 (+4%)
Rullan 2016[221]	Spain	HR+ HER2-	94% LN0-mic	Recomm	Decision	129	NR	NR	NR	NR	NR	119 (92%)	45 (35%)	-74 (-57%)
European studies: Decision to recommendation														
Hartmann 2012[219]	Germany	HR+ HER2-	LN0 LN1-3	Decision	Recomm	60	47	6	2	5	11 (18%)	7 (12%)	8 (13%)	+1 (+2%)

Study	Country	Population	Nodal status	Pre-test	Post-test	N	No CT unchanged	No CT changed to CT	CT unchanged	CT changed to no CT	Treatment change (%)	Pre-test CT (%)	Post-test CT (%)	Net change in CT (%)
N, number of patient; CT, chemotherapy; HR+, hormone receptor positive; HER2, human epidermal growth factor receptor; ER+, oestrogen receptor positive; LN, lymph node; Inter, intermediate; NR or -, not reported; Prosp, prospective; Retro, retrospective; CP, clinicopathological; Recomm, recommendation; MDT, multidisciplinary team; mic, micrometastases [a]Cusumano 2014: Each patient analysed 3 times at 3 different hospitals (in 3 countries) so n=151 patients but n=453 datapoints. [b]Drukker 2014: Each of 37 patients analysed by up to 12 physicians, giving 414 data points. [c]Kuijer 2016: Data presented here exclude 283 patients with pre-test CT decision recorded as “unsure”													

4.10 Anxiety and health-related quality of life

Six studies (7 publications)[76-82] reported outcomes relating to anxiety (including worry and distress) and HRQoL (Table 97). Studies reporting outcomes such as decision conflict and patient satisfaction did not meet the inclusion criteria for the review and were excluded.

Oncotype DX: Two studies[77, 79] reported data for Oncotype DX. Both adopted a pre-test/post-test design, and included LN+ or LN0 patients. Evans *et al*. 2016[77] used the Impact of Events Scale (IES),[223] and showed no difference between pre- and post-test values (p=0.09), and there were no differences by RS risk group (interaction tests reported as "not statistically significant"). Lo *et al*.[79] on the other hand, reported a statistically significant improvement in overall State-Trait Anxiety Inventory (STAI) score between pre- and post- tests values (p=0.007), but no difference in Trait anxiety (p=0.27). Results for State anxiety were not reported. Only Lo *et al*.[79] reported HRQoL using FACT-B (Functional Assessment of Cancer Therapy-Breast cancer) and FACT-G (Functional Assessment of Cancer Therapy-General) and reported no statistically significant change (p0.55 and p=0.49 respectively, Table 98).

MammaPrint: One study reported data for MammaPrint[81] (Table 97). The study recruited exclusively from patients who had been screened for eligibility in the MINDACT trial, but included both those eligible and those ineligible for MINDACT (due to being LN>3 or having a test failure). A modified version of Lynch's distress scale and one of Lerman's Cancer Worry Scale were used. Patients were separated out into seven subgroups according to their clinical risk, MammaPrint risk, whether they were assigned to chemotherapy or not, and whether the MammaPrint test result was missing (Table 98). Regression analyses adjusted for sociodemographics, understanding of genomic results, timing of test results, perceived risk and satisfaction with process showed higher distress where the genomic test failed, where the patient was high-risk by both clinical scoring and by MammaPrint and in patients with discordant results where the treatment matched the MammaPrint score (i.e. clinical low/genomic high, prescribed chemotherapy; clinical high/genomic low, not prescribed chemotherapy). Only patients with high clinical risk and no genomic test result had a statistically significant decrease in FACT-B HRQoL.

EndoPredict: One study[76] reported data for EndoPredict (Table 97). The study was a pre-test post-test design, and reported a statistically significant decrease in STAI for those whose treatment decision changed from chemotherapy to no chemotherapy on the basis of the EndoPredict (p<0.01), and an increase in STAI for those whose treatment decision changed from no chemotherapy to chemotherapy (p<0.001) (Table 98).

Prosigna: Two studies[80, 82] reported data for Prosigna (Table 97). Both adopted a pre-test post-test design and included only LN0 patients. In both studies there was no difference in Trait anxiety scores (p=0.858, p=0.431 respectively),[80, 82] and in both studies State anxiety changed significantly in low-risk (by Prosigna) patients (p<0.001, p=0.008 respectively) [80, 82] but not in the intermediate- or high-risk groups (Table 98). Both studies reported FACT-G; Martin *et al*.[80] reported no change in overall scores, whilst Wuerstlein *et al*.[82] reported a statistically significant ANOVA p-value for emotional and physical wellbeing (p=0.030, p=0.005 respectively).

Discussion

There were no data relating to the impact of IHC4 on anxiety or HRQoL. Other available data is limited in terms of study designs (pre-post test) and patient spectrum. The lack of a comparator makes it difficult to tell whether similar changes would have occurred were patients to have received a definitive decision based on their clinical risk factors alone. Across tests, and where reported, state anxiety decreased post-test, and total FACT-G generally stayed the same. Results for one study suggest that patients had higher distress where the genomic test failed, where the patient was high-risk by both clinical scoring and by genomic test and in patients with discordant results where the treatment matched the genomic score, though it was unclear if this was due to distress associated with change (in treatment decision) or a lack of trust in the genomic score.

Conclusions

Genomic testing may reduce state anxiety in some patients in some contexts, but generally there was little impact on HRQoL.

Table 97 Study and patient characteristics: Anxiety and HRQoL

Reference(s)	Test	Cohort(s)	Country	Study design	Details of test	Cut-offs	N	Population	Nodal status	Outcomes
Oncotype DX										
Evans, 2016[77]	O-DX	4 centres (Washington, Maryland and Florida)	USA	Pre-post test	NR	NR	193	ER+ Stage I&II	LN+/- (LN>3 NR)	IES[223]
Lo, 2010[79]	O-DX	NR	USA	Pre-post test	Genomic Health	NR	93	EBC HER2+ 7%	LN+/- (LN>3 NR)	STAI; FACT-B, FACT-G
MammaPrint										
Retel, 2013[81]	MMP	MINDACT (enrolled and ineligible pts)	Neths	Non-randomised clinical trial	NR	NR	347	EBC	LN+/-	Lynch's distress scale (adapted); Lerman's Cancer Worry Scale (adapated); FACT-B breast cancer subscale.
EndoPredict Clinical										
Bloomfield, 2017[76, 78]	EP Clin (EP+ NS + TS)	8 Hospitals	South east England	Pre-post test	NR	NR	149	ER+ HER2- EBC with equivocal indications for chemotherapy by Adjuvant! Online	NR	STAI
Prosigna										
Martin 2015[80]	Prosigna	15 centres	Spain	Pre-post test	Manufacturer's specifications	NR	200	ER+ HER2- EBC Stage I&II	LN0	STAI; FACT-G
Wuerstlein, 2016[82]	Prosigna	11 centres	Germany	Pre-post test	Manufacturer's specifications	40-60	198	ER+ LN0 Postmenopausal	LN0	STAI; FACT-G

N, number of patients, LN+, lymph node positive; LN0, lymph node negative; Neths, Netherlands, O-DX, Oncotype DX; MMP, MammaPrint; EP Clin, EndoPredict Clincial; NR, not reported, ER+, Oestrogen-receptor positive; HER2-, human epidermal growth factor receptor negative; EBS, early breast cancer; NS, nodal status; TS, tumour size; STAI, Spielberger's State/Trait Anxiety inventory; IES, Impact of Event Scale; EBC, early breast cancer; WSG BCIST West German Study Group Breast Cancer Intrinsic Subtype Study; FACT-B, function assessment of cancer therapy – breast cancer; FACT-G, Functional assessment of cancer therapy- General.

Table 98: Results: Anxiety and HRQoL

Reference(s)	Test	Country	Study design	Population	Nodal status	Anxiety	HRQoL
Oncotype DX							
Evans, 2016[77] N=193	O-DX	USA	PPT	ER+	LN+/- (LN>3 NR)	**IES** No change pre-post test, p=0.09. Not different by RS group (interaction tests not significant)	NR
Lo, 2010[79] N=93	O-DX	USA	PPT	EBC HER2+ 7%	LN+/- (LN>3 NR)	**STAI mean score (SD)** Pre: 39.6 (14.5) Immediately post: 36 (12.6) 12 months post: 34.0 (11.5), p=0.007 **Trait anxiety** Pre: 32.2 (14.5) Immediately post: 31.7 (13.3) 12 months post: 33.2 (11.0), p=0.27	**FACT-B mean score (SD)** Pre: 112.2 (17.4) 12 months post: 114.3 (18.6), p=0.55 **FACT-G mean score (SD)** Pre: 88.7 (12.3) 12 months post: 87.6 (14.9), p=0.49
MammaPrint							
Retel, 2013[81] N=347	MMP	Neths	Non-randomised clinical trial	EBC	LN+/-	**Lynch's distress scale (adapted):** Adjusted regression analysis: [a] C high/ G high: p<0.001 C-low/G high (no CT): p=0.043 C-low/G high (CT): p<0.001 C-high/G-low (no CT): p<0.001 C-high/G-low (CT):p=0.175 C-low/G-NA: p<0.001 C-high/G-NA: p<0.001 **Lerman's Cancer Worry Scale (adapated):** Adjusted regression analysis: [a] No risk group statistically significant (p ranged from 0.081 to 0.827)	**FACT-B, breast cancer subscale:** Adjusted regression analysis: [a] C high/ G high: p=0.013 C-low/G high (no CT): p=0.585 C-low/G high (CT): p=0.254 C-high/G-low (no CT): p=0.541 C-high/G-low (CT): p=0.296 C-low/G-NA: p=0.075 C-high/G-NA: p<0.001
EndoPredict Clinical							

Bloomfield 2017[76, 78] N=149	EP Clin	UK	PPT	ER+, HER2-, equivocal by AOL	NR	**STAI** Unchanged decision: STAI stable Change from CT to no CT: STAI lower (p<0.01) Change from no CT to CT: STAI increase (p<0.001)	NR
Prosigna							
Martin 2015[80] N=200	Prosigna	Spain	PPT	ER+, HER2-	LN0	**Trait anxiety, mean (SD) (n=180)** Pre: all:39.1 (11.1) Post: 39.2 (10.9) Difference: -0.1 (8.3), p=0.858 **State anxiety, mean (SD) (n=181)** Pre: 42.6 (12.5) Post: 39.8 (13.3) Difference: 2.8 (12.4), p=0.003; low p<0.001; inter p=0.2; high p=0.13	**FACT-G** Pre: 79.2 (15.6) Post: 79.6 (14.6) Difference: -0.4 (13.9), p=0.713
Wuerstlein, 2016[82] N=198	Prosigna	Germany	PPT	ER+ LN0 Post-menopausal	LN0	**State anxiety, mean difference (SD)** Low ROR: -4.3 (8.9), p=0.008 Intermediate ROR: 0.3 (8), p=0.639 High ROR: 0.9 (11.6), p=0.785 p=0.001[b] **Trait anxiety** No statistically significant difference in any group, p=0.431[b]	**FACT-G** No statistically significant differences in any group, for any subscale, except Emotional and Functional wellbeing. Physical wellbeing, p=0.969 [b] Social/family wellbeing, p=0.739 [b] Emotional wellbeing, p=0.030 [b] Functional wellbeing, p=0.005 [b]

N, number of patients, LN+, lymph node positive; EBS, early breast cancer; LN0, lymph node negative; Neths, Netherlands, O-DX, Oncotype DX; MMP, MammaPrint; EP Clin, EndoPredict Clincial; NR, not reported, ER+, Oestrogen-receptor positive; HER2-, human epidermal growth factor receptor negative; PPT, pre-test post-test design; NS, nodal status; TS, tumour size; STAI, Spielberger's State/Trait Anxiety inventory; FACT-B, function assessment of cancer therapy – breast cancer; FACT-G, Functional assessment of cancer therapy- General.

[a]adjusted for sociodemographic (age, marital status, children, education), information/knowledge levels, risk perception variables (understanding, results in 1st visit, knowledge, risk perception, satisfaction) and risk groups (as listed in table); [b] p-value ANOVA of Mean Differences

4.11 Time to test results

The only article identified relating to time to test results was the study by Losk et al. (2016)[224] which reported factors associated with delays in chemotherapy initiation (defined as 42 days or more from surgery to chemotherapy) in breast cancer patients at a US cancer centre in 2011-2013. Of 263 HR+ HER2- women receiving adjuvant chemotherapy, 82 had an Oncotype DX test ordered. Of those for whom an Oncotype test was ordered, 31% had a delay of at least 42 days to chemotherapy initiation, compared with 20% of patients for whom Oncotype DX was not ordered.

4.12 Comparison of TransATAC data to other study data (risk classification and prognosis)

TransATAC is an important study since it evaluates four of five in-scope tests; this study is used in the health economic analysis (see Section 5). It is therefore important to examine whether TransATAC results are consistent with those of other studies for prognostic ability of each test. A comparison of prognostic data is provided for LN0 patients across all studies (Table 7) and for TransATAC (Table 99), and for LN+ patients across all studies (Table 8) and for TransATAC (Table 100). MammaPrint and IHC4+C cannot be compared, since MammaPrint was not included in TransATAC and IHC4 only has data from TransATAC (further IHC4+C data were available but used different cut-points, hence these are not summarised here).

In LN0 patients (Table 7 and Table 99), the percentages categorised as low- or intermediate-risk appear relatively similar (or slightly higher) for TransATAC compared with other studies for Oncotype DX, ROR-PT and EPClin. In LN+ patients (Table 8 and Table 100), these percentages appear relatively similar between TransATAC and other studies for ROR-PT and EPClin; this comparison is difficult to make for Oncotype DX since all other LN+ studies are in populations receiving chemotherapy so may be higher-risk (percentage low-risk 36% to 39%; 3 studies[51, 89-91]) than in TransATAC (percentage low-risk [redacted]).

Among studies of both LN0 and LN+ patients, 10-year DRFS/DRFI rates in low-risk groups in TransATAC were very similar to those in other studies in ER+ patients receiving endocrine but no chemotherapy.

Table 99: Risk categorisation and prognostic ability in TransATAC: LN0

Test	Trial (refs)	Population	Nodal status	Endo / chemo	% categorised low-risk	% categorised int-risk	10yr DRFS/DRFI in low-risk groups	Significantly prognostic for DRFS/DRFI?
LN0, all ET, no CT								
Oncotype DX	TransATAC[43]	ER+ HER2-	LN0	All ET No CT	[redacted]	[redacted]	[redacted]	[redacted]
ROR-PT	TransATAC[43]	ER+ HER2-	LN0	All ET No CT	[redacted]	[redacted]	[redacted]	[redacted]
EPClin	TransATAC[43]	ER+ HER2-	LN0	All ET No CT	[redacted]	[redacted]	[redacted]	[redacted]
IHC4+C[a]	TransATAC[43]	ER+ HER2-	LN0	All ET No CT	[redacted]	[redacted]	[redacted]	[redacted]

CT, chemotherapy; DRFS/I, distant recurrence-free survival/interval; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor-2; LN, number of positive nodes; NR, not reported. [a]Only IHC4+C included in this summary table because there were few data divided by LN status for IHC alone.

Table 100: Risk categorisation and prognostic ability in TransATAC: LN+

Test	Trial (refs)	Population	Nodal status	Endo / chemo	% categorised low-risk	% categorised int-risk	10yr DRFS/DRFI in low-risk groups	Significantly prognostic for DRFS/DRFI?
LN+, all ET, no CT								
Oncotype DX	TransATAC[43]	ER+ HER2-	LN1-3	All ET No CT	[redacted]	[redacted]	[redacted]	[redacted]
ROR-PT	TransATAC[43]	ER+ HER2-	LN1-3	All ET No CT	[redacted]	[redacted]	[redacted]	[redacted]
EPClin	TransATAC[43]	ER+ HER2-	LN1-3	All ET No CT	[redacted]	[redacted]	[redacted]	[redacted]
IHC4+C[a]	TransATAC[43]	ER+ HER2-	LN1-3	All ET No CT	[redacted]	[redacted]	[redacted]	[redacted]

CT, chemotherapy; DRFS/I, distant recurrence-free survival/interval; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor-2; LN, number of positive nodes; NR, not reported. [a]For IHC alone, little data by LN status.

5. COST-EFFECTIVENESS

This chapter presents a systematic review of economic analyses of tumour profiling tests for early breast cancer published since NICE DG10,[21] a critique of economic analyses provided to the EAG by the manufacturers of Oncotype DX[113] and MammaPrint[121] and the chief investigator of the EndoPredict decision impact study,[225] and the methods and results of a *de novo* model-based health economic evaluation of each of the tumour profiling tests compared with current practice.

5.1 Review of existing economic analyses published since NICE DG10

5.1.1 Cost-effectiveness review - methods

Systematic searches were undertaken to identify existing economic evaluations of tumour profiling tests to guide treatment decisions in people with early breast cancer. Only those studies which were published since the previous appraisal of tumour profiling tests (NICE DG10[21]) were considered to be potentially relevant for inclusion in the review; a review and a critical appraisal of economic analyses published prior to this date is available in Ward *et al*.[18] The review was undertaken solely with the purpose of exploring methodological choices and their potential relevance to the current decision problem, rather than to assess the results of published economic evaluations or the potential sources of bias which might affect these.

A comprehensive search was undertaken to systematically identify economic evaluations of the five tumour profiling tests (EndoPredict, Oncotype DX, MammaPrint, IHC4 and Prosigna) and reviews of economic evaluations of tumour profiling tests for breast cancer.

Literature searching for economic evaluation studies was undertaken in March 2017 in the following electronic databases:

- MEDLINE Epub Ahead of Print, In-Process & Other Non-Indexed Citations: Ovid, 1946 to present
- EMBASE: Ovid, 1974 to present
- Health Technology Assessment Database (HTA): Centre for Reviews and Dissemination, 1995 to 2016
- NHS Economic Evaluation Database (NHS EED): Centre for Reviews and Dissemination, 1995 to March 2015
- Science Citation Index Expanded (SCI-E): Web of Science, 1900 to present
- Conference Proceedings Citation Index – Science (CPCI): Web of Science, 1990 to present.

The search strategies comprised MeSH or Emtree Thesauri terms and free-text synonyms for: (i) ‘tumour profiling tests’ and ‘breast cancer’ and (ii) ‘breast cancer’ only. Searches for Oncotype DX,

MammaPrint, IHC4, and Prosigna were limited by publication date from 2011 (the date cut-off for the previous appraisal), whereas no date limits were applied to EndoPredict. Searches were translated across databases and were not limited by language. The search strategies are presented in Appendix 1. Search filters designed to identify economic evaluations and reviews were used on MEDLINE and other databases, where appropriate. Reference and citation searching of included papers was undertaken.

In order to be considered potentially relevant for inclusion in the review, studies were required to meet all of the following criteria:

- Full economic evaluations comparing tumour profiling for breast cancer tests against other tests and/or current practice
- Published in English
- Available in full text format (studies which were available in abstract form only were excluded from the review)
- Relevant to the populations included within the final NICE scope.[22]

5.1.2 Cost-effectiveness review results - summary of studies identified

A total of 294 potentially includable studies (including potential duplicates) were identified by the searches. Of these, 59 studies were deemed to be potentially eligible for inclusion in the review and full texts were obtained, where available. A total of 26 unique studies met the inclusion criteria and were included in the review. The scope and methodological approaches adopted within the included studies are summarised in Table 101 and Table 102, respectively.

The models reported within the included studies were developed to assess the cost-effectiveness of tumour profiling tests across a variety of different countries including the UK, the US, Canada, Mexico, Japan, Austria, Germany, France and the Netherlands. Most studies compared Oncotype DX (eighteen studies) or MammaPrint (eight studies) against comparators such as AOL, the St Gallen guidelines, standard practice, or other conventional diagnostic tools. One included study (Blank *et al*,[226]) compared EndoPredict against a comparator which was comprised of a combination of three different guidelines. There was variation between the analyses with respect to the patient populations evaluated, their disease type and other patient characteristics. The models included populations with initial ages (where reported) ranging from 45 years to 64 years.

Across the breadth of included studies, there was a high level of consistency in terms of the general modelling approach and structure, and several studies were based on a previously published model. The majority of the included models adopted a Markov or hybrid decision tree - Markov approach,

with discrete nodes applied to estimate long-term costs and outcomes for patients assigned to different test risk classification categories. Two studies adopted a partitioned survival approach. One further study used a discrete event simulation (DES) approach. The structure of the model used in one study was not reported. The time horizons used in the economic models ranged from 10 years to the patient's remaining lifetime, with cycle lengths (where reported) ranging from one month to one year. Most of the models that evaluated Oncotype DX against current practice assumed that the test was associated with a predictive benefit of chemotherapy.

Most of the included studies that adopted a Markov structure included a common set of three health states: (i) alive and recurrence-free; (ii) alive with distant recurrence, and (iii) dead. However, several models also included other health states such as: local recurrence; disease-free after local recurrence; distant recurrence with response to treatment; distant recurrence with no response to treatment; progression of disease after distant recurrence; congestive heart failure (CHF); chronic myeloid leukaemia; acute myeloid leukaemia (AML)/myelodysplastic syndrome (MDS), and febrile neutropenia (FN) and chemotherapy-induced nausea and vomiting (CINV). One model, which was reported across two studies, used different health states for patients receiving endocrine therapy only (remission, local recurrence, distant recurrence and dead) and for patients receiving chemotherapy plus endocrine therapy (remission with chemotherapy, remission without chemotherapy, local recurrence, distant recurrence and dead).

Whilst many of the models identified by the review adopted a similar modelling approach, none included all of the relevant tests listed in the final NICE scope.[22] As such, none of the existing models included in the review were considered to be suitable for the current appraisal.

Table 101: Existing economic evaluations – analytic scope

Author	Population	Age	Intervention	Comparator	Country	Perspective	Time horizon	Discount rate
Bargallo-Rocha* (2015)[227]	HR+, LN0 or LN1-3 early-stage breast cancer	Baseline age 55.5 years	Oncotype DX	Current standard of care	Mexico	Instituto Mexicano del Seguro Social perspective	40 years	5%
Holt* (2013)[192]	LN0 or pNImi, ER-positive breast cancer in the UK	Mean age 60.55 years	Oncotype DX	Conventional diagnostic procedures (including AOL and NPI)	UK	NHS	30 years	3.5%
Davidson* (2013)[228]	ER+ LN0 breast cancer	Mean age 53 years	Oncotype DX	Conventional diagnostic procedures	Canada	Canadian health care system	Lifetime (up to maximum age 100 years)	5%
Jahn (2015)[229]	ER+ and/or PR+, HER-2/ neu negative, and LN0 breast cancer	Baseline age 50 years	Oncotype DX	AOL score	Austria	Societal perspective in line with the Austrian health care system	Lifetime	5%
Kondo (2011)[230]	ER+ early stage breast cancer	Baseline age 45 years	Oncotype DX	St Gallen	Japan	Societal	Lifetime (with assumptions about max survival after 10 1-year cycles)	3%
Lamond (2012)[231]	Early stage, endocrine-sensitive breast cancer undergoing adjuvant chemotherapy or no chemotherapy	Median age 50 years	Oncotype DX	Current practice (population-based study)	Canada	Canadian health care system perspective	25 years	3%
Paulden (2013)[232]	LN0, ER+ and/or PR+, (HER2-/neu) early breast cancer, who are candidates for adjuvant chemotherapy	Baseline age 50 years	Oncotype DX	AOL	Canada	Ontario Ministry of Health and Long-Term Care	Lifetime	5%

Author	Population	Age	Intervention	Comparator	Country	Perspective	Time horizon	Discount rate
Reed* (2013)[233]	LN0, ER+ breast cancer	Baseline age 55 years	Oncotype DX	No RS guided strategy	US	US health-system perspective and societal perspective	Lifetime	3%
Blank* (2015)[226]	ER+, HER2-negative breast cancer.	Median age appears to be 64 years	EndoPredict (EPClin) +/- 3 guidelines	3 guidelines (German S3, St Gallen, NCCN)	Germany	German health care system	Lifetime (50 years)	3%
Bonastre (2014)[234]	LN0 early breast cancer. Subgroup analysis of ER+ patients	Patients aged <61 years	MammaPrint	AOL, chemotherapy for all	France	French National Insurance Scheme	10-years	4%
Retel* (2012a)[235]	Early, operable, LN0, ER+ breast cancer	Baseline age 50 years	MammaPrint	Clinical-pathological guidelines (such as AOL)	Netherlands	Dutch health care perspective	20-years	4% costs, 1.5% health outcomes
Retel* (2012b)[236]	Early, LN0 breast cancer	Not reported	MammaPrint; Oncotype	AOL	Netherlands	Dutch health care perspective	20-years	4% costs, 1.5% health outcomes
Retel* (2013a)[237]	Early LN0 ER+ breast cancer after local therapy	Baseline age 50 years	MammaPrint 70G-FFT; MammaPrint 70G-PAR;	AOL	Netherlands	Societal perspective	20-years	4% costs, 1.5% health outcomes
Retel* (2013b)[238]	Reflective of RASTER population	Mean age 48 years	MammaPrint	AOL	Netherlands	Dutch health care perspective	20-years	4% costs, 1.5% health outcomes
Hall (2012)[239]	LN+, ER+ early-stage breast cancer	Baseline age 60 years	Oncotype DX	Standard care (chemotherapy for all)	UK	NHS	Lifetime (up to maximum age 100 years)	3.50%
Hannouf (2012)[240]	Early-stage ER+/PR+ axilliary LN0 breast cancer	Starting age unclear	Oncotype DX	Current practice (population-based study)	Canada	Canadian public health care system	Lifetime	5%
Hannouf (2014)[241]	Post-menopausal women with early-stage ER+/PR+ axillary lymph-node positive breast cancer	Mean age 61 years	Oncotype DX	Current practice	Canada	Canadian public health care system	Lifetime	5%

Author	Population	Age	Intervention	Comparator	Country	Perspective	Time horizon	Discount rate
Kondo (2012)[242]	HR+, LN0, HER2- early stage breast cancer	Baseline age 55 years	MammaPrint	St Gallen	Japan	Societal	10-years	3%
Mislick* (2014)[243]	Early-stage, LN0, ER+ breast cancer	Not reported	Mammostrat	Oncotype DX	US	Third-party payer perspective	10-years	3%
Stein (2016)[244]	ER+, HER2- early-stage breast cancer patients	Median age 58 years	Oncotype DX; MammaPrint/ Bluetest; Prosigna	Chemotherapy for all	UK	NHS	Lifetime (up to maximum age 100 years)	3.50%
Tiwana (2013) [245]	Women who are LN0, ER+ and/or PR+, HER2/neu negative early breast cancer, who are candidates for adjuvant chemotherapy	50 years	Oncotype DX	AOL	Canada	Not reported - appears to be payer perspective	Lifetime	5%
Vanderlaan* (2011) [246]	Minimally LN+, early-stage breast cancer	Mean age 62 years	Oncotype DX	Current care (US NCCN guidelines)	US	US payer (managed care) perspective	30-years	3%
Wong (2012)[247]	Women with LN+ HR+ breast cancer (1-3 nodes)	Reflective of RxPONDER[248]	Oncotype DX	Current care (US NCCN guidelines)	US	Payer	Lifetime (40 years)	3%
Ward (2013)[18]	Women with ER+ LN0, and HER2- early breast cancer	Mean age 58.3 years	Oncotype DX, IHC4, MammaPrint and Mammostrat	Current clinical practice (NPI and	UK	NHS and PSS	Lifetime (up to age 100 years)	3.5%
Yang* (2012)[249]	LN0, ER+ breast cancer	Not reported	Oncotype DX	MammaPrint	US	Third party payer	10 years	3%
Yamauchi * (2014)[250]	Women with ER+, LN0 (including micrometastases) ESBC who were eligible for treatment with adjuvant chemotherapy after having undergone surgery for primary tumour removal and lymph node dissection	Mean age 49.8 years	Oncotype DX	No RS guided strategy	Japan	Societal	Lifetime	3%

NCCN - National Comprehensive Cancer Network; RASTER - MicroarRAy-prognoSTics-in-breast-cancER; RxPONDER - Rx for Positive Node, Endocrine Responsive breast cancer; PSS – Personal Social Services

* known or potential conflict of interest declared

Table 102: Existing economic evaluations - modelling approach and assumptions regarding predictive benefit and chemotherapy

Author	Model approach	Cycle length	Model type	Does model claim predictive benefit for test?	Assumptions on chemotherapy use	Long-term health states
Bargallo-Rocha (2015)[227]	Markov	1-year	Classification to LR, IR, HR	Yes - RRR only in the high risk group, based on Paik	Proportion of all groups receive chemotherapy	3 states: (1) recurrence-free; (2) recurrence; (3) dead
Holt (2013)[192]	Markov	1-year	Classification to LR, IR, HR	Yes - RRRs in intermediate and high risk based on Paik	Change in chemotherapy use informed by decision conflict analysis (changes applied to all three risk groups)	3 states: (1) recurrence-free; (2) recurrence; (3) dead
Davidson (2013)[228]	Markov	1-year	Classification to LR, IR, HR	Yes - different RRRs between risk groups	Proportion of all groups receive chemotherapy	5 states: (1) RFS no chemo; (2) RFS chemo; (3) distant recurrence no chemo; (4) distant recurrence post-chemo; (5) dead.
Jahn (2015)[229]	DES	N/a	Sequential use of AOL and Oncotype DX - 8 test strategies considered	Yes	Chemotherapy provided to proportion of patients in all groups except which AOL low-risk and Oncotype DX low-risk	DES includes (1) recurrence-free; (2) distant recurrence and (3) death
Kondo (2011)[230]	Markov	Unclear - appears to be 1-year	Reclassification based on use of assay	Yes	Half of cases with no definitive indication undergo adjuvant chemotherapy and only cases with high RS undergo chemotherapy after the use of the assay based on the results of Japanese validation study	5 states: (1) ER+, ESBC after adjuvant therapy, (2) distant recurrence with response to treatment, (3) distant recurrence with no response to treatment, (4) progression of disease after distant recurrence and (5) death.
Lamond (2012)[231]	Markov	1-month	Classification to LR, IR, HR	Yes - only in low risk and high risk	For no test, based on Canadian population based study; for test, based on RS score. Usage in intermediate group assumed to be the	10 states: (1) chemotherapy; (2) CINV; (3) FN; (4) disease-free; (5) local relapse; (6) distant relapse; (7) treated local relapse; (8)

Author	Model approach	Cycle length	Model type	Does model claim predictive benefit for test?	Assumptions on chemotherapy use	Long-term health states
					same in both groups	AML/MDS; (9) CHF; (10) dead.
Paulden (2013)[232]	Markov	Appears to be monthly	Reclassification based on use of assay	Yes	Different regimens assumed for different risk groups. Different proportions of patients assumed to receive chemotherapy according to risk group (estimated by linear regression).	5 states: (1) risk classification; (2) adjuvant chemotherapy; (3) no distant recurrence; (4) distant recurrence; (5) dead.
Reed (2013)[233]	Markov	6-months	Classification based on RS	Yes - different RRR assumed in each risk group	No LR get chemotherapy, all IR and HR get chemotherapy	3 states: (1) disease-free; (2) distant recurrence; (3) dead.
Blank (2015) [226]	Markov	1-year	Based on sensitivity and specificity of test/guideline	No - same treatment effect applied to all groups irrespective of risk	No chemotherapy for low risk patients	3 states: (1) disease-free; (2) distant recurrence; (3) dead. LR modelled implicitly
Bonastre (2014)[234]	EEACT with partitioned survival	Unclear	Unclear	No – authors state there is no evidence to support predictive benefit for MammaPrint	For MammaPrint and AOL, only high risk patients were assumed to receive chemotherapy. For the all chemotherapy comparator, all patients receive chemotherapy irrespective of risk	4 states: (1) post-surgery with chemotherapy [disease-free]; (2) first year post-surgery without chemotherapy [disease-free]; (3) distant recurrence-free survival, and; (4) dead.
Retel (2012a)[235]	Markov	1-year	Based on sensitivity and specificity of test/guideline	Unclear	Chemotherapy used only in high-risk according to treatment guidelines	4 health states: (1) disease-free survival; (2) relapse (including local and regional recurrences, secondary primary and contralateral breast cancer); (3) metastasis, and; (4) dead.
Retel (2012b)[236]	Markov	1-year	Based on sensitivity and specificity of test/guideline	Unclear	High and intermediate groups combined - both assumed to receive	4 health states: (1) disease-free survival; (2) relapse (including local

Author	Model approach	Cycle length	Model type	Does model claim predictive benefit for test?	Assumptions on chemotherapy use	Long-term health states
					ET+CT	and regional recurrences, secondary primary and contralateral breast cancer); (3) metastasis, and; (4) dead.
Retel (2013a)[237]	Markov	1-year	Based on sensitivity and specificity of test/guideline	Unclear	Chemotherapy used only in high-risk according to treatment guidelines	Not reported but based on previous 4-state model reported by Retel *et al* (see above)
Retel (2013b)[238]	Markov	Not reported, but likely to be 1-year	Based on sensitivity and specificity of test/guideline	Unclear	Not reported but likely to be same as other Retel studies	4 health states: (1) disease-free survival; (2) relapse (including local and regional recurrences, secondary primary and contralateral breast cancer); (3) metastasis, and; (4) dead.
Hall (2012)[239]	Decision tree and modified Markov model	Not reported	Classification to LR or HR	Unclear - data contained within the appendices appear to suggest predictive benefit is modelled	All high risk patients receive chemotherapy	6 health states: (1) disease-free; (2) distant recurrence; (3) local recurrence; (4) disease-free after local recurrence; (5) CHF; (6) dead.
Hannouf (2012)[240]	Markov	1-month	Classification to LR, IR, HR with separate Markov nodes for chemotherapy+ET versus ET alone (accounting for chemotherapy-related AEs)	Unclear – appears to assume predictive benefit	Model assumes 50% IR patients receive chemotherapy	ET only model - 5 states: (1) remission; (2) local recurrence; (3) distant recurrence; (4) dead. CT+ET model - 5 states: (1) remission with chemotherapy SAEs; (2) remission without chemotherapy SAEs; (3) local recurrence; (4) distant recurrence; (5) dead.

Author	Model approach	Cycle length	Model type	Does model claim predictive benefit for test?	Assumptions on chemotherapy use	Long-term health states
Hannouf (2014)[241]	Markov	1-month	Classification to LR, IR, HR with separate Markov nodes for chemotherapy+ET versus ET alone (accounting for chemotherapy-related AEs)	Unclear – appears to assume predictive benefit	Model assumes 50% IR patients receive chemotherapy	ET only model - 5 states: (1) remission; (2) local recurrence; (3) distant recurrence; (4) dead. CT+ET model - 5 states: (1) remission with chemotherapy SAEs; (2) remission without chemotherapy SAEs; (3) local recurrence; (4) distant recurrence; (5) dead.
Kondo (2012)[242]	Markov	1-year	Classification to LR, HR	No	Chemotherapy applied to HR, ET only for low risk	5 states: (1) ER+, LN0, HER2-early state breast cancer after adjuvant chemotherapy; (2) distant recurrence responded to treatment; (3) distant recurrence not responded to treatment; (4) progression of disease after distant recurrence; (5) dead.
Mislick (2014)[243]	Markov	1-year	Classification to LR, IR, HR	Yes - for both Mammostrat and Oncotype	80% HR assumed to receive chemo; 10% LR assumed to receive chemotherapy; 50% IR assumed to receive chemotherapy	3 states: (1) no recurrence; (2) recurrence; (3) dead.
Stein (2016)[244]	Decision tree and modified Markov model	1-year	Classification to LR or HR.	Separate analyses undertaken including predictive benefit and assuming constant benefit across risk groups	All high risk patients receive chemotherapy	7 health states: (1) disease-free; (2) distant recurrence; (3) local recurrence; (4) disease-free after local recurrence; (5) CHF;

Author	Model approach	Cycle length	Model type	Does model claim predictive benefit for test?	Assumptions on chemotherapy use	Long-term health states
						(6) Chronic myeloid leukaemia; (7) dead.
Tiwana (2013) [245]	Not reported - appears to be Markov	Not reported	Classification to low-low risk, low-intermediate risk, low-high risk, low-none risk, intermediate-low risk, intermediate-intermediate risk intermediate-high risk, intermediate-none risk, high-low risk, high-intermediate risk, high-high risk or high-none risk. Based on a model constructed for the Ontario Heath Technology Assessment Committee	Yes - different recurrence rates modelled between groups and tests	Based on usage reported in Asad *et al*[251]	Not reported - appears to include 3 states: (1) relapse-free; (2) distant metastases; (3) dead.
Vanderlaan (2011) [246]	Appears to be Markov	Not reported	Classification to LR or HR. Original source model provided by Cedar Associates based in California USA	No - same recurrence rates for all HR patients	71% of women in usual care assumed to receive chemotherapy treatment	3 states: (1) non-progressed disease; (2) progressed disease; (3) death.
Wong (2012)[247]	Decision tree with partitioned survival approach to determine sojourn time	Not reported	For patients whose treatment decision was based on US NCCN criteria classification to LR or HR. For patients whose treatment was based on the Oncotype DX test results classification to LR, IR or HR	Yes – different treatment effects applied for each risk category	~55% women assumed to receive chemotherapy	Not clearly reported - appears to be 3 states: (1) disease-free; (2) relapsed; (3) dead.
Ward (2013)[18]	Markov	6-months	Classification to risk/prognosis group	No	Baseline chemotherapy use (without test) based on English cancer registry data. Use of chemotherapy conditional on test based on unpublished data	4 states: (1) recurrence-free; (2) distant recurrence; (3) long-term AEs (AML); (4) dead. Local recurrence included as event
Yang (2012)[249]	Markov	1-year	Classification to LR or HR using AOL and reclassification probabilities from the literature	Yes - different risk reductions applied between HR and LR	90% patients who were high risk according to both AOL and Oncotype DX/MammaPrint	3 states: (1) no recurrence; (2) recurrence; (3) dead.

Author	Model approach	Cycle length	Model type	Does model claim predictive benefit for test?	Assumptions on chemotherapy use	Long-term health states
					received chemotherapy, 90% of patients who were at low risk according to both AOL and Oncotype DX/MammaPrint did not receive chemotherapy. For patients who experienced a conflicting result between AOL and Oncotype DX/MammaPrint, 50% of the subpopulation received chemotherapy.	
Yamauchi (2014)[250]	Markov	Unclear - appears to be 1-year	Classification to LR, IR, HR	Yes - different risk reductions applied between risk groups	Based on empirical study (Yamauchi *et al* 2013[252])	3 states: (1) no recurrence; (2) recurrence; (3) dead.

LR – low risk; IR – intermediate-risk; HR – high-risk; RRR – relative risk reduction; RFS – recurrence-free survival; AML – acute myeloid leukaemia; MDS – myelodysplastic syndromes; CHF – congestive heart failure; CINV - chemotherapy-induced nausea and vomiting; FN – febrile neutropaenia; CT – chemotherapy; ET- endocrine therapy; RS – recurrence score

5.2 Review and critical appraisal of economic analyses provided by test manufacturers

Economic analyses were provided by the manufacturers of Oncotype DX (Genomic Health) and MammaPrint (Agendia) and the chief investigator of the EndoPredict (Myriad) decision impact study.[76, 113, 121, 225] The fully executable health economic models developed for the analyses of Oncotype DX and MammaPrint were made available to the EAG; the model referred to in the draft EndoPredict cost-effectiveness paper was not provided to the EAG. These three analyses are detailed and critically appraised in the following sections.

5.2.1 Agendia cost-effectiveness report – MammaPrint versus current practice[121]

Agendia provided a short unpublished paper detailing the methods and results of a *de novo* health economic evaluation of MammaPrint versus current practice for informing adjuvant chemotherapy decisions in women with early breast cancer in the UK.[121] The model is based principally on an analysis of the 5-year outcomes of the MINDACT trial.[134] The fully executable economic model used to undertake the analysis was made available to the EAG for review.

Agendia model scope

The Agendia model evaluates the cost-effectiveness of MammaPrint versus (modified) AOL or NPI over a 5-year time horizon from the perspective of the NHS. Cost-effectiveness is evaluated across the MINDACT ITT population and within clinical high-risk subgroups. The economic comparisons presented within the Agendia report are detailed in Table 103. Cost-effectiveness is expressed in terms of the incremental cost per quality-adjusted life year (QALY) gained for MammaPrint versus current practice. Health outcomes and costs were discounted at a rate of 3.5% per annum. A formal price year is not reported within the Agendia cost-effectiveness report.

Table 103: Economic comparisons presented in the Agendia cost-effectiveness report

Economic comparison	Population	Comparator	Definition of relapse
Analysis 1	Whole trial population	AOL	DMFS
Analysis 2	Whole trial population	AOL	DFS
Analysis 3	Clinical high-risk subgroup	AOL	DMFS
Analysis 4	Clinical high-risk ER+/HER2-/LN0 subgroup	AOL	DMFS
Analysis 5	Whole trial population	NPI	DMFS
Analysis 6	Clinical high-risk subgroup	NPI	DMFS
Analysis 7	Clinical high-risk ER+/HER2-/LN0 subgroup	NPI	DMFS
Analysis 8	UK subgroup	NPI	DMFS

DMFS – distant metastasis-free survival; DFS – disease-free survival

Agendia model structure

The Agendia model adopts a hybrid decision tree and Markov approach (see Figure 5 and Figure **6**). The decision tree component divides the total population into four sub-populations that are defined according to the patient's risk as determined by clinical practice and the MammaPrint test: (i) clinical low-risk, genomic low-risk; (ii) clinical low-risk, genomic high-risk; (iii) clinical high-risk, genomic low-risk, and (iv) clinical high-risk, genomic high-risk. Within each of the four sub-populations, the model assumes that adjuvant chemotherapy treatment decisions are determined exclusively by the test or by usual practice: patients who are deemed to be low-risk are assumed to not receive chemotherapy, whilst all patients who are deemed to be high-risk are assumed to receive chemotherapy. The differences in outcomes and costs between the groups are determined by different choices regarding the use of adjuvant chemotherapy in the sub-populations in which the two tests produce discordant results; costs and outcomes for concordant groups are assumed to be the same and therefore cancel out (see Table 104). Thus, with reference to Figure 5, the Agendia model compares MammaPrint Groups 1, 3, 5 and 6 against current practice Groups 1, 2, 4 and 6. A Markov node is attached to each of the four sub-populations; these Markov nodes are used to model clinical outcomes and costs over a 5-year horizon using an annual cycle length. A half-cycle correction is not applied. The Markov component of the model includes three health states: (i) relapse-free; (ii) distant metastases, and (iii) dead (see Figure 6). Health utilities are assumed to differ according to the presence/absence of distant metastases; an additional disutility associated with adjuvant chemotherapy is applied for 2 years from model entry. The health utilities applied within the model are not adjusted by age.

The treatment pathway for patients who are deemed to be low-risk (either by the MammaPrint test or according to clinical practice) is assumed to include monitoring, endocrine therapy, treatments for distant recurrence and end-of-life care. The treatment pathway for patients who are deemed to be high-risk (either by the MammaPrint test or usual clinical practice) is assumed to include adjuvant chemotherapy, trastuzumab (for a proportion of patients), G-CSF (for a proportion of patients) for the secondary prevention of febrile neutropenia, monitoring, endocrine therapy, treatments for distant recurrence and end-of-life care. The model also includes the costs associated with AML and CHF. The costs of the test are also included in the intervention group.

Figure 5: Agendia model structure – decision tree (reproduced from the Agendia model)

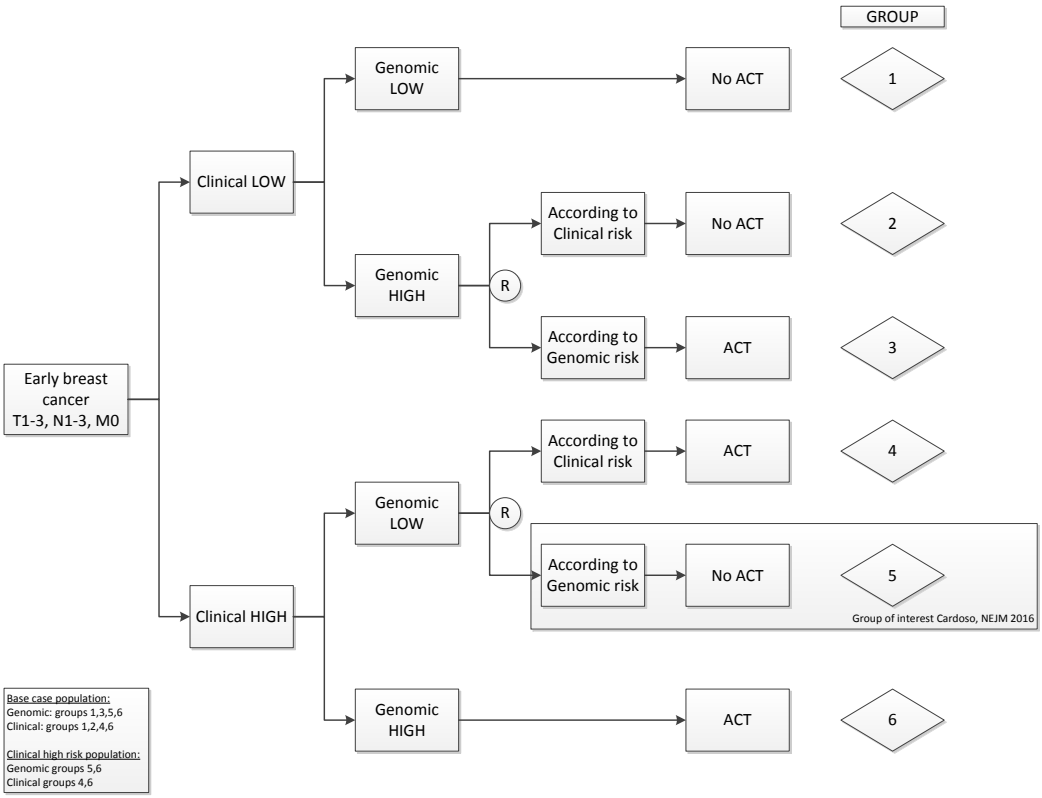


Figure 6: Agendia model structure – Markov component (reproduced from the Agendia model)

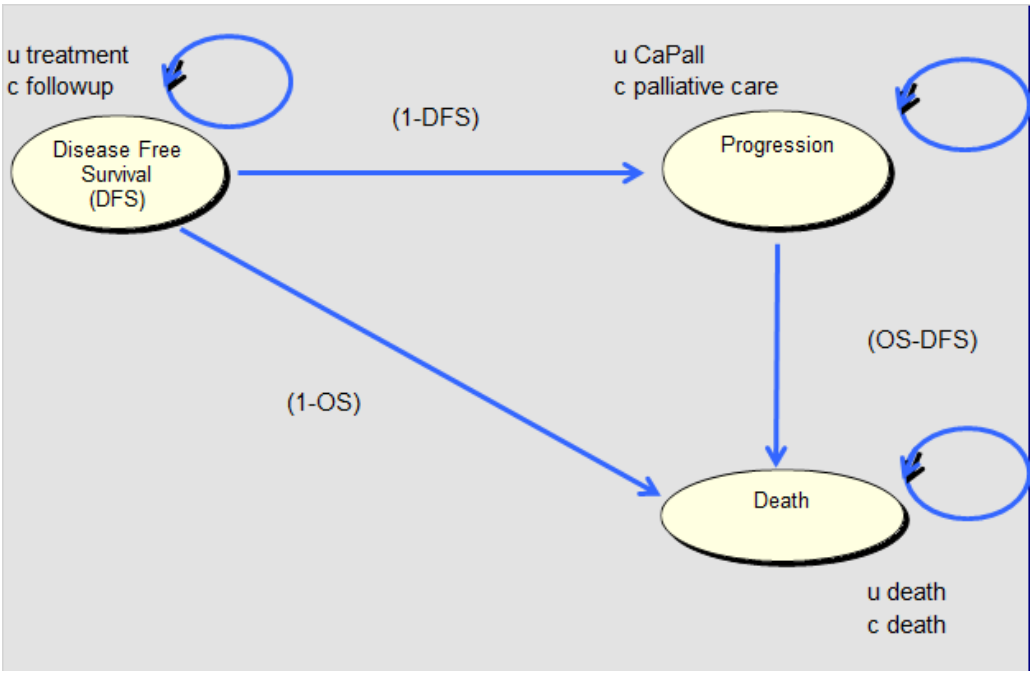


Table 104: Use of MINDACT trial subgroups in the Agendia model according to clinical and genomic risk

Clinical and genomic risk	Clinical practice	Genomic test
(1) Clinical low-risk, genomic low-risk	MINDACT C_low, G_low concordant group No chemotherapy (n=2,634)	
(2) Clinical low-risk, genomic high-risk	MINDACT C_low, G_high discordant group No chemotherapy (n=344)	MINDACT C_low, G_high discordant group Chemotherapy (n=346)
(3) Clinical high-risk, genomic low-risk	MINDACT C_high, G_low discordant group Chemotherapy (n=749)	MINDACT discordant high clinical risk group No chemotherapy group (n=748)
(4) Clinical high-risk, genomic high-risk	MINDACT C_high, G_high concordant group Chemotherapy (n=1,872)	

C_ - clinical; G_ - genomic

The Agendia model makes the following structural assumptions:

- The use of adjuvant chemotherapy (and other adjunctive treatments) is assumed to be determined solely by the results of the MammaPrint test/current practice (i.e. low-risk patients do not receive chemotherapy, whilst all high-risk patients receive chemotherapy)
- The model includes the possibility of only one relapse.
- Within all but one analyses, recurrence is based on DMFS rather than DFS.
- Health utilities are determined principally by the presence/absence of disease recurrence (distant metastases). An additional HRQoL decrement is applied only to patients receiving adjuvant chemotherapy for a period of 2 years.
- The costs and health impacts of local recurrence are not included in the model.
- All patients are assumed to receive endocrine therapy irrespective of clinical or genomic risk (although the EAG notes that a small proportion of patients in the MINDACT trial had ER-disease and therefore would not receive endocrine therapy).
- A proportion of patients receiving chemotherapy are assumed to develop a second primary tumour and a proportion of patients may develop CHF.
- Twenty five percent of patients receiving adjuvant chemotherapy also receive G-CSF for the secondary prevention of febrile neutropenia.

Evidence sources used to inform the Agendia model

The evidence sources used to inform the Agendia model are summarised in Table 105. As shown in the table, the majority of input parameters were derived from analyses of the MINDACT trial.[134] Additional evidence sources include the earlier HTA report by Ward *et al*[18] and the NICE Single Technology Appraisal of azacitidine for the treatment of myelodysplastic syndomes.[253]

Table 105: Evidence sources used in the Agendia model

Parameter	Source	EAG comments
Transition probabilities (DMFS/DFS and OS)	MINDACT[134]	Incorrect calculation of transition probabilities using DMFS and OS
Health utilities	*Relapse-free:* EQ-5D data collected within the MINDACT trial[134] *Distant metastases:* Ward *et al*[18] *Disutility for adjuvant chemotherapy:* Ward *et al*[18]	Minor discrepancy in disutility reported in Ward *et al*[18] and value used in model (0.037 versus 0.038). Disutility applied for 2 years.
Probability second primary tumour	MINDACT[134]	-
Probability patient receives trastuzumab	MINDACT[134]	-
Probability AML	MINDACT[134]	-
Probability CHF	MINDACT[134]	-
MammaPrint cost	Agendia BV[121]	-
AML cost	STA218[253]	-
All other costs	Ward *et al*[18]	Not uplifted to current price year

EQ-5D – Euroqol 5-Dimension

Results of the Agendia model

The deterministic results of the Agendia model are presented in Table **106**; these have been generated from the Agendia model made available to the EAG. Based on the deterministic version of the company's model, within the total MINDACT population (Analysis 1), the incremental cost-effectiveness ratio (ICER) for MammaPrint versus current practice (AOL) is estimated to be £369,397 per QALY gained. When the recurrence-free interval is determined according to DFS rather than DMFS (Analysis 2), the ICER is estimated to be £503,446 per QALY gained. Within both of these analyses, MammaPrint is associated with a small decrement in OS and a small gain in QALYs (due to less chemotherapy use) as well as an increase in costs (due to the test). Within the subgroup of patients with high clinical risk defined by AOL (Analysis 3), the model suggests that MammaPrint dominates current practice; the same conclusion was found for the subgroup of AOL high-risk patients with ER+/HER2- disease (Analysis 4). Within the comparisons of MammaPrint versus NPI (Analyses 5-8), the Agendia model suggests that MammaPrint dominates current practice. It should be noted that the EAG has concerns regarding the validity of these analyses due to the presence of programming errors in the model.

Table 106: Results of the Agendia model – MammaPrint versus clinical practice (deterministic)

Option	LYGs	QALYs	Costs	Inc. LYGs	Inc. QALYs	Inc. Costs	ICER (per QALY gained)
Analyses 1-4: MammaPrint versus AOL							
Analysis 1: DMFS total population							
MammaPrint	4.85	4.04	£14,497	-0.007	0.004	£1,512	**£369,397**
AO	4.85	4.03	£12,985	-	-	-	-
Analysis 2: DFS total population							
MammaPrint	4.85	3.90	£19,952	-0.007	0.003	£1,584	**£503,446**
AO	4.85	3.89	£18,368	-	-	-	-
Analysis 3: High clinical risk subgroup							
MammaPrint	4.79	3.88	£19,051	-0.002	0.029	-£907	**Dominating**
AO	4.79	3.85	£19,959	-	-	-	-
Analysis 4: High clinical risk ER+/HER2- subgroup							
MammaPrint	4.84	3.94	£14,243	-0.001	0.047	-£200	**Dominating**
AO	4.84	3.90	£14,444	-	-	-	-
Analyses 5-8: MammaPrint versus NPI							
Analysis 5: DMFS total population							
MammaPrint	4.86	4.04	£12,594	0.013	0.002	-£1,856	**Dominating**
NPI	4.85	4.04	£14,450	-	-	-	-
Analysis 6: DMFS-clinical high NPI (>3.4)							
MammaPrint	4.78	3.87	£19,236	0.039	0.066	-£1,038	**Dominating**
NPI	4.74	3.81	£20,274	-	-	-	-
Analysis 7: DMFS-clinical high NPI in ER+/HER2-							
MammaPrint	4.71	3.80	£18,777	0.079	0.112	-£2,367	**Dominating**
NPI	4.63	3.69	£21,144	-	-	-	-
Analysis 8: DMFS-clinical high NPI (>3.4) in UK population only (n=66)							
MammaPrint	4.81	3.90	£19,486	0.003	0.027	-£446	**Dominating**
NPI	4.81	3.87	£19,932	-	-	-	-

Inc. – incremental

Critical appraisal of the Agendia model

Box **1** summarises the main issues identified by the EAG's critical appraisal of the Agendia model. These concerns are discussed in more detailed below.

Box 1: Main issues relating to the Agendia model identified by the EAG

(1) Incorrect calculation of transition probabilities for all analyses
(2) Questionable assumption that risk exclusively determines whether patients receive adjuvant chemotherapy
(3) Use of potentially outdated cost estimates
(4) Short time horizon
(5) Potential bias in the redefinition of clinical risk by NPI
(6) Disutility associated with chemotherapy applied for 2 years
(7) Uncertainty surrounding UK clinical high risk analysis
(8) Other minor implementation issues

(1) Incorrect calculation of transition probabilities for all analyses

The Agendia model takes the form of a Markov model whereby the DMFS and OS curves from the MINDACT trial are used to estimate transition probabilities between the relapse-free, distant metastases and death states. However, the implementation of the model is subject to a substantial error that appears to derive from a misinterpretation of the Kaplan-Meier time-to-event curves (see Figure **7** and Figure **8**). Within the model, the transition probabilities between the health states are calculated by converting the annual cumulative survival probabilities (readings from various timepoints on the relevant Kaplan-Meier curves) for DMFS and OS to cumulative event probabilities (one minus the cumulative survival estimate at each timepoint); these are used to estimate the incidence of distant metastases and death. These cumulative event probabilities are then treated as annual event probabilities which are applied to the surviving cohort during each successive Markov cycle. However, the cumulative event probabilities derived from the Kaplan-Meier curves for DMFS and OS used in the model describe the probability of having experienced the relevant event(s) by time t_n rather than the probability of experiencing the event(s) at time t conditional on having not experienced the event at the previous timepoint t_{n-1}. As a consequence, the modelled health state populations are very different from the Kaplan-Meier curves used to inform them: given the adopted approach, which does not include any parametric curve-fitting, the model should replicate the observed cumulative survival probabilities exactly. The EAG considers this to be a fundamental problem which invalidates the results of the economic analyses contained within the Agendia model and accompanying cost-effectiveness report.

Figure 7: Comparison of observed and model-predicted values – clinical-high, genomic-low, DMFS

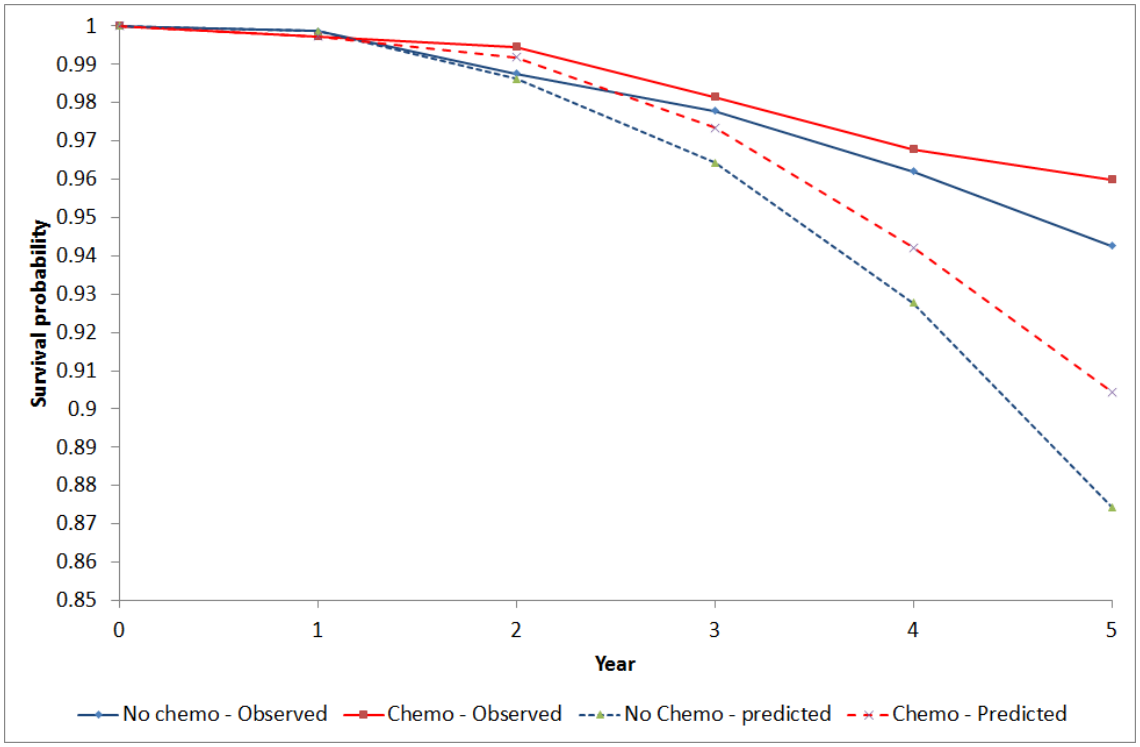
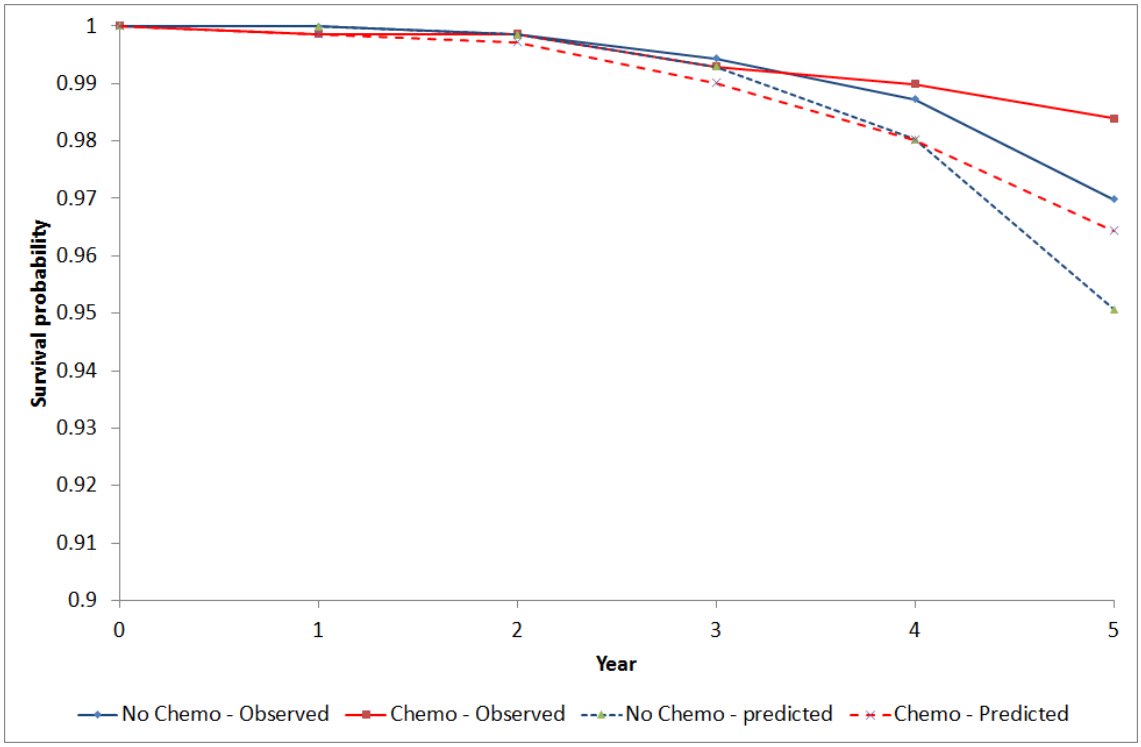


Figure 8: Comparison of observed and model-predicted values – clinical-high, genomic-low, OS



(2) Questionable assumption that risk exclusively determines treatment pathway

The Agendia model assumes that all patients who are deemed to be high-risk according to the MammaPrint test or clinical practice will receive chemotherapy, whilst none of the patients who are deemed to be low-risk are assumed to receive chemotherapy. This is unlikely to be a reasonable assumption as it implies that the decision to receive adjuvant chemotherapy is exclusively determined by the risk classification determined by the test or AOL. This is unlikely to reflect clinical practice in England whereby other factors may impact on the proportion of patients ultimately receiving adjuvant chemotherapy (for example, patient choice, patient fitness/co-morbidity, cultural attitudes, perceived clinician risk and expected survival gain). This view is reflected in the NCRAS dataset[254] and the UKBCG survey (see Section 5.3). The EAG considers that this assumption makes the interpretation of the Agendia analysis problematic.

(3) Use of potential outdated cost estimates

With the exceptions of the cost of the MammaPrint test which was sourced from the manufacturer, and the cost of treating AML which was derived from NICE STA218, all other resources and cost parameters were taken directly from Ward *et al*[18] without being uplifted to current prices.

(4) Short time horizon

The model adopts a 5-year time horizon and does not include any extrapolation of the available trial data beyond the observed period of the MINDACT trial[134] (although the EAG notes that any attempt to extrapolate may be hindered by the low numbers of events in the concordant and discordant groups). It is therefore likely that the company's model does not reflect all differences in health outcomes and costs between MammaPrint and standard care over a patient's lifetime. The impact of this issue is unclear.

(5) Potential bias in the redefinition of clinical risk by NPI

The Agendia model includes four sets of analyses in which current practice is assumed to be defined by the use of NPI rather than AOL (see Table 106, Analyses 5 to 8). The Agendia cost-effectiveness report does not provide any details on how this redefinition of clinical risk was undertaken. Given that within the MINDACT trial,[134] current practice was defined by a modified version of AOL, the redefinition of clinical risk by NPI breaks randomisation and creates an imbalance between the discordant clinical and genomic risk groups. As shown in Table 107, the redefinition of clinical risk by NPI changes the numbers of patients who receive chemotherapy in both the test and current practice groups, and increases the total number of patients in the MammaPrint group and reduces the total number of patients in the current practice group. The EAG considers that this redefinition of risk may produce bias in the company's results, although the magnitude and direction of this is unclear.

Table 107: Number of discordant risk patients allocated to chemotherapy/no chemotherapy in AO! and NPI analyses

	Number receiving chemotherapy/no chemotherapy	
	Clinical risk defined by AOL (as per MINDACT)	**Clinical risk redefined by NPI (Analyses 5-8)**
MammaPrint subgroups		
Clinical high; Genomic low – receive chemotherapy	748 (68%)	690 (61%)
Clinical low; Genomic high – no chemotherapy	344 (32%)	447 (39%)
Total population	1,092 (100%)	1,137 (100%)
Current practice subgroups		
Clinical high; Genomic low – receive chemotherapy	749 (68%)	606 (67%)
Clinical low; Genomic high – no chemotherapy	346 (32%)	300 (33%)
Total population	1,095 (100%)	906 (100%)

(6) Disutility associated with chemotherapy applied for 2 years

The Agendia model applies a disutility associated with chemotherapy for the first two annual cycles. Given that adjuvant chemotherapy is typically given for a period of 4-5 months, and significant long-term toxicity affects only a small minority of patients, the EAG considers that this is likely to represents a pessimistic assumption, which will produce a bias in favour of MammaPrint.

(7) Uncertainty surrounding UK clinical high-risk analysis

According to the Agendia cost-effectiveness report,[121] the UK-based clinical high risk analysis (see Table 106, Analysis 8) is reported to include only 66 patients. However, the model analysis includes only 49 patients of whom only 19 patients have discordant results (current practice group n=9, MammaPrint group n=10). Notwithstanding the other concerns raised by the EAG regarding the potential confounding in the Agendia NPI-based analyses, the EAG considers this analysis to be subject to considerable uncertainty.

(8) Other minor implementation issues

The Agendia model includes further less important implementation issues. The Markov trace is not half-cycle corrected, although the EAG does not consider that this will have a marked impact on the results of the analysis. In addition, the number of patients in the discordant sub-populations is modelled to reflect the populations of patients randomised to receive chemotherapy or no chemotherapy within these groups in the MINDACT trial.[134] As the randomisation procedure in the trial did not produce an equal number of patients receiving chemotherapy or no chemotherapy in either discordant group, the Agendia model also includes this slight imbalance. This reflects an artefact of the trial randomisation and recruitment procedures rather than a true difference between the

proportions of patients in each group. Given that the differences are very small, this will not have a major impact on the model results.

Corrected results for the Agendia MammaPrint model

Table 108 presents the results of the Agendia model including the correction of the errors in the formulae used to estimate health state occupancy. The EAG's corrected version of the model is a partitioned survival model in which the probability of being alive and free from distant recurrence at each time *t* is determined by the observed DMFS curves, the probability of being alive at time *t* is determined by the observed OS curves, and the probability of being alive post-recurrence is given by the difference between these two curves. As can be seen from the comparison of the uncorrected and corrected results presented in Table **106** and Table 108, respectively, this correction has a substantial impact on the results of several of the analyses. Based on the EAG's corrected model, within the total MINDACT population (Analysis 1), the ICER for MammaPrint is estimated to be £185,484 per QALY gained. When the probability of recurrence is based on DFS rather than DMFS (Analysis 2), the ICER is estimated to be £141,796 per QALY gained. Both these ICERs are considerably lower than those presented within the Agendia cost-effectiveness report. Within the total population in which clinical practice is assumed to be based on NPI (Analysis 5), the corrected ICER for MammaPrint is estimated to be £325,768 saved per QALY lost (a South-West quadrant ICER). MammaPrint remains dominant in the remaining analyses. As noted above, the interpretation of these corrected results remains problematic due to the assumption that treatment is determined exclusively by the test or AOL/NPI and the short time horizon.

Table 108: Corrected results of the Agendia model – MammaPrint versus current practice (deterministic)

Option	LYGs	QALYs	Costs	Inc. LYGs	Inc. QALYs	Inc. Costs	ICER (per QALY gained)
Analyses 1-4: MammaPrint versus AOL							
Analysis 1: DMFS total population							
MammaPrint	4.92	4.12	£13,926	-0.004	0.008	£1,477	£185,484
AO	4.93	4.11	£12,449	-	-	-	-
Analysis 2: DFS total population							
MammaPrint	4.92	4.06	£16,450	-0.004	0.010	£1,422	£141,796
AO	4.93	4.05	£15,029	-	-	-	-
Analysis 3: High clinical risk subgroup							
MammaPrint	4.89	4.00	£18,063	-0.004	0.030	-£1,013	Dominating
AO	4.89	3.97	£19,076	-	-	-	-
Analysis 4: High clinical risk ER+/HER2- subgroup							
MammaPrint	4.91	4.03	£13,303	-0.004	0.046	-£281	Dominating
AO	4.92	3.99	£13,584	-	-	-	-

Analyses 5-8: MammaPrint versus NPI							
Analysis 5: DMFS total population							
MammaPrint	4.93	4.12	£12,066	0.006	-0.006	-£1,810	£325,768*
NPI	4.93	4.12	£13,875	-	-	-	-
Analysis 6: DMFS clinical high NPI (>3.4)							
MammaPrint	4.89	3.99	£18,205	0.019	0.045	-£966	Dominating
NPI	4.87	3.95	£19,170	-	-	-	-
Analysis 7: DMFS clinical high NPI in ER+/HER2-							
MammaPrint	4.86	3.97	£17,361	0.032	0.070	-£2,383	Dominating
NPI	4.83	3.90	£19,744	-	-	-	-
Analysis 8: DMFS clinical high NPI (>3.4) in UK population only (n=66)							
MammaPrint	4.90	4.00	£18,548	-0.001	0.027	-£576	Dominating
NPI	4.90	3.98	£19,125	-	-	-	-

Inc. - incremental
* South-West quadrant ICER

5.2.2 Genomic Health dossier - Oncotype DX versus current practice

The Genomic Health dossier made available to NICE and the EAG includes a cost-effectiveness report detailing the methods and results of a *de novo* health economic evaluation of Oncotype DX versus current practice for early breast cancer in the UK.[113] The fully executable economic model was also made available to the EAG for scrutiny.

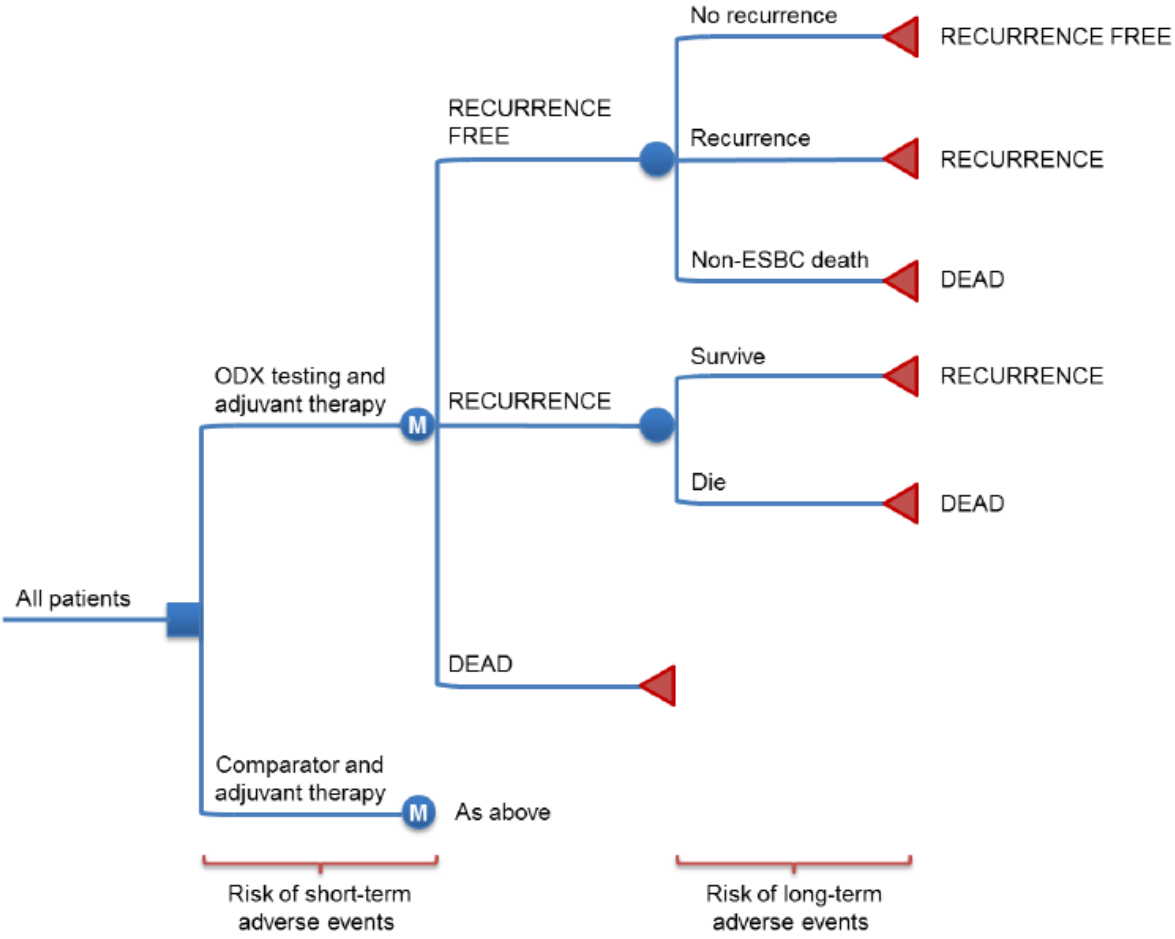
Genomic Health model scope

According to the Genomic Health dossier,[113] the model was based on the previous analysis reported by Ward *et al.*[18] The company's base case model evaluates the cost-effectiveness of Oncotype DX versus current practice in ER+ LN0 patients. The model also allows for the evaluation of Oncotype DX versus MammaPrint, EndoPredict and Prosigna as secondary analyses. The base case includes ER+, LN0 early breast cancer patients, with the option to evaluate LN+ patients as a secondary analysis. Cost-effectiveness results are expressed in terms of the incremental cost per QALY gained. Health outcomes and costs are discounted at a rate of 3.5% per annum. Costs are valued at 2016 prices and reflect an NHS and Personal Social Services (PSS) perspective.

Genomic Health model structure

The company's model is illustrated in Figure 9. The company's model is referred to as a Markov model in the Genomic Health dossier, but is more accurately described as a hybrid decision tree - Markov model. The decision tree portion of the model incorporates the decision to give adjuvant chemotherapy or not. Within the Oncotype DX group, this probability is driven by the Oncotype DX RS, whilst in the comparator group, this probability is driven by current clinical practice as recorded in the pre-Oncotype DX chemotherapy decision in the NHS England Oncotype DX Access Scheme Dataset.[255]

Figure 9: Genomic Health model structure (reproduced from Genomic Health dossier)



The Markov component of the model includes three health states: (i) recurrence free; (ii) distant recurrence, and (iii) dead. The model adopts a 30-year time horizon and a 6-month cycle length. The age of patients upon entry into the model 58.9 years, based on the mean age of patients in the NHS England Access Scheme Dataset.[255] Patients can die from breast cancer or from other causes. The model assumes that Oncotype DX is predictive of chemotherapy benefit, hence different treatment effects are applied according to the low, intermediate and high RS groups (applied to the Oncotype DX group only). Health utilities are assigned to the recurrence-free and distant recurrence states. A chemotherapy-related disutility is applied during each cycle for those who receive chemotherapy in either the test or no test group. A further QALY loss is applied for women who experience local recurrence. Separate health utility values are applied for patients who develop AML and for those in the final three months of life prior to death due to breast cancer. As the model does not contain separate health states for these two states, the health utility values for patients in the recurrence-free and distant recurrence health states are adjusted to account for the lower health utility values for patients with AML and for those dying from breast cancer.

The costs used in the Genomic Health model were based on Ward *et al*; these were uplifted to current values using the Hospital and Community Health Services (HCHS) pay and prices inflation index.[256] According to the Genomic Health dossier, all patients are assumed to receive endocrine therapy, based on the following assumptions:

- Tamoxifen for 5 years (40% of patients)
- Anastrozole for 5 years (20% of patients)
- Letrozole for 5 years (20% of patients)
- Tamoxifen for 2 years plus exemestane for 3 years (20% of patients)
- Tamoxifen for 5 years followed by letrozole for a further 3 years (half of patients completing tamoxifen for 5 years received an additional 3 years of letrozole, 10% of patients)

The model assumes that adjuvant chemotherapy consists of six cycles of FEC75 (5 fluorouracil [5-FU], epirubicin and cyclophosphamide). The cost of chemotherapy is included as a once-only cost and includes the costs of drug acquisition, administration, monitoring, and an echocardiogram for 25% patients undergoing chemotherapy (total chemotherapy cost=£4,678). The model includes costs associated with the following short-term AEs: anaemia (1.4%), thrombocytopenia (0.3%), neutropenic infection (1.6%), nausea/vomiting (24.2%) and stomatitis (4%). This cost is applied as a once-only cost of £315 for women receiving adjuvant chemotherapy. A proportion of women receiving chemotherapy (0.46%) are assumed to subsequently develop AML: this is included as a once-only cost of £13,123. Half of the annual cost of distant recurrence (£9,316/2) is applied to patients in the distant recurrence health state during each cycle. A once-only cost (£16,127) of treating local recurrence is applied to 10.5% of patients entering the distant recurrence state. The model also includes a cost of £4,608 to reflect end-of-life costs for women who die due to their breast cancer.

The Genomic Health model makes the following structural assumptions:

- The results of the Oncotype DX test are assumed to be predictive of the benefit deriving from subsequent chemotherapy use. Conversely, a common relative risk of recurrence is applied to all patients in the current practice group.
- Survival following distant recurrence is assumed to be 3.3 years (based on Thomas *et al* 2009[257])
- An HRQoL decrement associated with AEs is applied during every model cycle for the remaining lifetime of patients who receive adjuvant chemotherapy
- The costs of short-term AEs are included only in the first model cycle
- AML is included as a long-term complication of chemotherapy
- All patients are assumed to receive endocrine therapy.

Evidence sources used to inform the Genomic Health model

Table 109 summarises the evidence sources used to inform the Genomic Health model.

Table 109: Evidence sources used in the Genomic Health model

Parameter	Source	EAG comments
10-year risk of distant recurrence on endocrine therapy	Dowsett *et al*[35]	Figures presented in the Genomic Health dossier and model do not match the figures in the Dowsett *et al* paper
Oncotype DX RS classification	NHS England Access Scheme dataset[255]	Risk classification probabilities and risk of distant recurrence are not derived from the same source. This may produce a bias due to differences in the distribution of prognostic variables for patients within each RS category between the two sources. Risk reclassification is applied incorrectly in the model.
Relative risk reduction associated with chemotherapy	Paik *et al*[49]	This is applied incorrectly within the standard care group in a way which suggests that the same patient receiving the same treatment accrues a different level of benefit if they are tested with Oncotype DX.
Probability patient receives chemotherapy	NHS England Access Scheme dataset[255]	This source reflects the LN0 'intermediate-risk' group only
Health utilities	Ward *et al*[18]	Health losses due to chemotherapy-related AEs are applied incorrectly
Probability of short-term AEs during the first 6 months	Ward *et al*[18]	-
Probability of local recurrence	Ward *et al*[18]	-
Probability of AML	Ward *et al*[18]	-
Other-cause mortality rates	ONS[258]	-
Oncotype DX cost	Genomic Health[113]	-
All other costs	Ward *et al*[18]	-

AE - adverse event; AML – acute myeloid leukaemia; ONS – Office for National Statistics

[redacted]

[redacted]

Table 110: Corrected pre and post-Oncotype DX treatment decisions (provided by Genomic Health*)

Receiving chemotherapy (%)	Standard care	Oncotype DX	ODX - Standard care
All patients	[redacted]	[redacted]	[redacted]
Patients with low RS	[redacted]	[redacted]	[redacted]
Patients with intermediate RS	[redacted]	[redacted]	[redacted]
Patients with high RS	[redacted]	[redacted]	[redacted]

RS – recurrence score

* Provided in response to a request for clarification from the EAG

Risk of distant recurrence

The 10-year risk of distant recurrence according to Onctoype DX RS was taken from Dowsett *et al;*[35] the proportion of patients in each RS category is common to both modelled groups (see Table 111). Based on the assumptions employed in the model reported by Ward *et al,*[18] the risk of recurrence is tapered to be 50% of the estimated risk during years 11-15 and 0% thereafter. The RR of distant recurrence for chemotherapy versus no chemotherapy is taken from Paik *et al.*[49] Within the Oncotype DX group, it is assumed that the Oncotype DX test is predictive of chemotherapy benefit. As shown in Table 111, the relative risk applied is dependent on the Oncotype DX risk group, with the largest treatment effect applied in the high RS group. In contrast, within the standard care group, the model assumes that the relative risk associated with chemotherapy is constant across all patients.

Table 111: Risk of distant recurrence and the benefit (RR) of chemotherapy

Oncotype DX recurrence score risk groups	Risk of distant recurrence (no chemotherapy) (Dowsett *et al* (2010))	RR with chemotherapy (Standard care)	RR with chemotherapy (Oncotype DX)
Low RS	9%	82.7	1.00*
Intermediate RS	16%	82.7	0.61
High RS	23%	82.7	0.26

RS – recurrence score; RR – relative risk
* assumed value

Node-positive patients

The Genomic Health dossier includes a secondary analysis that explores the use of Oncotype DX in ER+ LN+ patients. The main differences between this analysis and the base case LN0 analysis are summarised in Table **112**. The proportion of patients receiving chemotherapy in the Oncotype DX group was based on Loncaster *et al*[196] which resulted in a 69.2% reduction in chemotherapy following the use of the test. It should be noted that unlike the base case, the analysis in the LN+ population did not use DRFS to estimate chemotherapy benefit; instead DFS rates were derived from Albain *et al*[68] (this same approach is used within the EAG's sensitivity analyses). The dossier states that only RRs for chemotherapy that were statistically significant were used: if this statement was accurate, this would result in an RR of 1.0 for the low RS and intermediate RS group and 0.59 in the high RS group. However, the Genomic Health model inputs do not reflect this: all reported RRs were used, irrespective of whether they were associated with a statistically significant difference (see **Table 112**).

Table 112: Parameter values in the LN+ analysis

Population	% receiving chemotherapy current practice group	% receiving chemotherapy Oncotype DX group	10-year cumulative probability of distant recurrence (no chemotherapy)	RR with chemotherapy (current practice)	RR with chemotherapy (Oncotype DX)
Low RS	100%	7.5%	40%	0.72	1.02
Intermediate RS	100%	63.2%	51%	0.72	0.72
High RS	100%	83.3%	57%	0.72	0.59

RS – recurrence score; RR – relative risk

Comparison of Oncotype DX versus other tests (MammaPrint, EP score, EPClin and Prosigna)

In order to estimate the proportion of patients receiving chemotherapy in the comparator group, data on concordance between Oncotype DX and the comparator tests were used to re-categorise patients in the NHS England Access Scheme dataset.[255] The proportions of patients receiving chemotherapy in each Oncotype DX RS group are shown in Table 113. For MammaPrint, concordance data from Shirvers *et al* (2013),[259] a US study with 135 patients were used. For EndoPredict, data from Varga *et al* (2013),[260] a small study of 24 patients in Germany and Switzerland were used. For Prosigna, a US study of 52 patients was used.[261] For MammaPrint, EndoPredict, and Prosigna, it was assumed that 100% of high-risk and 0% of low-risk patients in the comparator group would receive chemotherapy, whilst for Prosigna, 50% of the intermediate-risk group were assumed to receive chemotherapy. With the exception of the comparator test cost, all other parameters were held at the base case values.

Table 113: Parameter values for MammaPrint, EP score, EPClin, and Prosigna

Oncotype DX RS group	Proportion of patients	Proportion of patients receiving chemotherapy			
		MammaPrint	EP score	EPClin	Prosigna
Low RS	[redacted]	29%	40%	27%	26%
Intermediate RS	[redacted]	51%	80%	50%	38%
High RS	[redacted]	86%	100%	67%	50%

RS – recurrence score

Results of the Genomic Health model

The Genomic Health deterministic base case analysis indicates that Oncotype DX produces positive health gains (0.03 LYGs and 0.07 QALYs) at an additional cost of [redacted]; this corresponds to an ICER of [redacted] per QALY gained. The results are driven by an overall reduction in chemotherapy levels in women with a low or intermediate Oncotype DX RS (who benefit less from chemotherapy) and an increase in chemotherapy levels in those with a high Oncotype DX RS (who benefit more from chemotherapy). The company's probabilistic results indicate that the modelled estimates of incremental QALYs and costs are associated with considerable uncertainty. The cost-effectiveness plane (see Figure **10**) generated using the Genomic Health model shows a wide dispersion of results,

with a substantial number of samples being in the North-West quadrant (dominated) and the South-East quadrant (dominating). It should be noted that the cost-effectiveness plane presented in Figure 6-2 of the Genomic Health dossier appears very different to that generated by the EAG using the model; the reasons for this are unclear. The cost-effectiveness acceptability curve (CEAC) generated using the model (see Figure **11**) suggests that the probability that Oncotype DX produces more net benefit than current practice at willingness-to-pay (WTP) thresholds of £20,000 and £30,000 per QALY gained is approximately 0.51 and 0.52, respectively. The EAG has concerns regarding the robustness of the company's probabilistic ICER as different model runs produced very different results, ranging from less than £10,000 per QALY gained to more than £170,000 per QALY gained.

The results for the LN+ population and for the LN0 population comparing Oncotype DX against the other four tests are presented in Table 114. These analyses consistently indicate that, using mean values, Oncotype DX dominates the comparators. As with the LN+ analysis, the cost-effectiveness plane presented in the Genomic Health dossier (Figure 6-4) shows a wide dispersion of results, with a large proportion of samples in the North-West (dominated) and South-East (dominating) quadrants.

Table 114: Results of the Genomic Health model – Oncotype DX versus standard care and other comparator tests

Option	LYGs	QALYs	Costs	Inc. LYGs	Inc. QALYs	Inc. Costs	ICER (per QALY gained)
Analysis 1: LN0 – comparison versus standard care							
Oncotype DX	12.82	10.50	[redacted]	0.03	0.07	[redacted]	[redacted]
Standard care	12.80	10.43	[redacted]	-	-	-	
Analysis 2: LN+ comparison versus standard care							
Oncotype DX	12.95	10.60	[redacted]	-0.05	0.15	[redacted]	**Dominating**
Standard care	13.00	10.44	[redacted]	-	-	-	-
Analysis 3: LN0 - comparison with MammaPrint							
Oncotype DX	12.82	10.50	£6,319	0.02	0.07	-£1,272	**Dominating**
MammaPrint	12.80	10.43	£7,590	-	-	-	-
Analysis 4: LN0 - comparison versus EndoPredict EP score alone							
Oncotype DX	12.82	10.50	£6,139	0.01	0.08	-£762	**Dominating**
EndoPredict	12.82	10.41	£7,081	-	-	-	-
Analysis 5: LN0 - comparison versus EndoPredict EPClin score							
Oncotype DX	12.82	10.50	£6,319	0.03	0.06	-£532	**Dominating**
EndoPredict	12.80	10.44	£6,850	-	-	-	-
Analysis 6: LN0 - comparison versus Prosigna							
Oncotype DX	12.82	10.50	£6,319	0.03	0.06	-£655	**Dominating**
Prosigna	12.79	10.44	£6,974	-	-	-	-

Inc. - incremental

Figure 10: Cost-effectiveness plane – Oncotype DX versus current practice, LN0 population (generated by EAG using the Genomic Health model)

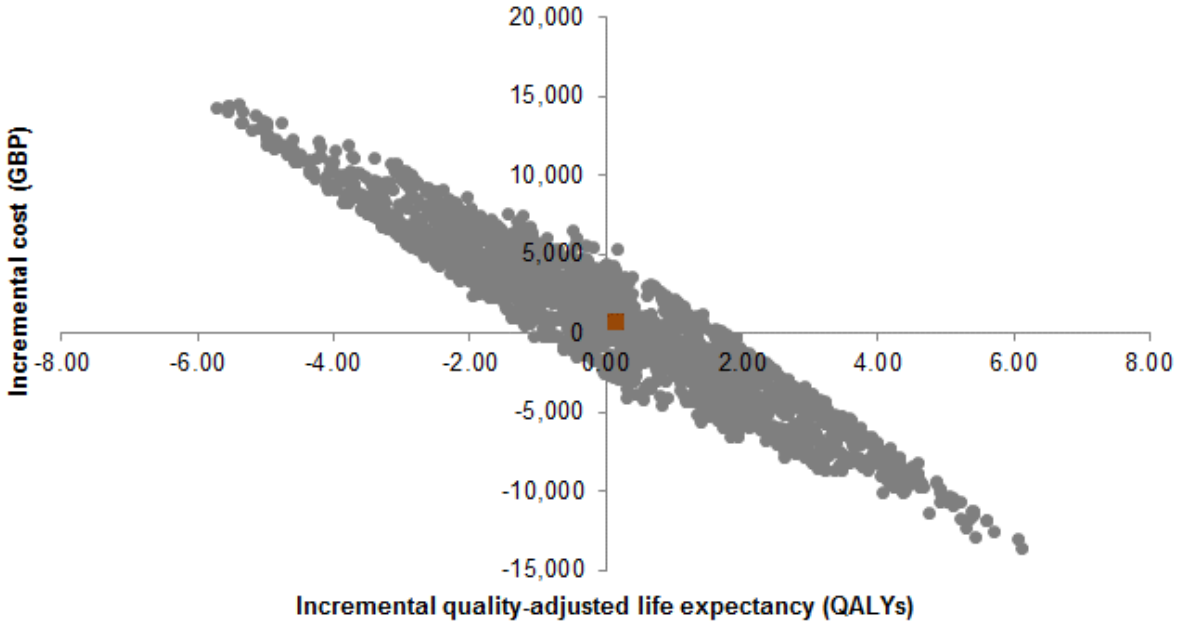


Figure 11: Cost-effectiveness acceptability curve – Oncotype DX versus current practice, LN0 population (generated by EAG using the Genomic Health model)

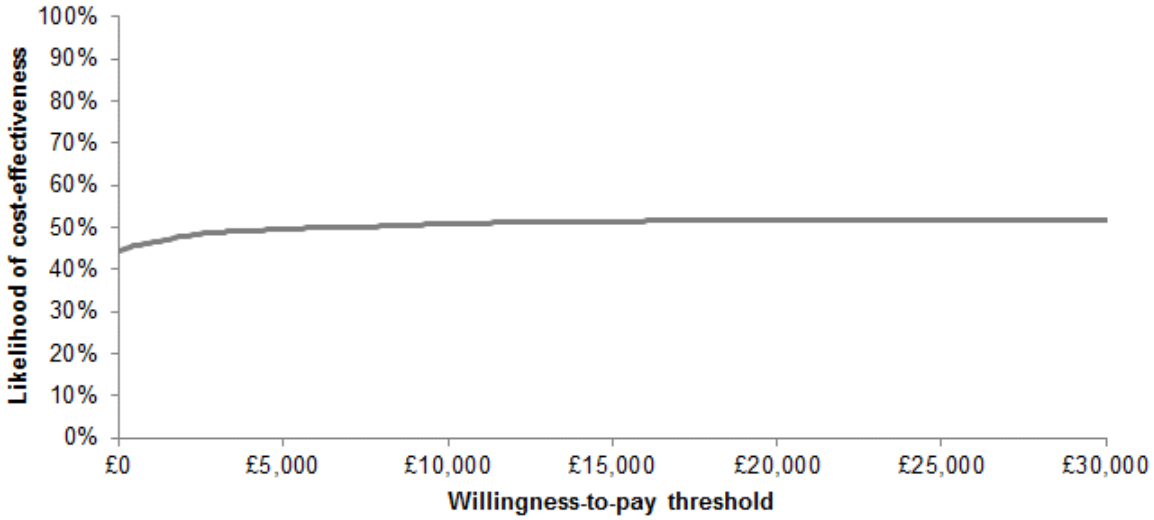


Table **115** presents the results of the company's one-way sensitivity analyses. These analyses indicate that the model results are sensitive to changes in several parameter values including: the time horizon, the discount rate, the disutility associated with chemotherapy, and current levels of chemotherapy use.

Table 115: Sensitivity analysis results of the Genomic Health model, Oncotype DX versus current practice, LN0 disease (adapted from Genomic Health dossier)

Sensitivity analysis	ICER (per QALY gained)
Time horizon	
10 years	£23,829
20 years	£14,319
Discount rates	
0% costs, 0% QALYs	£8,490
6% costs, 0% QALYs	£8,587
0% costs, 6% QALYs	£15,544
6% costs, 6% QALYs	£15,721
Cost	
Cost of chemotherapy based on 5 cycles (£3,901)	£13,578
Utilities	
Disutility associated with chemotherapy -0.07 (Peasgood *et al*. 2010)	£7,912
Disutility associated with chemotherapy -0.5 (Sime *et al*. 2001)	£1,335
Utility in recurrence state set to 0.51 (Milne *et al*. 2006)	£12,386
Clinical parameters	
25% of patients undecided about chemotherapy, not leaning either way, before Oncotype DX testing would have received chemotherapy under standard care	£16,910
75% of patients undecided about chemotherapy, not leaning either way, before Oncotype DX testing would have received chemotherapy	£9,271
Post-recurrence median survival set to 1.5 years (Remák and Brazil 20014)	£12,349
Net change in the use of chemotherapy guided by Manchester data (LN0 patients) (Loncaster *et al*. 2017)	£3,072
Relative risk reduction with chemotherapy in the low Recurrence Score group set to -1.1% (Paik *et al*. 2006)	£12,611
10-year risk of distant recurrence set to 3.2%, 0.1% and 39.5% in low, intermediate and high RS groups (Paik *et al*. 2006)	£9,078

The EAG notes that given that most of the probabilistic samples suggest either that Oncotype DX dominates or is dominated by current practice, it is surprising that none of the deterministic sensitivity analyses indicate this result.

Critical appraisal of the Genomic Health model

The EAG has several concerns regarding the Genomic Health model (Box 2). In particular, the EAG identified a number of programming errors within the model. As a consequence, the EAG does not consider the results of the Genomic Health model to be robust. The EAG's concerns are discussed in more detail below.

Box 2: Main issues relating to the Genomic Health model identified by the EAG

(1) Use of inappropriate structural assumptions which bias in favour of Oncotype DX
(2) Inappropriate application of chemotherapy-related disutility over remaining patient lifetime
(3) Risk classification probabilities and distant recurrence rates derived from separate sources
(4) Application of NHS England Access Scheme dataset to all LN0 patients
(5) Model errors

(1) Use of inappropriate structural assumptions which bias in favour of Oncotype DX

The Genomic Health model assumes that the Oncotype DX test is predictive of chemotherapy benefit. As shown in Table 111, this results in the RR for distant recurrence being dependent on the Oncotype DX RS category, with the greatest chemotherapy benefit being applied to the high RS group. In the model, the standard care arm mirrors the Oncotype DX arm in that patients are also split into the three Oncotype DX RS categories. The difference between the arms is the proportion of patients in each risk group who go on to receive adjuvant chemotherapy. However, as shown in Table 111, the relative risk used in the standard care arm is constant across all three risk groups and is based on the crude average of the three relative risks reported in the Paik *et al* study.[49] The distribution of patients between the three RS groups differs between the NHS England Access Scheme dataset[255] used in the Genomic Health model and the Paik *et al* study and therefore the crude mean RR from Paik *et al* does not represent the average RR for the population in the Genomic Health model. Furthermore, as Oncotype DX only identifies patients who may benefit from chemotherapy, the same RR of distant recurrence by RS category should be applied to both the modelled Oncotype DX and current practice groups (by RS score), as each group has exactly the same patient distribution across RS scores. If a patient is identified by Oncotype DX as being high-risk, the benefit they accrue from adjuvant chemotherapy should be identical to that accrued by the same patient who receives chemotherapy without the test.

(2) Inappropriate application of chemotherapy-related disutility over remaining patient lifetime

The QALY decrement resulting from the use of chemotherapy is applied during every cycle for the remainder of the modelled patients' lifetimes. The EAG considers it unlikely that patients would suffer the adverse effects of adjuvant chemotherapy years after they have completed their treatment. This represents a very pessimistic assumption which increases the overall reduction in QALYs associated with chemotherapy and overestimates the benefits associated with reducing overall chemotherapy use.

(3) Risk classification probabilities and distant recurrence rates derived from separate sources

The risk of distant recurrence and the proportion of patients in each Oncotype DX RS category were taken from two separate studies (Dowsett *et al*[35] and the NHS England Access Scheme dataset[255]).

The use of separate sources for these inputs may produce confounding due to differences in the characteristics of patients within each RS category between the two sources. In addition, the EAG notes that the risk of distant recurrence in the Genomic Health model does not match the 9-year risk of distant recurrence for LN0 patients presented in Dowsett *et al*[35] or the Genomic Health dossier (see Genomic Health dossier, Section 5.3.3.2).

(4) Application of NHS England Access Scheme dataset to all LN0 patients

The NHS England Access Scheme dataset[255] is only applicable to women who are at clinical intermediate-risk based on the NPI or other clinical indicators. However, it is unclear from the Genomic Health dossier whether the model applies only to this population, or whether the model is intended to reflect costs and outcomes of the Oncotype DX test across the whole LN0 population.

(5) Model errors

The EAG identified an error in the company's calculations relating to the proportion of patients in the low-, intermediate-, and high-risk groups who receive chemotherapy. The correct proportions are presented in the model; however, these are not applied directly but are instead incorrectly adjusted when used to calculate the traces for the Markov nodes. The correct proportions from the NHS England Access Scheme Dataset[255] and the incorrect values applied in the model for the Oncotype DX arm and the standard care arm are shown in Table **116**. This error leads to a substantial underestimate of the number of patients receiving chemotherapy in both the intermediate- and high-risk groups and has a significant impact on the model results.

Table 116: Correct proportions of patients receiving chemotherapy and those applied in the Genomic Health model (percentages reflect proportions of patients in entire group)

Oncotype DX (correct values)	**Chemotherapy**	**No chemotherapy**	**Total**
Low RS			
Intermediate RS			
High RS			
Oncotype DX (incorrect values applied in Genomic Health model)	**Chemotherapy**	**No chemotherapy**	**Total**
Low RS			
Intermediate RS			
High RS			
Standard care (correct values)	**Chemotherapy**	**No chemotherapy**	**Total**
Low RS			
Intermediate RS			
High RS			
Standard care (incorrect values applied in Genomic Health model)	**Chemotherapy**	**No chemotherapy**	**Total**
Low RS			
Intermediate RS			
High RS			

In addition, the results reported for the node-positive patients in the Genomic Health model could not be replicated by the EAG using the data described in the Genomic Health dossier. In order to replicate the results, two different sets of data were required. For the risk of distant recurrence, the Dowsett *et al* study[35] used in the base case analysis had to be selected (rather than the appropriate Albain *et al*[68] study). In addition, the results in the dossier use Paik *et al*[49] (rather than the appropriate Albain *et al*[68] study).

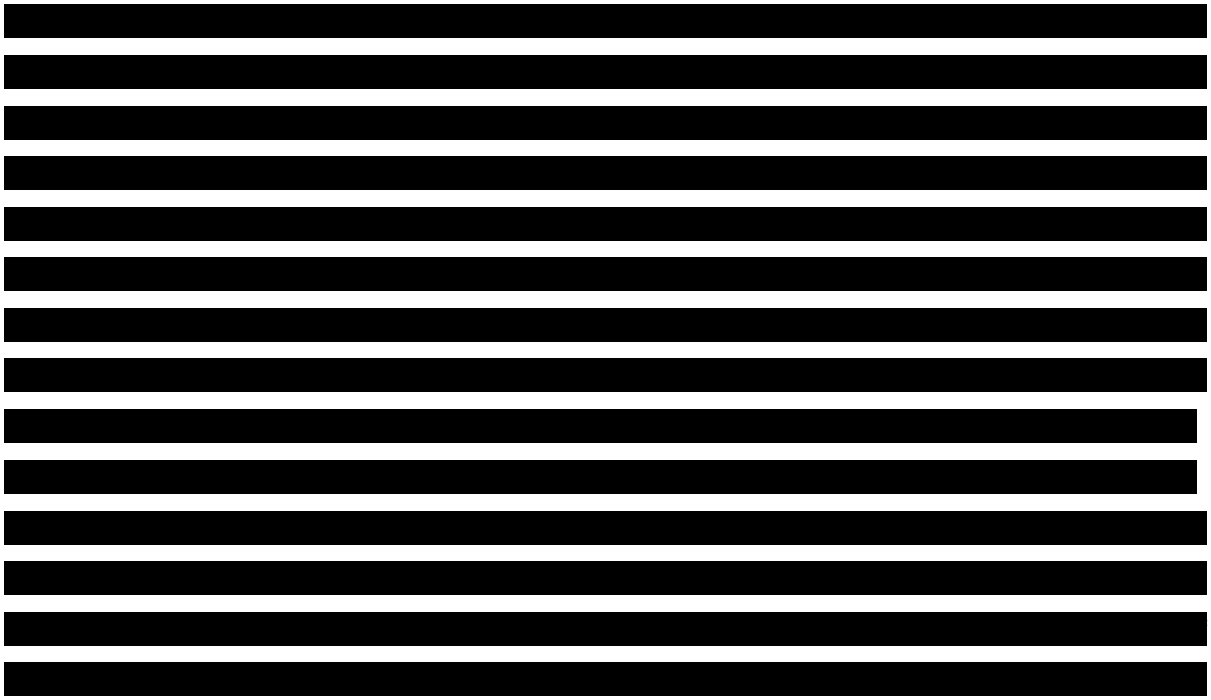
The impact of correcting the major errors in the Genomic Health model is explored further through comparison with the EAG model in Section 5.3.7.

5.2.3 EndoPredict draft cost-effectiveness paper (Myriad)[225]

Model scope

The chief investigator of the EndoPredict decision impact study[76] made available a draft manuscript which outlines the methods and results of an economic analysis comparing of EPClin+AOL versus AOL alone in women with ER+, HER2- early breast cancer, having had an intermediate-risk score using AOL. The EAG notes that the AOL risk interval is not explicitly defined. The executable model was not made available, hence the EAG was unable to verify whether it has been implemented appropriately.

The manuscript presents two sets of analyses: (i) a short-term cost minimisation analysis of EPClin versus usual practice (including only chemotherapy acquisition costs and the costs of providing the EPClin test), and (ii) a cost-effectiveness analysis of EndoPredict plus AOL versus AOL alone from the perspective of the NHS over a lifetime horizon.



[redacted]

Model structure

The cost-minimisation analysis includes the proportion of patients receiving chemotherapy with and without the EndoPredict test as well as the intensity and type of chemotherapy prescribed.

[redacted]

[redacted]

[redacted]

Results of the Myriad model

***[redacted]

Table 118 presents the results of the cost-minimisation analysis of EndoPredict plus AOL versus AOL alone. The results indicate that based on the observed IPD, EndoPredict led to a small but non-statistically significant increase in the mean per patient cost of acquisition and provision of chemotherapy to the NHS of £149 per patient. [redacted]

[redacted]

Table 118: Results of the Myriad cost-minimisation analysis – EndoPredict plus AOL versus AOL (adapted from draft cost-effectiveness paper)

Outcome	EndoPredict plus AOL	AOL	Total cost difference (per patient average, *p*-value)
Cost of chemotherapy acquisition and delivery per treated patient – mean (SD)	[redacted]	[redacted]	[redacted] £149, [redacted]
Total short-term cost of chemotherapy plus EndoPredict to all follow-up - mean (SD)	[redacted]	[redacted]	[redacted]

SD – standard deviation

[redacted]; this corresponds to an expected ICER of £26,836 per QALY gained.

Table 119: Results of the Myriad cost-effectiveness analysis – EndoPredict plus AOL versus AOL (adapted from draft cost-effectiveness paper)

Option	LYGs	QALYs	Costs	Inc. LYGs	Inc. QALYs	Inc. Costs	ICER (per QALY gained)
EndoPredict plus AOL	NR	[redacted]	[redacted]	NR	[redacted]	[redacted]	**£26,836**
AOL	NR	[redacted]	[redacted]	-	-	-	-

Inc. - incremental

[redacted]

[redacted]

[redacted]

The EAG has concerns regarding the economic evidence for Oncotype DX, MammaPrint and EndoPredict made available to the EAG. In particular, the Genomic Health model for Oncotype DX includes a number of errors and, in the opinion of the EAG, unreasonable assumptions. The Agendia model for MammaPrint includes correctable errors; [redacted] The EAG did not receive a model for EndoPredict and therefore cannot comment fully on the reliability of the results presented. No economic evidence was provided by the manufacturers of Prosigna or IHC4.

5.3 Independent economic evaluation

5.3.1 Scope of the EAG economic analysis

The EAG developed a *de novo* model to assess the cost-effectiveness of Oncotype DX, Prosigna, IHC4+C, EPClin, and MammaPrint versus current practice alone. The scope of the EAG model is summarised in Table **120**. The model assesses the health outcomes and costs associated with each strategy over a lifetime horizon from the perspective of the UK NHS and PSS. All costs and health outcomes are discounted at a rate of 3.5% per annum. Unit costs are valued at 2015/16 prices. The principal sources of evidence used to inform the analyses of Oncotype DX, Prosigna, IHC4+C and EPClin are the TransATAC study[43] and the NHS England Access Scheme dataset.[255] As the TransATAC study does not include the MammaPrint test, the MINDACT study[134] was instead used as the basis for estimating classification probabilities and conditional DMFS probabilities for MammaPrint. Additional studies identified within the clinical evidence review (see Chapter 4) which provide alternative relevant data on test risk classification probabilities, 10-year DMFS probabilities conditional on risk classification and post-test chemotherapy probabilities (decision impact) are explored within the sensitivity analyses.

Table 120: Scope of the EAG economic analysis

Population	Women with ER+, HER2-, early breast cancer (LN0-3). For Oncotype DX, Prosigna, IHC4+C and EPClin, analyses are presented for three discrete patient subgroups: • LN0 NPI≤3.4 (clinical low-risk) • LN0 NPI>3.4 (clinical intermediate-risk) • LN+ (1-3 nodes) For the evaluation of MammaPrint, the modelled population reflects the ITT population of the MINDACT trial.[134] Additional analyses are also presented for the mAOL clinical high-risk subgroup and the mAOL clinical low-risk subgroups separately.
Interventions	(1) Oncotype DX* (cut-offs: low <18, intermediate 18-30, high ≥31) (2) Prosigna (cut-offs LN0: low 0-40, intermediate 41-60, high 61-100; cut-offs LN+: low 0-15, intermediate 16-40, high 41-100) (3) IHC4+C (cut-offs: low <10%, intermediate 10-20%, high >20%) (4) EPClin (cut-off: 3.3) (5) MammaPrint (cut-off as per MINDACT trial[134])
Comparator	The comparator for all analyses is current practice (including a mix of risk prediction tools and diagnostic guidelines). For MammaPrint, current practice is based on mAOL, as per the design of the MINDACT trial.[134] Due to evidence limitations,† the competing tests were not compared incrementally against one another.
Primary health economic outcome	Incremental cost per QALY gained
Perspective	NHS and PSS
Time horizon	Lifetime
Discount rate	3.5% per annum
Price year	2015/16

* RSPC (Oncotype DX including clinico-pathological factors) is considered within the sensitivity analyses
† MammaPrint data derived from different source than the other tests; TransATAC analysis based on non-restricted dataset with different numbers of samples available for each test

Population

The population reflected in the model relates to women with ER+, HER2- early breast cancer with 0-3 nodes. For Oncotype DX, IHC4+C, Prosigna and EPClin, the economic analysis is presented for three discrete subgroups: (1) women with node-negative disease and an NPI≤3.4 (clinical low-risk); (2) women with node-negative disease and an NPI>3.4 (clinical intermediate-risk), and (3) node-positive (1-3 nodes). The modelled population for these four tests reflects that of the TransATAC study,[43] as this is used as the source of data on risk classification for each test and the 10-year DMFS probability conditional on each risk classification. Within the LN0 population, an NPI cut-off of 3.4 was chosen as a means of distinguishing between clinical low-risk and clinical intermediate-risk for Oncotype

DX, IHC4+C, Prosigna and EPClin, as data by NPI score were available from the TransATAC trial[43] and the NCRAS cancer registration dataset.[254] PREDICT scores were not available in either dataset, therefore this tool could not be used to define clinical risk.

MammaPrint was not included in the TransATAC study, hence an alternative source was required. The economic analysis of MammaPrint was instead largely based on data reported within the original paper and supplementary material of the MINDACT trial publication.[134] As the randomisation schedule within the MINDACT trial was performed using a modified version of AOL (mAOL) and sufficient data were not presented separately for patients with 1-3 lymph nodes, the population of the primary analysis largely reflects the MINDACT ITT population.[134] Additional analyses are also presented for the mAOL high-risk subgroup and the mAOL low-risk subgroups.

Interventions

The EAG's economic analysis includes all five tests included in the final NICE scope[22] (see Table 5). The tests are modelled in line with how their manufacturers state that they will be used in clinical practice: IHC4 and EndoPredict are assumed to be applied together with clinico-pathological factors (IHC4+C and EPClin, respectively). RSPC (Oncotype DX in conjunction with clinico-pathological factors) is considered separately within the sensitivity analyses but is not included in the EAG's base case. The EAG's economic analysis also assumes that all pathology analysis is undertaken centrally; local pathology analysis is not considered within the EAG's base case.

Comparator

The most commonly used tools for predicting the risk of recurrence after surgery to guide the use of adjuvant chemotherapy for breast cancer in England are PREDICT and NPI. AOL is currently being updated and has been temporarily disabled. As noted above, a modified version of AOL was used to inform the randomisation schedule for the discordant clinical and genomic risk groups within the MINDACT trial.[134] As such, the comparator for the analysis of MammaPrint is current practice using mAOL.

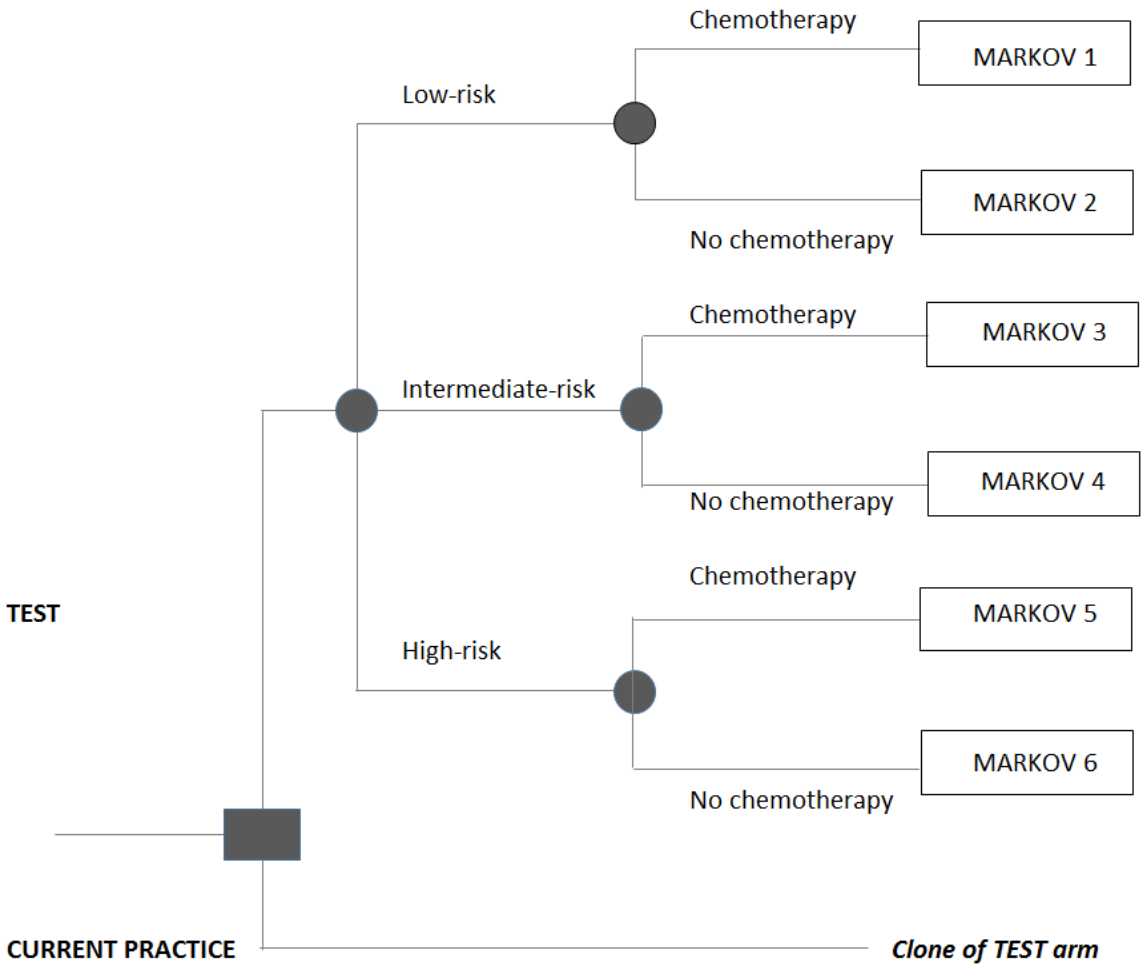
Owing to the use of a different evidence source for MammaPrint[134] compared with the other four tumour profiling tests, and the use of the unrestricted TransATAC trial dataset,[43] each test is compared only against current practice; tests were not assessed incrementally against each other.

5.3.2 Model structure

The general structure of the EAG model is based on the model previously developed by Ward *et al*[18] to inform NICE DG10.[21] This is also broadly consistent with the majority of studies identified within the review of published economic evaluations (see Section 5.1). The EAG model takes the form of a

hybrid decision-tree - Markov model (see Figure **13** and Figure 14). The decision tree component of the model classifies patients in the current practice (no test) group and the tumour profiling test group into high-, intermediate- and low-risk categories based on the results of the test. For EPClin and MammaPrint, the intermediate-risk category is not relevant as these tests provide results in terms of high- and low-risk only. The treatment group (test or no test) and the risk level predicted by the test determines the probability that the patient will subsequently receive adjuvant chemotherapy. Within both the test group and the current practice group, the decision tree determines the probability that a patient will be assigned to one of six groups: (i) low-risk, chemotherapy; (ii) low-risk, no chemotherapy; (iii) intermediate-risk, chemotherapy; (iv) intermediate-risk, no chemotherapy; (v) high-risk, chemotherapy, and (vi) high-risk, no chemotherapy (for the analyses of EPClin and MammaPrint, four branches are used due to the absence of an intermediate-risk category). Each of the branches is then linked to a Markov model that predicts lifetime QALYs and costs according to the patient's risk of distant recurrence and whether or not they receive chemotherapy.

Figure 13: EAG model - decision tree component*

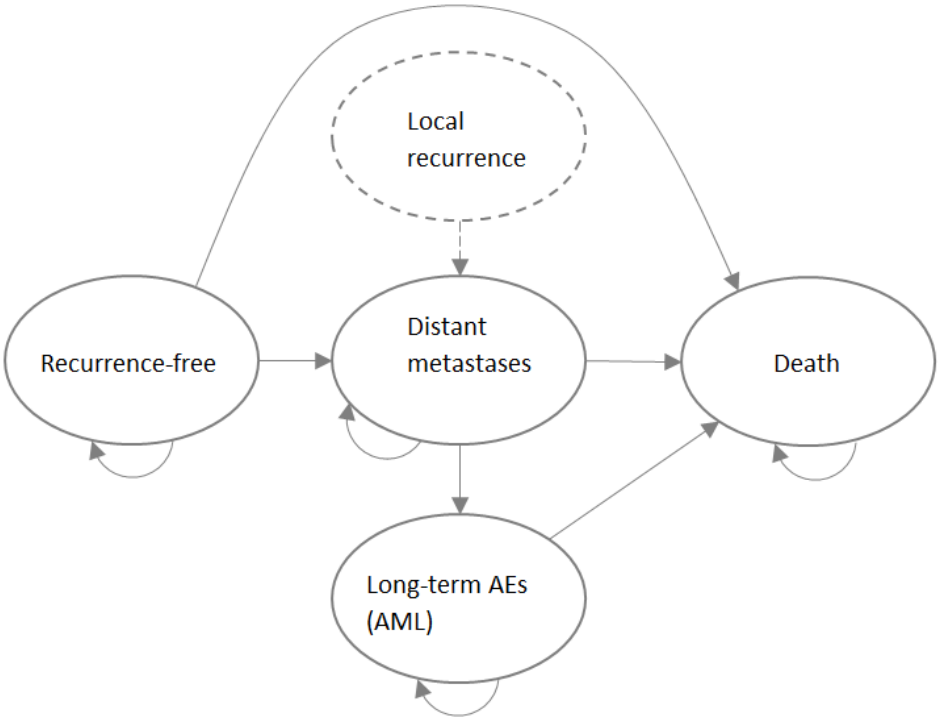


* For EPClin and MammaPrint, four branches are used due to the absence of an intermediate-risk category for these tests

Figure 14 illustrates the Markov nodes of the model. Each Markov node is evaluated over eighty-four 6-month cycles (42 years): patients are assumed to enter the model aged 58 years and the evaluation is continued until the cohort has reached age 100 years. Each Markov node includes four health states: (1) recurrence-free; (2) distant recurrence; (3) long-term AEs (AML), and (4) dead. Each Markov node differs with respect to the patient's risk of distant metastases (determined by their risk classification and whether or not they receive adjuvant chemotherapy). For all Markov nodes, patients enter the model in the recurrence-free health state. During any 6-month cycle, patients who are recurrence-free can remain in their current health state, transit to the long-term AEs state, develop distant metastases or die. Patients in the distant metastases state can remain in their current health state, transit to the long-term AEs (AML) state or die. Patients in the long-term AEs (AML) state are assumed to remain in this state until death (if free from breast cancer recurrence, they cannot subsequently develop distant metastases due to their breast cancer). Patients may die due to breast cancer, AML or other causes. An HRQoL decrement is applied during the first model cycle for patients receiving adjuvant chemotherapy to account for health losses associated with short-term chemotherapy-related AEs. The benefit of adjuvant chemotherapy is modelled using a RR of distant recurrence within each risk classification group. The impact of each test is therefore captured in the model only by changing the probability that patients with each test risk classification receive adjuvant chemotherapy. In the evaluation of Oncotype DX, a sensitivity analysis is included in which the test is assumed to provide a predictive benefit of chemotherapy, hence different RRs of developing distant metastases (for chemotherapy versus no chemotherapy) are applied across the low-, intermediate- and high-risk groups.

Different health utilities are applied to each of the modelled health states. The model assumes that a proportion of patients who experience distant recurrence will also have previously developed local recurrence: this is assumed to be associated with additional costs and a once-only QALY loss. The model includes costs associated with the tumour profiling test (in the intervention group only), adjuvant chemotherapy acquisition and administration and associated toxicity, endocrine therapy (all patients), routine follow-up visits and tests, additional therapies (zoledronic acid and G-CSF), and treatments for local recurrence and treatments for distant metastases. The costs and health outcomes for each Markov node differ due to the different risks of recurrence associated with each tumour profiling test and whether chemotherapy is given (together with its associated benefits, AEs and costs).

Figure 14: EAG model - state transition model component



Key EAG model assumptions

The EAG model makes the following structural assumptions:

- Within the base case analysis, the proportion of patients who receive chemotherapy under current practice (no test) is assumed to be the same for each test risk classification (low-, intermediate- and high-risk). This proportion is however assumed to differ between subgroups defined according to clinical risk (LN0 NPI≤3.4, LN0 NPI>3.4, LN+ (1-3 nodes), MINDACT ITT, MINDACT mAOL low-risk, and MINDACT mAOL high-risk).
- The model assumes that clinicians interpret each of the 3-level tests in the same way (e.g. an Oncotype DX high-risk score would lead to the same chemotherapy decision as a Prosigna high-risk score). The model also assumes that clinicians interpret each of the 2-level tests in the same way (a MammaPrint high-risk score would lead to the same chemotherapy decision as an EPClin high-risk score).
- Within the base case analysis, the relative benefit of adjuvant chemotherapy is assumed to be the same across all risk score categories for all tests (the same RR is applied to all patients, irrespective of test risk score). The impact of assuming a predictive benefit for Oncotype DX, which is applied by assuming different RRs between test risk score categories, is explored within the sensitivity analyses.
- The impact of the tests is modelled by changing which patients receive adjuvant chemotherapy.

- A proportion of patients who develop distant metastases are assumed to have previously developed local recurrence. Local recurrence is not modelled as a separate event or health state. QALY losses and costs associated with local recurrence are applied once only (upon entry into the distant metastases state).
- A disutility associated with short-term AEs related to adjuvant chemotherapy is applied once during the first model cycle only (whilst the patient is receiving treatment).
- Patients can enter the long-term AEs (AML) health state from either the recurrence-free state or the distant metastases state. The prognosis of patients with AML and the costs and QALYs accrued within the AML state are assumed to be independent of whether the patient has previously developed distant metastases due to their breast cancer. Once a patient develops AML, the model assumes that this alone determines their survival prognosis. Whilst CHF is also a potentially relevant long-term AE associated with chemotherapy, this was excluded from the model due to a lack of evidence on the joint survival impact of CHF and metastatic breast cancer.
- Costs associated with endocrine therapy, bisphosphonates (zoledronic acid) follow-up appointments and mammograms are assumed to differ according to time since model entry.
- Across all three analysis subgroups, patients are assumed to enter the model aged 58 years, based on the mean age of patients in the NHS England Access dataset[255] (rounded down to an integer value).
- The model includes both pre- and post-menopausal women. However, the TransATAC study relates only to post-menopausal women.

5.3.3 Evidence sources used to inform the model parameters

Table **121** summarises the evidence sources used to inform the parameters of the EAG model. The individual parameter values are discussed in further detail in the subsequent sections.

Table 121: Evidence sources used in the model

Parameter group	Source
Patient age	Based on the NHS England Access Scheme Dataset[255]
Risk classification probabilities for Oncotype DX, EPClin, Prosigna, IHC4+C	TransATAC bespoke data request.[43] Analysed by subgroup (LN0 NPI≤3.4, LN0 NPI>3.4 and LN+ [1-3 nodes]).
Risk classification probabilities for MammaPrint	MINDACT.[134] Analysed according to ITT trial population and mAOL low-risk and mAOL high-risk subgroups.
Distant recurrence rates (10 years) conditional on test risk classification (Oncotype DX, EPClin, Prosigna, IHC4+C)	TransATAC bespoke data request.[43] Analysed by subgroup (LN0 NPI≤3.4, LN0 NPI>3.4 and LN+ [1-3 nodes]).
Distant recurrence rates (10 years) conditional on test risk classification (MammaPrint)	MINDACT.[134] Analysed according to ITT trial population and mAOL low-risk and mAOL high-risk subgroups. All analyses involve extrapolation from the 5-year data.

Baseline probability of receiving adjuvant chemotherapy (current practice)	*LN0 NPI≤3.4 subgroup* NCRAS bespoke data request[254] *LN0 NPI>3.4 subgroup* NHS England Access Scheme dataset[255] *LN+ (1-3 nodes) subgroup* NCRAS bespoke data request[254] *MINDACT population (MammaPrint only)* Clinical judgement (Professor Rob Stein, UCL, estimates weighted by proportion with LN0 and LN+ disease and mAOL low-risk/high-risk).
Probability of receiving chemotherapy conditional on results of test (3-level tests – Oncotype DX, IHC4+C and Prosigna)	*LN0 NPI≤3.4 subgroup* UKBCG survey (see Appendix 5) *LN0 NPI>3.4 subgroup* NHS England Access Scheme dataset[255] *LN+ (1-3 nodes) subgroup* Loncaster *et al*[196]
Probability of receiving chemotherapy conditional on results of test (2-level tests – EPClin and MammaPrint)	Bloomfield *et al.*[76] Applied to all analysis subgroups.
10-year relative risk of recurrence chemotherapy versus no chemotherapy	EBCTCG 2012 meta-analysis[274]
Predictive chemotherapy benefit - Oncotype DX (applied in sensitivity analysis only)	*LN0 subgroups* Paik *et al*[49] *LN+ (1-3 nodes) subgroup* Albain *et al*[68]
Probability of death following distant recurrence	Thomas *et al*[257]
Probability of local recurrence	De Bock *et al*[275]
Probability of AML	Wolff *et al*[276]
Probability of death following onset of AML	Edlin *et al*[253]
Other-cause mortality (life tables)	ONS[258]
HRQoL	*Utility - recurrence-free and distant recurrence* Lidgren *et al*[265] *Utility AML* Younis *et al*[277] *HRQoL decrement - local recurrence and AEs related to adjuvant chemotherapy* Campbell *et al*[263]
Tumour profiling test costs	Test manufacturers
Costs - adjuvant chemotherapy	Hall *et al*[278]
Costs - endocrine therapy	BNF[279]
Costs – G-CSF	BNF[279] and PSSRU[256]
Costs - routine follow-up	NHS Reference Costs 2015/16[280] and Campbell *et al*[263]
Costs – bisphosphonates (zoledronic acid)	BNF[279] and NHS Reference Costs 2015/16[280]
Costs – local recurrence	Karnon *et al*[269]
Costs – distant metastases	Thomas *et al*[257]

NCRAS - National Cancer Registration and Analysis Service; UCL – University College London; BNF – British National Formulary; PSSRU – Personal Social Services Research Unit

Patient age

Mean age was assumed to be 58 years of age, based on the NHS England Access Scheme dataset[255] (rounded down to an integer value).

Risk classification probabilities using each test – Oncotype DX, Prosigna, IHC4+C, EPClin

Data relating to risk classification probabilities for each test were obtained from a bespoke analysis of the TransATAC trial provided by the trial investigators[43] (see Table 122). As discussed in Section 4, the ATAC trial evaluated the efficacy and safety of anastrozole vs tamoxifen. The TransATAC trial tested tumour blocks from post-menopausal patients who had been included in the monotherapy arms of the ATAC trial[281] in order to determine whether the tests could provide independent information on the risk of distant recurrence. Separate data analyses were provided by the trial investigators for ER+, HER2- patients for the three modelled subgroups (LN0 NPI≤3.4, LN0 NPI>3.4, and LN+ [1-3 nodes]). In order to maximise the information available for each test, data were not restricted only to those with information on all four tests. The EAG considers that the use of this study has particular value as: (a) it includes the use of four of the five tests included in the final NICE scope (Oncotype DX, Prosigna, IHC4+C and EPClin) within the same patient population; (b) it provides a source of data on 10-year DMFI probabilities conditional on test risk classification, thereby avoiding confounding due to the use of different evidence sources for these parameters, and (c) TransATAC is a large UK study. However, two caveats should be noted with respect to the choice of this data source. Firstly, the non-restricted TransATAC dataset was used for the analysis – this maximises the sample size for each individual test; however, as each additional test was analysed, more tissue was required and for some samples, insufficient tissue was left. This reduces the number of patients with available data and may introduce bias comparing across tests. In addition, whilst the ATAC trial included only post-menopausal women, the economic analysis assumes that the risk classification and DMFI probabilities obtained from the TransATAC analysis can be translated to a pre-menopausal population; this assumption introduces an additional degree of uncertainty with respect to the generalisability of the analysis.

Table 122: Risk classification probabilities using Oncotype DX, Prosigna, IHC4+C and EPClin (TransATAC)

Test (number of samples)	Proportion of patients with risk classification*		
	Low-risk	Intermediate-risk	High-risk
LN0 NPI≤3.4			
Oncotype DX (541)	0.72	0.24	0.04
Prosigna (410)	0.72	0.24	0.03
IHC4+C (510)	0.88	0.11	0.01
EPClin (423)	0.90	-	0.10
LN0 NPI>3.4			
Oncotype DX (284)	0.50	0.31	0.19

Prosigna (253)	0.27	0.38	0.35
IHC4+C (279)	0.36	0.38	0.25
EPClin (254)	0.47	-	0.53
LN+ (1-3 nodes)			
Oncotype DX (219)	0.57	0.32	0.11
Prosigna (192)	0.08	0.32	0.60
IHC4+C (213)	0.28	0.34	0.38
EPClin (198)	0.24	-	0.76

* Values may not sum to 1.0 due to rounding errors

Risk classification probabilities - MammaPrint

The evaluation of MammaPrint was based on the MINDACT trial.[134] This study was selected for inclusion in the analysis for three reasons: (a) the trial publication and supplementary material provide sufficient information to estimate risk classification probabilities and DMFS probabilities conditional on risk classification within the same patient populations; (b) it includes a large sample size, and (c) the study allows for the estimation of the benefit of chemotherapy between discordant groups.

Risk classification probabilities for MammaPrint were obtained from the trial publication of the MINDACT trial[134] and the accompanying supplementary material (see Table **123**).

Table 123: Risk classification probabilities using MammaPrint (MINDACT)

Population	**Proportion of patients with risk classification**	
	MammaPrint low-risk	**MammaPrint high-risk**
MINDACT ITT population (n=6,693)	0.64	0.36
MINDACT mAOL clinical high-risk subgroup (n=3,370)	0.46	0.54
MINDACT mAOL clinical high-risk subgroup (n=3,324)	0.82	0.18

Probability of developing distant metastases (without chemotherapy) – Oncotype DX, Prosigna, IHC4+C, EPClin

The probability of developing distant metastases was based on 10-year DMFI/DMFS outcomes for each test risk classification. For Oncotype DX, Prosigna, IHC4+C and EPClin, these were obtained from a bespoke data analysis of the TransATAC study[43] (see Table **124**).

Table 124: 10-year distant recurrence rates by risk classification for Oncotype DX, Prosigna, IHC4+C and EPClin

Population	10-year distant metastasis-free interval (95% CI)			
	Oncotype DX*	Prosigna	IHC4+C	EPClin
LN0, NPI≤3.4, low-risk	0.983 (0.963-0.992)	0.986 (0.962-0.995)	0.975 (0.954-0.987)	0.971 (0.947-0.984)
LN0, NPI≤3.4, intermediate-risk	0.931 (0.867-0.965)	0.933 (0.857-0.969)	0.878 (0.747-0.943)	n/a
LN0, NPI≤3.4, high-risk	0.838 (0.577-0.945)	0.636 (0.297-0.845)	0.800 (0.204-0.969)	0.870 (0.714-0.944)
LN0, NPI>3.4, low-risk	0.854 (0.776-0.907)	0.923 (0.825-0.967)	0.873 (0.787-0.926)	0.848 (0.761-0.905)
LN0, NPI>3.4, intermediate-risk	0.798 (0.694-0.869)	0.796 (0.687-0.870)	0.788 (0.688-0.859)	n/a
LN0, NPI>3.4, high-risk	0.749 (0.598-0.851)	0.699 (0.584-0.788)	0.769 (0.645-0.855)	0.774 (0.688-0.83.8)
LN+ (1-3 nodes), low-risk	0.818 (0.727-0.880)	1 (n/a)	0.961 (0.851-0.990)	0.95 (0.811-0.988)
LN+ (1-3 nodes), intermediate-risk	0.754 (0.630-0.842)	0.807 (0.679-0.889)	0.758 (0.635-0.845)	n/a
LN+ (1-3 nodes), high-risk	0.686 (0.447-0.839)	0.707 (0.604-0.788)	0.672 (0.546-0.771)	0.716 (0.629-0.785)

* Equivalent data relating to RPSC (Oncotype DX plus clinico-pathological factors) were also provided by the study investigators. The cost-effectiveness of this option is explored within the sensitivity analyses

The 10-year DMFI probability was converted to a cumulative probability of recurrence for each test within each risk classification category (1-DMFI) and converted to a 6-month probability of distant recurrence assuming a constant rate.

Probability of developing distant metastases (without chemotherapy) – MammaPrint

Cardoso *et al*[134] report 5-year DMFS probabilities for patients who did, or did not, receive adjuvant chemotherapy in the discordant risk groups in the MINDACT trial.[134] Additional information is also provided on chemotherapy use and 5-year DMFS in the concordant risk groups. For the economic analysis based on the MINDACT ITT population, it was necessary to estimate DMFS probabilities for all concordant and discordant groups according to clinical and genomic risk classification and whether patients received chemotherapy. This was done as follows (refer to data presented in Table 125):

- 10-year DMFS outcomes were estimated for all concordant and discordant clinical and genomic risk groups according to whether patients received adjuvant chemotherapy or not

(EAG group labels A-H) based on 5-year DMFS outcomes, assuming a constant event rate. The proportions of patients who received chemotherapy were obtained from the supplementary material of the Cardoso *et al* trial publication.[134] An adjustment was made to the mAOL high-risk MammaPrint high-risk group to estimate counterfactual 10-year DMFS for patients not receiving chemotherapy (EAG group label H); this was done by estimating the 10-year DMFS probability for this group (with chemotherapy) and multiplying this value by the reciprocal of the estimated 10-year RR of distant metastases for chemotherapy versus no chemotherapy for the overall discordant population (relative risk=0.77, adjusted 10-year DMFS for group=0.766).

- 10-year DMFS outcomes for the MammaPrint low-risk group (without adjuvant chemotherapy) were estimated by weighting the estimated 10-year DMFS probabilities for the MammaPrint low-risk no chemotherapy groups (EAG group labels B and D) according to the number of mAOL low-risk and mAOL high-risk patients in these groups.
- 10-year DMFS outcomes for the MammaPrint high-risk group (without adjuvant chemotherapy) were estimated by weighting the estimated 10-year DMFS probabilities for the genomic high-risk no chemotherapy groups (EAG group labels F and H, including the adjustment described above) according to the number of mAOL low-risk and mAOL high-risk patients in these groups.

Table 125: Calculation of 5-year DMFS probabilities by clinical/genomic risk group and chemotherapy use

Randomised group	Treatment	EAG group label	N randomised before genomic correction	N randomised after genomic correction	N in group*	Percent	5-year DMFS	5-year cumulative DMFS probability	Rate (year)	10- year DMFS probability for group	6-month recurrence probability for group
mAOL low, MMP low	Chemotherapy	A	2634	2745	37	0.55%	97.60%	2.40%	0.005	0.953	0.002
	No chemotherapy	B			2708	40.46%	97.60%	2.40%	0.005	0.953	0.002
mAOL high, MMP low	Chemotherapy	C	1497	1550	793	11.85%	95.90%	4.10%	0.008	0.920	0.004
	No chemotherapy	D			757	11.31%	94.40%	5.60%	0.012	0.891	0.006
mAOL low, MMP high	Chemotherapy	E	690	592	296	4.42%	95.80%	4.20%	0.009	0.918	0.004
	No chemotherapy	F			296	4.42%	95.00%	5.00%	0.010	0.903	0.005
mAOL high, MMP high	Chemotherapy	G	1873	1806	1735	25.92%	90.60%	9.40%	0.020	0.821	0.010
	No chemotherapy	H			71	1.06%	90.60%	9.40%	0.020	0.821†	0.010

***** Based on Cardoso *et al*[134] supplementary material, Table S11

†Adjusted 10-year DMFS without chemotherapy estimated to be 0.766

Tapering of risk of recurrence over time

The EAG notes that there is uncertainty with respect to the long-term risk of distant recurrence. The EAG model makes the same assumptions regarding long-term distant metastasis risk as the previous model reported by Ward *et al*.[18] The model assumes that the risk of distant metastases between 10 and 15 years is equal to half the risk during the preceding period (0-10 years); beyond 15-years, the risk of distant recurrence is assumed to be zero. The EAG notes that this is a simplification. This general decrease in the hazard of recurrence can be seen in the 10-15 year control arm annualised recurrence data reported in the 2005 EBCTCG meta-analysis.[262] Whilst there is some evidence which suggests that for some patients with particular disease subtypes, recurrence rates remain approximately constant between 5 and 20-years,[282] there is also uncertainty surrounding the duration over which the benefit of chemotherapy is sustained, hence constraining recurrence at 15-years reduces the likelihood of overestimating this benefit of chemotherapy. The impact of removing this assumption of recurrence risk tapering is explored within the sensitivity analyses.

Probability of receiving chemotherapy in the current practice group

The EAG identified two empirical sources which could be used to inform the probability that a patient receives chemotherapy without tumour profile testing: (i) the NCRAS dataset,[254] and (ii) the NHS England Access Scheme dataset (intermediate clinical risk only).[255] These alternative sources are discussed briefly below.

NCRAS dataset

A bespoke data request was placed with the NCRAS to obtain aggregate data relating to the use of adjuvant chemotherapy in women with early breast cancer in England (see Table **126**). The NCRAS cancer registration datasets were used to estimate current levels of chemotherapy use in each of the three model subgroups (LN0 NPI≤3.4; LN0 NPI>3.4 and LN+ [1-3 nodes]). An age restriction of 55-75 years was applied with the intention of only selecting those patients who were sufficiently fit to undergo chemotherapy and therefore may benefit from tumour profile testing, whilst also removing younger patients who are more likely to receive chemotherapy and are less reflective of the populations used to estimate risk classification probabilities and distant recurrence risk.[43] An additional data analysis on chemotherapy use for the whole population aged <75 years of age was also obtained. As shown in Table 126, within the age 55-75 years group, the proportion of women receiving chemotherapy is 7.19%, 40.01% and 62.72% in the LN0 NPI≤3.4, LN0 NPI>3.4 and LN+ (1-3 nodes) subgroups, respectively. As expected, the proportion of women receiving chemotherapy is higher in the broader age ≤75 years population.

Table 126: Baseline chemotherapy probabilities by risk group (provided by NCRAS)

Group	Age 55-75 years			Age ≤75 years		
	ACT	No ACT	Percentage	ACT	No ACT	Percentage
LN0, NPI≤3.4	329	4248	7.19%	964	6,008	13.83%
LN0, NPI>3.4	1388	2081	40.01%	3,265	2,897	52.99%
LN+ (1-3 nodes)	1849	1099	62.72%	4,557	1,526	74.91%

ACT – adjuvant chemotherapy

It should be noted that the NCRAS dataset reflects an unselected population who are not necessarily eligible for tumour profile testing; this may increase the size of the denominator, hence, in reality, the proportion of women who are eligible for testing who go on to receive adjuvant chemotherapy may be greater than the estimates generated using this dataset.

NHS England Oncotype DX Access dataset[255]

The NHS England Access Dataset[255] (previously described in Section 5.2) contains data on the pre-test chemotherapy decision for women who received the Oncotype DX test in England. It should be noted that this dataset relates only to women who were deemed to be at intermediate clinical risk, hence the data may not provide a good reflection of pre- and post-test chemotherapy decision-making for women with LN0 disease and an NPI score ≤3.4 or for women with LN+ disease. [redacted] the pre-test probability of receiving adjuvant chemotherapy is 0.431 [redacted] This estimate is only slightly higher than the estimate generated using the NCRAS dataset[254] (probability = 0.40).

Within the EAG base case analysis, the following selections were made:

- For the LN0 NPI≤3.4 subgroup, the NCRAS dataset[254] was used as these are the only data on baseline chemotherapy use available for this patient subgroup.
- For the LN+ (1-3 nodes) subgroup, the NCRAS dataset[254] was used as these are the only data on baseline chemotherapy use available for this patient subgroup.
- For the LN0 NPI>3.4 subgroup, the NHS England Access Scheme dataset[255] was used. This source was selected on the basis of consistency: the same dataset is used to inform the post-test probabilities of receiving chemotherapy conditional on risk score. It should also be noted

that the collection of these data was requested by the NICE Diagnostics Appraisal Committee in NICE DG10.[21]

- For the MammaPrint analyses, the EAG is not aware of any empirical evidence source which provides estimates of baseline chemotherapy use (without testing) for patients who are mAOL high-risk or mAOL low-risk. For this reason, these parameters were informed by expert opinion (personal communication: Professor Rob Stein, UCL). The following estimates were applied in the model, based on the modified version of AOL applied in the MINDACT trial:
    - LN0, mAOL high-risk, baseline chemotherapy probability = 70%
    - LN+, mAOL high-risk, baseline chemotherapy probability = 90%
    - LN0, mAOL low-risk, baseline chemotherapy probability = 15%
    - LN+, mAOL low-risk, baseline chemotherapy probability = 30%.

  These estimates were then weighted according to the proportion of women with LN0 and LN+ disease within the overall MINDACT population and within the mAOL high-risk and low-risk subgroups. This leads to baseline probabilities of 0.46, 0.77 and 0.16 for the MINDACT overall trial population, the mAOL high-risk subgroup and the mAOL low-risk subgroup, respectively.

Where appropriate, the source not selected for inclusion in the EAG base case was tested in the sensitivity analyses.

Probability of receiving chemotherapy conditional on test risk classification

Based on the review of decision impact studies presented in Section 4.9, five UK-based sources relating largely to the three analysis subgroups (LN0 NPI≤3.4; LN0 NPI>3.4; LN+ [1-3 nodes]) were identified as providing potentially usable data relating to the probability that a patient receives adjuvant chemotherapy conditional on the risk score given by the tumour profiling test. Evidence selection for these parameters was focussed on UK-based studies as these are more likely to reflect how clinicians will use the tests in England, although European studies were considered where the UK-based evidence was particularly limited (specifically for the 2-level tests). The five UK-based studies identified are: (i) the NHS England Access Dataset;[255] (ii) Holt *et al*;[283] (iii) Loncaster *et al*[196]; (iv) Bloomfield *et al*[76]; and (v) the UKBCG survey (see Appendix 5). The advantages and disadvantages of using each of these studies is summarised in Table 127.

Table 127: Studies available to inform chemotherapy use conditional on test results

Study	Disease type	EAG comments
NHS England Access Dataset[255]	LN0, intermediate clinical risk	This dataset was described previously in Section 5.2. This data collection exercise was requested by the NICE Diagnostics Appraisal Committee within the guidance for NICE DG10.[21] The dataset includes only patients with intermediate clinical risk and is likely to be relevant only to patients with LN0 disease and NPI>3.4. The data relate to the actual chemotherapy decision rather than a recommendation. [redacted]
Holt *et al*[283]	LN0 or pN1mic (micrometastasis)	Prospective UK clinical study on the impact of Oncotype DX on adjuvant treatment decisions and risk classification by NPI and Oncotype DX RS. Results were available for 74 patients. The data relate to the chemotherapy recommendation rather than the final decision. The EAG notes that this study has been published only in abstract form and few details are available regarding the methods.
Bloomfield *et al*[76]	Unclear	UK study of decision impact of EndoPredict. Fourteen oncologists in 8 UK hospitals saw 149 patients judged by clinical teams to have equivocal indications for chemotherapy. Patients and oncologists discussed provisional treatment decisions based on conventional prognostic factors. Initial decisions were reconsidered when EndoPredict results were available. The data appear to relate to the final decision rather than recommendations. The EAG notes that this is the only available UK study which relates to decision impact with a 2-level tumour profiling test. The population relates to patients for whom there was no clear decision on whether chemotherapy should be given. This study is unlikely to accurately represent the use of chemotherapy in women with LN+ disease.
Loncaster *et al*[196]	LN0 and LN+	Prospective UK pilot study designed to evaluate the clinical value of Oncotype DX testing. Testing was performed in 201 women with newly diagnosed, ER+, HER-2-, invasive breast cancer who underwent breast surgery with curative intent. Separate estimates are provided for the LN0 and LN+ subgroups. The data appear to relate to recommendations rather than the final decision. The EAG notes that patients enrolled in this study had already been recommended chemotherapy, therefore the use of this study may exaggerate the proportion of women for whom the final decision was to receive chemotherapy.
UKBCG survey	LN0 NPI≤3.4, LN0 NPI>3.4 and	The UKBCG network disseminated a bespoke unfunded survey designed by the EAG to their members (see Appendix

	LN+ (1-3 nodes)	5). Respondents were asked *"Based on your own subjective opinion, please estimate the probability that a woman in each of these subgroups and with each genomic/immunohistochemical test result would go on to receive adjuvant chemotherapy."* Responses were requested for 2-level and 3-level tests for the LN0 NPI≤3.4, LN0 NPI>3.4 and LN+ (1-3 nodes) subgroups. Eleven usable responses were received from participating oncologists. The results indicate considerable variation in practice. Several respondents noted uncertainty with respect to the 2-level tests as they do not currently have access to these technologies.

UKBCG – UK Breast Cancer Group

Estimates of the use of adjuvant chemotherapy conditional on test risk classification based on these alternative sources are summarised in Table 128.

Table 128: Summary of post-test chemotherapy probabilities conditional on risk classification

Source	**Proportion of patients receiving adjuvant chemotherapy conditional on test risk classification**			
	Population	**Low-risk**	**Intermediate-risk**	**High-risk**
NHS England Access dataset[255]	LN0, intermediate clinical risk	0.01	0.33	0.89
Holt *et al*[283]	LN0 or pN1mic	0.07	0.59	0.91
Bloomfield *et al*[76]	Unclear	0.07	n/a	0.77
Loncaster *et al*[196]	LN0	0.02	0.51	0.85
	LN+	0.08	0.63	0.83
UKBCG survey (3-level tests)	LN0, NPI≤3.4	0.00	0.17	0.74
	LN0, NPI>3.4	0.04	0.41	0.92
	LN+ (1-3 nodes)	0.46	0.76	0.95
UKBCG survey (2-level tests)	LN0, NPI≤3.4	0.01	n/a	0.71
	LN0, NPI>3.4	0.15	n/a	0.92
	LN+ (1-3 nodes)	0.40	n/a	0.97

UKBCG – UK Breast Cancer Group

With respect to the EAG base case, the following study selections were made:

- For the use of the 3-level tests (Oncotype DX, Prosigna and IHC4+C) in the LN0 NPI≤3.4 subgroup, the UKBCG survey data were used. This selection was made due to the absence of any published UK evidence on the decision impact of tumour profiling tests in this patient subgroup.
- For the use of the 3-level tests (Oncotype DX, Prosigna and IHC4+C) in the LN0 NPI>3.4 subgroup, the NHS England Access Scheme dataset[255] was used. This selection was made for two reasons: (1) this source is consistent with the source used to inform the baseline chemotherapy probabilities without testing, and (2) the EAG considers that this source

provides the best reflection of the way in which 3-level tumour profiling tests are used in clinical practice in England.

- For the use of the 3-level tests (Oncotype DX, Prosigna and IHC4+C) in the LN+ (1-3 nodes) subgroup, the Loncaster *et al* LN+ estimates[196] were selected as this is the only published UK evidence on decision impact which specifically relates to this patient subgroup.
- For the 2-level tests (EPClin and MammaPrint), the Bloomfield *et al* study[76] was selected for use in all three analysis subgroups as this is the only available published UK study which relates to a 2-level tumour profiling test. Given the limited UK-based evidence relating to the impact of 2-level tests, two additional European studies are explored in the sensitivity analyses.[213, 216]

The other sources not selected for inclusion in the EAG base case were included in the sensitivity analyses.

Adjuvant chemotherapy treatment effect on distant recurrence - Oncotype DX, Prosigna, IHC4+C, EPClin

As noted in Section 4.3.3, the evidence relating to the predictive benefit of Oncotype DX is limited to two re-analyses of RCTs[49, 68] which do not provide consistent conclusions regarding this aspect of the value of the test across the range of analyses reported. Within the base case analysis, all tests are assumed to be associated with prognostic benefit only (the relative benefit of chemotherapy is assumed to be the same across all test risk classification groups). For the analyses of Oncotype DX, Prosigna, IHC4+C and EPClin, the RR of recurrence for chemotherapy versus no chemotherapy was derived from a meta-analysis reported by the EBCTCG (2012).[274] Based on data presented in the supplementary material (see EBCTCG publication[274] extra web material, page 12, any anthracycline-based regimen versus no chemotherapy, distant recurrence), the 10-year risk of distant recurrence for chemotherapy and no chemotherapy was estimated by projecting forward the annualised risk of distant metastases (3.3%/year for chemotherapy, 4.6%/year for no chemotherapy). The RR for chemotherapy versus no chemotherapy was then calculated based on the difference between the projected 10-year DMFS probabilities for the two groups: this gives a 10-year RR of 0.76. The same RR was assumed to apply to the LN0 and LN+ subgroups. The impact of assuming higher and lower RRs for distant recurrence are explored in the sensitivity analyses. Further sensitivity analyses were also undertaken to explore the impact of assuming a predictive benefit of chemotherapy associated with Oncotype DX, based on the studies reported by Paik *et al*[49] (LN0) and Albain *et al*[68] (LN+). Within the model, this is implemented by applying different RRs to each of the risk classification groups, based on these two studies.

Adjuvant chemotherapy treatment effect on distant recurrence - MammaPrint

Within the analysis of MammaPrint, the benefit of adjuvant chemotherapy was estimated using data reported within the MINDACT trial publication,[134] rather than from an external source. The 10-year RR of relapse for adjuvant chemotherapy versus no adjuvant chemotherapy was estimated based only on the discordant clinical and genomic risk group data (see Table 125, EAG group labels C, D, E and F), extrapolated beyond the study endpoint. RRs of chemotherapy versus no chemotherapy were calculated for each of the two discordant groups (clinical low, genomic high and clinical high, genomic low) based on estimated 10-year DMFS; these were then weighted according to the number of patients in each discordant group. The weighted RR for the discordant populations was estimated to be 0.77. Within the mAOL low-risk and mAOL high-risk subgroups, the RRs of recurrence for each subgroup were based only on the discordant population relevant to that subgroup (mAOL low-risk RR = 0.84, mAOL high-risk RR = 0.74). The impact of assuming higher and lower RRs for distant recurrence are explored in the sensitivity analyses.

The relative risks of recurrence applied in the EAG base case analysis and sensitivity analyses are summarised in Table 129.

Table 129: Estimates of adjuvant chemotherapy benefit applied in the EAG model

	10-year RR of distant recurrence – chemotherapy versus no chemotherapy		
	EAG base case - Oncotype DX, Prosigna, EPClin, IHC4+C, non-predictive (EBCTCG[262])	**EAG sensitivity analysis - Oncotype DX, predictive benefit (Paik *et al*[49] and Albain *et al*[68])**	**EAG base case - MammaPrint MINDACT population, non-predictive (Cardoso *et al*[134])**
LN0 subgroups (NPI≤3.4 and NPI>3.4)			
Low-risk	0.76	1.31*	-
Intermediate-risk	0.76	0.61*	-
High-risk	0.76	0.26*	-
LN+ (1-3 nodes) subgroup			
Low-risk	0.76	1.02*†	-
Intermediate-risk	0.76	0.72*†	-
High-risk	0.76	0.59*†	-
MINDACT ITT population			
MMP low-risk	-	-	0.77
MMP high-risk	-	-	0.77
MINDACT mAOL low-risk			
MMP low-risk	-	-	0.84
MMP high-risk	-	-	0.84
MINDACT mAOL high-risk			
MMP low-risk	-	-	0.74
MMP high-risk	-	-	0.74

MMP - MammaPrint

* Deterministic values applied in the sensitivity analyses are also adjusted by half of the variance, derived from reported 95% confidence intervals

† HRs treated as relative risks

It should be noted that the model translates 10-year DMFS probabilities (without chemotherapy) into 6-month event probabilities assuming a constant rate. As an RR relates only to the specified timepoint of the analysis, it is inappropriate to apply this directly to the 6-month probability of recurrence. Instead, the EAG model applies a conversion by: (i) estimating the 10-year DMFS probability with chemotherapy based on the 10-year DMFS probability without chemotherapy and the 10-year RR of recurrence for chemotherapy versus no chemotherapy; (ii) estimating the HR for the DMFS outcomes at 10-years for chemotherapy versus no chemotherapy, assuming a constant event rate; (iii) applying the estimated HR to the 6-month DMFS probability for the no chemotherapy group, and (iv) converting this HR-adjusted 6-month DMFS probability with chemotherapy to a 6-month distant recurrence probability. This approach ensures that the relative distance between the predicted chemotherapy group and the observed no chemotherapy group is maintained at 10-years.

Survival following onset of distant metastases

The survival prognosis of patients with distant metastases was based on analysis of complete hospital and community records of 77 women randomly selected from 232 women who had relapsed breast cancer between 2000 and 2005 (Thomas *et al*[257]). The population included in this study had an average age of 62.3 years and included patients who had originally been diagnosed with LN+ disease (44%) and LN0 (56%). Forty-five percent of women were ER+ and 21% of women were HER2+. Median survival was reported to be 40.1 months following distant recurrence. The 6-month probability of death was estimated by fitting an exponential distribution with a median of 40.1 months; based this approach, the 6-month probability of death following distant recurrence was estimated to be 0.098, assuming a constant rate. The model assumes that the rate of death due to distant metastases is constant across the different model subgroups and across each test risk classification group due to a lack of population or risk group specific data.

Probability of local recurrence

The model assumes that 10.5% of patients entering the distant recurrence health state have previously experienced a local recurrence. This estimate was based on a study by de Bock *et al*[275] which analysed 3,601 women enrolled in three EORTC trials. The study included both LN0 and LN+ women who had been treated for early breast cancer. Of the 1,224 women who developed distant metastases, 129 women (10.54%) experienced a previous loco-regional recurrence. The model does not take into account the time spent alive with local recurrence; instead, the impact of local recurrence is applied crudely in the model as a once-only cost and QALY loss.

Probability of developing AML

The probability of developing AML following chemotherapy was taken from an analysis of 20,063 patients with Stage I-III breast cancer treated at US academic centers between 1998 and 2007 (Wolff

et al[276]). Within the cohort of 3,227 patients, the estimated 10-year risk of developing AML was reported to be 0.49%. The 6-month probability of developing AML was estimated to be 0.00025, assuming a constant event rate.

Survival following onset of AML

Survival following the onset of AML was estimated from the NICE STA of azacitidine for myelodysplastic syndromes (MDS).[284] Within this appraisal, the manufacturer estimated mean survival following the onset of AML to be approximately 8 months; assuming a constant event rate, this gives a 6-month probability of death following AML of 0.53.

Health utilities associated with relapse-free and distant metastases

Systematic searches were undertaken to identify studies reporting on HRQoL associated with different health states for women with breast cancer. Searches were undertaken in July 2017 in the following electronic databases:

- MEDLINE Epub Ahead of Print, In-Process & Other Non-Indexed Citations: Ovid, 1946 to present
- EMBASE: Ovid, 1974 to 2017 July 07
- Science Citation Index Expanded (SCI-E): Web of Science, 1900 to present
- Conference Proceedings Citation Index – Science (CPCI): Web of Science, 1990 to present.

The searches focussed specifically on studies which reported HRQoL estimates for health states which were measured and valued using the EQ-5D. The search strategy comprised sensitive MeSH or Emtree Thesauri terms and free-text synonyms for 'breast cancer' combined with free-text synonyms for 'EQ-5D'. The search strategies are presented in Appendix 1. Studies were considered potentially relevant if they reported EQ-5D valuations in both patients with non-metastatic and relapsed disease, thereby reflecting key health states in the model. Studies which reported disutilities associated with AEs resulting from the use of chemotherapy were retained for separate consideration. Studies were sifted according to their titles and abstracts; full texts were retrieved for studies which potentially met the inclusion criteria on the basis of the information provided in their title and abstract. HRQoL estimates for other modelled health states and events were not based on new searches; instead, these were derived through consideration of estimates which have been identified from a previous systematic review of health utilities (Peasgood *et al*[285]).

The EAG's searches identified a total of 227 studies. Of these, only four studies reported EQ-5D valuations for both non-metastatic and metastatic breast cancer states (see Table 130). Three of the identified studies were reported as full papers, whilst the fourth study was reported only in abstract

form. None of the identified studies were undertaken in UK patients: the studies were undertaken in Finland (Farkkila *et al*[286]), Sweden (Lidgren *et al*[265]), Iran (Yousefi *et al*[287]) and Canada (Naik *et al*[288]).

The study reported by Lidgren *et al* was selected for use in the EAG base case analysis on the basis that this population was most likely to best reflect the population of ER+ women with breast cancer who are treated in England. This study reported values for the recurrence-free (receiving endocrine therapy) health state and distant recurrence health state of 0.824 and 0.685, respectively. This same study was used to inform the health state utility estimates within the earlier model reported by Ward *et al*[18] and the Myriad model.[225]

Table 130: Summary of EQ-5D health state valuations in identified studies

Author	Publication type	Country	Population	Health state description	EQ-5D valuation for health state
Farkkila *et al*[286]	Abstract	Finland	778 breast cancer patients aged 31-90 in the Hospital District of Helsinki and Uusimaa	Baseline	0.818 (SD 0.228)
				First year of remission	0.860 (SD 0.178)
				Following years after remission	0.843 (SD 0.189)
				Metastatic disease	0.746 (SD 0.251)
				Palliative patients	0.514 (SD 0.300)
Lidgren *et al*[265]	Paper	Sweden	361 consecutive breast cancer patients attending the breast cancer outpatient clinic at Karolinska University hospital Solna for outpatient visits between April and May 2005	First year after primary breast cancer*	0.696 (95% CI 0.634–0.747)
				First year after recurrence	0.779 (95% CI 0.700–0.849)
				Second and following years after primary breast cancer / recurrence	0.779 (95% CI 0.745–0.811)
				Metastatic disease	0.685 (95% CI 0.620–0.735)
Naik *et al*[288]	Paper	Canada	1,759 ambulatory cancer survivors at the Princess Margaret Cancer Centre (mixed cohort with various cancer types, 282 patients with breast cancer)	Breast local/regional	0.82 (SE 0.01)
				Breast distant/metastatic	0.75 (SE 0.03)
Yousefi *et al*[287]	Paper	Iran	163 patients with breast cancer who attended the breast cancer subspecialty clinic affiliated with the Breast Cancer Research Center (BCRC), in Tehran, Iran	First year after primary breast cancer	0.674 (SD 0.201)
				First year after recurrence	0.718 (SD 0.139)
				Second and following years after primary breast cancer / recurrence	0.730 (0.221)
				Metastatic disease	0.552 (0.227)

SD – standard deviation; SE – standard error

* Lidgren *et al* also report EQ-5D utility score for patients receiving adjuvant hormone therapy of 0.824 (n = 79, CI: 0.785–0.857)

Health utilities associated with other model health states and events

The disutility associated with local recurrence was taken from a published model of first, second, and third generation adjuvant chemotherapy regimens for breast cancer reported by Campbell *et al*.[263] Within this study, the 6-month disutility associated with local recurrence was estimated to be 0.108 (SE=0.04). The HRQoL impact of chemotherapy-related AEs was also taken from Campbell *et al*;[263] the model assumes a disutility of 0.04 (assumed SE=0.004) during the first 6-month model cycle. The health utility associated with AML was assumed to be 0.26 based on a previous economic evaluation.[277]

Health utility estimates applied in the EAG model

Table 131 summarises the health utilities assumed in the EAG's base case analysis.

Table 131: Health utilities applied in the EAG model

Health state / event	Duration applied in model	Mean	Standard error	Source
Recurrence-free	Indefinite	0.824	0.018	Lidgren *et al*[265]
Distant metastases	Indefinite	0.685	0.029	
Disutility distant metastases	Indefinite	-0.14	0.11	Calculated using difference method[289]
Local recurrence disutility	Once-only QALY loss applied on transition to distant recurrence state	-0.108	0.04 (assumed)	Campbell *et al*[263]
Chemotherapy AEs disutility	6-months	-0.038	0.004 (assumed)	
AML	Indefinite	0.26	0.04 (assumed)	Younis *et al*[277]

Resource use and costs

The model includes the following cost components:

(i) Costs associated with the tumour profiling test

(ii) Costs of adjuvant chemotherapy acquisition and administration (including chemotherapy-related toxicity)

(iii) Costs associated with endocrine therapy

(iv) Costs of routine follow-up visits and tests

(v) Costs of other therapies (zoledronic acid and G-CSF)

(vi) Costs of treating local recurrence (once-only cost)

(vii) Costs associated with treating distant metastases.

Test costs

The costs of the tumour profiling tests were sourced from information provided to NICE by the manufacturers as part of the appraisal process. The cost of Oncotype DX includes the price discount offered through the Patient Access Scheme (PAS) for this product. The manufacturers of Oncotype DX, MammaPrint and EndoPredict submitted a cost for testing of samples in each of their centralised laboratories. IHC4 and Prosigna have no established centralised laboratory system. The manufacturers provided prices for conducting these two tests in NHS laboratories as outlined in Table 132. NanoString submitted a cost of £1,970; this is in line with the £1,596 (2013 prices) cost of the Prosigna test estimated as part of the OPTIMA prelim trial.[244] EndoPredict can also be conducted within an NHS laboratory; the impact of assuming a lower cost is considered within the sensitivity analyses.

Table 132: Test costs assumed in EAG analysis

Test	Cost	Comments
Oncotype DX (excluding PAS)	£2,580	Tests carried out in Genomic Health laboratory in US. Cost includes sample handling and customer service.
Prosigna*	£1,970	NanoString submitted a cost per Prosigna test based on conducting the test in an NHS laboratory which included the laboratory costs (£240), the list price for Prosigna kits (£1,650), cost of the nCounter System (£194,600) and was based on 2,500 sampled per lifetime of the nCounter System
EndoPredict*	£1,500	Tests carried out in Myriad's laboratory in Munich
IHC4	£203	IHC4 submitted a document outlining the time and equipment necessary to perform the test in 2014 prices. The total cost of the test (£198) was uplifted using the HCHS index.[256]
MammaPrint	£2,326	Converted from Euros to UK Pounds Sterling assuming exchange rate of 1.15.

PAS – Patient Access Scheme
* Alternative costs per test due to NHS testing explored in sensitivity analyses

Costs of adjuvant chemotherapy acquisition and administration (including toxicity)

The costs associated with adjuvant chemotherapy were obtained from a previous costing analysis undertaken to inform the economic analysis of the OPTIMA prelim trial (Hall *et al,*[278] see Table 133). The fully executable spreadsheet developed to inform the OPTIMA prelim analysis was made available to the EAG by the study authors (personal communication: Professor Robert Stein, UCL). Within this analysis, standard supportive medication, procurement, laboratory, pharmacy and administration costs were taken from the drugs and pharmaceutical electronic market information tool (eMIT),[290] the British National formulary (BNF)[279] and NHS Reference Costs 2013/14.[291] Unit costs associated with the management of chemotherapy-related Grade 3/4 toxicity were based on NHS Reference Costs 2013/14.[291] Within the original costing analysis, all costs were valued at 2013/14 prices; within the EAG analysis, these costs were uplifted to current values using the HCHS index.[256]

The EAG model assumes that women with ER+, HER2-, early breast cancer with 0-3 nodes typically receive one of four adjuvant chemotherapy regimens: (1) FEC100-T (3+3 cycles, assumed to be given to 25% patients); (2) TC (4 cycles, assumed to be given to 20% patients); (3) FEC75 (6 cycles, assumed to be given to 45% patients) and FEC100-Pw (3+3 cycles, assumed to be given to 10% patients). The weighted mean cost of adjuvant chemotherapy acquisition, delivery and toxicity was estimated to be £3,145.19 per course. All adjuvant chemotherapy costs are applied during the first model cycle. The EAG notes that the choice and proportionate use of alternative chemotherapy regimens may differ between centres; for this reason, the use of alternative chemotherapy cost assumptions are explored in the sensitivity analyses.

Table 133: Adjuvant chemotherapy costs applied in the EAG model

Regimen	Proportion of women receiving regimen	Central line costs	Drug costs	Delivery costs	Supportive meds costs	Medical oncology costs	Specialist nurse review	Blood tests	Toxicity costs	Total cost
FEC100-T (3+3 cycles)	0.25	£18.17	£306.84	£1,284.58	£435.64	£450.03	£613.10	£62.32	£378.20	£3,548.88
TC (4 cycles)	0.20	£18.17	£52.80	£856.39	£15.91	£310.81	£408.74	£41.55	£144.83	£1,849.19
FEC75 (6 cycles)	0.45	£18.17	£346.38	£1,284.58	£80.77	£310.81	£613.10	£62.32	£245.91	£2,962.05
FEC100-Pw (3+3 cycles)	0.10	£18.17	£274.53	£2,569.16	£435.89	£450.03	£613.10	£124.64	£378.20	£4,863.72

FEC100-T – fluorouracil, epirubicin, cyclophosphamide and docetaxel; TC - docetaxel and cyclophosphamide; FEC75 - fluorouracil, epirubicin and cyclophosphamide; FEC100-Pw - fluorouracil, epirubicin, cyclophosphamide and weekly paclitaxel

Costs of endocrine therapy

The model assumes that all surviving patients receive endocrine therapy for a period of between 5 and 8 years. The costs associated with endocrine therapy were based on the assumptions employed within the previous economic analysis reported by Ward *et al*.[18] The model assumes that patients may receive one of four endocrine therapy regimens: (1) tamoxifen for 5 years; (2) anastrozole for 5 years; (3) letrozole for 5 years or (vi) tamoxifen for 2 years then exemestane for 3 years. The proportion of patients receiving each regimen was taken from Ward *et al*[18] (tamoxifen – 40% patients; anastrozole – 20% patients; letrozole – 20% patients; tamoxifen then exemestane 20% patients, see Table 134). In line with the previous model reported by Ward *et al,*[18] 10% of patients are assumed to receive extended letrozole for 3 further years (years 6-8).

Table 134: Endocrine therapy costs applied in the EAG model

Endocrine therapy	Proportion of patients	Dosage (per day)	Product	Price per pack	Annual cost	Source
Tamoxifen	0.40	20mg	30 x 20mg tablet (various manufacturers)	2.88	£35.06	BNF[279]
Anastrozole	0.20	1mg	28 x 1mg tablet (various manufacturers)	£1.08 (NHS Drug Tariff price)	£14.09	
Letrozole	0.20	2.5mg	28 x 2.5mg tablet (Alliance Healthcare)	2.52	£32.87	
Exemestane	0.20	25mg	30 x 25mg tablet (various manufacturers)	£5.71 (NHS Drug Tariff price)	£69.52	

Costs of additional treatments (zoledronic acid)

The model assumes that 30% of women with early breast cancer will receive 4mg bisphosphonates (zoledronic acid) every 6 months by i.v. infusion for up to 3 years (cost per 36-month course = £58.50). Treatment is assumed to be given in a day case setting, based on the cost of delivering simple parenteral chemotherapy (unit cost = £199.94, based on NHS Reference Costs 2015/16,[280] outpatient, currency code SB12Z).

Follow-up costs

The model assumes that all patients receive two routine follow-up visits during the first year following surgery, with annual visits thereafter for a period of 5 years. Patients are also assumed to undergo a routine annual mammogram for up to 5 years. The cost of a routine follow-up visit was taken from the NHS Reference Costs 2015/16[280] (mean cost = £162.84, SE = £6.48, consultant-led, non-admitted, face-to-face attendance, follow-up, medical oncology, service code 370). The cost of a mammogram was not available within the NHS Reference Costs 2015/16 tariff: this unit cost was

instead taken from Campbell *et al*[263] (mean cost = £46.37, SE = £9.27) and uplifted to current values using the HCHS index.[256]

Costs of treatments for local and distant recurrence

The costs associated with treating local recurrence were taken from a UK-based patient-level costing analysis of breast cancer recurrence reported by Karnon *et al*.[269] This cost estimate was uplifted to current prices using the HCHS index[256] (uplifted mean cost = £13,912.92, assumed SE = £2,010.20). This is applied as a once-only cost upon the incidence of distant recurrence.

The costs associated with treating distant metastases were derived from the study reported by Thomas *et al*[257]). Costs included those associated with visits, drugs, pharmacy, hospital admission and intervention, imaging, radiotherapy, pathology and transport. Cost components specifically associated with terminal care were excluded. The 6-monthly cost of treating metastatic breast cancer was estimated to be £4,082. This estimate was uplifted to current prices using the HCHS index[256] (uplifted mean cost = £4,541, assumed SE = £908.13).

5.3.4 Methods for model evaluation

The incremental health outcomes and costs of each test versus standard care were evaluated in a pairwise fashion; the cost-effectiveness of each test was not compared against the other tests. Central estimates of cost-effectiveness were based on the expectation of the mean. Uncertainty was evaluated using probabilistic sensitivity analysis (PSA) and deterministic sensitivity analyses (DSA). PSA was undertaken using simple Monte Carlo sampling methods (10,000 samples). The choice of distribution assumed for each group of parameters in the model is summarised in Table 135. The results of the PSA are presented as CEACs and in tabular form. DSAs were undertaken to explore the impact of alternative assumptions and evidence sources regarding the probability of receiving chemotherapy with and without the tests, risk classification probabilities, recurrence rates, the potential predictive benefit of Oncotype DX, the magnitude of chemotherapy benefit, HRQoL estimates, and costs.

Table 135: Distributions used in EAG probabilistic analyses

Model parameter group	Distribution	EAG comments
Classification probabilities with/without test	Dirichlet	-
Chemotherapy use (conditional on test result)	Beta	-
Recurrence rates (conditional on test result)	Beta	Distribution parameters fitted to 95% CI around 10-year RFS data from TransATAC[43] or based on number of patients in treatment/risk group in MINDACT[134]
RR chemotherapy versus no chemotherapy	Log normal	SE assumed
Distant recurrence risk taper parameters	Fixed	-
OS rate following distant recurrence	Beta	SE estimated using 95% CI of Kaplan-Meier curve in Thomas *et al*[257]
Probability local recurrence	Beta	-
Probability AML	Beta	-
OS rate following incidence of AML	Beta	-
HRQoL	Beta	-
Chemotherapy costs	Normal	SE assumed to reflect uncertainty in delivery costs only
Endocrine therapy costs	Fixed	-
Zoledronic acid costs	Normal	SE for delivery costs estimated from NHS Reference Costs 2015/16[280] using interquartile ranges and number of submissions
Mammogram costs	Normal	SE taken from Campbell *et al*[263]
Follow-up/visit costs	Normal	SE estimated from NHS Reference Costs 2015/16[280] using interquartile range and number of submissions
Local recurrence cost	Normal	SE assumed equal to 20% of mean
Distant recurrence cost	Normal	SE assumed equal to 20% of mean
AML cost (one off)	Normal	SE assumed equal to 20% of mean
Test costs	Fixed	-

SE – standard error

The model was subjected to a number of DSAs: these are listed in Table 136. Additional input data applied in these sensitivity analyses are presented in Appendix 6.

.

Table 136: List of deterministic sensitivity analyses undertaken for each test

DSA description	DSA undertaken for test?				
	Oncotype DX	IHC4+C	Prosigna	EPClin	MammaPrint
Deterministic base case analysis	Yes	Yes	Yes	Yes	Yes
Post-test chemotherapy probabilities based on NHS England Access Scheme dataset[255] (clinical intermediate-risk, Table 128)	Yes	Yes	Yes	No	No
Post-test chemotherapy probabilities based on Holt *et al*[283] (Table 128)	Yes	Yes	Yes	No	No
Post-test chemotherapy probabilities based on Loncaster *et al*[196] (Table 128)	Yes	Yes	Yes	No	No
Post-test chemotherapy probabilities based on UKBCG survey (Table 128)	Yes	Yes	Yes	Yes	Yes
Post-test chemotherapy probabilities based on NHS England Access dataset[255] (Table 128); baseline chemotherapy probabilities from NCRAS[254] (Table 126)	Yes	Yes	Yes	No	No
Post-test chemotherapy probabilities based on Holt *et al*[283] (Table 128); baseline chemotherapy probabilities from NCRAS[254] (Table 126)	Yes	Yes	Yes	No	No
Post-test chemotherapy probabilities based on Loncaster *et al*[196] (Table 128); baseline chemotherapy probabilities from NCRAS[254] (Table 126)	Yes	Yes	Yes	No	No
Post-test chemotherapy probabilities based on UKBCG survey (Table 128); baseline chemotherapy probabilities from NCRAS[254] (Table 126)	Yes	Yes	Yes	No	No
Post-test chemotherapy probabilities based on Cusumano *et al*[216] (see Appendix 6, Table 1)	No	No	No	Yes	Yes
Post-test chemotherapy probabilities based on Penault-Llorca *et al*[213] (LN0 only, NPI>3.4 assumed, see Appendix 6, Table 2)	No	No	No	Yes	Yes
Baseline chemotherapy probabilities adjusted by Oncotype DX risk score (see Appendix 6, Table 3)	Yes	No	No	No	No
Chemotherapy assumptions (with and without test) based on Ward *et al*[18] (LN0, NPI>3.4 only)	Yes	Yes	Yes	No	No
Oncotype DX benefit assumed to be associated with predictive benefit. (LN0 RRs based on Paik *et al*[49] – low-risk 1.31; intermediate-risk 0.61; high-risk 0.26; LN+ RRs based on Albain *et al*[68] - low-risk 1.02; intermediate-risk 0.72; high-risk 0.59)	Yes	No	No	No	No
Risk classification and distant metastases probabilities based on Oncotype RPSC[43] (LN0 only, see Appendix 6, Table 4)	Yes	No	No	No	No
Prosigna risk classification and distant metastases probabilities derived from Gnant and Filipits[54] (LN+ only, see Appendix 6, Table 5)	No	No	Yes	No	No
EPClin risk classification and distant metastases probabilities derived from Dubsky *et al*[57]	No	No	No	Yes	No

(LN+ only, see Appendix 6, Table 6)					
MammaPrint risk classification and distant metastases probabilities derived from Van' t Veer *et al*[292] (LN0 only, see Appendix 6, Table 7)	No	No	No	No	Yes
Subgroup analysis in ER+, HER2-, LN+ subgroup	No	No	No	No	Yes
Assume MammaPrint low-risk receive no chemotherapy, MammaPrint high-risk receive 100% chemotherapy	No	No	No	No	Yes
10% lower cost per test due to increased efficiency (local NHS testing)	No	No	Yes	Yes	No
10% higher cost per test due to decreased efficiency (local NHS testing)	No	No	Yes	No	No
Baseline chemotherapy use halved	No	No	No	No	Yes
Start age based on TransATAC[35] (64 years)	Yes	Yes	Yes	Yes	Yes
Relative risk of distant metastases for chemotherapy versus no chemotherapy = 0.70	Yes	Yes	Yes	Yes	Yes
Relative risk of distant metastases for chemotherapy versus no chemotherapy = 0.80	Yes	Yes	Yes	Yes	Yes
Removal of distant metastases risk tapering	Yes	Yes	Yes	Yes	Yes
Utilities derived from Farkkila *et al*[286] (RFS=0.818, DM=0.746)	Yes	Yes	Yes	Yes	Yes
Distant metastases death rate doubled	Yes	Yes	Yes	Yes	Yes
Distant metastases death rate halved	Yes	Yes	Yes	Yes	Yes
AML removed from model	Yes	Yes	Yes	Yes	Yes
Chemotherapy cost doubled	Yes	Yes	Yes	Yes	Yes
Chemotherapy cost halved	Yes	Yes	Yes	Yes	Yes
Endocrine therapy costs doubled	Yes	Yes	Yes	Yes	Yes
Endocrine therapy costs halved	Yes	Yes	Yes	Yes	Yes
Local and distant metastases costs doubled	Yes	Yes	Yes	Yes	Yes
Local and distant metastases costs halved	Yes	Yes	Yes	Yes	Yes

UKBCG – UK Breast Cancer Group; NCRAS – National Cancer Registration and Analysis Service; RFS – recurrence-free survival; DM – distant metastasis

5.3.5 Model verification and validation

The EAG undertook a number of measures to ensure the credibility of the model.

- Peer review of the economic analysis by a modeller not involved in the assessment
- Verification and scrutiny of the executable model by two model developers
- Double-programming of the deterministic version of the model for all pairwise comparisons presented in the EAG base case
- Double-checking of the accuracy of all model inputs against sources
- Comparison of model results using point estimates of parameters and the expectation of the mean
- Comparison of mean of all probabilistic parameter samples against point estimates of parameters
- Examination of all identified sources of discrepancy
- Model testing using sensitivity analysis and use of extreme parameter values.

5.3.6 Cost-effectiveness results

Oncotype DX versus current practice

Central estimates of cost-effectiveness - (probabilistic)

Central estimates of cost-effectiveness for Oncotype DX versus current practice are presented in Table 137. All estimates are based on the probabilistic version of the EAG model. Within the LN0 NPI≤3.4 subgroup, Oncotype DX is expected to produce 0.01 additional QALYs at an additional cost of £1,313 per woman tested compared with current practice; this corresponds to an ICER of £122,725 per QALY gained. Within the LN0 NPI>3.4 subgroup, Oncotype DX is expected to produce 0.01 less QALYs at an additional cost of £881 per woman tested compared with current practice; within this subgroup, Oncotype DX is expected to be dominated. Within the LN+ (1-3 nodes) subgroup, Oncotype DX is expected to produce 0.07 less QALYs at an additional cost of £687 per woman tested compared with current practice; within this subgroup, Oncotype DX is again expected to be dominated. As shown in Table 138, the PSA indicates that the probability that Oncotype DX produces more net benefit than current practice at WTP thresholds of £20,000 and £30,000 per QALY gained is 0.04 or lower across all three subgroups. The results for the LN0 NPI>3.4 subgroup and the LN+ (1-3 nodes) subgroup are primarily driven by the lower use of chemotherapy (and the benefits forgone) with Oncotype DX compared with current practice (see Appendix 7 for impact of all tests).

Table 137: Central estimates of cost-effectiveness – Oncotype DX versus current practice, probabilistic model

Option	QALYs	Costs	Inc. QALYs	Inc. costs	ICER (per QALY gained)
LN0 NPI≤3.4					
Oncotype DX	13.89	£5,474	0.01	£1,313	**£122,725**
No test	13.88	£4,161	-	-	-
LN0 NPI>3.4					
Oncotype DX	12.73	£11,806	-0.01	£881	**Dominated**
No test	12.74	£10,925	-	-	-
LN+ (1-3 nodes)					
Oncotype DX	12.48	£13,212	-0.07	£687	**Dominated**
No test	12.55	£12,525	-	-	-

Inc. - incremental

Table 138: Probability of optimality – Oncotype DX versus current practice

Subgroup	Probability (λ=£20,000 per QALY gained)		Probability (λ=£30,000 per QALY gained)	
	Oncotype DX	Current practice	Oncotype DX	Current practice
LN0 NPI≤3.4	0.00	1.00	0.00	1.00
LN0 NPI>3.4	0.01	0.99	0.04	0.96
LN+ (1-3 nodes)	0.00	1.00	0.01	0.99

Deterministic sensitivity analysis

The results of the DSAs for Oncotype DX are presented in Table 139 (shaded cells reflect analyses which are unchanged from the EAG's base case). The DSAs indicate the following results across the three subgroups:

- *LN0 NPI≤3.4:* The ICER for Oncotype DX versus current practice remains in excess of £34,000 per QALY gained across all scenarios. The only analysis in which the ICER is below £70,000 per QALY gained relates to the scenario in which Oncotype DX is assumed to be predictive of chemotherapy benefit, based on RRs reported by Paik *et al*.[49]
- *LN0 NPI>3.4:* Oncotype is either dominated or has an ICER in excess of £35,000 per QALY gained across almost all scenarios. The only exception is the scenario in which Oncotype DX is assumed to be predictive of chemotherapy benefit, based on RRs reported by Paik *et al*.[49] Within this analysis, Oncotype DX dominates current practice.
- *LN+ (1-3 nodes):* Oncotype DX remains dominated across the majority of scenarios tested. The exceptions are: (i) the scenario in which Oncotype DX is assumed to be predictive of chemotherapy benefit, based on treatment effect estimates reported by Albain *et al*,[68] and (ii) the scenario in which the cost of adjuvant chemotherapy is doubled.

Table 139: Deterministic sensitivity analyses – Oncotype DX versus current practice

Oncotype DX	LN0 NPI≤3.4			LN0 NPI>3.4			LN+ (1-3 nodes)		
Scenario	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**
Base case (deterministic)	0.01	£1,317	**£120,144**	-0.02	£869	**Dominated**	-0.07	£647	**Dominated**
LN0 NPI≤3.4 post-test P(chemo) NHSE	0.01	£1,458	**£117,326**	-0.02	£869	**Dominated**	-0.07	£647	**Dominated**
LN0 NPI≤3.4 post-test P(chemo) Holt *et al*	0.01	£1,849	**£173,680**	-0.02	£869	**Dominated**	-0.07	£647	**Dominated**
LN0 NPI≤3.4 post-test P(chemo) Loncaster *et al*	0.01	£1,640	**£129,527**	-0.02	£869	**Dominated**	-0.07	£647	**Dominated**
LN0 NPI>3.4 post-test P(chemo) Holt *et al*	0.01	£1,317	**£120,144**	0.02	£1,138	**£60,831**	-0.07	£647	**Dominated**
LN0 NPI>3.4 post-test P(chemo) Loncaster *et al*	0.01	£1,317	**£120,144**	0.00	£999	**£651,857**	-0.07	£647	**Dominated**
LN0 NPI>3.4 post-test P(chemo) UKBCG survey	0.01	£1,317	**£120,144**	0.00	£978	**Dominated**	-0.07	£647	**Dominated**
LN0 NPI>3.4 baseline P(chemo) NCRAS, post-test P(chemo) Holt *et al*	0.01	£1,317	**£120,144**	0.03	£1,207	**£44,817**	-0.07	£647	**Dominated**
LN0 NPI>3.4 baseline P(chemo) NCRAS, post-test P(chemo) Loncaster LN0	0.01	£1,317	**£120,144**	0.01	£1,069	**£109,429**	-0.07	£647	**Dominated**
LN0 NPI>3.4 baseline P(chemo) NCRAS, post-test P(chemo) UKBCG survey	0.01	£1,317	**£120,144**	0.01	£1,048	**£161,535**	-0.07	£647	**Dominated**
LN+ post-test P(chemo) UKBCG survey	0.01	£1,317	**£120,144**	-0.02	£869	**Dominated**	0.00	£1,155	**Dominated**
Baseline P(chemo) adjusted by Oncotype DX RS score	0.01	£1,317	**£120,144**	-0.02	£888	**Dominated**	-0.07	£647	**Dominated**
Ward *et al* scenario - baseline P(chemo) WMCIU, post-test P(chemo) Holt *et al*	0.01	£1,317	**£120,144**	0.04	£1,268	**£35,782**	-0.07	£647	**Dominated**
Oncotype predictive benefit	0.04	£1,211	**£34,245**	0.27	-£364	**Dominating**	0.09	-£68	**Dominating**
Oncotype RSPC LN0	0.02	£1,146	**£70,435**	-0.02	£847	**Dominated**	-0.07	£647	**Dominated**

Chemotherapy disutility doubled	0.01	£1,317	**£121,879**	-0.01	£869	**Dominated**	-0.06	£647	**Dominated**
Chemotherapy disutility halved	0.01	£1,317	**£119,294**	-0.02	£869	**Dominated**	-0.08	£647	**Dominated**
Start age based on TransATAC (64 years)	0.01	£1,319	**£156,971**	-0.01	£867	**Dominated**	-0.05	£638	**Dominated**
Farkkila utilities (RFS=0.818, DM=0.746)	0.01	£1,317	**£125,021**	-0.01	£869	**Dominated**	-0.07	£647	**Dominated**
Chemotherapy RR=0.70	0.01	£1,305	**£94,920**	-0.02	£905	**Dominated**	-0.10	£759	**Dominated**
Chemotherapy RR=0.80	0.01	£1,325	**£145,102**	-0.01	£845	**Dominated**	-0.05	£573	**Dominated**
No risk tapering	0.01	£1,292	**£92,613**	-0.03	£974	**Dominated**	-0.10	£870	**Dominated**
Distant metastases death rate doubled	0.01	£1,339	**£106,090**	-0.02	£803	**Dominated**	-0.09	£443	**Dominated**
Distant metastases death rate halved	0.01	£1,282	**£154,090**	-0.01	£974	**Dominated**	-0.05	£972	**Dominated**
AML removed	0.01	£1,318	**£119,771**	-0.03	£879	**Dominated**	-0.09	£663	**Dominated**
Chemotherapy cost doubled	0.01	£1,330	**£121,322**	-0.02	£374	**Dominated**	-0.07	-£266	**£3,700**
Chemotherapy cost halved	0.01	£1,311	**£119,554**	-0.02	£1,116	**Dominated**	-0.07	£1,103	**Dominated**
Endocrine therapy costs doubled	0.01	£1,317	**£120,149**	-0.02	£869	**Dominated**	-0.07	£646	**Dominated**
Endocrine therapy costs halved	0.01	£1,317	**£120,141**	-0.02	£869	**Dominated**	-0.07	£647	**Dominated**
Local and distant recurrence costs doubled	0.01	£1,268	**£115,630**	-0.02	£1,017	**Dominated**	-0.07	£1,106	**Dominated**
Local and distant recurrence costs halved	0.01	£1,342	**£122,400**	-0.02	£795	**Dominated**	-0.07	£417	**Dominated**

NHSE – NHS England; UKBCG – UK Breast Cancer Group; RSPC - Recurrence score pathology-clinical; NCRAS - National Cancer Registration and Analysis Service; WMCIU - West Midlands Cancer Intelligence Unit; RFS – recurrence-free survival; DM – distant metastasis; P(chemo) – probability of receiving chemotherapy

IHC4+C versus current practice

Central estimates of cost-effectiveness - (probabilistic)

Central estimates of cost-effectiveness for IHC4+C versus current practice are presented in Table 140. All estimates are based on the probabilistic version of the EAG model. Within the LN0 NPI≤3.4 subgroup, IHC4+C is expected to produce 0.01 additional QALYs at an additional cost of £22 per woman tested compared with current practice; this corresponds to an ICER of £2,654 per QALY gained. Within the LN0 NPI>3.4 subgroup, IHC4+C is expected to produce 0.01 additional QALYs and cost savings of £90 per woman tested compared with current practice; within this subgroup, IHC4+C is expected to dominate current practice. Within the LN+ (1-3 nodes) subgroup, IHC4+C is expected to produce 0.05 additional QALYs and cost savings of £287 per woman tested compared with current practice; within this subgroup, IHC4+C is expected to dominate current practice. As shown in Table **141**, the PSA indicates that within the LN0 NPI≤3.4 subgroup, the probability that IHC4+C produces more net benefit than current practice at WTP thresholds of £20,000 and £30,000 per QALY gained is 0.95 and 0.97, respectively. Within the LN0 NPI>3.4 subgroup, the probability that IHC4+C produces more net benefit than current practice at WTP thresholds of £20,000 and £30,000 per QALY gained is 0.69 and 0.67, respectively. Within the LN+ (1-3 nodes) subgroup, the probability that IHC4+C produces more net benefit than current practice at these WTP thresholds is 0.94 or higher.

Table 140: Central estimates of cost-effectiveness – IHC4+C versus current practice, probabilistic model

Option	QALYs	Costs	Inc. QALYs	Inc. costs	ICER (per QALY gained)
LN0 NPI≤3.4					
IHC4+C	13.86	£4,291	0.01	£22	**£2,654**
No test	13.86	£4,269	-	-	-
LN0 NPI>3.4					
IHC4+C	12.73	£10,941	0.01	-£90	**Dominating**
No test	12.72	£11,031	-	-	-
LN+ (1-3 nodes)					
IHC4+C	12.59	£12,268	0.05	-£287	**Dominating**
No test	12.54	£12,554	-	-	-

Inc. – incremental

Table 141: Probability of optimality – IHC4+C versus current practice

Subgroup	Probability (λ=£20,000 per QALY gained)		Probability (λ=£30,000 per QALY gained)	
	IHC4+C	Current practice	IHC4+C	Current practice
LN0 NPI≤3.4	0.95	0.05	0.97	0.03
LN0 NPI>3.4	0.69	0.31	0.67	0.33
LN+ (1-3 nodes)	0.95	0.05	0.94	0.06

Deterministic sensitivity analysis

The results of the DSAs for IHC4+C are presented in Table 142 (shaded cells reflect analyses which are unchanged from the EAG's base case). The DSAs indicate the following:

- *LN0 NPI≤3.4:* The ICER for IHC4+C versus current practice remains below £16,000 per QALY gained across all scenarios, except the analysis in which post-test chemotherapy probabilities are derived from Holt *et al*.[283] IHC4+C dominates current practice in the scenario in which the cost of adjuvant chemotherapy is doubled.
- *LN0 NPI>3.4:* IHC4+C dominates current practice or has an ICER below £6,000 per QALY gained across all scenarios.
- *LN+ (1-3 nodes):* IHC4+C dominates current practice across all scenarios except the analysis in which the probability of receiving chemotherapy conditional on IHC4+C risk level is based on the UKBCG survey; within this analysis, the ICER is estimated to be £1,929 per QALY gained.

Table 142: Deterministic sensitivity analyses – IHC4+C versus current practice

IHC4+C	LN0 NPI≤3.4			LN0 NPI>3.4			LN+ (1-3 nodes)		
Scenario	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**
Base case (deterministic)	0.01	£22.43	**£2,752**	0.01	-£89.12	**Dominating**	0.05	-£269.39	**Dominating**
LN0 NPI≤3.4 post-test P(chemo) NHSE	0.01	£94.18	**£9,265**	0.01	-£89.12	**Dominating**	0.05	-£269.39	**Dominating**
LN0 NPI≤3.4 post-test P(chemo) Holt *et al*	0.01	£390.39	**£36,259**	0.01	-£89.12	**Dominating**	0.05	-£269.39	**Dominating**
LN0 NPI≤3.4 post-test P(chemo) Loncaster *et al*	0.01	£195.20	**£15,875**	0.01	-£89.12	**Dominating**	0.05	-£269.39	**Dominating**
LN0 NPI>3.4 post-test P(chemo) Holt *et al*	0.01	£22.43	**£2,752**	0.05	£194.16	**£4,147**	0.05	-£269.39	**Dominating**
LN0 NPI>3.4 post-test P(chemo) Loncaster *et al*	0.01	£22.43	**£2,752**	0.03	£52.99	**£1,864**	0.05	-£269.39	**Dominating**
LN0 NPI>3.4 post-test P(chemo) UKBCG survey	0.01	£22.43	**£2,752**	0.02	£23.00	**£1,040**	0.05	-£269.39	**Dominating**
LN0 NPI>3.4 baseline P(chemo) NCRAS, post-test P(chemo) Holt *et al*	0.01	£22.43	**£2,752**	0.06	£262.95	**£4,760**	0.05	-£269.39	**Dominating**
LN0 NPI>3.4 baseline P(chemo) NCRAS, post-test P(chemo) Loncaster LN0	0.01	£22.43	**£2,752**	0.04	£121.78	**£3,305**	0.05	-£269.39	**Dominating**
LN0 NPI>3.4 baseline P(chemo) NCRAS, post-test P(chemo) UKBCG survey	0.01	£22.43	**£2,752**	0.03	£91.80	**£3,005**	0.05	-£269.39	**Dominating**
LN+ post-test P(chemo) UKBCG survey	0.01	£22.43	**£2,752**	0.01	-£89.12	**Dominating**	0.09	£167.12	**£1,929**
Ward *et al* scenario - baseline P(chemo) WMCIU, post-test P(chemo) Holt *et al*	0.01	£22.43	**£2,752**	0.06	£325.33	**£5,160**	0.05	-£269.39	**Dominating**
Chemotherapy disutility doubled	0.01	£22.43	**£2,304**	0.01	-£89.12	**Dominating**	0.05	-£269.39	**Dominating**
Chemotherapy disutility	0.01	£22.43	**£3,049**	0.01	-£89.12	**Dominating**	0.05	-£269.39	**Dominating**

halved									
Start age based on TransATAC (64 years)	0.01	£23.21	**£3,542**	0.01	-£88.85	**Dominating**	0.04	-£264.75	**Dominating**
Farkkila utilities (RFS=0.818, DM=0.746)	0.01	£22.43	**£2,802**	0.01	-£89.12	**Dominating**	0.05	-£269.39	**Dominating**
Chemotherapy RR=0.70	0.01	£19.09	**£2,138**	0.01	-£86.90	**Dominating**	0.06	-£314.00	**Dominating**
Chemotherapy RR=0.80	0.01	£24.62	**£3,223**	0.01	-£90.60	**Dominating**	0.05	-£240.14	**Dominating**
No risk tapering	0.01	£19.17	**£2,221**	0.00	-£51.24	**Dominating**	0.06	-£282.61	**Dominating**
Distant metastases death rate doubled	0.01	£28.48	**£3,309**	0.01	-£93.23	**Dominating**	0.06	-£188.59	**Dominating**
Distant metastases death rate halved	0.01	£12.78	**£1,722**	0.01	-£82.69	**Dominating**	0.04	-£398.33	**Dominating**
AML removed	0.00	£26.13	**£5,560**	0.00	-£83.05	**Dominating**	0.04	-£260.08	**Dominating**
Chemotherapy cost doubled	0.01	-£108.78	**Dominating**	0.01	-£326.21	**Dominating**	0.05	-£499.21	**Dominating**
Chemotherapy cost halved	0.01	£88.03	**£10,803**	0.01	£29.42	**£4,056**	0.05	-£154.48	**Dominating**
Endocrine therapy costs doubled	0.01	£22.45	**£2,755**	0.01	-£89.10	**Dominating**	0.05	-£269.11	**Dominating**
Endocrine therapy costs halved	0.01	£22.41	**£2,751**	0.01	-£89.14	**Dominating**	0.05	-£269.53	**Dominating**
Local and distant recurrence costs doubled	0.01	£8.80	**£1,079**	0.01	-£79.87	**Dominating**	0.05	-£451.35	**Dominating**
Local and distant recurrence costs halved	0.01	£29.24	**£3,588**	0.01	-£93.75	**Dominating**	0.05	-£178.41	**Dominating**

NHSE – NHS England; UKBCG – UK Breast Cancer Group; NCRAS - National Cancer Registration and Analysis Service; WMCIU - West Midlands Cancer Intelligence Unit; RFS – recurrence-free survival; DM – distant metastasis; P(chemo) – probability of receiving chemotherapy

Prosigna versus current practice

Central estimates of cost-effectiveness - (probabilistic)

Central estimates of cost-effectiveness for Prosigna versus current practice are presented in Table 143. All estimates are based on the probabilistic version of the EAG model. Within the LN0 NPI≤3.4 subgroup, Prosigna is expected to produce 0.02 additional QALYs at an additional cost of £1,884 per woman tested compared with current practice; this corresponds to an ICER of £91,028 per QALY gained. Within the LN0 NPI>3.4 subgroup, Prosigna is expected to produce 0.06 additional QALYs at an additional cost of £1,686 per woman tested compared with current practice; the corresponding ICER is £26,058 per QALY gained. Within the LN+ (1-3 nodes) subgroup, Prosigna is expected to produce 0.07 additional QALYs at an additional cost of £1,936 per woman tested compared with current practice; the corresponding ICER is £28,731 per QALY gained. As shown in Table 144, the PSA indicates that within the LN0 NPI≤3.4 subgroup, the probability that Prosigna produces more net benefit than current practice at WTP thresholds of £20,000 and £30,000 per QALY gained is approximately zero. Within the LN0 NPI>3.4 subgroup, the probability that Prosigna produces more net benefit than current practice at these WTP thresholds is 0.24 and 0.60, respectively. Within the LN+ (1-3 nodes) subgroup, the probability that Prosigna produces more net benefit than current practice at these WTP thresholds is 0.24 and 0.55, respectively.

Table 143: Central estimates of cost-effectiveness – Prosigna versus current practice, probabilistic model

Option	QALYs	Costs	Inc. QALYs	Inc. costs	ICER (per QALY gained)
LN0 NPI≤3.4					
Prosigna	13.87	£6,201	0.02	£1,884	**£91,028**
No test	13.84	£4,318	-	-	-
LN0 NPI>3.4					
Prosigna	12.65	£13,330	0.06	£1,686	**£26,058**
No test	12.59	£11,644	-	-	-
LN+ (1-3 nodes)					
Prosigna	12.47	£15,172	0.07	£1,936	**£28,731**
No test	12.40	£13,236	-	-	-

Inc. - incremental

Table 144: Probability of optimality – Prosigna versus current practice

Subgroup	Probability (λ=£20,000 per QALY gained)		Probability (λ=£30,000 per QALY gained)	
	Prosigna	Current practice	Prosigna	Current practice
LN0 NPI≤3.4	0.00	1.00	0.00	1.00
LN0 NPI>3.4	0.24	0.76	0.60	0.40
LN+ (1-3 nodes)	0.24	0.76	0.55	0.45

Deterministic sensitivity analysis

The results of the DSAs for Prosigna are presented in Table 145 (shaded cells reflect analyses which are unchanged from the EAG's base case). The DSAs indicate the following:

- *LN0 NPI≤3.4:* The ICER for Prosigna versus current practice is estimated to be greater than £71,000 per QALY gained across all scenarios.
- *LN0 NPI>3.4:* The ICER for Prosigna versus current practice is estimated to be below £30,000 per QALY gained across most scenarios. The DSAs indicate that the ICER for Prosigna versus current practice is greater than £30,000 per QALY gained for scenarios in which: (i) an older start age is assumed, and (ii) the RR of distant metastases for chemotherapy versus no chemotherapy is set equal to 0.80.
- *LN+ (1-3 nodes):* The ICER for Prosigna versus current practice is estimated to be consistently below £38,000 per QALY gained across all analyses. Less favourable ICERs were estimated for scenarios in which: (i) the disutility associated with chemotherapy-related AEs is doubled; (ii) an older cohort age is assumed; (iii) the RR of distant metastases for chemotherapy versus no chemotherapy is set equal to 0.80, (iv) the cost of chemotherapy is doubled; (v) the costs of treating local and distant recurrence is halved; (vi) the mortality rate for distant metastases is halved, and (vii) the cost per test is assumed to be increased due to lower efficiency. The analysis in which risk classification probabilities and associated DMFS probabilities were taken from Gnant and Filipits[54] was not evaluable as no events occurred at 10-years within the low-risk Prosigna category.

Table 145: Deterministic sensitivity analyses – Prosigna versus current practice

Prosigna	LN0 NPI≤3.4			LN0 NPI>3.4			LN0 (1-3 nodes)		
Scenario	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**
Base case (deterministic)	0.02	£1,891.35	**£89,693**	0.07	£1,712.67	**£25,857**	0.07	£1,966.54	**£28,666**
LN0 NPI≤3.4 post-test P(chemo) NHSE	0.02	£2,025.87	**£84,090**	0.07	£1,712.67	**£25,857**	0.07	£1,966.54	**£28,666**
LN0 NPI≤3.4 post-test P(chemo) Holt *et al*	0.02	£2,421.22	**£109,620**	0.07	£1,712.67	**£25,857**	0.07	£1,966.54	**£28,666**
LN0 NPI≤3.4 post-test P(chemo) Loncaster *et al*	0.02	£2,213.71	**£93,938**	0.07	£1,712.67	**£25,857**	0.07	£1,966.54	**£28,666**
LN0 NPI>3.4 post-test P(chemo) Holt *et al*	0.02	£1,891.35	**£89,693**	0.10	£1,991.89	**£19,356**	0.07	£1,966.54	**£28,666**
LN0 NPI>3.4 post-test P(chemo) Loncaster *et al*	0.02	£1,891.35	**£89,693**	0.09	£1,993.10	**£21,216**	0.07	£1,966.54	**£28,666**
LN0 NPI>3.4 post-test P(chemo) UKBCG survey	0.02	£1,891.35	**£89,693**	0.08	£1,820.85	**£22,420**	0.07	£1,966.54	**£28,666**
LN0 NPI>3.4 baseline P(chemo) NCRAS, post-test P(chemo) Holt *et al*	0.02	£1,891.35	**£89,693**	0.11	£2,056.25	**£18,288**	0.07	£1,966.54	**£28,666**
LN0 NPI>3.4 baseline P(chemo) NCRAS, post-test P(chemo) Loncaster LN0	0.02	£1,891.35	**£89,693**	0.09	£1,922.48	**£20,971**	0.07	£1,966.54	**£28,666**
LN0 NPI>3.4 baseline P(chemo) NCRAS, post-test P(chemo) UKBCG survey	0.02	£1,891.35	**£89,693**	0.09	£1,885.20	**£20,774**	0.07	£1,966.54	**£28,666**
LN+ post-test P(chemo) UKBCG survey	0.02	£1,891.35	**£89,693**	0.07	£1,712.67	**£25,857**	0.11	£2,227.53	**£20,427**
Ward *et al* scenario - baseline P(chemo) WMCIU, post-test P(chemo) Holt *et al*	0.02	£1,891.35	**£89,693**	0.13	£2,109.68	**£16,568**	0.07	£1,966.54	**£28,666**
Risk classification and DMFS probabilities from Gnant and Filipits LN+	0.02	£1,891.35	**£89,693**	0.07	£1,712.67	**£25,857**	Not evaluable		

10% lower cost per test due to increased efficiency (local NHS testing)	0.02	£1,694	**£80,348**	0.07	£1,516	**£22,882**	0.07	£1,769	**£25,794**
10% higher cost per test due to decreased efficiency (local NHS testing)	0.02	£2,088	**£99,038**	0.07	£1,910	**£28,832**	0.07	£2,164	**£31,539**
Chemotherapy disutility doubled	0.02	£1,891.35	**£90,123**	0.07	£1,712.67	**£25,935**	0.07	£1,966.54	**£30,026**
Chemotherapy disutility halved	0.02	£1,891.35	**£89,480**	0.07	£1,712.67	**£25,818**	0.07	£1,966.54	**£28,032**
Start age based on TransATAC (64 years)	0.02	£1,893.35	**£115,741**	0.05	£1,718.65	**£33,348**	0.05	£1,973.37	**£37,480**
Farkkila utilities (RFS=0.818, DM=0.746)	0.02	£1,891.35	**£93,183**	0.06	£1,712.67	**£26,854**	0.07	£1,966.54	**£29,913**
Chemotherapy RR=0.70	0.03	£1,869.14	**£71,107**	0.08	£1,643.59	**£19,926**	0.09	£1,884.89	**£21,508**
Chemotherapy RR=0.80	0.02	£1,905.92	**£107,875**	0.06	£1,757.96	**£31,645**	0.06	£2,020.11	**£36,018**
No risk tapering	0.02	£1,870.73	**£78,043**	0.07	£1,681.49	**£23,298**	0.08	£1,874.66	**£23,138**
Distant metastases death rate doubled	0.02	£1,931.61	**£80,059**	0.08	£1,837.80	**£24,281**	0.08	£2,114.58	**£26,505**
Distant metastases death rate halved	0.02	£1,827.15	**£112,523**	0.05	£1,513.02	**£29,575**	0.05	£1,730.53	**£34,081**
AML removed	0.02	£1,892.76	**£91,182**	0.06	£1,717.28	**£26,432**	0.07	£1,965.75	**£26,851**
Chemotherapy cost doubled	0.02	£1,899.68	**£90,088**	0.07	£1,729.27	**£26,107**	0.07	£2,223.61	**£32,414**
Chemotherapy cost halved	0.02	£1,887.19	**£89,496**	0.07	£1,704.37	**£25,731**	0.07	£1,838.01	**£26,793**
Endocrine therapy costs doubled	0.02	£1,891.48	**£89,699**	0.07	£1,713.06	**£25,863**	0.07	£1,966.96	**£28,673**
Endocrine therapy costs halved	0.02	£1,891.29	**£89,690**	0.07	£1,712.47	**£25,854**	0.07	£1,966.33	**£28,663**
Local and distant recurrence costs doubled	0.02	£1,800.69	**£85,393**	0.07	£1,430.87	**£21,602**	0.07	£1,633.14	**£23,806**
Local and distant recurrence costs halved	0.02	£1,936.69	**£91,843**	0.07	£1,853.57	**£27,984**	0.07	£2,133.24	**£31,096**

NHSE – NHS England; UKBCG – UK Breast Cancer Group; NCRAS - National Cancer Registration and Analysis Service; WMCIU - West Midlands Cancer Intelligence Unit; RFS – recurrence-free survival; DM – distant metastasis; P(chemo) – probability of receiving chemotherapy

EPClin versus current practice

Central estimates of cost-effectiveness - (probabilistic)

Central estimates of cost-effectiveness for EPClin versus current practice are presented in Table 146. All estimates are based on the probabilistic version of the EAG model. Within the LN0 NPI≤3.4 subgroup, EPClin is expected to produce 0.01 additional QALYs at an additional cost of £1,679 per woman tested compared with current practice; this corresponds to an ICER of £147,419 per QALY gained. Within the LN0 NPI>3.4 subgroup, EPClin is expected to produce 0.03 additional QALYs at an additional cost of £1,388 per woman tested compared with current practice; the corresponding ICER is £46,788 per QALY gained. Within the LN+ (1-3 nodes) subgroup, EPClin is expected to produce 0.05 additional QALYs at an additional cost of £1,164 per woman tested compared with current practice; the corresponding ICER is £21,458 per QALY gained. As shown in Table 147, the PSA indicates that within the LN0 NPI≤3.4 subgroup, the probability that EPClin produces more net benefit than current practice at WTP thresholds of £20,000 and £30,000 per QALY gained is zero. Within the LN0 NPI>3.4 subgroup, the probability that EPClin produces more net benefit than current practice at these WTP thresholds is 0.09 and 0.26, respectively. Within the LN+ (1-3 nodes) subgroup, the probability that EPClin produces more net benefit than current practice at these WTP thresholds is 0.44 and 0.73, respectively.

Table 146: Central estimates of cost-effectiveness – EPClin versus current practice, probabilistic model

Option	QALYs	Costs	Inc. QALYs	Inc. costs	ICER (per QALY gained)
LN0 NPI≤3.4					
EPClin	13.85	£6,034	0.01	£1,679	**£147,419**
No test	13.84	£4,355	-	-	-
LN0 NPI>3.4					
EPClin	12.71	£12,612	0.03	£1,388	**£46,788**
No test	12.68	£11,224	-	-	-
LN+ (1-3 nodes)					
EPClin	12.52	£14,080	0.05	£1,164	**£21,458**
No test	12.46	£12,916	-	-	-

Inc. - incremental

Table 147: Probability of optimality – EPClin versus current practice

Subgroup	Probability (λ=£20,000 per QALY gained)		Probability (λ=£30,000 per QALY gained)	
	EPClin	Current practice	EPClin	Current practice
LN0 NPI≤3.4	0.00	1.00	0.00	1.00
LN0 NPI>3.4	0.09	0.91	0.26	0.74
LN0 (1-3 nodes)	0.44	0.56	0.73	0.27

Deterministic sensitivity analysis

The results of the DSAs for EPClin are presented in Table 148 (shaded cells reflect analyses which are unchanged from the EAG's base case). The DSAs indicate the following:

- *LN0 NPI≤3.4:* The ICER for EPClin versus current practice remains in excess of £91,000 per QALY gained across all scenarios.
- *LN0 NPI>3.4:* The ICER for EPClin versus current practice remains in excess of £30,000 per QALY gained across almost all of the analyses. The exceptions are the scenarios in which: (i) the UKBCG survey is used to inform the probability of receiving chemotherapy conditional on the EPClin test result, and (ii) Cusumano *et al* is used to inform the probability of receiving chemotherapy conditional on the EPClin test result.
- *LN+ (1-3 nodes):* The ICER for EPClin versus current practice remains below £30,000 per QALY gained across all scenarios.

Table 148: Deterministic sensitivity analyses – EPClin versus current practice

EPClin	LN0 NPI≤3.4			LN0 NPI>3.4			LN+ (1-3 nodes)		
Scenario	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**
Base case (deterministic)	0.01	£1,685.68	**£141,848**	0.03	£1,400.62	**£46,482**	0.06	£1,184.94	**£21,489**
Post-test P(chemo) UKBCG survey	0.01	£1,470.85	**£101,514**	0.06	£1,630.80	**£25,250**	0.12	£1,632.35	**£13,132**
Chemotherapy disutility doubled	0.01	£1,685.68	**£181,242**	0.03	£1,400.62	**£46,938**	0.06	£1,184.94	**£21,140**
Chemotherapy disutility halved	0.01	£1,685.68	**£127,943**	0.03	£1,400.62	**£46,257**	0.05	£1,184.94	**£21,667**
Risk classification and DMFS Dubsky LN+	0.01	£1,685.68	**£141,848**	0.03	£1,400.62	**£46,482**	0.05	£1,179.22	**£21,450**
Post-test P(chemo) Penault-Llorca *et al* LN0	0.02	£1,515.12	**£91,800**	0.04	£1,425.80	**£33,212**	0.06	£1,184.94	**£21,489**
Post-test P(chemo) Cusumano *et al*	0.02	£1,673.61	**£109,964**	0.06	£1,532.67	**£26,689**	0.14	£1,668.00	**£12,205**
10% lower cost per test due to increased efficiency (local NHS testing)	0.01	£1,536	**£129,225**	0.03	£1,251	**£41,504**	0.06	£1,035	**£18,768**
Start age based on TransATAC (64 years)	0.01	£1,687.17	**£194,520**	0.02	£1,403.45	**£60,061**	0.04	£1,190.15	**£27,705**
Farkkila utilities (RFS=0.818, DM=0.746)	0.01	£1,685.68	**£150,858**	0.03	£1,400.62	**£48,314**	0.05	£1,184.94	**£22,275**
Chemotherapy RR=0.70	0.02	£1,664.55	**£99,445**	0.04	£1,368.43	**£36,317**	0.07	£1,130.67	**£16,663**
Chemotherapy RR=0.80	0.01	£1,699.55	**£195,508**	0.03	£1,421.74	**£56,485**	0.05	£1,220.54	**£26,089**
No risk tapering	0.02	£1,638.80	**£94,376**	0.03	£1,380.99	**£41,242**	0.06	£1,146.32	**£18,707**
Distant metastases death rate doubled	0.01	£1,724.04	**£116,644**	0.03	£1,458.96	**£42,242**	0.06	£1,283.29	**£20,510**
Distant metastases death rate halved	0.01	£1,624.61	**£223,409**	0.02	£1,307.57	**£56,592**	0.04	£1,028.09	**£23,745**
AML removed	0.02	£1,681.60	**£99,734**	0.03	£1,402.34	**£46,797**	0.05	£1,190.90	**£22,954**
Chemotherapy cost doubled	0.01	£1,899.47	**£159,838**	0.03	£1,424.85	**£47,286**	0.06	£1,109.61	**£20,122**
Chemotherapy cost halved	0.01	£1,578.79	**£132,853**	0.03	£1,388.51	**£46,080**	0.06	£1,222.61	**£22,172**

Endocrine therapy costs doubled	0.01	£1,685.77	**£141,855**	0.03	£1,400.80	**£46,488**	0.06	£1,185.25	**£21,494**
Endocrine therapy costs halved	0.01	£1,685.64	**£141,844**	0.03	£1,400.53	**£46,479**	0.06	£1,184.79	**£21,486**
Local and distant recurrence costs doubled	0.01	£1,599.30	**£134,579**	0.03	£1,269.24	**£42,122**	0.06	£963.46	**£17,472**
Local and distant recurrence costs halved	0.01	£1,728.87	**£145,482**	0.03	£1,466.31	**£48,662**	0.06	£1,295.69	**£23,497**

UK Breast Cancer Group; NCRAS - National Cancer Registration and Analysis Service; RFS – recurrence-free survival; DM – distant metastasis; P(chemo) – probability of receiving chemotherapy

MammaPrint versus current practice (mAOL)

Central estimates of cost-effectiveness - (probabilistic)

Central estimates of cost-effectiveness for MammaPrint versus current practice (mAOL) are presented in Table 149. Estimates are based on the probabilistic version of the EAG model. Within the overall MINDACT population, MammaPrint is expected to produce 0.01 additional QALYs at an additional cost of £1,760 per woman tested compared with current practice; this corresponds to an ICER of £131,482 per QALY gained. Within the mAOL high-risk subgroup, MammaPrint is expected to produce 0.04 less QALYs at an additional cost of £1,413; within this subgroup, MammaPrint is expected to be dominated by current practice. Within the mAOL low-risk subgroup, MammaPrint is expected to generate an additional 0.01 QALYs at an additional cost of £2,410; this corresponds to an expected ICER of £414,202 per QALY gained. The PSA indicates that within the overall MINDACT population and both subgroups, the probability that MammaPrint produces more net benefit than current practice at WTP thresholds of £20,000 and £30,000 per QALY gained is approximately zero (see Table 150).

Table 149: Central estimates of cost-effectiveness – MammaPrint versus current practice (mAOL), probabilistic model

Option	QALYs	Costs	Inc. QALYs	Inc. costs	ICER (per QALY gained)
MINDACT ITT population					
MammaPrint	13.51	£9,151	0.01	£1,760	**£131,482**
No test	13.49	£7,391	-	-	-
MINDACT mAOL high-risk subgroup					
MammaPrint	12.86	£12,727	-0.04	£1,413	**Dominated**
No test	12.90	£11,313	-	-	-
MINDACT mAOL low-risk subgroup					
MammaPrint	13.70	£7,777	0.01	£2,410	**£414,202**
No test	13.69	£5,366	-	-	-

Inc. - incremental

Table 150: Probability of optimality – MammaPrint versus current practice (mAOL)

Subgroup	Probability (λ=£20,000 per QALY gained)		Probability (λ=£30,000 per QALY gained)	
	MammaPrint	Current practice	MammaPrint	Current practice
MINDACT overall population	0.00	1.00	0.00	1.00
mAOL high-risk subgroup	0.00	1.00	0.00	1.00
mAOL low-risk subgroup	0.00	1.00	0.00	1.00

Deterministic sensitivity analysis

The results of the DSAs for MammaPrint are presented in Table 151 (shaded cells reflect analyses which are unchanged from the EAG's base case). The DSAs indicate the following:

- Within the overall MINDACT population, the ICER for MammaPrint versus current practice is estimated to be greater than £76,000 per QALY gained across all scenarios.
- Within the mAOL high-risk subgroup, MammaPrint is dominated by current practice across almost all scenarios. The most favourable ICER relates to the scenario in which the probability of receiving chemotherapy under current practice is halved.
- Within the mAOL low-risk subgroup, the ICER for MammaPrint versus current practice is greater than £161,000 per QALY gained across all analyses.

Table 151: Deterministic sensitivity analyses – MammaPrint versus current practice (mAOL)

MammaPrint	MINDACT ITT population			MINDACT mAOL high-risk subgroup			MINDACT mAOL low-risk subgroup		
Scenario	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**
Base case (deterministic)	0.01	£1,756.58	**£134,059**	-0.04	£1,380.11	**Dominated**	0.01	£2,415.05	**£399,182**
Risk classification and DMFS probabilities Van 't Veer *et al*	0.01	£1,609.52	**£169,183**	-0.04	£1,380.11	**Dominated**	0.01	£2,415.05	**£399,182**
ER+, HER2-, LN0 subgroup	0.01	£1,756.58	**£134,059**	-0.04	£1,400.94	**Dominated**	0.01	£2,415.05	**£399,182**
Post-test P(chemo) Penault-Llorca *et al* LN0	0.02	£1,724.59	**£97,939**	-0.03	£1,386.99	**Dominated**	0.01	£2,291.88	**£257,484**
Post-test P(chemo) Cusumano *et al*	0.02	£1,874.42	**£91,453**	-0.01	£1,492.18	**Dominated**	0.01	£2,454.55	**£336,904**
Post-test P(chemo) UKBCG	0.01	£1,610	**£130,970**	-0.01	£1,601	**Dominated**	-0.01	£3,421	**Dominated**
MP low-risk receive no chemotherapy; MP high-risk all receive chemotherapy	0.02	£1,846.54	**£76,201**	0.00	£1,497.09	**£375,444**	0.01	£2,350.50	**£242,895**
Baseline chemotherapy probabilities halved	0.03	£2,512.88	**£96,782**	0.07	£2,243.31	**£32,800**	0.00	£2,704.26	**£903,528**
Chemotherapy disutility doubled	0.02	£1,756.58	**£93,877**	-0.03	£1,380.11	**Dominated**	0.00	£2,415.05	**£503,351**
Chemotherapy disutility halved	0.01	£1,756.58	**£170,560**	-0.05	£1,380.11	**Dominated**	0.01	£2,415.05	**£361,750**
Start age based on TransATAC (64 years)	0.01	£1,757.57	**£158,110**	-0.03	£1,374.22	**Dominated**	0.00	£2,415.84	**£547,979**
Farkkila utilities (RFS=0.818, DM=0.746)	0.01	£1,756.58	**£133,215**	-0.04	£1,380.11	**Dominated**	0.01	£2,415.05	**£423,893**
Chemotherapy RR=0.70	0.01	£1,762.13	**£148,424**	-0.06	£1,431.22	**Dominated**	0.01	£2,377.25	**£161,338**
Chemotherapy RR=0.80	0.01	£1,753.86	**£127,971**	-0.03	£1,296.55	**Dominated**	0.01	£2,403.58	**£276,670**
No risk tapering	0.01	£1,784.24	**£173,280**	-0.07	£1,591.64	**Dominated**	0.01	£2,391.23	**£270,639**
Distant metastases death rate doubled	0.01	£1,747.73	**£140,551**	-0.06	£1,218.99	**Dominated**	0.01	£2,434.08	**£325,055**
Distant metastases death rate halved	0.01	£1,770.62	**£125,010**	-0.03	£1,636.40	**Dominated**	0.00	£2,384.61	**£636,029**

AML removed	0.00	£1,768.44	**£1,353,592**	-0.07	£1,401.41	**Dominated**	0.01	£2,413.62	**£291,353**
Chemotherapy cost doubled	0.01	£1,292.39	**£98,632**	-0.04	£351.31	**Dominated**	0.01	£2,518.68	**£416,311**
Chemotherapy cost halved	0.01	£1,988.67	**£151,772**	-0.04	£1,894.51	**Dominated**	0.01	£2,363.24	**£390,617**
Endocrine therapy costs doubled	0.01	£1,756.59	**£134,060**	-0.04	£1,379.77	**Dominated**	0.01	£2,415.09	**£399,189**
Endocrine therapy costs halved	0.01	£1,756.57	**£134,058**	-0.04	£1,380.28	**Dominated**	0.01	£2,415.03	**£399,178**
Local and distant recurrence costs doubled	0.01	£1,776.51	**£135,580**	-0.04	£1,743.07	**Dominated**	0.01	£2,372.26	**£392,109**
Local and distant recurrence costs halved	0.01	£1,746.61	**£133,298**	-0.04	£1,198.63	**Dominated**	0.01	£2,436.45	**£402,719**

UK Breast Cancer Group; NCRAS - National Cancer Registration and Analysis Service; RFS – recurrence-free survival; DM – distant metastasis; P(chemo) – probability of receiving chemotherapy

5.3.7 Comparison between the Genomic Health model, the current EAG model and the previous EAG model (LN0, clinical intermediate-risk subgroup)

There are notable differences between the cost-effectiveness estimates for Oncotype DX versus current practice generated using the current EAG model and those produced using the Genomic Health model[113] and the earlier EAG model reported by Ward *et al*:[18]

- The current EAG model indicates that within the ER+ LN0 NPI>3.4 subgroup, Oncotype DX is expected to be dominated by current practice. This finding contrasts sharply with the findings of the Genomic Health model and the previous EAG model.
- The Genomic Health model[113] produces a base case ICER of [redacted] per QALY gained, assuming that the test is used for women with ER+, LN0 early breast cancer who are deemed to be at clinical intermediate-risk.
- The previous EAG model (Ward *et al*[18]) produced a base-case ICER for Oncotype DX versus current practice of £22,572 per QALY gained, assuming that the test is given to women with ER+, LN0 early breast cancer with NPI>3.4 (deemed to be clinical intermediate-risk).

In order to understand the differences between these results, it is important to consider the differences between the key parameters and structural assumptions between the three models (see Table **152**):

- The general modelling approach is very similar between the three models, although the Ward *et al* model defined test risk classification according to both Oncotype DX RS and IHC4, rather than Oncotype DX RS only.
- Within the original and current EAG models, data on risk reclassification (the proportion of patients with a low, intermediate and high RS) were taken from analyses of the TransATAC trial[43] (albeit using different datasets). Conversely, the Genomic Health model derives these proportions from the NHS England Access Scheme dataset.[255]
- Data on the risk of distant recurrence in the absence of chemotherapy were taken from the ATAC trial in all three models.[281] The updated EAG model uses newer data from the ATAC trial.[43]
- The proportions of women who are assumed to receive chemotherapy conditional on the Oncotype DX risk score were taken from the NHS England Access Scheme dataset[255] in both the updated EAG model and the Genomic Health model. Ward *et al* used unpublished data (Holt *et al* 2013) to estimate the probability of receiving chemotherapy conditional on Oncotype DX RS.
- The proportion of patients receiving chemotherapy in the standard care arm was taken from the NHS England Access Scheme dataset[255] in both the updated EAG model and the Genomic Health model. Conversely, Ward *et al* derived estimates of these proportions from English cancer registry datasets.

- Within both the current and earlier EAG models, the benefit of chemotherapy was assumed to be constant across all Oncotype DX RS classifications (non-predictive); the RR of distant recurrence was taken from EBCTCG meta-analyses. The current EAG model uses a different mathematical approach to apply this RR which ensures that modelled treatment effect at 10-years is maintained within the Markov trace.
- The Genomic Health model assumes a predictive benefit and uses different treatment effects across the low, intermediate and high RS score classifications, based on Paik *et al*.[49] These differential effects are applied only to the Oncotype DX testing group; a constant treatment effect is applied in the current practice group.
- The current and earlier EAG models both apply an HRQoL decrement associated with short-term chemotherapy-related AEs in the first model cycle. In contrast, the Genomic Health model applies a decrement during every model cycle; the implicit assumption is that patients who receive adjuvant chemotherapy remain less well, relative to those do not receive adjuvant chemotherapy, for the remainder of their lives.
-

Table 152: Summary of structural assumptions and evidence sources

	Current EAG model	**Genomic Health model[113]**	**Original EAG model (Ward *et al*[18])**
Approach	Risk classification based on Oncotype DX RS	Risk classification based on Oncotype DX RS	Risk classification based on Oncotype DX RS and IHC4
Data on risk classification	TransATAC	NHS England Access Scheme dataset	TransATAC
Data on risk of recurrence	TransATAC (updated)	TransATAC	TransATAC
Proportion of people receiving chemotherapy in the oncotype arm	NHS England Access Scheme dataset	NHS England Access Scheme dataset	Holt *et al* (2013)
Proportion of people receiving chemotherapy in the standard care arm	NHS England Access Scheme dataset	NHS England Access Scheme dataset	Registry data
Benefit of adjuvant chemotherapy	No predictive effect (based on EBCTCG meta-analysis)	Predictive effect only in the Oncotype DX group. No predictive effect assumed in people with same risk score in current practice group.	No predictive effect (based on EBCTCG meta-analysis)
HRQoL decrement associated with chemotherapy	Applied to the first cycle only	Applied to all model cycles over patients' remaining lifetime	Applied to the first cycle only

As described in Section 5.2, the EAG identified several errors within the Genomic Health model. Three key errors are corrected here:

(a) *The application of the risk reclassification in the model.* Whilst Genomic Health use data from the NHS England Access Scheme dataset for the risk reclassification, this is applied incorrectly in the model. This can be seen by examining the proportion of women receiving chemotherapy predicted by the model. [redacted]

(b) *The application of the HRQoL decrement associated with chemotherapy-related AEs.* Page 159 of the Genomic Health dossier[113] states that: "*utilities in the present analysis were the same as those used by Ward et al. (2013)*", and Table 6-4 (page 159) of the dossier states that the disutility associated with chemotherapy is -0.038. However, the Genomic Health model applies this decrement for women receiving chemotherapy during every model cycle, including decades after the adjuvant treatment has been discontinued. This overestimates the health losses associated with chemotherapy and is therefore favourable to Oncotype DX, as the test is estimated to reduce the proportion of women receiving chemotherapy.

(c) *Predictive chemotherapy benefit:* The Genomic Health model assumes that women with a low, intermediate and high Oncotype DX RS experience different benefits of chemotherapy in the modelled Oncotype DX group compared with the same patients across these RS classifications in the modelled current practice group. Irrespective of whether Oncotype DX is predictive of chemotherapy benefit, the modelling approach adopted by the company is illogical, as the benefits of chemotherapy for women within these RS classifications will be identical irrespective of whether the test is used to classify that level of risk or not (they are exactly the same patients).

In order to understand the differences between the results of the three models, the errors identified above were corrected by the EAG. In addition, the Genomic Health model was modified to assume a prognostic benefit only, thereby making it consistent with the current EAG base case model. The earlier EAG model (Ward *et al*[18]) was also modified to include the chemotherapy probabilities (with and without the test) from the NHS England Access Scheme dataset. Whilst there are other differences between the models, these are either more difficult to align across the models and/or are expected to have only a negligible impact on results. The results of the current EAG model, the amended Ward *et al*[18] model and the corrected Genomic Health model are presented in Table 153 (assuming no predictive benefit) and Table 154 (assuming predictive benefit).

Table 153: ICER assuming no predictive effect (LN0 NPI>3.4 subgroup)

Model	QALYs		Costs		Inc. QALYs	Inc. costs	ICER
	Oncotype DX	No test	Oncotype DX	No test			
Current EAG model (no predictive effect)	12.68	12.70	£11,249	£10,380	-0.02	£869	**Dominated**
Uncorrected Genomic Health model[113] (with predictive effect)	10.50	10.43	[redacted]	[redacted]	0.07	[redacted]	[redacted]
Corrected Genomic health model (no predictive effect)	10.59	10.62	[redacted]	[redacted]	- 0.03	[redacted]	[redacted]
Ward *et al*[18] model (no predictive effect)	12.85	12.80	£10,172	£8,897	0.06	£1,275	**£22,572**
Ward *et al* model, including NHS England Access Scheme dataset for proportion of people who receive chemotherapy (no predictive effect)	12.83	12.83	£9,861	£9,253	- 0.00	£608	**Dominated**

Inc. - incremental

Table 154: ICER assuming predictive effect (LN0 NPI>3.4 subgroup)

Model	QALYs		Costs		Inc. QALYs	Inc. costs	ICER
	Oncotype DX	No test	Oncotype DX	No test			
Current EAG model (predictive effect)	12.87	12.60	£10,457	£10,822	0.27	-£364	**Dominating**
Uncorrected Genomic Health model (predictive effect) [113]	10.50	10.43	[redacted]	[redacted]	0.07	[redacted]	[redacted]
Corrected Genomic health model (predictive effect)	10.74	10.69	[redacted]	[redacted]	0.05	[redacted]	[redacted]
Ward *et al* model (predictive effect)	13.06	12.83	£9,681	£8,816	0.23	£865	**£3,720**
Ward *et al* model, including NHS England Access Scheme dataset for proportion of people who receive chemotherapy (including predictive effect)	13.02	12.91	£9,412	£9,078	0.11	£334	**£2,917**

Inc. - incremental

In the scenario in which all three models use pre- and post-test chemotherapy probabilities from the NHS England Access Scheme dataset and no predictive benefit is assumed (see Table 153), all three models produce the same economic conclusion: Oncotype DX is dominated by current practice. When a predictive effect is incorporated into these versions of the models (see Table 154), these three models consistently suggest that Oncotype DX has an ICER which is below £7,000 per QALY gained.

5.4 Discussion

The EAG undertook a systematic review of existing economic evaluations of tumour profiling tests to guide treatment decisions in people with early breast cancer. Only those studies which were published since the previous appraisal of tumour profiling tests (NICE DG10[21]) were included in the review. The review suggests a high level of consistency in terms of the general modelling approach and structure; the majority of published models adopted a decision tree - Markov approach based on test risk classification and DMFS outcomes conditional on test risk classification probabilities. None of the published analyses included all relevant tumour profiling tests listed in the final NICE scope.

Two manufacturers provided economic evidence to inform the appraisal (Agendia[121] and Genomic Health[113]). The models developed to inform these two analyses were made available to the EAG for scrutiny. In addition, the chief investigator of the EndoPredict UK decision impact study provided a draft cost-effectiveness paper which compares EPClin versus AOL.[225] The model supporting this analysis was not made available to the EAG.

[redacted]

Genomic Health provided a model which compares Oncotype DX versus current practice in patients with LN0 early breast cancer. The EAG notes that the model includes a number of errors. Based on the uncorrected model, the Genomic Health submission presents a base case

ICER for Oncotype DX versus current practice of [redacted] per QALY gained. Three errors were corrected by the EAG (see Section 5.3.7); these relate to: (i) the incorrect application of risk classifications; (ii) the application of health losses associated with short-term chemotherapy-related AEs during every model cycle, and (iii) the inconsistent handling of predictive benefits of chemotherapy between the test and no test groups. The EAG's corrected version of the model suggests that under the assumption of no predictive benefit of chemotherapy, Oncotype DX is dominated by current practice. When the test is assumed to be predictive of chemotherapy benefit, the ICER for Oncotype DX versus current practice is estimated to be [redacted] per QALY gained. The EAG notes that other errors may remain within the company's model.

The draft cost-effectiveness paper assessing EPClin versus AOL suggests that the expected ICER for EPClin versus AO! is £26,836 per QALY gained. The EAG has some concerns regarding this analysis, in particular, the use of separate evidence sources to estimate test risk classification probabilities and DMFS probabilities conditional on test risk classification.

The EAG developed a *de novo* health economic model to assess the cost-effectiveness of Oncotype DX, MammaPrint, Prosigna, EPClin and IHC4+C, each versus current practice. The health economic analysis was undertaken from the perspective of the NHS and PSS and was largely based on the model developed to inform NICE DG10.[18] The EAG model adopts a hybrid decision tree – Markov structure. The model parameters were informed by a number of sources including a bespoke analysis of the TransATAC trial,[43] the MINDACT trial,[134] a bespoke analysis of the NCRAS dataset,[254] a bespoke survey disseminated by the UKBCG, the NHS England Access Scheme Database,[255] standard costing sources and other literature. The EAG's base case model suggests the following results.

Oncotype DX: Within the LN0 NPI≤3.4 subgroup, the ICER for Oncotype DX versus current practice is expected to be £122,725 per QALY gained (£34,245 per QALY gained assuming a predictive benefit). Within the LN0 NPI>3.4 and LN+ (1-3 nodes) subgroups, Oncotype DX is expected to be dominated by current practice (conversely, Oncotype DX dominates current practice if a predictive benefit is assumed). The results generated using the EAG model are primarily driven by the modelled reduction in the use of adjuvant chemotherapy using the Oncotype DX test.

IHC4+C: Within the LN0 NPI≤3.4 subgroup, the ICER for IHC4+C versus current practice is expected to be £2,654 per QALY gained. Within the LN0 NPI>3.4 and LN+ (1-3 nodes) subgroups, IHC4+C is expected to dominate current practice.

Prosigna: Within the LN0 NPI≤3.4 subgroup, the ICER for Prosigna versus current practice is expected to be £91,028 per QALY gained. Within the LN0 NPI>3.4 and LN+ (1-3 nodes) subgroups, the ICERs for Prosigna versus current practice are £26,058 and £28,731 per QALY gained, respectively.

EPClin: Within the LN0 NPI≤3.4 subgroup, the ICER for EPClin versus current practice is expected to be £147,419 per QALY gained. Within the LN0 NPI>3.4 subgroup, the ICER for EPClin versus current practice is expected to be £46,788 per QALY gained. Within the LN+ (1-3 nodes) subgroup, the ICER for EPClin versus current practice is expected to be £21,458 per QALY gained.

MammaPrint: Within the overall MINDACT population, the ICER for MammaPrint versus current practice is expected to be £131,482 per QALY gained. Within the mAOL high-risk subgroup, MammaPrint is expected to be dominated by current practice. Within the mAOL low-risk subgroup, the ICER for MammaPrint versus current practice is expected to be £414,202 per QALY gained.

The EAG model is subject to the following strengths:

- The model structure is consistent with the general approach used in a number of previous economic analyses of tumour profiling tests for early breast cancer.
- For all tests, test risk classification probabilities and DMFS probabilities are derived from the same source – this maintains correlation between these parameters and avoids the potential for spectrum bias to produce spurious results.
- Within the LN0 intermediate-risk subgroup (NPI>3.4, analysis of 3-level tests), the probability of receiving chemotherapy with and without the test is based on the same source – the NHS England Access Scheme dataset.[255] The EAG takes the view that this source is likely to best reflect how the 3-level tumour profiling tests would be used in clinical practice in England. However, this evidence source relates only to the clinical intermediate-risk group; the UK-specific evidence surrounding decision impact within the LN0 NPI≤3.4 and LN+ subgroups is considerably weaker.
- When based on the same test risk classification probabilities, recurrence rates and the same estimates of pre- and post-test chemotherapy use, the EAG model produces similar results to the previous model reported by Ward *et al*[18] and the Genomic Health model.
- A large number of scenarios have been considered to explore the impact of alternative evidence choices and assumptions on the cost-effectiveness of the alternative tests.

The EAG model is also subject to a number of limitations and uncertainties:

- Test risk classification probabilities and DMFS probabilities for Oncotype DX, Prosigna, IHC4+C and EPClin are based on a post-menopausal population only (TransATAC). It is expected that the tumour profiling tests may also be used in pre-menopausal women.
- The subgroups employed within the analysis are defined according to NPI. In practice, other tools may be used to define risk, for example, PREDICT. The EAG explored the possibility of framing the analyses around PREDICT, however this was not possible as PREDICT scores were not available within either the TransATAC dataset or the NCRAS dataset, nor was an analysis presented by PREDICT within the publication of the MINDACT trial.[134] It may be possible to calculate PREDICT scores within these datasets in the future, however this would require access to the individual patient-level data.
- The analysis of MammaPrint using the MINDACT trial compares the test only against mAOL and may therefore not reflect current practice in England. This issue is particularly relevant to determining the baseline level of chemotherapy use for the current practice group within this population.
- Within the current practice group of the EAG model, the probability of receiving chemotherapy is assumed to be the same irrespective of test risk score. This is unlikely to be realistic, as those with higher test risk scores may already be more likely to receive adjuvant chemotherapy, whilst those with lower test risk scores may already be less likely to receive adjuvant chemotherapy. It was possible to explore this assumption for the evaluation of Oncotype DX within the sensitivity analyses (and the conclusions were unchanged), however there were insufficient data available to undertake similar analyses for the other four tests.
- The TransATAC trial was the derivation study for IHC4+C. This means that there is potential for the overestimation of prognostic performance; this leads to additional uncertainty around the likely cost-effectiveness profile of this test.
- The MINDACT trial used to inform the analyses of MammaPrint is limited as this study does not provide information regarding predictive benefit. In addition, the follow-up period for this study was limited to a duration of 5-years.
- Across all analyses, it is clear that the model results are dependent on assumptions about pre- and post-test chemotherapy use. This aspect of the evidence base is subject to considerable uncertainty. In particular, there is only one UK-based decision impact study relating to a 2-level tumour profiling test (Bloomfield *et al*[76]); the characteristics of patients enrolled into this study, and their relevance to the modelled

subgroups, are unclear. As shown in the DSAs, the use of alternative European studies[213, 216] and the UKBCG survey appear to lead to generally more favourable cost-effectiveness estimates for EPClin and MammaPrint. In addition, the use of the Loncaster *et al* study[196] to estimate chemotherapy use in the LN+ population may be biased as this study included a pre-selected population for whom chemotherapy had already been recommended.

- As Nanostring does not offer a centralised testing service for Prosigna, the cost per test will depend on the efficiency of local testing centres and the number of tests undertaken within each centre. This may affect the cost-effectiveness estimates presented here.
- The model does not include CHF as a long-term AE associated with adjuvant chemotherapy; this was excluded from the model due to a lack of evidence on the joint survival impact of CHF and metastatic breast cancer. Whilst CHF is a more common event than AML, the development of cancer is likely to have more serious consequences and is expected to be associated with a greater impact on health care resources.
- There is uncertainty surrounding whether Oncotype DX is predictive of chemotherapy benefit; based on the current EAG model, the inclusion of this potential test characteristic has a marked impact on the conclusions drawn from the analysis. Whilst the ongoing TAILORx study may generate additional evidence to inform this, the cut-offs used within this trial differ from those employed within the TransATAC analysis.
- Overall, there remains uncertainty regarding the cost-effectiveness of all tests. It is noteworthy that the inclusion of additional data collected through the NHS England Access Scheme Dataset has a significant impact upon the conclusions drawn from the Oncotype DX analysis within NICE DG10 (moving from an ICER of £22,572 per QALY gained to a situation in which Oncotype DX is dominated in the LN0 NPI>3.4 subgroup). The EAG considers that additional UK-based data collection relating to pre- and post-test chemotherapy use for EPClin, IHC4+C, Prosigna and MammaPrint may be important in reducing existing uncertainty surrounding the cost-effectiveness of these tests.

6 DISCUSSION AND CONCLUSIONS

6.1 Statement of principal findings

6.1.1 Clinical effectiveness – principal findings

The review included 153 studies across all five tests and across all outcomes listed in the NICE scope.

Among studies of LN0 patients receiving endocrine monotherapy, percentages categorised as high-risk ranged from 9-33% across all five tests. In LN+ patients, three tests (Prosigna/ROR-PT, EPClin [EndoPredict Clinical] and IHC4+C [IHC4 + clinical score]) categorised far more ([redacted] lymph node positive (LN+) than lymph node negative (LN0) patients as high-risk among studies of endocrine monotherapy, whilst Oncotype-DX categorised a similar number as high-risk in LN0 and LN+ groups. However, Oncotype DX categorised more patients as low-risk in LN+ than other tests (57% in Oncotype DX versus 4% to [redacted]% in other tests), but with worse 10-year distant-recurrence free survival/interval (DRFS/DRFI) outcomes (82% in Oncotype DX versus 95% to 100% in other tests).

In terms of prognostic performance, all tests had statistically significant prognostic power in unadjusted analyses in LN0 and LN+ populations. However, recurrence score pathology-clinical (RSPC) was only validated in LN0 patients, and unadjusted analyses using clinical cut-offs were not reported in the validation sets for IHC4 or IHC4+C. All tests provided additional prognostic information over most commonly used clinicopathological factors and over clinical treatment score (CTS) and Nottingham Prognostic Index (NPI) in LN0. Results were more varied in LN+ patients.

There was some evidence of differential chemotherapy benefit between risk groups for Oncotype DX as shown by significant interaction tests between risk group and chemotherapy treatment in unadjusted analyses, but interaction tests sometimes became non-significant when clinicopathological factors were adjusted for. Oncotype DX RSPC (Oncotype DX plus age, tumour size and grade) was prognostic but not statistically significantly predictive for chemotherapy benefit, indicating that the incorporation of CP factors to Oncotype DX may reduce prediction of chemotherapy benefit.

Evidence relating to the ability of MammaPrint to predict benefit from chemotherapy was extremely limited. Although the effect of chemotherapy was significant in high-risk groups and not in low-risk groups, interaction tests between risk groups and chemotherapy treatment were not significant, suggesting no statistically significant difference in effect of chemotherapy between risk groups.

For Oncotype DX and MammaPrint, evidence from observational, non-comparative studies assessing the impact of the test used prospectively in clinical practice suggested that recurrence/survival outcomes in low-risk groups were acceptable even with low rates of chemotherapy. There was no similar evidence relating to the other tests.

The MINDACT randomised controlled trial (RCT) for MammaPrint reported that for patients who were high-mAOL, low-MammaPrint risk, chemotherapy gave an absolute benefit of 1.5% in 5 year DRFI. This raises the possibility of avoiding chemotherapy in these patients. In patients who were low-mAOL, high-MammaPrint risk, chemotherapy gave an absolute benefit of 0.8%. This could be interpreted to mean MammaPrint would not be a useful test in mAOL low-risk patients, as it would not alter treatment decisions.

Decision impact studies from the UK and Europe reported that the percentage of patients with any change in chemotherapy recommendation or decision pre-/post-test ranged from 27% to 49% across UK studies (included Oncotype DX, EndoPredict and IHC4+C) and from 5% to 70% across European studies (included all tests except IHC4). The net change in the percentage of patients with a chemotherapy recommendation or decision pre-/post-test ranged from an increase of 1% to a decrease of 23% among UK studies, and a decrease of 0% to 64% across European studies.

Concordance between tests was not fully reviewed, but one UK study (OPTIMA prelim) which compared Oncotype DX, MammaPrint, Prosigna and IHC4 concluded that whilst tests assigned similar proportions of patients to low/intermediate and high-risk categories, test results for an individual patient could differ markedly depending on which test was used.

Data relating to anxiety and health-related quality of life (HRQoL) was limited as most studies did not include a comparator, instead adopting a pre-test/post-test design. Anxiety generally reduced post-test, but it is unclear if this would occur equally after a treatment decision made according to clinical factors. HRQoL improved in some analyses.

Microarray studies support conclusions from studies using the commercial versions of the assays in suggesting that Oncotype DX, MammaPrint and EndoPredict can discriminate between high- and low-risk patients regardless of LN status (there were no relevant microarray studies for EndoPredict or IHC4).

6.1.2 Cost-effectiveness – principal findings

The EAG developed a *de novo* health economic model to assess the cost-effectiveness of Oncotype DX, MammaPrint, Prosigna, EPClin and IHC4+C, each versus current practice. The health economic analysis was undertaken from the perspective of the NHS and PSS and was largely based on the model developed to inform NICE DG10. The EAG model adopts a hybrid decision tree – Markov structure. The model parameters were informed by a number of sources including a bespoke analysis of the TransATAC trial, the MINDACT trial, a bespoke analysis of the NCRAS dataset, a bespoke survey disseminated by the UKBCG, the NHS England Access Scheme Database, standard costing sources and other literature. The EAG's base case model suggests the following results.

Oncotype DX: Within the LN0 NPI≤3.4 subgroup, the ICER for Oncotype DX versus current practice is expected to be £122,725 per QALY gained (£34,245 per QALY gained assuming a predictive chemotherapy benefit). Within the LN0 NPI>3.4 and LN+ (1-3 nodes) subgroups, Oncotype DX is expected to be dominated by current practice (conversely, Oncotype DX dominates current practice if a predictive chemotherapy benefit is assumed). The results generated using the EAG model are primarily driven by the modelled reduction in the use of adjuvant chemotherapy using the Oncotype DX test.

IHC4+C: Within the LN0 NPI≤3.4 subgroup, the ICER for IHC4+C versus current practice is expected to be £2,654 per QALY gained. Within the LN0 NPI>3.4 and LN+ (1-3 nodes) subgroups, IHC4+C is expected to dominate current practice.

Prosigna: Within the LN0 NPI≤3.4 subgroup, the ICER for Prosigna versus current practice is expected to be £91,028 per QALY gained. Within the LN0 NPI>3.4 and LN+ (1-3 nodes) subgroups, the ICERs for Prosigna versus current practice are £26,058 and £28,731 per QALY gained, respectively.

EPClin: Within the LN0 NPI≤3.4 subgroup, the ICER for EPClin versus current practice is expected to be £147,419 per QALY gained. Within the LN0 NPI>3.4 subgroup, the ICER for EPClin versus current practice is expected to be £46,788 per QALY gained. Within the LN+ (1-3 nodes) subgroup, the ICER for EPClin versus current practice is expected to be £21,458 per QALY gained.

MammaPrint: Within the overall MINDACT population, the ICER for MammaPrint versus current practice is expected to be £131,482 per QALY gained. Within the mAOL high-risk subgroup, MammaPrint is expected to be dominated by current practice. Within the mAOL

low-risk subgroup, the ICER for MammaPrint versus current practice is expected to be £414,202 per QALY gained.

6.2 Strengths and limitations of the assessment

6.2.1 Strengths and limitations in the clinical evidence base

The clinical review benefitted from a comprehensive search strategy and a high quality, prospectively designed systematic review methodology.

The evidence base was large, and included multiple reanalyses of RCTs, which are generally considered to be a high quality source of data. However, nearly all studies excluded patients who did not have enough tissue sample (though this is unavoidable in retrospective analyses), which leaves the evidence base at potential risk of spectrum bias, as patients with smaller tumours (who may be systematically different to those with large tumours) will likely be under-represented.

There were some key gaps in the literature for IHC4+C and RSPC. Notably, the IHC4+C algorithm has only been validated in one cohort (in an adjusted analysis), and RSPC has only been validated in one cohort (in an unadjusted analysis, and for chemotherapy benefit). In both cases, the validation study was conducted as part of the derivation study. The IHC4/IHC4+C evidence base was also limited in that most of the data related to the IHC4 score alone, without the clinical score, and most studies used tertiles and quartiles to define low-, intermediate- and high-risk patients, which may not be useful in a clinical setting. In addition, there are known problems with conducting the analyses required for IHC4, and whilst a number of studies report methodologies that are largely compliant with the original methodology, it is unclear whether the absolute IHC4 values obtained would be similar across centres.

Much of the evidence base relates to unadjusted analyses, which do not assess the crucial question of whether a test has additional value over clinicopathological factors. Where adjusted analyses were performed, the clinicopathological variables included were not always consistent, and it is unclear if all important factors (including stratification factors from the original RCT studies where applicable) were included in all analyses.

There were relatively limited data relating to the ability of Oncotype DX and MammaPrint to predict benefit from chemotherapy, and some of the analyses conducted were also subject to criticisms relating to adjustment for all relevant variables.

Data relating to the ability of the test to affect patient outcomes (such as recurrence and survival) through the prospective use of the test to guide treatment decisions were also limited. Most studies were observational in nature, and the selection of patients on the basis of them having received a test may have introduced spectrum bias, and as such these studies may not match the decision problem. They also do not, by their nature, include a comparator arm, and it is difficult to draw any firm conclusions about the effect of the test in real clinical practice.

Similar observational study designs reported data relating to chemotherapy effects in different risk groups. Some of these studies are subject to the same limitations in terms of spectrum bias, but also from confounding whereby patients who received chemotherapy are likely to be systematically different in terms of known (and potentially unknown) prognostic variables (e.g. age) and treatment effect modifiers to those who did not, leading to a high risk of confounding.

Retrospective observational studies (where patients were treated according to usual practice without the tests) reporting data relating to prognostic performance are also at risk of confounding in that chemotherapy rates per risk group may differ (and thus affect estimates of prognostic performance). Observational studies which excluded patients who received chemotherapy, in order to obtain a group of patients unaffected by treatment, are likely to be subject to spectrum bias, as patients who receive chemotherapy are likely to be systematically different to those who do not, and this may also affect estimates of prognostic performance. These problems were particularly relevant to the MammaPrint evidence base, where most studies were observational in nature rather than reanalyses of RCTs. MammaPrint was also unusual, in that many of the included studies pooled multiple cohorts, and as such it was not possible to gauge the degree of double counting of patients. The overall sample size was also low (total N=1,805) compared to the evidence base for most other tests.

The evidence base relating to impact of tests on treatment decisions (decision impact studies) was limited in that use of chemotherapy differs across countries and there were no UK studies for two tests (MammaPrint; Prosigna), and only one UK study for 2 tests (EndoPredict; IHC4+C).

6.2.2 Strengths and limitations relating to the health economic analysis

The EAG model has a number of strengths; in particular: (i) for all tests, risk classification and DMFS probabilities are derived from the same source (TransATAC or MINDACT); (ii) within the LN0 intermediate-risk subgroup (NPI>3.4, analysis of 3-level tests), the probability

of receiving chemotherapy with and without the test is based on the NHS England Access Scheme dataset – this is likely to best reflect how the 3-level tumour profiling tests would be used in clinical practice in England; (iii) the model structure is consistent with that of other published models of tumour profiling tests - when similar data inputs are used, the EAG model produces similar results to the previous EAG model and the Genomic Health model, and (iv) extensive deterministic sensitivity analyses have been conducted to explore the impact of uncertainty on the model results.

However, the model is also subject to several limitations, most of which stem from uncertainties in the evidence base. The main limitations and uncertainties relating to the cost-effectiveness analysis are: (i) with the exception of Oncotype DX in the LN0 NPI>3.4 group (clinical intermediate risk), the evidence surrounding the pre- and post-test chemotherapy probabilities is subject to considerable uncertainty – this has the propensity to influence the conclusions regarding the cost-effectiveness of all tests; (ii) there is uncertainty regarding whether Oncotype DX and MammaPrint are predictive of chemotherapy benefit – the inclusion of such effects are likely to strongly influence economic conclusions drawn from the analysis; (iii) the analysis of MammaPrint is based on a different data source than the other four tests; and (iv) the TransATAC study used to estimate test risk classification and DMFS probabilities was the derivation study for IHC4 – as such, there is potential for the overestimation of prognostic performance for this test.

6.3 Uncertainties

Due to time and data constraints, it was not possible to perform a thorough analysis of how the baseline CP characteristics of patients (e.g. tumour grade, stage, age) affect prognostic performance.

The evidence relating to the impact on patient outcomes where the test is used in clinical practice remains largely unanswered, and is impeded by the long-term follow-up required, the large sample sizes required, and ethical problems with withholding chemotherapy from clinically high-risk patients.

Evidence relating to key subgroups defined in the scope were largely lacking. Data relating specifically to micrometastases were rarely reported, there were no data at all in male-only subgroups or cohorts, and data relating to people of different ethnicities were difficult to interpret due to differences in treatment practices in different countries. A more detailed consideration of the available evidence base may have allowed some observations to be drawn regarding pre- and post-menopausal patients, but time constraints prevented this.

IHC4 is known to have implementation issues in terms of conducting the test in other laboratories, especially local laboratories. The precise details are beyond the expertise of the EAG. It is uncertain if these could be overcome. Furthermore, it is somewhat unclear what cut-off values should be used for IHC4 and IHC4+C.

6.4 Generalisability

The EAG notes that there may be issues relating to the generalisability of the evidence contained within this report. In particular, the classification of risk by NPI will not reflect current practice across all centres. In addition, the TransATAC study used to inform test risk classification and DMFS probabilities for Oncotype DX, Prosigna, IHC4+C and EPClin relates only to a post-menopausal population only; it is expected that the tumour profile testing may also be used in pre-menopausal women.

6.5 Implications for service provision

The per test costs for Prosigna provided by NanoString (used in the EAG economic analyses) are based on an efficient level of throughput. This may not hold if centres do not undertake the anticipated number of tests (for example, in smaller centres, or if multiple tumour profiling tests are available). Furthermore, as NanoString does not offer a centralised testing service, local testing services will need to be established.

IHC4 is not currently commercially available. Standardisation of IHC4 and quality assurance programs are required before this test may be used routinely within the NHS.

6.6 Suggested research priorities

- There is uncertainty regarding whether Oncotype DX and MammaPrint are predictive of chemotherapy benefit. Further studies are required which adjust for all relevant clinico-pathological factors.
- There is limited evidence demonstrating long-term impacts resulting from the use of the five tumour profiling tests. Future studies assessing the comparative long-term impact of the tests compared with risk prediction tools commonly used in clinical practice would be valuable.
- There is uncertainty regarding the cost-effectiveness of all five tests included in the NICE scope. It is noteworthy that under the assumption of no predictive chemotherapy benefit, the inclusion of additional data collected through the NHS England Access Scheme Dataset has a significant impact upon the conclusions

previously drawn from the Oncotype DX analysis within NICE DG10 (moving from an ICER of £22,572 per QALY gained to a situation in which Oncotype DX is dominated in the LN0 NPI>3.4 subgroup). Additional UK-based data collection relating to pre- and post-test chemotherapy use for EPClin, IHC4+C, Prosigna and MammaPrint may be important in reducing existing uncertainty surrounding the cost-effectiveness of these tests.

8 REFERENCES

1. Cancer Research UK. Statistics by cancer type. 2017, http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/survival#heading-Three.
2. Dumitrescu RG, Cotarla I. Understanding breast cancer risk -- where do we stand in 2005? *Journal of Cellular & Molecular Medicine* 2005;9:208-21.
3. 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.
4. Ziegler RG, Hoover RN, Pike MC, Hildesheim A, Nomura AM, West DW*, et al.* Migration patterns and breast cancer risk in Asian-American women. *Journal of the National Cancer Institute* 1993;85:1819-27.
5. Brush J, Boyd K, Chappell F, Crawford F, Dozier M, Fenwick E*, et al.* The value of FDG positron emission tomography/computerised tomography (PET/CT) in pre-operative staging of colorectal cancer: A systematic review and economic evaluation. *Health Technology Assessment* 2011;15:i-142.
6. Lagerros YT, Hsieh SF, Hsieh CC. Physical activity in adolescence and young adulthood and breast cancer risk: a quantitative review. *European Journal of Cancer Prevention* 2004;13:5-12.
7. 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.
8. National Institute for Health and Care Excellence. Early and locally advanced breast cancer: diagnosis and treatment. Clinical guideline [CG80]. 2017, https://www.nice.org.uk/guidance/cg80.
9. Office for National Statistics. Survival Rates in England, patients diagnosed 2001-2006 followed up to 2007. In; 2007: http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=14007.
10. Cancer Research UK. Breast Cancer (C50), Five-Year Net Survival by Age, Women, England, 2009-2013. 2017, http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/survival#heading-One.
11. Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham Prognostic Index in primary breast cancer. *Breast Cancer Research and Treatment* 1992;22:207-19.
12. Office for National Statistics. Cancer Registration Statistics, England. 2017, https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancerregistrationstatisticscancerregistrationstatisticsengland.
13. Cancer Research UK. National Cancer Intelligence Network. Cancer Incidence and Survival By Major Ethnic Group, England, 2002 - 2006. 2009, http://publications.cancerresearchuk.org/WebRoot/crukstoredb/CRUK_PDFs/CR-UKNCIN_ETHNIC.pdf.
14. National Cancer Registration and Analysis Service. Cancer by Deprivation in England 1996 - 2011. 2017, http://www.ncin.org.uk/about_ncin/cancer_by_deprivation_in_england.
15. Massat NJ, Dibden A, Parmar D, Cuzick J, Sasieni PD, Duffy SW. Impact of Screening on Breast Cancer Mortality: The UK Program 20 Years On. *Cancer Epidemiology Biomarkers & Prevention* 2016;25:455-62.
16. American Joint Committee on Cancer. AJCC Cancer Staging Manual: Springer International Publishing; 2017.
17. Cancer AJCo. AJCC Cancer Staging Atlas, 2nd Edition,. Chicago, US.: Springer; 2010.
18. Ward S, Scope A, Rafia R, Pandor A, Harnan S, Evans P*, et al.* Gene expression profiling and expanded immunohistochemistry tests to guide the use of adjuvant

chemotherapy in breast cancer management: a systematic review and cost-effectiveness analysis. *Health Technology Assessment* 2013;17:1-302.
19. National Institute for Clinical Excellence. Improving Outcomes in Breast Cancer NICE; 2011.
20. Early Breast Cancer Trialists' Collaborative G. 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.
21. National Institute for Health and Care Excellence. Gene expression profiling and expanded immunohistochemistry tests for guiding adjuvant chemotherapy decisions in early breast cancer management: MammaPrint, Oncotype DX, IHC4 and Mammostrat. Diagnostics Guidance 10. London; 2013.
22. National Institute for Health and Care Excellence. Tumour profiling tests to guide adjuvant chemotherapy decisions in people with breast cancer (update of DG10) - final scope. London; 2017.
23. Bartlett JM, Christiansen J, Gustavson M, Rimm DL, Piper T, van de Velde CJ*, et al.* Validation of the IHC4 breast cancer prognostic algorithm using multiple approaches on the multinational TEAM clinical trial. *Archives of Pathology and Laboratory Medicine* 2016;140:66-74.
24. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E*, 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. *Journal of Clinical Oncology* 2011;29:4273-8.
25. Centre for Reviews Dissemination. Systematic review: CRD's guidance for undertaking reviews in health care. CRD, University of York; 2009. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]; 2011.
26. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA*, et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339.
27. National Institute for Health and Care Excellence. Diagnostics Assessment Programme Manual. London; 2011.
28. Waintraub SE, McNamara DM, Graham DA, Pecora A, Min J, Wu T*, et al.* Early economic benefits of gene expression profiling using a 21-gene panel among patients with early stage, lymph node negative, hormone receptor positive, her2-neu oncogene negative breast cancer. *Journal of Clinical Oncology Conference* 2016;34.
29. Hudis CA, Barlow WE, Costantino JP, Gray RJ, Pritchard KI, Chapman J-AW*, et al.* Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *Journal of Clinical Oncology* 2007;25:2127-32.
30. Higgins J. Cochrane Handbook for Systematic Reviews of Interventions. Chapter 8: Assessing risk of bias in included studies; 2011.
31. Michiels S, Ternès N, Rotolo F. Statistical controversies in clinical research: prognostic gene signatures are not (yet) useful in clinical practice. *Annals of Oncology* 2016;27:2160-7.
32. Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N*, et al.* The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: methods of the EGAPP Working Group. *Genetics in Medicine* 2009;11:3-14.
33. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG*, et al.* Critical appraisal and data extraction for systematic reviews of prediction modelling studies: The CHARMS checklist. *PLOS Medicine* 2014;11:e1001744.
34. Buus R, Sestak I, Kronenwett R, Denkert C, Dubsky P, Krappmann K*, et al.* Comparison of EndoPredict and EPclin with Oncotype DX recurrence score for prediction of risk of distant recurrence after endocrine therapy. *Journal of the National Cancer Institute* 2016;108.

35. 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. *Journal of Clinical Oncology* 2010;28:1829-34.
36. Dowsett M, Sestak I, Lopez-Knowles E, Sidhu K, Dunbier AK, Cowens JW, *et al.* Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *Journal of Clinical Oncology* 2013;31:2783-90.
37. Sestak I, Buus R, Cuzick J, Dubsky P, Kronenwett R, Ferree S, *et al.* Comprehensive comparison of prognostic signatures for breast cancer recurrence in TransATAC. San Antonio Breast Cancer Symposium, abstract no. 5773.
38. Sestak I, Cuzick J, Dowsett M, Lopez-Knowles E, Filipits M, Dubsky P, *et al.* Prediction of late distant recurrence after 5 years of endocrine treatment: a combined analysis of patients from the Austrian breast and colorectal cancer study group 8 and arimidex, tamoxifen alone or in combination randomized trials using the PAM50 risk of recurrence score. *Journal of Clinical Oncology* 2015;33:916-22.
39. Sestak I, Dowsett M, Zabaglo L, Lopez-Knowles E, Ferree S, Cowens JW, *et al.* Factors predicting late recurrence for estrogen receptor-positive breast cancer. *Journal of the National Cancer Institute* 2013;105:1504-11.
40. Sestak I, Zhang Y, Schroeder BE, Schnabel CA, Dowsett M, Cuzick J, *et al.* Cross-Stratification and Differential Risk by Breast Cancer Index and Recurrence Score in Women with Hormone Receptor-Positive Lymph Node-Negative Early-Stage Breast Cancer. *Clinical Cancer Research* 2016b;22:5043-8.
41. Sgroi DC, Sestak I, Cuzick J, Zhang Y, Schnabel CA, Schroeder B, *et al.* Prediction of late distant recurrence in patients with oestrogen-receptor-positive breast cancer: a prospective comparison of the breast-cancer index (BCI) assay, 21-gene recurrence score, and IHC4 in the TransATAC study population. *Lancet Oncology* 2013;14:1067-76.
42. Tang G, Cuzick J, Costantino JP, Dowsett M, Forbes JF, Crager M, *et al.* Risk of recurrence and chemotherapy benefit for patients with node-negative, estrogen receptor-positive breast cancer: recurrence score alone and integrated with pathologic and clinical factors. *Journal of Clinical Oncology* 2011b;29:4365-72.
43. Sestak I, Dowsett M, Cuzick J. NICE request - TransATAC data analysis (academic-in-confidence, data held on file). In; 2017.
44. Statistics How to. http://www.statisticshowto.com/c-statistic/; 17/10/2017).
45. 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. *New England Journal of Medicine* 2004;351:2817-26.
46. van 't Veer LJ, Dai H, van de Vijver M, He YD, Hart AA, Mao M, *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530-6.
47. van de Vijver MJ, He Y, van't Veer LJ, Dai H, Hart AA, Voskuil DW, *et al.* A gene-expression signature as a predictor of survival in breast cancer. *New England Journal of Medicine* 2002;347:1999-2009.
48. Filipits M, Rudas M, Jakesz R, Dubsky P, Fitzal F, Singer CF, *et al.* A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clinical Cancer Research* 2011;17:6012-20.
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. *Journal of Clinical Oncology* 2006;24:3726-34.
50. Tang G, Shak S, Paik S, Anderson SJ, Costantino JP, Geyer CE, Jr., *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* 2011a;127:133-42.

51. Wolmark N, Mamounas EP, Baehner FL, Butler SM, Tang G, Jamshidian F*, et al.* Prognostic impact of the combination of recurrence score and quantitative estrogen receptor expression (ESR1) on predicting late distant recurrence risk in estrogen receptor-positive breast cancer after 5 years of tamoxifen: Results from NRG Oncology/National Surgical Adjuvant Breast and Bowel Project B-28 and B-14. *Journal of Clinical Oncology* 2016;34:2350-8.
52. 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;116:3112-8.
53. van 't Veer LJ, Yau C, Yu NY, Benz CC, Nordenskjold B, Fornander T*, et al.* Tamoxifen therapy benefit for patients with 70-gene signature high and low risk. *Breast Cancer Research and Treatment* 2017; 10.1007/s10549-017-4428-9.
54. Gnant M, Filipits M, Greil R, Stoeger H, Rudas M, Bago-Horvath Z*, et al.* Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk: using the PAM50 Risk of Recurrence score in 1478 postmenopausal patients of the ABCSG-8 trial treated with adjuvant endocrine therapy alone. *Annals of Oncology* 2014;25:339-45.
55. Filipits M, Nielsen TO, Rudas M, Greil R, Stoger H, Jakesz R*, et al.* The PAM50 risk-of-recurrence score predicts risk for late distant recurrence after endocrine therapy in postmenopausal women with endocrine-responsive early breast cancer. *Clinical Cancer Research* 2014;20:1298-305.
56. Laenkholm AV, Jensen MB, Eriksen JO, Kiboll T, Rasmussen BB, Knoop AS*, et al.* Prediction of 10yr distant recurrence (DR) using the Prosigna (PAM50) assay in a Danish Breast Cancer Cooperative Group (DBCG) cohort of post-menopausal Danish women with hormone receptor-positive (HR+) early breast cancer (EBC) allocated to 5yr of endocrine therapy (ET) alone. *Journal of Clinical Oncology Conference* 2015;33:Abstract 546.
57. Dubsky P, Filipits M, Jakesz R, Rudas M, Singer CF, Greil R*, et al.* EndoPredict improves the prognostic classification derived from common clinical guidelines in ER-positive, HER2-negative early breast cancer. *Annals of Oncology* 2013;24:640-7.
58. Dubsky P, Brase JC, Jakesz R, Rudas M, Singer CF, Greil R*, et al.* The EndoPredict score provides prognostic information on late distant metastases in ER+/HER2- breast cancer patients. *British Journal of Cancer* 2013;109:2959-64.
59. Myriad Genetics. EPclin in ABCSG-6+8: Subgroup analyses. 2015.
60. Mook S, Schmidt MK, 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 Research and Treatment* 2009;116:295-302.
61. Drukker CA, van Tinteren H, Schmidt MK, Rutgers EJ, Bernards R, van de Vijver MJ*, et al.* Long-term impact of the 70-gene signature on breast cancer outcome. *Breast Cancer Res Treat* 2014;143:587-92.
62. Stephen J, Murray G, Cameron DA, Thomas J, Kunkler IH, Jack W*, et al.* Time dependence of biomarkers: non-proportional effects of immunohistochemical panels predicting relapse risk in early breast cancer. *British Journal of Cancer* 2014;111:2242-7.
63. 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 Research and Treatment* 2009;117:483-95.
64. 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. *Journal of the National Cancer Institute* 2006;98:1183-92.

65. Mook S. The 70-gene prognosis signature predicts early metastasis in breast cancer patients between 55 70 years of age. *Annals of Oncology* 2010;21:717-22.
66. Wittner BS, Sgroi DC, Ryan PD, Bruinsma TJ, Glas AM, Male A*, et al.* Analysis of the MammaPrint breast cancer assay in a predominantly postmenopausal cohort. *Clinical Cancer Research* 2008;14:2988-93.
67. Drukker CA, van Tinteren H, Schmidt MK, Rutgers EJ, Bernards R, van de Vijver MJ*, et al.* Long-term impact of the 70-gene signature on breast cancer outcome. *Breast Cancer Research and Treatment* 2014;143:587-92.
68. Albain K, Barlow W, Shak S, Hortobagyi G, Livingston R, 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 Oncology* 2010;11:55-65.
69. Barcenas CH, Raghavendra A, Sinha AK, Syed MP, Hsu L, Patangan MG, Jr.*, et al.* Outcomes in patients with early-stage breast cancer who underwent a 21-gene expression assay. *Cancer* 2017;15:15.
70. Le Du F, Gonzalez-Angulo AM, Park M, Liu DD, Hortobagyi GN, Ueno NT. Effect of 21-gene RT-PCR assay on adjuvant therapy and outcomes in patients with stage I breast cancer. *Clinical Breast Cancer* 2015;15:458-66.
71. Stemmer S, Steiner M, Rizel S, Soussan-Gutman L, Ben-Baruch N, Bareket-Samish A*, et al.* Clinical outcomes in early-stage breast cancer treated according to the recurrence score results: Evidence from a large prospectively-designed cohort study. In. Academic-In-Confidence; 2017.
72. Stemmer SM, Steiner M, Rizel S, Geffen D, Nisenbaum B, Peretz T*, et al.* First prospectively-designed outcome study in estrogen receptor (ER)+ breast cancer (BC) patients (pts) with N1mi or 1-3 positive nodes in whom treatment decisions in clinical practice incorporated the 21-gene recurrence score (RS) result. *Annals of Oncology Conference: 41st European Society for Medical Oncology Congress,* ***ESMO*** 2016;27.
73. Petkov VI, Miller DP, Howlader N, Gliner N, Howe W, Schussler N*, et al.* Breast-cancer-specific mortality in patients treated based on the 21-gene assay: a SEER population-based study. *NPJ Breast Cancer* 2016;2:16017.
74. Roberts M, Petkov VI, Miller DP, Shak S, Howlader N, Cronin K*, et al.* Breast cancer specific survival in patients with node-positive hormone receptor positive invasive breast cancer and Oncotype DX recurrence score results in the SEER database. *Journal of Clinical Oncology Conference* 2016;34.
75. Bartlett JM, Bayani J, Marshall A, Dunn JA, Campbell A, Cunningham C*, et al.* Comparing breast cancer multiparameter tests in the OPTIMA prelim trial: No test is more equal than the others. *Journal of the National Cancer Institute* 2016;108.
76. Bloomfield DJ, Arbon A, Cox J, Hack B, Hall J, Harper-Wynne C. Patient/oncologist decisions about adjuvant chemotherapy in ER+ve, HER2-ve early breast cancer following EndoPredict testing. *American Society of Clinical Oncology (ASCO) conference* 2017.
77. Evans CN, Brewer NT, Vadaparampil ST, Boisvert M, Ottaviano Y, Lee MC*, et al.* Impact of genomic testing and patient-reported outcomes on receipt of adjuvant chemotherapy. *Breast Cancer Res Treat* 2016;156:549-55.
78. Hinde S, Theriou C, May S, Matthews L, Arbon A, Fallowfield L*, et al.* The cost-effectiveness of EndoPredict to inform adjuvant chemotherapy decisions in early breast cancer. *Annals of Oncology* 2017;28:mdx375.012-mdx375.012.
79. 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. *Journal of Clinical Oncology* 2010;28:1671-6.
80. Martin M, Gonzalez-Rivera M, Morales S, de la Haba-Rodriguez J, Gonzalez-Cortijo L, Manso L*, et al.* Prospective study of the impact of the Prosigna assay on adjuvant clinical decision-making in unselected patients with estrogen receptor positive,

human epidermal growth factor receptor negative, node negative early-stage breast cancer. *Current Medical Research & Opinion* 2015;31:1129-37.
81. Retel VP, Groothuis-Oudshoorn CG, Aaronson NK, Brewer NT, Rutgers EJ, van Harten WH. Association between genomic recurrence risk and well-being among breast cancer patients. *BMC Cancer* 2013;13:295.
82. Wuerstlein R, Sotlar K, Gluz O, Otremba B, von Schumann R, Witzel I*, et al.* The West German Study Group Breast Cancer Intrinsic Subtype study: a prospective multicenter decision impact study utilizing the Prosigna assay for adjuvant treatment decision-making in estrogen-receptor-positive, HER2-negative early-stage breast cancer. *Current Medical Research & Opinion* 2016;32:1217-24.
83. Martin M, Brase JC, Ruiz A, Prat A, Kronenwett R, Calvo L*, et al.* Prognostic ability of EndoPredict compared to research-based versions of the PAM50 risk of recurrence (ROR) scores in node-positive, estrogen receptor-positive, and HER2-negative breast cancer. A GEICAM/9906 sub-study. *Breast Cancer Res Treat* 2016;156:81-9.
84. Wuerstlein R, Sotlar K, Gluz O, Otremba B, von Schumann R, Witzel I*, et al.* The West German Study Group Breast Cancer Intrinsic Subtype study: a prospective multicenter decision impact study utilizing the Prosigna assay for adjuvant treatment decision-making in estrogen-receptor-positive, HER2-negative early-stage breast cancer. *Current Medical Research and Opinion* 2016;32:1217-24.
85. Gong C, Tan W, Chen K, You N, Zhu S, Liang G*, et al.* Prognostic value of a BCSC-associated microRNA signature in hormone receptor-positive HER2-negative breast cancer. *EBioMedicine* 2016;11:199-209.
86. 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. *Japanese Journal of Clinical Oncology* 2010;40:508-12.
87. Rohan TE, Xue X, Lin HM, D'Alfonso TM, Ginter PS, Oktay MH*, et al.* Tumor microenvironment of metastasis and risk of distant metastasis of breast cancer. *Journal of the National Cancer Institute* 2014;106.
88. Viale G, Speirs V, Bartlett JM, Mousa K, Kalaitzaki E, Palmieri C*, et al.* Pr prognostic and predictive value of IHC4 and erb1 in the intergroup exemestane study (IES)-on behalf of the pathies investigators. *Annals of Oncology* 2013;24:iii29-iii30.
89. Sun B, Zhang F, Wu SK, Guo X, Zhang LL, Jiang ZF*, et al.* Gene expression profiling for breast cancer prognosis in Chinese populations. *Breast Journal* 2011;17:172-9.
90. Mamounas EP, Tang G, Paik S, Baehner FL, Liu Q, Jeong JH*, et al.* Prognostic impact of the 21-gene recurrence score (RS) on disease-free and overall survival of node-positive, ER-positive breast cancer patients (pts) treated with adjuvant chemotherapy: Results from NSABP B-28. *Journal of Clinical Oncology Conference* 2012;30.
91. Penault-Llorca FM, Filleron T, Asselain B, Baehner FL, Fumoleau P, Lacroix-Triki M*, et al.* Prediction of recurrence with the Oncotype DX recurrence score in node-positive, HR-positive, breast cancer patients treated with adjuvant chemotherapy: Results from PACS01 trial. *Journal of Clinical Oncology Conference* 2014;32.
92. Martin M, Brase JC, Calvo L, Krappmann K, Ruiz-Borrego M, Fisch K*, et al.* Clinical validation of the EndoPredict test in node-positive, chemotherapy-treated ER+/HER2- breast cancer patients: results from the GEICAM 9906 trial. *Breast Cancer Research* 2014;16:R38.
93. Gluz O, Liedtke C, Huober J, Peyro-Saint-Paul H, Kates RE, Kreipe HH*, et al.* Comparison of prognostic and predictive impact of genomic or central grade and immunohistochemical subtypes or IHC4 in HR+/HER2- early breast cancer: WSG-AGO EC-Doc Trial. *Annals of Oncology* 2016;27:1035-40.
94. Martín M, Rodríguez-Lescure Á, Ruiz A, Alba E, Calvo L, Ruiz-Borrego M*, et al.* Randomized phase 3 trial of fluorouracil, epirubicin, and cyclophosphamide alone or followed by Paclitaxel for early breast cancer. *Journal of the National Cancer Institute* 2008;100:805-14.

95. Cobleigh M, Bitterman P, Baker J, Cronin M, Liu M, Borchik R*, et al.* Tumor gene expression predicts distant disease-free survival (DDFS) in breast cancer patients with 10 or more positive nodes: High throughput RT-PCR assay of paraffin-embedded tumor tissues. Proc Am Soc Clin Oncol, abstract no. 6369, p. 2003.
96. Esteban J, Baker J, Cronin M, Liu M, Llamas M, Walker M*, et al.* Tumor gene expression and prognosis in breast cancer: multi-gene RT-PCR assay of paraffin-embedded tissue. Proc Am Soc Clin Oncol, abstract no. 6368, p. 2003.
97. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M*, et al.* Multi-gene RT-PCR assay for predicting recurrence in node negative breast cancer patients-NSABP studies B-20 and B-14. *Breast Cancer Research and Treatment* 2003;82:S10-S1.
98. Goldstein LJ, Gray R, Badve S, Childs BH, 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. *Journal of Clinical Oncology* 2008;26:4063-71.
99. Sparano JA, O'Neill A, Gray RJ, Perez EA, Shulman LN, Martino S*, et al.* 10-year update of E2197: Phase III doxorubicin/docetaxel (AT) versus doxorubicin/cyclophosphamide (AC) adjuvant treatment of LN+ and high-risk LN- breast cancer and the comparison of the prognostic utility of the 21-gene recurrence score (RS) with clinicopathologic features. *Journal of Clinical Oncology Conference* 2012;30.
100. Russell S, Shivers SC, Blumencranz L. Long-term follow-up of early stage breast cancer patients with results of MammaPrint, Oncotype DXand MammoStrat risk classif ication assays. *San Antonio Breast Cancer Symposium (SABCS)* 2016.
101. Sestak I, Dowsett M, Cuzick J. NICE request - transATAC data. In: Group EA, ed.; 2017.
102. Mamounas EP, Tang G, Paik S, Baehner FL, Liu Q, Jeong JH*, et al.* Association between the 21-gene recurrence score (RS) and benefit from adjuvant paclitaxel (Pac) in node-positive (N+), ER-positive breast cancer patients (pts): Results from NSABP B-28. *Cancer Research Conference: 35th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2012;72.
103. Petkov VI, Miller DP, Howlader N, Gliner N, Howe W, Schussler NC*, et al.* Outcome disparities by age and 21-gene recurrence score (RS) in hormone receptor positive (HR+) breast cancer (BC). *Journal of Clinical Oncology Conference* 2016;34.
104. Roberts MC, Miller DP, Shak S, Petkov VI. Breast cancer-specific survival in patients with lymph node-positive hormone receptor-positive invasive breast cancer and Oncotype DX Recurrence Score results in the SEER database. *Breast Cancer Research and Treatment* 2017;163:303-10.
105. Stemmer SM, Steiner M, Rizel S, Soussan-Gutman L, Geffen DB, Nisenbaum B*, et al.* Real-life analysis evaluating 1594 N0/Nmic breast cancer patients for whom treatment decisions incorporated the 21-gene recurrence score result: 5-year KM estimate for breast cancer specific survival with recurrence score results <30 is >98%. *Cancer Research Conference: 38th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2016;76.
106. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF*, et al.* Prospective validation of a 21-gene expression assay in breast cancer. *New England Journal of Medicine* 2015;373:2005-14.
107. Gluz O, Nitz U, Christgen M, Kates R, Clemens M, Kraemer S*, et al.* Prospective WSG phase III PlanB trial: Clinical outcome at 5 year follow up and impact of 21 Gene Recurrence Score result, central/local-pathological review of grade, ER, PR and Ki67 in HR+/HER2-high risk node-negative and-positive breast cancer. *European Journal of Cancer* 2016d;57:S6.
108. Gluz O, Nitz U, Chrlstgen M, Kates RE, Clemens M, Kraemer S*, et al.* Prognostic impact of 21 gene recurrence score, IHC4, and central grade in high-risk HR+/HER2-

early breast cancer (EBC): 5-year results of the prospective Phase III WSG PlanB trial. *Journal of Clinical Oncology Conference* 2016a;34.

109. Gluz O, Nitz UA, Christgen M, Kates RE, Shak S, Clemens M*, et al.* West German Study Group Phase III PlanB trial: First prospective outcome data for the 21-gene recurrence score assay and concordance of prognostic markers by central and local pathology assessment. *Journal of Clinical Oncology* 2016b;34:2341-9.
110. Wen HY, Krystel-Whittemore M, Patil S, Pareja F, Bowser ZL, Dickler MN*, et al.* Breast carcinoma with an Oncotype Dx recurrence score <18: Rate of distant metastases in a large series with clinical follow-up. *Cancer* 2017;123:131-7.
111. Nitz U, Gluz O, Christgen M, Kates RE, Clemens M, Malter W*, et al.* Reducing chemotherapy use in clinically high-risk, genomically low-risk pN0 and pN1 early breast cancer patients: five-year data from the prospective, randomised phase 3 West German Study Group (WSG) PlanB trial. *Breast Cancer Research and Treatment* 2017; 10.1007/s10549-017-4358-6.
112. Stemmer S, Steiner M, Rizel S, Ben-Baruch N, Soussan-Gutman L, Rosengarten O*, et al.* First prospective outcome data in 930 patients with more than 5 year median follow up in whom treatment decisions in clinical practice have been made incorporating the 21-Gene Recurrence Score. *European Journal of Cancer* 2015;51:S321.
113. Genomic Health. Oncotype DX company submission. 2017.
114. Van't Veer LJ, Dai H, van de Vijver M, He YD, Hart AA, Mao M*, et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530-6.
115. 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-.
116. Beumer I, Witteveen A, Delahaye L, Wehkamp D, Snel M, Dreezen C*, et al.* Equivalence of MammaPrint array types in clinical trials and diagnostics. *Breast Cancer Res Treat* 2016;156:279-87.
117. Kok M, Koornstra RH, Mook S, Hauptmann M, Fles R, Jansen MP*, et al.* Additional value of the 70-gene signature and levels of ER and PR for the prediction of outcome in tamoxifen-treated ER-positive breast cancer. *Breast* 2012;21:769-78.
118. Yao K, Goldschmidt R, Turk M, Wesseling J, Stork-Sloots L, de Snoo F*, et al.* Molecular subtyping improves diagnostic stratification of patients with primary breast cancer into prognostically defined risk groups. *Breast Cancer Research and Treatment* 2015;154:81-8.
119. Esserman LJ, Thompson CK, Yau C, Van't Veer LJ, Borowsky AD, Tobin NP*, et al.* Identification of tumors with an indolent disease course: MammaPrint ultralow signature validation in a retrospective analysis of a Swedish randomized tamoxifen trial. *Cancer Research Conference: 38th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2016;76.
120. Lindstrom LS, Benz CC, Yau C, van't Veer LJ, Thompson CK, Esserman LJ. MammaPrint accurately predicts long-term survival (25 years) and adjuvant tamoxifen therapy benefit in lymph node negative patients. *Cancer Research Conference: 37th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2015;75.
121. Agendia. MammaPrint - evidence provided to NICE. 2017.
122. 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. *Annals of Surgical Oncology* 2010;17:1406-13.
123. Bueno-de-Mesquita JM, van Harten WH, Retel VP, van 't Veer LJ, van Dam FS, 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.

124. Drukker CA, Bueno-de-Mesquita JM, Retel VP, van Harten WH, van Tinteren H, Wesseling J*, et al.* A prospective evaluation of a breast cancer prognosis signature in the observational RASTER study. *International Journal of Cancer* 2013;133:929-36.
125. Drukker CA, Nijenhuis MV, Bueno-de-Mesquita JM, Retel VP, van Harten WH, van Tinteren H*, et al.* Optimized outcome prediction in breast cancer by combining the 70-gene signature with clinical risk prediction algorithms. *Breast Cancer Res Treat* 2014;145:697-705.
126. 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 Research and Treatment* 2010;120:655-61.
127. Bueno-de-Mesquita JM, Sonke GS, van de Vijver MJ, Linn SC. Additional value and potential use of the 70-gene prognosis signature in node-negative breast cancer in daily clinical practice. *Annals of Oncology* 2011;22:2021-30.
128. Beumer IJ, Persoon M, Witteveen A, Dreezen C, Chin SF, Sammut SJ*, et al.* Prognostic value of MammaPrint in invasive lobular breast cancer. *Biomark Insights* 2016;11:139-46.
129. Michaut M, Chin SF, Majewski I, Severson TM, Bismeijer T, de Koning L*, et al.* Integration of genomic, transcriptomic and proteomic data identifies two biologically distinct subtypes of invasive lobular breast cancer. *Sci Rep* 2016;6:18517.
130. Yao K, Goldschmidt R, Turk M, Wesseling J, Stork-Sloots L, de Snoo F*, et al.* Molecular subtyping improves diagnostic stratification of patients with primary breast cancer into prognostically defined risk groups. *Breast Cancer Res Treat* 2015;154:81-8.
131. Beumer IJ, Persoon M, Witteveen A, Dreezen C, Chin SF, Sammut SJ*, et al.* Prognostic Value of MammaPrint in Invasive Lobular Breast Cancer. *Biomark Insights* 2016;11:139-46.
132. Agendia. MammaPrint company submission. 2017.
133. Kok M, Koornstra RH, Margarido TC, Fles R, Armstrong NJ, Linn SC*, et al.* Mammosphere-derived gene set predicts outcome in patients with ER-positive breast cancer. *J Pathol* 2009;218:316-26.
134. Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S*, et al.* 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *New England Journal of Medicine* 2016;375:717-29.
135. Vliek SB, Retel V, Drukker C, Rutgers E, van Tinteren H, Van de Vijver MJ. 10 years follow up of the RASTER study; implementing a genomic signature in daily practice (poster). *ESMO conference* 2017a.
136. Linn SC, Retel V, Drukker C, Bueno-de-Mesquita JM, Rutgers E, van Tinteren H. 10 years follow up of the RASTER study; implementing a genomic signature in daily practice (abstract). *ESMO conference* 2017.
137. VvlK. KvdGC. Adjuvante Systemische Therapie voor het Operabel Mammacarcinoom. Behandeling van het Mammacarcinoom.; 2004.
138. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T*, et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology* 2009;27:1160-7.
139. Wallden B, Storhoff J, Nielsen T, Dowidar N, Schaper C, Ferree S*, et al.* Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC Medical Genomics* 2015;8:54.
140. Liu MC, Pitcher B, Mardis E, Davies S, Friedman PN, Snider J. PAM50 gene signatures and breast cancer prognosis with adjuvant anthracycline- and taxane-based chemotherapy: correlative analysis of C9741 (Alliance). *Breast Cancer* 2016;2.
141. Liu S, Chapman JA, Burnell MJ, Levine MN, Pritchard KI, Whelan TJ*, et al.* Prognostic and predictive investigation of PAM50 intrinsic subtypes in the NCIC CTG MA.21 Phase III chemotherapy trial. *Breast Cancer Research and Treatment* 2015;149:439-48.

142. 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. *Clinical Cancer Research* 2012;18:4465-72.
143. Ejlertsen B, Jensen MB, Eriksen JO, Kiboll T, Rasmussen BB, Knoop AS*, et al.* Validation of prediction of distant recurrence (DR) by Prosigna (PAM50) in subgroups of a Danish Breast Cancer Cooperative Group (DBCG) cohort of node-positive (N1), hormone receptor positive (HR+), post-menopausal early breast cancer (EBC) patients allocated 5yr of endocrine therapy (ET). *Journal of Clinical Oncology Conference* 2015;33.
144. Laenkholm A, Jensen M, Eriksen J. Prediction of late distant recurrence (DR) using the Prosigna (PAM50) test in a Danish Breast Cancer Cooperative Group (DBCG) cohort of postmenopausal women diagnosed with hormone receptor-positive (HR+) early breast cancer (EBC) allocated to 5yr of endocrine therapy (ET). *J Clin Oncol* 2015;33:Abstract 544.
145. Laenkholm AV, Jensen MB, Buckingham W. Prediction of 10yr distant recurrence (DR) using the Prosigna® (PAM50) assay in histological subgroups of a Danish breast cancer group (DBCG) cohort of postmenopausal Danish women with hormone receptor-positive (HR+) early breast cancer (EBC) allocated to 5yr of endocrine therapy (ET) alone. *SABCS* 2016.
146. 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. *Clinical Cancer Research* 2010;16:5222-32.
147. Liu S, Chapman JA, Burnell MJ, Levine MN, Pritchard KI, Whelan TJ*, et al.* Prognostic and predictive investigation of PAM50 intrinsic subtypes in the NCIC CTG MA.21 phase III chemotherapy trial. *Breast Cancer Res Treat* 2015;149:439-48.
148. Wallden B, Storhoff J, Nielsen T, Dowidar N, Schaper C, Ferree S*, et al.* Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC Medical Genomics [Electronic Resource]* 2015;8:54.
149. Dowsett M, Allred C, Knox J, Quinn E, Salter J, Wale C*, et al.* Relationship between quantitative estrogen and progesterone receptor expression and human epidermal growth factor receptor 2 (HER-2) status with recurrence in the Arimidex, Tamoxifen, Alone or in Combination trial. *Journal of Clinical Oncology* 2008;26:1059-65.
150. Zabaglo L, Salter J, Anderson H, Quinn E, Hills M, Detre S*, et al.* Comparative validation of the SP6 antibody to Ki67 in breast cancer. *Journal of Clinical Pathology* 2010;63:800-4.
151. Christiansen J, Bartlett JMS, Gustavson M, Rimm D, Robson T, Van De Velde CJH*, et al.* Validation of IHC4 algorithms for prediction of risk of recurrence in early breast cancer using both conventional and quantitative IHC approaches. *Journal of Clinical Oncology Conference* 2012;30.
152. Lin CH, Chen IC, Huang CS, Hu FC, Kuo WH, Kuo KT*, et al.* TP53 mutational analysis enhances the prognostic accuracy of IHC4 and PAM50 assays. *Scientific Reports* 2015;5:17879.
153. Vincent-Salomon A, Benhamo V, Gravier E, Rigaill G, Gruel N, Robin S*, et al.* Genomic instability: a stronger prognostic marker than proliferation for early stage luminal breast carcinomas. *PLoS ONE [Electronic Resource]* 2013;8:e76496.
154. Prat A, Cheang MC, Martin M, Parker JS, Carrasco E, Caballero R*, et al.* Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically defined luminal A breast cancer. *Journal of Clinical Oncology* 2013;31:203-9.
155. Van de Velde CJ, Rea D, Seynaeve C, Putter H, Hasenburg A, Vannetzel J-M*, et al.* Adjuvant tamoxifen and exemestane in early breast cancer (TEAM): a randomised phase 3 trial. *The Lancet* 2011;377:321-31.
156. Coombes RC, Hall E, Gibson LJ, Paridaens R, Jassem J, Delozier T*, et al.* A randomized trial of exemestane after two to three years of tamoxifen therapy in

postmenopausal women with primary breast cancer. *New England Journal of Medicine* 2004;350:1081-92.
157. Harbeck N, Gluz O, Clemens MR, Malter W, Reimer T, Nuding B*, et al.* Prospective WSG phase III PlanB trial: Final analysis of adjuvant 4xEC→4x doc vs. 6x docetaxel/cyclophosphamide in patients with high clinical risk and intermediate-to-high genomic risk HER2-negative, early breast cancer. *Journal of Clinical Oncology* 2017;35:504-.
158. Nitz U, Gluz O, Huober J, Kreipe H, Kates R, Hartmann A*, et al.* Final analysis of the prospective WSG-AGO EC-Doc versus FEC phase III trial in intermediate-risk (pN1) early breast cancer: efficacy and predictive value of Ki67 expression. *Annals of Oncology* 2014;25:1551-7.
159. Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes RC, Cuzick J*, et al.* Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. *Journal of the National Cancer Institute* 2011;103:1656-64.
160. Bartlett JM, Brookes CL, Robson T, van de Velde CJ, Billingham LJ, Campbell FM*, et al.* Estrogen receptor and progesterone receptor as predictive biomarkers of response to endocrine therapy: a prospectively powered pathology study in the Tamoxifen and Exemestane Adjuvant Multinational trial. *Journal of Clinical Oncology* 2011;29:1531-8.
161. Faratian D, Kay C, Campbell F, Robson T, Grant M, Rea D*, et al.* Automated image analysis for high-throughput quantitative detection of ER and PgR expression levels in large-scale clinical studies: The TEAM trial experience. *Breast Cancer Research and Treatment* 2007;106:S215.
162. Hammond MEH, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S*, et al.* American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Archives of pathology & laboratory medicine* 2010;134:e48-e72.
163. Wolff AC, Hammond MEH, Schwartz JN, Hagerty KL, Allred DC, Cote RJ*, et al.* American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Journal of Clinical Oncology* 2006;25:118-45.
164. Kirkegaard T, Edwards J, Tovey S, McGlynn L, Krishna S, Mukherjee R*, et al.* Observer variation in immunohistochemical analysis of protein expression, time for a change? *Histopathology* 2006;48:787-94.
165. Onkologie AG. Aktuelle Empfehlungen zur Diagnostik und Therapie primärer und metastasierter Mammakarzinome der Kommission MAMMA in der AGO. www.ago-online.de.; 2006.
166. National Institute for H, Clinical E. Gene expression profiling and expanded immunohistochemistry tests to guide selection of chemotherapy regimes in breast cancer management. *Diagnostics Assessment Programme* 2011;Final Scope.
167. Howell A. Results of the ATAC (Arimidex, Tamoxifen, Alone or in Combination) trial after completion of 5 years' adjuvant treatment for breast cancer. *The Lancet* 2005;365:60.
168. Montemurro F, Prat A, Rossi V, Valabrega G, Sperinde J, Peraldo-Neia C*, et al.* Potential biomarkers of long-term benefit from single-agent trastuzumab or lapatinib in HER2-positive metastatic breast cancer. *Cancer Research Conference: 36th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2013;73.
169. Gene Expression Omnibus. In. https://www.ncbi.nlm.nih.gov/geo/: National Centre for Biotechnology Information; 2017.
170. Ahn SG, Lee HM, Lee HW, Lee SA, Lee SR, Leem SH*, et al.* Prognostic discrimination using a 70-gene signature among patients with estrogen receptor-

positive breast cancer and an intermediate 21-gene recurrence score. *International Journal of Molecular Sciences* 2013;14:23685-99.
171. Bianchini G, Pusztai L, Karn T, Iwamoto T, Rody A, Kelly C*, et al.* Proliferation and estrogen signaling can distinguish patients at risk for early versus late relapse among estrogen receptor positive breast cancers. *Breast Cancer Research* 2013;15:R86.
172. Cockburn JG, Hallett RM, Gillgrass AE, Dias KN, Whelan T, Levine MN*, et al.* The effects of lymph node status on predicting outcome in ER+ HER2- tamoxifen treated breast cancer patients using gene signatures. *BMC Cancer* 2016;16:555.
173. Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB*, et al.* Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006;355:560-9.
174. Finetti P, Guille A, Adelaide J, Birnbaum D, Chaffanet M, Bertucci F. ESPL1 is a candidate oncogene of luminal B breast cancers. *Breast Cancer Res Treat* 2014;147:51-9.
175. Gyorffy B, Karn T, Sztupinszki Z, Weltz B, Muller V, Pusztai L. Dynamic classification using case-specific training cohorts outperforms static gene expression signatures in breast cancer. *International Journal of Cancer* 2015;136:2091-8.
176. Jonsdottir K, Assmus J, Slewa A, Gudlaugsson E, Skaland I, Baak JP*, et al.* Prognostic value of gene signatures and proliferation in lymph-node-negative breast cancer. *PLoS ONE [Electronic Resource]* 2014;9:e90642.
177. 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 Research and Treatment* 2009;113:231-7.
178. Loi S, Haibe-Kains B, Desmedt C, Lallemand F, Tutt AM, Gillet C*, et al.* Definition of clinically distinct molecular subtypes in estrogen receptor–positive breast carcinomas through genomic grade. *Journal of Clinical Oncology* 2007;25:1239-46.
179. Naoi Y, Kishi K, Tsunashima R, Shimazu K, Shimomura A, Maruyama N*, et al.* Comparison of efficacy of 95-gene and 21-gene classifier (Oncotype DX) for prediction of recurrence in ER-positive and node-negative breast cancer patients. *Breast Cancer Res Treat* 2013;140:299-306.
180. Prat A, Parker JS, Fan C, Cheang MC, Miller LD, Bergh J*, et al.* Concordance among gene expression-based predictors for ER-positive breast cancer treated with adjuvant tamoxifen. *Annals of Oncology* 2012;23:2866-73.
181. Tobin NP, Lindstrom LS, Carlson JW, Bjohle J, Bergh J, Wennmalm K. Multi-level gene expression signatures, but not binary, outperform Ki67 for the long term prognostication of breast cancer patients. Erratum appears in Mol Oncol. 2014 Sep 12;8(6):1159. *Molecular Oncology* 2014;8:741-52.
182. Vollan HK, Rueda OM, Chin SF, Curtis C, Turashvili G, Shah S*, et al.* A tumor DNA complex aberration index is an independent predictor of survival in breast and ovarian cancer. *Molecular Oncology* 2015;9:115-27.
183. Xu W, Jia G, Cai N, Huang S, Davie JR, Pitz M*, et al.* A 16 Yin Yang gene expression ratio signature for ER+/node- breast cancer. *International Journal of Cancer* 2017;140:1413-24.
184. Yang THO, Cheng WY, Zheng T, Maurer MA, Anastassiou D. Breast cancer prognostic biomarker using attractor metagenes and the FGD3-SUSD3 metagene. *Cancer Epidemiology Biomarkers and Prevention* 2014;23:2850-6.
185. Yin ZQ, Liu JJ, Xu YC, Yu J, Ding GH, Yang F*, et al.* A 41-gene signature derived from breast cancer stem cells as a predictor of survival. *Journal of Experimental & Clinical Cancer Research* 2014;33:49.
186. Zemmour C, Bertucci F, Finetti P, Chetrit B, Birnbaum D, Filleron T*, et al.* Prediction of early breast cancer metastasis from DNA microarray data using high-dimensional cox regression models. *Cancer Informatics [Electronic Resource]* 2015;14:129-38.

187. Zhao X, Rodland EA, Sorlie T, Vollan HK, Russnes HG, Kristensen VN*, et al.* Systematic assessment of prognostic gene signatures for breast cancer shows distinct influence of time and ER status. *BMC Cancer* 2014;14:211.
188. School of Health and Related Research. Tumour profiling tests to guide adjuvant chemotherapy decisions in people with breast cancer (update of DG10) - protocol. London; 2017.
189. Bartlett J, Canney P, Campbell A, Cameron D, Donovan J, Dunn J*, et al.* Selecting breast cancer patients for chemotherapy: The opening of the UK OPTIMA trial. *Clinical Oncology* 2013;25:109-16.
190. Hassan S, Ooi J, Garnsey C, Walker P. Does oncotype DX testing prove to be both cost effective and clinically beneficial in patients with early breast cancer? A single unit observational study. *British Journal of Surgery* 2015;102:32-.
191. Hassan S, Ooi J, Welsh R, Pearson M. Oncotype DX testing: Our experience at the Royal Bolton Hospital. *European Journal of Surgical Oncology* 2015;41 (6):S28.
192. Holt S, Bertelli G, Humphreys I, Valentine W, Durrani S, Pudney D*, et al.* A decision impact, decision conflict and economic assessment of routine Oncotype DX testing of 146 women with node-negative or pNImi, ER-positive breast cancer in the U.K. *British Journal of Cancer* 2013;108:2250-8.
193. Albanell J, Svedman C, Gligorov J, Holt SD, Bertelli G, Blohmer JU*, et al.* Pooled analysis of prospective European studies assessing the impact of using the 21-gene Recurrence Score assay on clinical decision making in women with oestrogen receptor-positive, human epidermal growth factor receptor 2-negative early-stage breast cancer. *European Journal of Cancer* 2016;66:104-13.
194. Kiernan T, Olsson-Brown AC, Innes H, Holcombe C, Thorp N, O'Hagan J*, et al.* Knowledge of oncotype Dx recurrence score increases confidence and concordance in adjuvant decisions of U.K. oncologists. *Cancer Research Conference: 38th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2016;76.
195. Kuchel A, Robinson T, Comins C, Shere M, Varughese M, Sparrow G*, et al.* The impact of the 21-gene assay on adjuvant treatment decisions in oestrogen receptor-positive early breast cancer: a prospective study. *British Journal of Cancer* 2016;114:731-6.
196. Loncaster J, Armstrong A, Howell S, Wilson G, Welch R, Chittalia A*, et al.* Impact of Oncotype DX breast Recurrence Score testing on adjuvant chemotherapy use in early breast cancer: Real world experience in Greater Manchester, UK. *European Journal of Surgical Oncology* 2017;09:09.
197. Albanell J, Gonzalez A, Ruiz-Borrego M, Alba E, Garcia-Saenz JA, Corominas JM*, et al.* Prospective transGEICAM study of the impact of the 21-gene Recurrence Score assay and traditional clinicopathological factors on adjuvant clinical decision making in women with estrogen receptor-positive (ER+) node-negative breast cancer. *Annals of Oncology* 2012;23:625-31.
198. Bodmer A, Hugli A, Diebold Berger S, Favet L, Guetty Alberto M, Exquis B. Usefulness of the 21-gene assay to guide adjuvant chemotherapy decision-making: Geneva experience. *Breast* 2015;24:S108.
199. De San Vicente BL, Novas P, Calvo EG, Arrazubi V, Sala MA, Fernandez S*, et al.* Impact of the intermediate Oncotype DX Recurrence Score results on adjuvant treatment recommendations in hormone receptor-positive early breast cancer in a single center. *Journal of Clinical Oncology Conference* 2015;33.
200. Dieci MV, Guarneri V, Mion M, Tortora G, Morandi P, Gori S*, et al.* First prospective multicenter Italian study on the impact of the 21-gene recurrence score (RS) in adjuvant clinical decisions for ER + HER2-early breast cancer patients. *Annals of Oncology Conference: 41st European Society for Medical Oncology Congress, ESMO* 2016;27.

201. Dreyfus C, Ballester M, Gligorov J, Agranat P, Antoine M, Tengher I*, et al.* Impact of the 21-gene assay in decision-making during multidisciplinary breast meeting: A French experience. *Gynecologie, Obstetrique & Fertilite* 2015;43:780-5.
202. Eiermann W, Rezai M, Kummel S, Kuhn T, Warm M, Friedrichs K*, et al.* The 21-gene recurrence score assay impacts adjuvant therapy recommendations for ER-positive, node-negative and node-positive early breast cancer resulting in a risk-adapted change in chemotherapy use. *Annals of Oncology* 2013;24:618-24.
203. Gligorov J, Pivot XB, Jacot W, Naman HL, Spaeth D, Misset JL*, et al.* Prospective clinical utility study of the use of the 21-gene assay in adjuvant clinical decision making in women with estrogen receptor-positive early invasive breast cancer: Results From the SWITCH study. *Oncologist* 2015;20:873-9.
204. Hejduk K, Petrakova K, Petruzelka L, Bielcikova Z, Kolarova I, Finek J*, et al.* Economic assessment of routine oncotype DX testing of estrogen receptor positive (ER+) early breast cancer (EBC) patients with grade 2 tumours in Czech Republic. *Value in Health* 2016;19 (7):A720.
205. Petrakova K, Petruzelka L, Holanek M, Svoboda T, Palacova M, Bielcikova Z*, et al.* The impact of the 21-gene assay in the Czech Republic on adjuvant chemotherapy (CT) recommendations and costs in estrogen receptor positive (ER+) early stage breast cancer (ESBC) patients with grade 2 tumors and risk factors. *Annals of Oncology Conference: 41st European Society for Medical Oncology Congress, ESMO* 2016;27.
206. Petrakova K, Petruzelka LB, Holanek M, Svoboda T, Palacova M, Kolarova I*, et al.* The impact of the 21-gene assay in the Czech Republic on adjuvant chemotherapy (CT) recommendations and costs in estrogen receptor positive (ER+) early stage breast cancer (ESBC) patients with grade 2 tumors and risk factors. *Journal of Clinical Oncology Conference* 2016;34.
207. Mouysset J, Laplaige P, Fignon A, Jallais L, Lafuma A, Michaud P*, et al.* The 21-gene assay in the decision impact assessment of ER+, HER2-breast cancer: A French real life prospective study. *Value in Health* 2016;19 (7):A685.
208. Novas P, Calvo EG, Gonzalez MAS, Arteaga JFA, San Vicente BLD, Arrazubi V*, et al.* Impact of Onco-type DX recurrence score in the management of pN1mic early breast cancer. *Journal of Clinical Oncology Conference* 2016;34.
209. Pestalozzi BC, Tausch C, Dedes KJ, Rochlitz C, Zimmermann S, Von Moos R*, et al.* Adjuvant treatment recommendations for ER+ early breast cancer patients by Swiss tumor boards (SAKK 26/10). *European Journal of Cancer* 2015;51:S313.
210. Wassermann J, Toledano A, Gligorov J, Conforti R, Grapin JP, Benothman S*, et al.* Routine practice use of Oncotype Dx assay in French breast cancer patients. *European Journal of Cancer* 2015;51:S323.
211. Ettl J, Lackmann KG, Hapfelmeier A, Klein E, Paepke S, Petry C*, et al.* Prospective comparison of conventional clinicopathological factors, uPA/PAI-1 and EndoPredict-clin score (EPclin) for adjuvant clinical decision making in ER-positive, HER2-negative breast cancer: Progesterone receptor expression is strongly associat. *Cancer Research Conference: 37th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2015;75.
212. Muller BM, Keil E, Lehmann A, Winzer KJ, Richter-Ehrenstein C, Prinzler J*, et al.* The EndoPredict Gene-Expression Assay in Clinical Practice - Performance and Impact on Clinical Decisions. *PLoS ONE [Electronic Resource]* 2013;8:e68252.
213. Penault-Llorca FM, Kwiatkovski F, Grenier J. A prospective multicenter non-randomized trial evaluating the effect of EndoPredict® (EPclin®) clinicogenomic test on treatment decision making among patients with intermediate clinical risk. *SABCS poster* 2016.
214. Yeo B, Zabaglo L, Hills M, Dodson A, Smith I, Dowsett M. Clinical utility of the IHC4+C score in oestrogen receptor-positive early breast cancer: a prospective decision impact study. *British Journal of Cancer* 2015;113:390-5.

215. Van Asten K, Neven P, Puthod V, Ghali N, Lintermans A, Jongen L*, et al.* Concordance between PAM50 and clinico-pathological prognostic markers when deciding on adjuvant chemotherapy in early breast cancer. *European Journal of Cancer* 2016;57:S146.
216. Cusumano PG, Generali D, Ciruelos E, Manso L, Ghanem I, Lifrange E*, et al.* European inter-institutional impact study of MammaPrint. *Breast* 2014;23:423-8.
217. Drukker CA, van den Hout HC, Sonke GS, Brain E, Bonnefoi H, Cardoso F*, et al.* Risk estimations and treatment decisions in early stage breast cancer: agreement among oncologists and the impact of the 70-gene signature. *European Journal of Cancer* 2014;50:1045-54.
218. Exner R, Bago-Horvath Z, Bartsch R, Mittlboeck M, Retel VP, Fitzal F*, et al.* The multigene signature MammaPrint impacts on multidisciplinary team decisions in ER+, HER2- early breast cancer. *British Journal of Cancer* 2014;111:837-42.
219. Hartmann S, Gerber B, Elling D, Heintze K, Reimer T. The 70-Gene Signature as Prognostic Factor for Elderly Women with Hormone Receptor-Positive, HER2-Negative Breast Cancer. *Breast Care* 2012;7:19-24.
220. Kuijer A, Straver M, Van Bommel AC. Impact of the 70-gene signature on adjuvant chemotherapy decisions in ER+ early breast cancer patients: results of a prospective cohort study. *San Antonio Breast Cancer Symposium* 2016.
221. Rullan AJ, Pernas S, Margeli M, Joan D, Quiroga V, Stradella A*, et al.* Use of Mammaprint (c) (MMP) genetic signature in early breast cancer patients. Economic analisys of a 129 patient cohort treated in three Spanish hospitals. *Cancer Research* 2016;76.
222. Wuerstlein R, Gluz O, Kates R. Results of multigene assay (MammaPrint) and molecular subtyping (BluePrint) substantially impact treatment decision making in early breast cancer: Final analysis of the WSG PRIMe Decision Impact Study. *San Antonio Breast Cancer Symposium (SABCS)* 2016.
223. Horowitz M, Wilner N, Alvarez W. Impact of Event Scale: a measure of subjective stress. *Psychosomatic medicine* 1979;41:209-18.
224. Losk K, Vaz-Luis I, Camuso K, Batista R, Lloyd M, Tukenmez M*, et al.* Factors Associated With Delays in Chemotherapy Initiation Among Patients With Breast Cancer at a Comprehensive Cancer Center. *Journal of the National Comprehensive Cancer Network* 2016;14:1519-26.
225. Myriad Genetics. Company submission - EndoPredict; 2017.
226. Blank PR, Filipits M, Dubsky P, Gutzwiller F, Lux MP, Brase JC*, et al.* Cost-effectiveness analysis of prognostic gene expression signature-based stratification of early breast cancer patients. *Pharmacoeconomics* 2015;33:179-90.
227. Bargallo-Rocha JE, Lara-Medina F, Perez-Sanchez V, Vazquez-Romo R, Villarreal-Garza C, Martinez-Said H*, et al.* Cost-effectiveness of the 21-gene breast cancer assay in Mexico. *Advances in Therapy* 2015;32:239-53.
228. Davidson JA, Cromwell I, Ellard SL, Lohrisch C, Gelmon KA, Shenkier T*, et al.* A prospective clinical utility and pharmacoeconomic study of the impact of the 21-gene Recurrence Score assay in oestrogen receptor positive node negative breast cancer. *European Journal of Cancer* 2013;49:2469-75.
229. Jahn B, Rochau U, Kurzthaler C, Hubalek M, Miksad R, Sroczynski G*, et al.* Cost effectiveness of personalized treatment in women with early breast cancer: the application of OncotypeDX and Adjuvant! Online to guide adjuvant chemotherapy in Austria. *Springerplus* 2015;4:752.
230. Kondo M, Hoshi SL, Yamanaka T, Ishiguro H, Toi M. Economic evaluation of the 21-gene signature (Oncotype DX) in lymph node-negative/positive, hormone receptor-positive early-stage breast cancer based on Japanese validation study (JBCRG-TR03). *Breast Cancer Research and Treatment* 2011;127:739-49.
231. Lamond NW, Skedgel C, Rayson D, Lethbridge L, Younis T. Cost-utility of the 21-gene recurrence score assay in node-negative and node-positive breast cancer. *Breast Cancer Res Treat* 2012;133:1115-23.

232. Paulden M, Franek J, Pham B, Bedard PL, Trudeau M, Krahn M. Cost-effectiveness of the 21-gene assay for guiding adjuvant chemotherapy decisions in early breast cancer. *Value in Health* 2013;16:729-39.
233. Reed SD, Dinan MA, Schulman KA, Lyman GH. Cost-effectiveness of the 21-gene recurrence score assay in the context of multifactorial decision making to guide chemotherapy for early-stage breast cancer. *Genetics in Medicine* 2013;15:203-11.
234. Bonastre J, Marguet S, Lueza B, Michiels S, Delaloge S, Saghatchian M. Cost effectiveness of molecular profiling for adjuvant decision making in patients with node-negative breast cancer. *Journal of Clinical Oncology* 2014;32:3513-9.
235. Retel VP, Joore MA, Linn SC, Rutgers EJ, van Harten WH. Scenario drafting to anticipate future developments in technology assessment. *BMC Research Notes* 2012;5:442.
236. Retel VP, Joore MA, van Harten WH. Head-to-head comparison of the 70-gene signature versus the 21-gene assay: cost-effectiveness and the effect of compliance. *Breast Cancer Res Treat* 2012;131:627-36.
237. Retel VP, Grutters JP, van Harten WH, Joore MA. Value of research and value of development in early assessments of new medical technologies. *Value in Health* 2013;16:720-8.
238. Retel VP, Joore MA, Drukker CA, Bueno-de-Mesquita JM, Knauer M, van Tinteren H*, et al.* Prospective cost-effectiveness analysis of genomic profiling in breast cancer. *European Journal of Cancer* 2013;49:3773-9.
239. Hall PS, McCabe C, Stein RC, Cameron D. Economic evaluation of genomic test-directed chemotherapy for early-stage lymph node-positive breast cancer. *Journal of the National Cancer Institute* 2012;104:56-66.
240. Hannouf MB, Xie B, Brackstone M, Zaric GS. Cost-effectiveness of a 21-gene recurrence score assay versus Canadian clinical practice in women with early-stage estrogen- or progesterone-receptor-positive, axillary lymph-node negative breast cancer. *BMC Cancer* 2012;12:447.
241. Hannouf MB, Xie B, Brackstone M, Zaric GS. Cost effectiveness of a 21-gene recurrence score assay versus Canadian clinical practice in post-menopausal women with early-stage estrogen or progesterone-receptor-positive, axillary lymph-node positive breast cancer. *Pharmacoeconomics* 2014;32:135-47.
242. Kondo M, Hoshi SL, Ishiguro H, Toi M. Economic evaluation of the 70-gene prognosis-signature (MammaPrint) in hormone receptor-positive, lymph node-negative, human epidermal growth factor receptor type 2-negative early stage breast cancer in Japan. *Breast Cancer Research and Treatment* 2012;133:759-68.
243. Mislick K, Schonfeld W, Bodnar C, Tong KB. Cost-effectiveness analysis of Mammostrat compared with Oncotype DX to inform the treatment of breast cancer. *Clinicoeconomics & Outcomes Research* 2014;6:37-47.
244. Stein RC, Dunn JA, Bartlett JM, Campbell AF, Marshall A, Hall P*, et al.* OPTIMA prelim: a randomised feasibility study of personalised care in the treatment of women with early breast cancer. *Health Technology Assessment* 2016;20:xxiii-xxix, 1-201.
245. Tiwana SK, Smith A, Leggett L, MacKean G, Lorenzetti D, Clement F. Use of Oncotype DX for guiding adjuvant chemotherapy decisions in early stage invasive breast cancer patients in Alberta. 2013, http://www.crd.york.ac.uk/CRDWeb/ShowRecord.asp?ID=32015000433.
246. Vanderlaan BF, Broder MS, Chang EY, Oratz R, Bentley TG. Cost-effectiveness of 21-gene assay in node-positive, early-stage breast cancer. *American Journal of Managed Care* 2011;17:455-64.
247. Wong WB, Ramsey SD, Barlow WE, Garrison LP, Jr., Veenstra DL. The value of comparative effectiveness research: projected return on investment of the RxPONDER trial (SWOG S1007). *Contemporary Clinical Trials* 2012;33:1117-23.
248. Ramsey SD, Barlow WE, Gonzalez-Angulo AM, Tunis S, Baker L, Crowley J*, et al.* Integrating comparative effectiveness design elements and endpoints into a phase III, randomized clinical trial (SWOG S1007) evaluating oncotypeDX-guided

management for women with breast cancer involving lymph nodes. *Contemporary Clinical Trials* 2013;34:1-9.
249. Yang M, Rajan S, Issa AM. Cost effectiveness of gene expression profiling for early stage breast cancer: a decision-analytic model. *Cancer* 2012;118:5163-70.
250. Yamauchi H, Nakagawa C, Yamashige S, Takei H, Yagata H, Yoshida A*, et al.* Societal cost-effectiveness analysis of the 21-gene assay in estrogen-receptor-positive, lymph-node-negative early-stage breast cancer in Japan. *BMC Health Services Research* 2014;14:372.
251. Asad J, Jacobson AF, Estabrook A, Smith SR, Boolbol SK, Feldman SM*, et al.* Does oncotype DX recurrence score affect the management of patients with early-stage breast cancer? *American Journal of Surgery* 2008;196:527-9.
252. Yamauchi H, Nakagawa C, Takei H, Chao C, Yoshizawa C, Yagata H*, et al.* Prospective study of the effect of the 21-gene assay on adjuvant clinical decision-making in Japanese women with estrogen receptor-positive, node-negative, and node-positive breast cancer. *Clinical Breast Cancer* 2014;14:191-7.
253. Edlin R, Connock M, Tubeuf S, Round J, Fry-Smith A, Hyde C*, et al.* Azacitidine for the treatment of myelodysplastic syndrome, chronic myelomonocytic leukaemia and acute myeloid leukaemia. *Health technology assessment (Winchester, England)* 2010;Suppl1:69-74.
254. National Cancer Registration and Analysis Service. Bespoke data request - chemotherapy use amongst women with early breast cancer (data held on file); 2017.
255. NHS England. NHS England Oncotype DX Access Scheme dataset (unpublished, data held on file by Genomic Health); 2016.
256. Curtis L, Burns A. Unit costs of health and social care 2016. Kent; 2016.
257. Thomas RJ, Williams M, Marshall C, Glen J, Callam M. The total hospital and community UK costs of managing patients with relapsed breast cancer. *British Journal of Cancer* 2009;100:598-600.
258. Office for National Statistics. National life tables, UK: 2013–2015; 2016.
259. Shivers SC, Clark L, Esposito N, Howard N, King J, Acs G*, et al.* Direct comparison of risk classification between MammaPrint, Oncotype DX and MammoStrat assays in patients with early stage breast cancer. *Cancer Research Conference: 36th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2013;73.
260. Varga Z, Sinn P, Fritzsche F, von Hochstetter A, Noske A, Schraml P*, et al.* Comparison of EndoPredict and Oncotype DX test results in hormone receptor positive invasive breast cancer. Erratum appears in PLoS One. 2013;8(10). doi:10.1371/annotation/f715f38e-7aee-4d2b-8bbf-da0411dc6ef3. *PLoS ONE [Electronic Resource]* 2013;8:e58483.
261. Alvarado MD, Prasad C, Rothney M, Cherbavaz DB, Sing AP, Baehner FL*, et al.* A prospective comparison of the 21-gene recurrence score and the PAM50-based Prosigna in estrogen receptor-positive early-stage breast cancer. *Advances in Therapy* 2015;32:1237-47.
262. 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* 2005;365:1687-717.
263. 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. *European Journal of Cancer* 2011;47:2517-30.
264. Chang J, Clark GM, Allred DC, Mohsin S, Chamness G, Elledge RM. Survival of patients with metastatic breast carcinoma. *Cancer* 2003;97:545-53.
265. Lidgren M, Wilking N, Jonsson B, Rehnberg C. Health related quality of life in different states of breast cancer. *Quality of Life Research* 2007;16:1073-81.
266. Kind PHG, Macran S. UK Population Norms for EQ-5D York. University of York; 1999.

267. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. *Value in Health* 2010;13:509-18.
268. National Institute for Health and Care Excellence. MIB27: The Prosigna gene expression profiling assay for assessing longterm risk of breast cancer recurrence. London; 2015.
269. 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. *British Journal of Cancer* 2007;97:479-85.
270. National Institute for Health and Care Excellence. Early and locally advanced breast cancer: diagnosis and treatment. London; 2009.
271. Robertson C, Arcot Ragupathy SK, Boachie C, Dixon JM, Fraser C, Hernández R. The clinical effectiveness and cost-effectiveness of different surveillance mammography regimens after the treatment for primary breast cancer: systematic reviews, registry database analyses and economic evaluation. *Health technology assessment (Winchester, England)* 2011;15:1-322.
272. Jakesz R, Jonat W, Gnant M, Mittlboeck M, Greil R, Tausch C*, et al.* Switching of postmenopausal women with endocrine-responsive early breast cancer to anastrozole after 2 years' adjuvant tamoxifen: combined results of ABCSG trial 8 and ARNO 95 trial. *Lancet* 2005;366.
273. Schmid M, Jakesz R, Samonigg H, Kubista E, Gnant M, Menzel C*, et al.* Randomized trial of tamoxifen versus tamoxifen plus aminoglutethimide as adjuvant treatment in postmenopausal breast cancer patients with hormone receptor-positive disease: Austrian breast and colorectal cancer study group trial 6. *J Clin Oncol* 2003;21:984-90.
274. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Comparisons between diff erent polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100 000 women in 123 randomised trials. *Lancet* 2012;379:432-44.
275. de Bock GH, Putter H, Bonnema J, van der Hage JA, Bartelink H, van de Velde CJ. 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.
276. Wolff AC, Blackford AL, Visvanathan K, Rugo HS, Moy B, Goldstein LJ. Risk of marrow neoplasms after adjuvant breast cancer therapy: the National Comprehensive Cancer Network experience. *Journal of Clinical Oncology* 2015;33:340-9.
277. Younis T, Rayson D, Skedgel C. Adjuvant chemotherapy in breast cancer: is TC a cost-effective regimen compared to AC? *Cancer Research* 2008;CTRC-AACR San Antonio Breast Cancer Symposium: 2008 Abstracts.
278. Hall P., Smith A, Hulme C, Vargas-Palacios A, Makris A, Hughes-Davies L*, et al.* Value of information analysis of multiparameter tests for chemotherapy in early breast cancer: the OPTIMA-prelim trial. *Value in Health (epub ahead of print)* 2017.
279. Joint Formulary Committee. British National Formulary. British National Formulary; 2017.
280. Department of Health. NHS Reference Costs 2015/16. London; 2017.
281. Cuzick J, Sestak I, Baum M, Buzdar A, Howell A, Dowsett M*, et al.* Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 10-year analysis of the ATAC trial. *The Lancet Oncology* 2010;11:1135-41.
282. Pan H, Gray R, Davies C, Peto R, Bergh J, Pritchard KI*, et al.* Long-term recurrence risks after use of endocrine therapy for only 5 years: Relevance of breast tumour characteristics. ASCO Annual Meeting 2016, abstract no. 6362.
283. Holt S, Bertelli G, Brinkworth E, Durrani S, Jones S, Khawaja S*, et al.* Results from a prospective clinical study on the impact of Oncotype DX on adjuvant treatment decision making in a cohort of 142 UK patients. *Cancer Research Conference: 34th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2011;71.

284. National institute for Health and Care Excellence. Azacitidine for the treatment of myelodysplastic syndromes, chronic myelomonocytic leukaemia and acute myeloid leukaemia. Technology Appraisal Guidance 218. London; 2011.
285. Peasgood T, Ward SE, Brazier J. Health state utility values in breast cancer: A review and metaanalysis. *Value in Health* 2011;14 (7):A462.
286. Farkkila N, Roine R, Jahkola T, Sintonen H, Hanninen J, Taari K*, et al.* Health state utilities in breast cancer. *Value in Health* 2011;14 (7):A459.
287. Yousefi M, Najafi S, Ghaffari S, Mahboub-Ahari A, Ghaderi H. Comparison of SF-6D and EQ-5D Scores in Patients With Breast Cancer. *Iranian Red Crescent Medical Journal* 2016;18:e23556.
288. Naik H, Howell D, Su S, Qiu X, Brown MC, Vennettilli A*, et al.* EQ-5D Health Utility Scores: Data from a Comprehensive Canadian Cancer Centre. *Patient* 2017;10:105-15.
289. Ren S, Minton J, Whyte S, Latimer N, Stevenson M. How to Sample Ordered Parameters in Probabilistic Sensitivity Analysis. ISPOR; Glasgow, abstract no. 4033.
290. Department of Health. Drugs and pharmaceutical electronic market information (eMit). Leeds; 2017.
291. Department of Health. NHS Reference Costs 2013 to 2014. 2015. https://www.gov.uk/government/publications/nhs-reference-costs-2013-to-2014 (Accessed
292. van't Veer LJ, Rutgers EJT, Piccart M, et al. 70-gene signature in early-stage breast cancer. *New England Journal of Medicine* 2016;375:2200-1.

9 APPENDICES

Appendix 1: Literature search strategies

CLINICAL EFFECTIVENESS

Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present
27th February 2017

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

35	pam50.mp.
36	50-gene.mp.
37	gene50.mp.
38	gene?fifty.mp.
39	fifty?gene.mp.
40	breast bioclassifier.mp.
41	ihc4.mp.
42	or/10-14
43	or/15-41
44	9 and 42
45	9 and 43
46	limit 45 to yr="2011 -Current"
47	44 or 46

Embase 1974 to 2017 February 24
27th February 2017

#	**Searches**
1	exp breast tumor/
2	exp breast/
3	exp neoplasm/
4	2 and 3
5	(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.
6	(mammar* adj5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)).mp.
7	1 or 4 or 5 or 6
8	EndoPredict.mp.
9	myriad genetics.mp.
10	sividon diagnostics.mp.
11	ep score.mp.
12	epclin score.mp.
13	MammaPrint.mp.
14	70-gene.mp.
15	gene70.mp.
16	gene?seventy.mp.
17	seventy?gene.mp.
18	amsterdam profile.mp.
19	oncotype.mp.
20	oncotype dx.mp.
21	21-gene.mp.
22	gene21.mp.
23	gene?twentyone.mp.
24	twentyone?gene.mp.
25	ghi recurrence score.mp.

26	ghi-rs.mp.
27	92-gene.mp.
28	gene92.mp.
29	gene?ninetytwo.mp.
30	ninetytwo?gene.mp.
31	(rct-pcr adj5 '21').mp.
32	prosigna.mp.
33	pam50.mp.
34	50-gene.mp.
35	gene50.mp.
36	gene?fifty.mp.
37	fifty?gene.mp.
38	breast bioclassifier.mp.
39	ihc4.mp.
40	or/8-12
41	or/13-39
42	7 and 40
43	7 and 41
44	limit 43 to yr="2011 -Current"
45	42 or 44

Web of Science® Core Collection
Science Citation Index Expanded (1900-)
Conference Proceedings Citation Index - Science (1990-)
27th February 2017

#	**Searches**
# 1	TOPIC: ((breast* NEAR/5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullary)))
# 2	TOPIC: ((mammar* NEAR/5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)))
# 3	#2 OR #1
# 4	TOPIC: (EndoPredict) OR TOPIC: (myriad genetics) OR TOPIC: (sividon diagnostics) OR TOPIC: (ep score) OR TOPIC: (epclin score)
# 5	TS=(MammaPrint) OR TS=(70-gene) OR TS=(gene70) OR TS=(gene?seventy) OR TS=(seventy?gene) OR TS=(amsterdam profile)
# 6	TS=(oncotype) OR TS=(oncotype dx) OR TS=(21-gene) OR TS=(gene21) OR TS=(gene?twentyone) OR TS=(twentyone?gene) OR TS=(ghi recurrence score) OR TS=(ghi-rs) OR TS=(92-gene) OR TS=(gene92) OR TS=(gene?ninetytwo) OR TS=(ninetytwo?gene) OR TS=((rct-pcr NEAR/5 '21'))
# 7	TOPIC: (prosigna) OR TOPIC: (pam50) OR TOPIC: (50-gene) OR TOPIC: (gene50) OR TOPIC: (gene?fifty) OR TOPIC: (fifty?gene) OR TOPIC: (breast bioclassifier)
# 8	TOPIC: (ihc4)
# 9	#8 OR #7 OR #6 OR #5
# 10	#9 AND #3 Indexes=SCI-EXPANDED, CPCI-S Timespan=2011-2017

# 11	#4 AND #3
# 12	#11 OR #10

Cochrane Database of Systematic Reviews (CDR): Wiley Interscience.
Cochrane Central Register of Controlled Trials (CENTRAL): Wiley Interscience.
Health Technology Assessment Database (HTA): Wiley Interscience. 1995-2016
Database of Abstracts of Reviews of Effects (DARE)): Wiley Interscience. 1995-2015
NHS Economic Evaluation Database (NHS EED): Wiley Interscience. 1995-2015
28th February 2017

#	Searches
#1	MeSH descriptor: [Breast Neoplasms] explode all trees
#2	MeSH descriptor: [Neoplasms, Ductal, Lobular, and Medullary] explode all trees
#3	MeSH descriptor: [Breast] explode all trees
#4	MeSH descriptor: [Neoplasms] explode all trees
#5	#3 and #4
#6	(breast* near/5 (neoplasm* or cancer* or tumor* or tumour* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullary))
#7	(mammar* near/5 (neoplasm* or cancer* or tumor* or tumour* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar))
#8	#1 or #2 or #5 or #6 or #7
#9	EndoPredict
#10	myriad genetics
#11	sividon diagnostics
#12	ep score
#13	epclin score
#14	MammaPrint
#15	70-gene
#16	gene70
#17	gene*seventy
#18	seventy*gene
#19	amsterdam profile
#20	oncotype
#21	oncotype dx
#22	21-gene
#23	gene21
#24	gene*twentyone
#25	twentyone*gene
#26	ghi recurrence score
#27	ghi-rs
#28	92-gene
#29	gene92
#30	gene*ninetytwo
#31	ninetytwo*gene
#32	(rct-pcr near/5 '21')

#33	prosigna
#34	pam50
#35	50-gene
#36	gene50
#37	gene*fifty
#38	fifty*gene
#39	breast bioclassifier
#40	ihc4
#41	(or #9-#13)
#42	27-#40
#43	#8 and #41
#44	#8 and #42 Publication Year from 2011
#45	#43 or #44

WHOICTRP
28th February 2017

#	**Searches**
1	EndoPredict or MammaPrint or Oncotype or IHC4 or Prosigna

COST-EFFECTIVENESS STUDIES OF TUMOUR PROFILING TESTS

Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present
6th March 2017

#	Searches
1	exp Breast Neoplasms/
2	exp mammary neoplasms/
3	exp "Neoplasms, Ductal, Lobular, and Medullary"/
4	exp breast/
5	exp neoplasms/
6	4 and 5
7	(breast* adj5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullary)).mp.
8	(mammar* adj5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)).mp.
9	1 or 2 or 3 or 6 or 7 or 8
10	EndoPredict.mp.
11	myriad genetics.mp.
12	sividon diagnostics.mp.
13	ep score.mp.
14	epclin score.mp.
15	MammaPrint.mp.
16	70-gene.mp.

17	gene70.mp.
18	gene?seventy.mp.
19	seventy?gene.mp.
20	amsterdam profile.mp.
21	oncotype.mp.
22	oncotype dx.mp.
23	21-gene.mp.
24	gene21.mp.
25	gene?twentyone.mp.
26	twentyone?gene.mp.
27	ghi recurrence score.mp.
28	ghi-rs.mp.
29	92-gene.mp.
30	gene92.mp.
31	gene?ninetytwo.mp.
32	ninetytwo?gene.mp.
33	(rct-pcr adj5 '21').mp.
34	prosigna.mp.
35	pam50.mp.
36	50-gene.mp.
37	gene50.mp.
38	gene?fifty.mp.
39	fifty?gene.mp.
40	breast bioclassifier.mp.
41	ihc4.mp.
42	or/10-14
43	or/15-41
44	9 and 42
45	9 and 43
46	limit 45 to yr="2011 -Current"
47	44 or 46
48	exp "Costs and Cost Analysis"/
49	Economics/
50	exp Economics, Hospital/
51	exp Economics, Medical/
52	Economics, Nursing/
53	exp models, economic/
54	Economics, Pharmaceutical/
55	exp "Fees and Charges"/
56	exp Budgets/
57	budget$.tw.
58	ec.fs.
59	cost$.ti.
60	(cost$ adj2 (effective$ or utilit$ or benefit$ or minimi$)).ab.

61	(economic$ or pharmacoeconomic$ or pharmaco-economic$).ti.
62	(price$ or pricing$).tw.
63	(financial or finance or finances or financed).tw.
64	(fee or fees).tw.
65	(value adj2 (money or monetary)).tw.
66	quality-adjusted life years/
67	(qaly or qalys).af.
68	(quality adjusted life year or quality adjusted life years).af.
69	or/48-68
70	47 and 69

Embase 1974 to 2017 March 03
6th March 2017

#	**Searches**
1	exp breast tumor/
2	exp breast/
3	exp neoplasm/
4	2 and 3
5	(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.
6	(mammar* adj5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)).mp.
7	1 or 4 or 5 or 6
8	EndoPredict.mp.
9	myriad genetics.mp.
10	sividon diagnostics.mp.
11	ep score.mp.
12	epclin score.mp.
13	MammaPrint.mp.
14	70-gene.mp.
15	gene70.mp.
16	gene?seventy.mp.
17	seventy?gene.mp.
18	amsterdam profile.mp.
19	oncotype.mp.
20	oncotype dx.mp.
21	21-gene.mp.
22	gene21.mp.
23	gene?twentyone.mp.
24	twentyone?gene.mp.
25	ghi recurrence score.mp.
26	ghi-rs.mp.
27	92-gene.mp.

28	gene92.mp.
29	gene?ninetytwo.mp.
30	ninetytwo?gene.mp.
31	(rct-pcr adj5 '21').mp.
32	prosigna.mp.
33	pam50.mp.
34	50-gene.mp.
35	gene50.mp.
36	gene?fifty.mp.
37	fifty?gene.mp.
38	breast bioclassifier.mp.
39	ihc4.mp.
40	or/8-12
41	or/13-39
42	7 and 40
43	7 and 41
44	limit 43 to yr="2011 -Current"
45	42 or 44
46	Socioeconomics/
47	Cost benefit analysis/
48	Cost effectiveness analysis/
49	Cost of illness/
50	Cost control/
51	Economic aspect/
52	Financial management/
53	Health care cost/
54	Health care financing/
55	Health economics/
56	Hospital cost/
57	(fiscal or financial or finance or funding).tw.
58	Cost minimization analysis/
59	(cost adj estimate$).mp.
60	(cost adj variable$).mp.
61	(unit adj cost$).mp.
62	or/46-61
63	45 and 62

Web of Science® Core Collection
Science Citation Index Expanded (1900-)
Conference Proceedings Citation Index - Science (1990-)
6th March 2017

#	Searches
# 1	TOPIC: ((breast* NEAR/5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullary)))
# 2	TOPIC: ((mammar* NEAR/5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)))
# 3	#2 OR #1
# 4	TOPIC: (EndoPredict) OR TOPIC: (myriad genetics) OR TOPIC: (sividon diagnostics) OR TOPIC: (ep score) OR TOPIC: (epclin score)
# 5	TS=(MammaPrint) OR TS=(70-gene) OR TS=(gene70) OR TS=(gene?seventy) OR TS=(seventy?gene) OR TS=(amsterdam profile)
# 6	TS=(oncotype) OR TS=(oncotype dx) OR TS=(21-gene) OR TS=(gene21) OR TS=(gene?twentyone) OR TS=(twentyone?gene) OR TS=(ghi recurrence score) OR TS=(ghi-rs) OR TS=(92-gene) OR TS=(gene92) OR TS=(gene?ninetytwo) OR TS=(ninetytwo?gene) OR TS=((rct-pcr NEAR/5 '21'))
# 7	TOPIC: (prosigna) OR TOPIC: (pam50) OR TOPIC: (50-gene) OR TOPIC: (gene50) OR TOPIC: (gene?fifty) OR TOPIC: (fifty?gene) OR TOPIC: (breast bioclassifier)
# 8	TOPIC: (ihc4)
# 9	#8 OR #7 OR #6 OR #5
# 10	#9 AND #3 Indexes=SCI-EXPANDED, CPCI-S Timespan=2011-2017
# 11	#4 AND #3
# 12	#11 OR #10
# 13	TS=(cost* and (effective* or utilit* or benefit* or minimi*)) OR TS=(cost*) OR TI=(economic* or pharmacoeconomic* or pharmaco-economic*) OR TS=(price* or pricing*) OR TS=(financial or finance or finances or financed) OR TS=(fee or fees) OR TS=(value and (money or monetary)) OR TS=(economic*) OR TS=(economic* and (hospital or medical or nursing or pharmaceutical)) OR TS=("quality adjusted life year" or "quality adjusted life years") OR TS=(qaly or qalys) OR TS=(budget*)
# 14	#13 AND #12

Health Technology Assessment Database (HTA): Wiley Interscience. 1995-2016
NHS Economic Evaluation Database (NHS EED): Wiley Interscience. 1995-2015
7th March 2017

#	Searches
1	MeSH DESCRIPTOR Breast Neoplasms EXPLODE ALL TREES
2	MeSH DESCRIPTOR Neoplasms, Ductal, Lobular, and Medullary EXPLODE ALL TREES
3	MeSH DESCRIPTOR Breast EXPLODE ALL TREES
4	MeSH DESCRIPTOR Neoplasms EXPLODE ALL TREES
5	#3 AND #4
6	((breast* ADJ5 (neoplasm* or cancer* or tumor* or tumour* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullary)))

7	((mammar* ADJ5 (neoplasm* or cancer* or tumor* or tumour* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)))
8	#1 OR #2 OR #5 OR #6 OR #7
9	(EndoPredict or myriad genetics or sividon diagnostics or ep score or epclin score)
10	(MammaPrint or 70-gene or gene70 or gene*seventy or seventy*gene or amsterdam profile)
11	(oncotype or oncotype dx or 21-gene or gene21 or gene*twentyone or twentyone*gene or ghi recurrence score or ghi-rs or 92-gene or gene92 or gene*ninetytwo or ninetytwo*gene or (rct-pcr ADJ5 '21'))
12	(prosigna or pam50 or 50-gene or gene50 or gene*fifty or fifty*gene or breast bioclassifier)
13	(ihc4)
14	#8 AND #9
15	#10 OR #11 OR #12 OR #13
16	(#8 AND #15) FROM 2011 TO 2017
17	(#8 AND #15) IN HTA FROM 2011 TO 2017
18	(#8 AND #15) IN NHSEED FROM 2011 TO 2017

COST-EFFECTIVENESS REVIEWS FOR BREAST CANCER

Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present
7th March 2017

#	**Searches**
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)).ti.
8	(mammar* adj5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)).ti.
9	1 or 2 or 3 or 6 or 7 or 8
10	exp "Costs and Cost Analysis"/
11	Economics/
12	exp Economics, Hospital/
13	exp Economics, Medical/
14	Economics, Nursing/
15	exp models, economic/
16	Economics, Pharmaceutical/
17	exp "Fees and Charges"/
18	exp Budgets/

19	budget$.tw.
20	ec.fs.
21	cost$.ti.
22	(cost$ adj2 (effective$ or utilit$ or benefit$ or minimi$)).ab.
23	(economic$ or pharmacoeconomic$ or pharmaco-economic$).ti.
24	(price$ or pricing$).tw.
25	(financial or finance or finances or financed).tw.
26	(fee or fees).tw.
27	(value adj2 (money or monetary)).tw.
28	quality-adjusted life years/
29	(qaly or qalys).af.
30	(quality adjusted life year or quality adjusted life years).af.
31	or/10-30
32	9 and 31
33	meta-analysis/
34	meta-analysis as topic/
35	(meta analy* or metanaly* or metaanaly*).ti,ab.
36	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
37	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
38	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
39	(search* adj4 literature).ab.
40	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or scie nce citation index or bids or cancerlit).ab.
41	cochrane.jw.
42	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
43	or/33-42
44	32 and 43
45	limit 44 to yr="2011 -Current"

Embase 1974 to 2017 March 06
7th March 2017

#	Searches
1	exp breast tumor/
2	exp breast/
3	exp neoplasm/
4	2 and 3
5	(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)).ti.
6	(mammar* adj5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)).ti.
7	1 or 4 or 5 or 6
8	Socioeconomics/

9	Cost benefit analysis/
10	Cost effectiveness analysis/
11	Cost of illness/
12	Cost control/
13	Economic aspect/
14	Financial management/
15	Health care cost/
16	Health care financing/
17	Health economics/
18	Hospital cost/
19	(fiscal or financial or finance or funding).tw.
20	Cost minimization analysis/
21	(cost adj estimate$).mp.
22	(cost adj variable$).mp.
23	(unit adj cost$).mp.
24	or/8-23
25	7 and 24
26	systematic review/
27	meta-analysis/
28	(meta analy* or metanaly* or metaanaly*).ti,ab.
29	((systematic or evidence) adj3 (review* or overview*)).ti,ab.
30	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
31	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
32	(search* adj4 literature).ab.
33	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
34	cochrane.jw.
35	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
36	or/26-35
37	25 and 36
38	limit 37 to yr="2011 -Current"

Web of Science® Core Collection
Science Citation Index Expanded (1900-)
Conference Proceedings Citation Index - Science (1990-)
7th March 2017

#	Searches
# 1	TI=((breast* NEAR/5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullary)))
# 2	TI=((mammar* NEAR/5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)))
# 3	#2 OR #1

# 4	TS=(cost* and (effective* or utilit* or benefit* or minimi*)) OR TS=(cost*) OR TI=(economic* or pharmacoeconomic* or pharmaco-economic*) OR TS=(price* or pricing*) OR TS=(financial or finance or finances or financed) OR TS=(fee or fees) OR TS=(value and (money or monetary)) OR TS=(economic*) OR TS=(economic* and (hospital or medical or nursing or pharmaceutical)) OR TS=("quality adjusted life year" or "quality adjusted life years") OR TS=(qaly or qalys) OR TS=(budget*)
# 5	#4 AND #3
# 6	TS=(meta-analysis or meta analy* or metaanaly*) OR TS=("review literature" or "literature review") OR TS=("systematic review*" or "systematic overview*") OR TS=(cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or cinhal or science citation index or bids or cancerlit) OR TS=("reference list*" or bibliograph* or hand-search* or "relevant journals" or "manual search*") OR TS=(("selection criteria" or "data extraction") and review)
# 7	#6 AND #5 Indexes=SCI-EXPANDED, CPCI-S Timespan=2011-2017

Health Technology Assessment Database (HTA): Centre for Reviews and Dissemination. 1995-2016
NHS Economic Evaluation Database (NHS EED): Centre for Reviews and Dissemination. 1995-2015
7th March 2017

#	Searches
1	MeSH DESCRIPTOR Breast Neoplasms EXPLODE ALL TREES
2	MeSH DESCRIPTOR Neoplasms, Ductal, Lobular, and Medullary EXPLODE ALL TREES
3	MeSH DESCRIPTOR Breast EXPLODE ALL TREES
4	MeSH DESCRIPTOR Neoplasms EXPLODE ALL TREES
5	#3 AND #4
6	((breast* ADJ5 (neoplasm* or cancer* or tumor* or tumour* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullary)))
7	((mammar* ADJ5 (neoplasm* or cancer* or tumor* or tumour* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)))
8	#1 OR #2 OR #5 OR #6 OR #7
9	(EndoPredict or myriad genetics or sividon diagnostics or ep score or epclin score)
10	(MammaPrint or 70-gene or gene70 or gene*seventy or seventy*gene or amsterdam profile)
11	(oncotype or oncotype dx or 21-gene or gene21 or gene*twentyone or twentyone*gene or ghi recurrence score or ghi-rs or 92-gene or gene92 or gene*ninetytwo or ninetytwo*gene or (rct-pcr ADJ5 '21'))
12	(prosigna or pam50 or 50-gene or gene50 or gene*fifty or fifty*gene or breast bioclassifier)
13	(ihc4)
14	#8 AND #9
15	#10 OR #11 OR #12 OR #13
16	(#8 AND #15) FROM 2011 TO 2017
17	(#8 AND #15) IN HTA FROM 2011 TO 2017
18	(#8 AND #15) IN NHSEED FROM 2011 TO 2017

QUALITY OF LIFE REVIEWS FOR BREAST CANCER

Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present
7th March 2017

#	Searches
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)).ti.
8	(mammar* adj5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)).ti.
9	1 or 2 or 3 or 6 or 7 or 8
10	"Quality of Life"/
11	(qol or (quality adj2 life)).ab,ti.
12	(value adj2 (money or monetary)).tw.
13	value of life/
14	quality adjusted life year/
15	quality adjusted life.tw.
16	(qaly$ or qald$ or qale$ or qtime$).tw.
17	disability adjusted life.tw.
18	daly$.tw.
19	health status indicators/
20	(sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shorform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).tw.
21	(sf 6 or sf6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).tw.
22	(sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).tw.
23	(sf6D or sf 6D or short form 6D or shortform 6D or sf six D or sfsixD or shortform six D or short form six D).tw.
24	(sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty).tw.
25	(euroqol or euro qol or eq5d or eq 5d).tw.
26	(hql or hqol or h qol or hrqol or hr qol).tw.
27	(hye or hyes).tw.
28	health$ year$ equivalent$.tw.
29	health utilit$.tw.
30	(hui or hui1 or hui2 or hui3).tw.
31	disutilit$.tw.

32	rosser.tw.
33	(quality adj2 wellbeing).tw.
34	qwb.tw.
35	(willingness adj2 pay).tw.
36	standard gamble$.tw.
37	time trade off.tw.
38	time tradeoff.tw.
39	tto.tw.
40	letter.pt.
41	editorial.pt.
42	comment.pt.
43	40 or 41 or 42
44	or/10-39
45	44 not 43
46	9 and 45
47	meta-analysis/
48	meta-analysis as topic/
49	(meta analy* or metanaly* or metaanaly*).ti,ab.
50	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
51	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
52	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
53	(search* adj4 literature).ab.
54	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or scie nce citation index or bids or cancerlit).ab.
55	cochrane.jw.
56	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
57	or/47-56
58	46 and 57
59	limit 58 to yr="2011 -Current"

Embase 1974 to 2017 March 06
7th March 2017

#	Searches
1	exp breast tumor/
2	exp breast/
3	exp neoplasm/
4	2 and 3
5	(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)).ti.
6	(mammar* adj5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)).ti.
7	1 or 4 or 5 or 6

8	socioeconomics/
9	quality adjusted life year/
10	quality adjusted life.tw.
11	(qaly$ or qald$ or qale$ or qtime$).tw.
12	disability adjusted life.tw.
13	daly$.tw.
14	health survey/
15	(sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).tw.
16	(sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).tw.
17	(sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).tw.
18	(sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortform sixteen or short form sixteen).tw.
19	(sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty).tw.
20	(euroqol or euro qol or eq5d or eq 5d).tw.
21	(hql or hqol or h qol or hrqol or hr qol).tw.
22	(hye or hyes).tw.
23	health$ year$ equivalent$.tw.
24	health utilit$.tw.
25	(hui or hui1 or hui2 or hui3).tw.
26	disutili$.tw.
27	rosser.tw.
28	quality of wellbeing.tw.
29	qwb.tw.
30	willingness to pay.tw.
31	standard gamble$.tw.
32	time trade off.tw.
33	time tradeoff.tw.
34	tto.tw.
35	or/8-34
36	7 and 35
37	systematic review/
38	meta-analysis/
39	(meta analy* or metanaly* or metaanaly*).ti,ab.
40	((systematic or evidence) adj3 (review* or overview*)).ti,ab.
41	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
42	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
43	(search* adj4 literature).ab.
44	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.

45	cochrane.jw.
46	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
47	or/37-46
48	36 and 47
49	limit 48 to yr="2011 -Current"

Web of Science® Core Collection
Science Citation Index Expanded (1900-)
Conference Proceedings Citation Index - Science (1990-)
7th March 2017

#	Searches
# 1	TI=((breast* NEAR/5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullary)))
# 2	TI=((mammar* NEAR/5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)))
# 3	#2 OR #1
# 4	TS=(qol or "quality of life" or "quality adjusted life" or qaly* or qald* or qale* or qtime* or "disability adjusted life" or daly*)OR TS=(sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shorform thirtysix or shortform thirty six or short form thirtysix or short form thirty six) OR TS=(sf 6 or sf6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six) OR TS=(sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve) OR TS=(sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortfrom sixteen or short form sixteen) OR TS=(sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty) OR TS=(euroqol or euro qol or eq5d or eq 5d or hql or hqol or h qol or hrqol or hr qol or disutilit* or rosser "quality of wellbeing" or qwb or "willingness to pay" or "standard gamble*" or "time trade off" or "time tradeoff" or tto)
# 5	#4 AND #3
# 6	TS=(meta-analysis or meta analy* or metaanaly*) OR TS=("review literature" or "literature review") OR TS=("systematic review*" or "systematic overview*") OR TS=(cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or cinhal or science citation index or bids or cancerlit) OR TS=("reference list*" or bibliograph* or hand-search* or "relevant journals" or "manual search*") OR TS=(("selection criteria" or "data extraction") and review)
# 7	#6 AND #5
	Indexes=SCI-EXPANDED, CPCI-S Timespan=2011-2017

Health Technology Assessment Database (HTA): Centre for Reviews and Dissemination.
1995-2016
NHS Economic Evaluation Database (NHS EED): Centre for Reviews and Dissemination.
1995-2015
7th March 2017

#	Searches
1	MeSH DESCRIPTOR Breast Neoplasms EXPLODE ALL TREES
2	MeSH DESCRIPTOR Neoplasms, Ductal, Lobular, and Medullary EXPLODE ALL TREES

3	MeSH DESCRIPTOR Breast EXPLODE ALL TREES
4	MeSH DESCRIPTOR Neoplasms EXPLODE ALL TREES
5	#3 AND #4
6	(((breast* ADJ5 (neoplasm* or cancer* or tumor* or tumour* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullary)))):TI
7	(((mammar* ADJ5 (neoplasm* or cancer* or tumor* or tumour* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)))):TI
8	#1 OR #2 OR #5 OR #6 OR #7
9	(#8) IN HTA FROM 2011 TO 2017
10	(#8) IN NHSEED FROM 2011 TO 2017

EQ-5D AND BREAST CANCER

Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present
10th July 2017

#	Searches
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)).ti.
8	(mammar* adj5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)).ti.
9	1 or 2 or 3 or 6 or 7 or 8
10	(euroqol or euro qol or eq5d or eq 5d).tw.
11	9 and 10

Embase 1974 to 2017 July 07
10th July 2017

#	Searches
1	exp breast tumor/
2	exp breast/
3	exp neoplasm/
4	2 and 3
5	(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)).ti.
6	(mammar* adj5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)).ti.

7	1 or 4 or 5 or 6
8	(euroqol or euro qol or eq5d or eq 5d).tw.
9	7 and 8

Web of Science® Core Collection
Science Citation Index Expanded (1900-)
Conference Proceedings Citation Index - Science (1990-)
10th July 2017

#	Searches
# 1	TI=((breast* NEAR/5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullary)))
# 2	TI=((mammar* NEAR/5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)))
# 3	#2 OR #1
# 4	TOPIC: (euroqol or euro qol or eq5d or eq 5d)
# 5	#4 AND #5

Appendix 2: Table of excluded studies with rationale

Population not relevant	->3 lymph nodes -advanced breast cancer -neoadjuvant setting -not breast cancer -non-European (for decision impact studies)	40 references[1-40]
Intervention not relevant	-not in-scope test	27 references[41-67]
Comparator not relevant	-not in-scope comparator	3 references[68-70]
Outcome not relevant	-no outcomes of interest -follow-up <5 years -insufficient data -pooled analysis (where individual studies included) -correlation only -analytic validity only	146 references[71-216]
Study type not relevant	-not English Language -editorial or comment -Systematic review -modelling -review -retrospective use of test	34 references[217-250]
Other reasons for exclusion	could not obtain full text	2 references[251 252]
	no novel data (secondary reference to other study)	128 references[253-380]

References for excluded studies

Population not relevant

1. Ademuyiwa FO, Miller A, O'Connor T, et al. The effects of oncotype DX recurrence scores on chemotherapy utilization in a multi-institutional breast cancer cohort. *Breast Cancer Research & Treatment* 2011;126(3):797-802. doi: https://dx.doi.org/10.1007/s10549-010-1329-6
2. Bargallo JE, Lara F, Shaw-Dulin R, et al. A study of the impact of the 21-gene breast cancer assay on the use of adjuvant chemotherapy in women with breast cancer in a Mexican public hospital. *Journal of Surgical Oncology* 2015;111(2):203-07. doi: https://dx.doi.org/10.1002/jso.23794
3. Basaran G, Saglam S, Tansan S, et al. The impact of 21-gene recurrence score (RS) scores on treatment decisions: Retrospective evaluation in Turkish early-stage breast cancer (BC) patients. *Journal of Clinical Oncology Conference: ASCO Annual Meeting* 2011;29(15 SUPPL. 1)
4. Bayewitz AY, Schneider JG, Ogden L, et al. Discrepancies in three commercial prognostic breast cancer assays used in adjuvant chemotherapy decisions. *Journal of Clinical Oncology Conference* 2014;32(26 SUPPL. 1)
5. Biroschak JR, Schwartz GF, Palazzo JP, et al. Impact of Oncotype DX on treatment decisions in ER-positive, node-negative breast cancer with histologic correlation. *Breast Journal* 2013;19(3):269-75. doi: https://dx.doi.org/10.1111/tbj.12099
6. Cheung PS, Tong AC, Leung RC, et al. Initial experience with the Oncotype DX assay in decision-making for adjuvant therapy of early oestrogen receptor-positive breast

cancer in Hong Kong. *Hong Kong Medical Journal* 2014;20(5):401-06. doi: https://dx.doi.org/10.12809/hkmj134140
7. Chin-Lenn L, Segelov E, De Boer R, et al. Indications for, and impact of oncotype DX on adjuvant treatment recommendations when third party funding is unavailable. *Cancer Research Conference: 38th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2016;76(4 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS15-P5-15-02
8. Crumbaker M, Rutovitz J, Chin R, et al. Effect of oncoytpe DX testing on adjuvant treatment recommendations in hormone sensitive early breast cancer. *Asia-Pacific Journal of Clinical Oncology* 2015;11:52. doi: http://dx.doi.org/10.1111/ajco.12398
9. Davidson JA, Cromwell I, Ellard SL, et al. A prospective clinical utility and pharmacoeconomic study of the impact of the 21-gene Recurrence Score assay in oestrogen receptor positive node negative breast cancer. *European Journal of Cancer* 2013;49(11):2469-75. doi: https://dx.doi.org/10.1016/j.ejca.2013.03.009
10. de Boer RH, Baker C, Speakman D, et al. The impact of a genomic assay (Oncotype DX) on adjuvant treatment recommendations in early breast cancer. *Medical Journal of Australia* 2013;199(3):205-08.
11. Denkert C, Weber K, Krappmann K, et al. Risk assessment after neoadjuvant chemotherapy in luminal breast cancer: A prospectively planned validation of gene expression based and clinical prognostic scores in 428 residual tumor samples from two neoadjuvant clinical trials. *Journal of Clinical Oncology Conference* 2016;34(no pagination)
12. Fried G, Moskovitz M. Treatment decisions in estrogen receptor-positive early breast cancer patients with intermediate oncotype DX recurrence score results. *Springerplus* 2014;3:71. doi: https://dx.doi.org/10.1186/2193-1801-3-71
13. Geffen DB, Abu-Ghanem S, Sion-Vardy N, et al. The impact of the 21-gene recurrence score assay on decision making about adjuvant chemotherapy in early-stage estrogen-receptor-positive breast cancer in an oncology practice with a unified treatment policy. *Annals of Oncology* 2011;22(11):2381-86. doi: https://dx.doi.org/10.1093/annonc/mdq769
14. Jaafar H, Bashir MA, Taher A, et al. Impact of Oncotype DX testing on adjuvant treatment decisions in patients with early breast cancer: a single-center study in the United Arab Emirates. *Asia-Pacific Journal of Clinical Oncology* 2014;10(4):354-60. doi: https://dx.doi.org/10.1111/ajco.12259
15. Jinyin Y, Xiaoyun Z, Ning L, et al. The use of 21-gene breast cancer assay on adjuvant chemotherapy choice in women with breast cancer. *Basic and Clinical Pharmacology and Toxicology* 2015;117:5. doi: http://dx.doi.org/10.1111/bcpt.12422
16. Joh JE, Esposito NN, Kiluk JV, et al. The effect of Oncotype DX recurrence score on treatment recommendations for patients with estrogen receptor-positive early stage breast cancer and correlation with estimation of recurrence risk by breast cancer specialists. *Oncologist* 2011;16(11):1520-26. doi: https://dx.doi.org/10.1634/theoncologist.2011-0045
17. Jorgensen CL, Nielsen TO, Bjerre KD, et al. PAM50 breast cancer intrinsic subtypes and effect of gemcitabine in advanced breast cancer patients. *Acta Oncologica* 2014;53(6):776-87. doi: https://dx.doi.org/10.3109/0284186X.2013.865076
18. Lee MH, Han W, Lee JE, et al. The clinical impact of 21-gene recurrence score on treatment decisions for patients with hormone receptor-positive early breast cancer in Korea. *Cancer Research & Treatment* 2015;47(2):208-14. doi: https://dx.doi.org/10.4143/crt.2013.223
19. Leitzin L, Sikorsky N, Leviov M, et al. Oncotype Dx; St. Gallen risk group and Adjuvant!online nomogram as prognostic predictor for lymph nodes positive breast cancer patients. *Breast* 2011;20:S43.
20. Leung RC, Yau TC, Chan MC, et al. The impact of the Oncotype DX breast cancer assay on treatment decisions for women with estrogen receptor-positive, node-negative

breast carcinoma in Hong Kong. *Clinical Breast Cancer* 2016;16(5):372-78. doi: https://dx.doi.org/10.1016/j.clbc.2016.03.002

21. Levine MN, Julian JA, Bedard PL, et al. Prospective evaluation of the 21-gene recurrence score assay for breast cancer decision-making in Ontario. *Journal of Clinical Oncology* 2016;34(10):1065-71. doi: https://dx.doi.org/10.1200/JCO.2015.62.8503
22. Li Y, Kurian AW, Bondarenko I, et al. The influence of 21-gene recurrence score assay on chemotherapy use in a population-based sample of breast cancer patients. *Breast Cancer Research and Treatment* 2017;161(3):587-95. doi: https://dx.doi.org/10.1007/s10549-016-4086-3 https://dx.doi.org/ 10.1002/cam4.1016
23. Lo SS, Mumby PB, Norton 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. *Journal of Clinical Oncology* 2010;28(10):1671-76.
24. Marshall DA, Deal K, Bombard Y, et al. How do women trade-off benefits and risks in chemotherapy treatment decisions based on gene expression profiling for early-stage breast cancer? A discrete choice experiment. *Bmj Open* 2016;6(6) doi: 10.1136/bmjopen-2015-010981
25. Millien J, Edwards C, Kaltman R, et al. The 21-gene recurrence score influences treatment recommendations for patients with node-positive breast cancer. *Annals of Surgical Oncology* 2014;21:83-84. doi: http://dx.doi.org/10.1245/s10434-014-3672-z
26. Oratz R, Kim B, Chao C, et al. Physician survey of the effect of the 21-gene recurrence score assay results on treatment recommendations for patients with lymph node-positive, estrogen receptor-positive breast cancer. *Journal of oncology practice/American Society of Clinical Oncology* 2011;7(2):94-99. doi: https://dx.doi.org/10.1200/JOP.2010.000046
27. Ozmen V, Atasoy A, Gokmen E, et al. Impact of Oncotype DX Recurrence Score on Treatment Decisions: Results of a Prospective Multicenter Study in Turkey. *Cureus* 2016;8(3):e522. doi: https://dx.doi.org/10.7759/cureus.522
28. Partin JF, Mamounas EP. Impact of the 21-gene recurrence score assay compared with standard clinicopathologic guidelines in adjuvant therapy selection for node-negative, estrogen receptor-positive breast cancer. *Annals of Surgical Oncology* 2011;18(12):3399-406. doi: https://dx.doi.org/10.1245/s10434-011-1698-z
29. Pohl H, Kotze MJ, Grant KA, et al. Impact of MammaPrint on Clinical Decision-Making in South African Patients with Early-Stage Breast Cancer. *Breast Journal* 2016;22(4):442-46. doi: https://dx.doi.org/10.1111/tbj.12605
30. Rayhanabad JA, Difronzo LA, Haigh PI, et al. Changing paradigms in breast cancer management: introducing molecular genetics into the treatment algorithm. *The American surgeon* 2008;74(10):887-90.
31. Saghatchian M, Mook S, Pruneri G, et al. Additional prognostic value of the 70-gene signature (MammaPrint()) among breast cancer patients with 4-9 positive lymph nodes. *Breast* 2013;22(5):682-90. doi: https://dx.doi.org/10.1016/j.breast.2012.12.002
32. Schneider JG, Khalil DN. Why does Oncotype DX recurrence score reduce adjuvant chemotherapy use? *Breast Cancer Research and Treatment* 2012;134(3):1125-32. doi: https://dx.doi.org/10.1007/s10549-012-2134-1
33. Smith JS, Kass R, Kauffman G. Changes in treatment recommendations based on the 21-gene recurrence score even in node-positive patients. *Annals of Surgical Oncology* 2012;1):105-06. doi: http://dx.doi.org/10.1245/s10434-012-2344-06p
34. Torres S, Trudeau M, Gandhi S, et al. Prospective evaluation of the impact of the 21-gene recurrence score assay on adjuvant treatment decisions for women with node-positive breast cancer in Ontario, Canada. *Annals of Oncology Conference: 41st European Society for Medical Oncology Congress, ESMO* 2016;27(no pagination) doi: http://dx.doi.org/10.1093/annonc/mdw364.39
35. Tsai ML, Untch S, Treece T, et al. The 70-gene signature to provide risk stratification and treatment guidance for patients classified as intermediate by the 21-gene assay. *Journal of Clinical Oncology Conference* 2016;34(no pagination)

36. Turnbull A, Lee Y, Pearce D, et al. A test utilising diagnostic and on-treatment biomarkers to improve prediction of response to endocrine therapy in breast cancer. *Journal of Clinical Oncology Conference* 2016;34(no pagination)
37. Williams SA, Tatarian T, Teal CB, et al. Clinical utility and therapeutic implications of oncotype analysis in patients with breast cancer. *Journal of Clinical Oncology Conference* 2012;30(15 SUPPL. 1)
38. Wolf DM, Daemen A, Yau C, et al. MammaPrint ultra-high risk score is associated with response to neoadjuvant chemotherapy in the I-SPY 1 TRIAL (CALGB 150007/150012; ACRIN 6657). *Cancer Research Conference: 36th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2013;73(24 SUPPL. 1) doi: http://dx.doi.org/10.1158/0008-5472.SABCS13-P1-08-01
39. Yamauchi H, Nakagawa C, Takei H, et al. Prospective study of the effect of the 21-gene assay on adjuvant clinical decision-making in Japanese women with estrogen receptor-positive, node-negative, and node-positive breast cancer. *Clinical Breast Cancer* 2014;14(3):191-97. doi: https://dx.doi.org/10.1016/j.clbc.2013.10.017
40. Zhang YN, Zhou YD, Mao F, et al. Impact of the 21-Gene Recurrence Score Assay in adjuvant chemotherapy selection for node-negative, hormone receptor-positive breast cancer in the Chinese population. *Neoplasma* 2015;62(4):658-65. doi: https://dx.doi.org/10.4149/neo_2015_079

Intervention not relevant

41. Afentakis M, Dowsett M, Sestak I, et al. Immunohistochemical BAG1 expression improves the estimation of residual risk by IHC4 in postmenopausal patients treated with anastrazole or tamoxifen: a TransATAC study. *Breast Cancer Research and Treatment* 2013;140(2):253-62. doi: https://dx.doi.org/10.1007/s10549-013-2628-5
42. Budczies J, Bockmayr M, Denkert C, et al. Classical pathology and mutational load of breast cancer - integration of two worlds. *The Journal of Pathology Clinical Research* 2015;1(4):225-38. doi: https://dx.doi.org/10.1002/cjp2.25
43. Chang JM, McCullough AE, Dueck AC, et al. Back to basics: Traditional Nottingham grade mitotic counts alone are significant in predicting survival in invasive breast carcinoma. *Annals of Surgical Oncology* 2015;22 Suppl 3:S509-15. doi: https://dx.doi.org/10.1245/s10434-015-4616-y
44. Cheang MC, Voduc KD, Tu D, et al. Responsiveness of intrinsic subtypes to adjuvant anthracycline substitution in the NCIC.CTG MA.5 randomized trial. *Clinical Cancer Research* 2012;18(8):2402-12. doi: https://dx.doi.org/10.1158/1078-0432.CCR-11-2956
45. D'Alfonso TM, van Laar RK, Vahdat LT, et al. BreastPRS is a gene expression assay that stratifies intermediate-risk Oncotype DX patients into high- or low-risk for disease recurrence. *Breast Cancer Research & Treatment* 2013;139(3):705-15. doi: https://dx.doi.org/10.1007/s10549-013-2604-0
46. Dowsett M, Sestak I, Buus R, et al. Oestrogen module of 21-gene recurrence score predicts increased late recurrence for ER+HER2-breast cancer. *European Journal of Cancer* 2014;50:S105-S06. doi: http://dx.doi.org/10.1016/S0959-8049%2814%2970087-5
47. Dowsett M, Sestak I, Buus R, et al. Estrogen receptor expression in 21-gene recurrence score predicts increased late recurrence for estrogen-positive/HER2-negative breast cancer. *Clinical Cancer Research* 2015;21(12):2763-70. doi: https://dx.doi.org/10.1158/1078-0432.CCR-14-2842
48. Fumagalli D, Jose V, Salgado R, et al. Effect of sample preservation method and transportation duration on tumor gene expression profiling in breast cancer. *Cancer Research Conference: 35th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2012;72(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS12-P1-07-08

49. Jones LW, Kwan ML, Weltzien E, et al. Exercise and prognosis on the basis of clinicopathologic and molecular features in early-stage breast cancer: The LACE and Pathways studies. *Cancer Research* 2016;76(18):5415-22. doi: https://dx.doi.org/10.1158/0008-5472.CAN-15-3307
50. Kuniyoshi RK, Gehrke FD, Alves BCA, et al. Gene profiling and circulating tumor cells as biomarker to prognostic of patients with locoregional breast cancer. *Tumor Biology* 2015;36(10):8075-83. doi: 10.1007/s13277-015-3529-5
51. Kurshumliu F, Gashi-Luci L, Kadare S, et al. Classification of patients with breast cancer according to Nottingham Prognostic Index highlights significant differences in immunohistochemical marker expression. *World Journal of Surgical Oncology* 2014;12 doi: 10.1186/1477-7819-12-243
52. Lisette SS, Yao K, Turk M, et al. Molecular subtyping using MammaPrint and BluePrint as an outcome predictor in 180 U.S. breast cancer (BC) patients. *Journal of Clinical Oncology Conference* 2012;30(15 SUPPL. 1)
53. Marin M, Migliosi G, Pagliari C, et al. Translating gene expression profiling of breast cancer into clinical practice: A multidisciplinary study in an italian hospital. *American Journal of Pathology* 2014;1):S48.
54. Marin MG, Baiocco R, Fontana P, et al. Translating gene expression profiling of breast cancer into clinical practice: Experience in an italian breast cancer hospital unit. *Biochimica Clinica* 2013;37:S180.
55. Martin M, Prat A, Rodriguez-Lescure A, et al. PAM50 proliferation score as a predictor of weekly paclitaxel benefit in breast cancer. *Breast Cancer Research and Treatment* 2013;138(2):457-66. doi: https://dx.doi.org/10.1007/s10549-013-2416-2
56. Martin M, Rodriguez-Lescure A, Stijleman IJ, et al. PAM50 proliferation index predicts response to weekly adjuvant paclitaxel in node-positive operable breast cancer. *Cancer Research Conference: 34th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2011;71(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS11-P1-06-04
57. Milburn M, Rosman M, Mylander C, et al. Is oncotype DX recurrence score (RS) of prognostic value once HER2-positive and. low-ER expression patients are removed? *Breast Journal* 2013;19(4):357-64. doi: https://dx.doi.org/10.1111/tbj.12126
58. Milioli HH, Vimieiro R, Riveros C, et al. The discovery of novel biomarkers improves breast cancer intrinsic subtype prediction and reconciles the labels in the METABRIC data set. *PLoS ONE* 2015;10(7):e0129711. doi: https://dx.doi.org/10.1371/journal.pone.0129711
59. Milioli HH, Vimieiro R, Tishchenko I, et al. Iteratively refining breast cancer intrinsic subtypes in the METABRIC dataset. *BioData Mining [electronic resource]* 2016;9:2. doi: https://dx.doi.org/10.1186/s13040-015-0078-9
60. Muniz J, Kidwell KM, Henry NL. Associations between metabolic syndrome, breast cancer recurrence, and the 21-gene recurrence score assay. *Breast Cancer Research & Treatment* 2016;157(3):597-603. doi: https://dx.doi.org/10.1007/s10549-016-3846-4
61. Muranen TA, Blomqvist C, Dork T, et al. Patient survival and tumor characteristics associated with CHEK2:p.I157T - findings from the Breast Cancer Association Consortium. *Breast Cancer Research* 2016;18(1):98. doi: https://dx.doi.org/10.1186/s13058-016-0758-5
62. Neapolitan RE, Jiang X. Study of integrated heterogeneous data reveals prognostic power of gene expression for breast cancer survival. *PLoS ONE* 2015;10(2):e0117658. doi: https://dx.doi.org/10.1371/journal.pone.0117658
63. Netanely D, Avraham A, Ben-Baruch A, et al. Expression and methylation patterns partition luminal-A breast tumors into distinct prognostic subgroups. Erratum appears in Breast Cancer Res. 2016 Nov 28;18(1):117; PMID: 27894324. *Breast Cancer Research* 2016;18(1):74. doi: https://dx.doi.org/10.1186/s13058-016-0724-2
64. Nishio M, Naoi Y, Tsunashima R, et al. 72-gene classifier for predicting prognosis of estrogen receptor-positive and node-negative breast cancer patients using formalin-

fixed, paraffin-embedded tumor tissues. *Clinical Breast Cancer* 2014;14(3):e73-80. doi: https://dx.doi.org/10.1016/j.clbc.2013.11.006

65. Rakha EA, Agarwal D, Green AR, et al. Prognostic stratification of oestrogen receptor-positive HER2-negative lymph node-negative class of breast cancer. *Histopathology* 2017;70(4):622-31. doi: https://dx.doi.org/10.1111/his.13108
66. Stavridi F, Kalogeras KT, Pliarchopoulou K, et al. Comparison of the Ability of Different Clinical Treatment Scores to Estimate Prognosis in High-Risk Early Breast Cancer Patients: A Hellenic Cooperative Oncology Group Study. *PLOS One* 2016;11(10) doi: 10.1371/journal.pone.0164013
67. Yamamoto-Ibusuki M, Yamamoto Y, Yamamoto S, et al. Comparison of prognostic values between combined immunohistochemical score of estrogen receptor, progesterone receptor, human epidermal growth factor receptor 2, Ki-67 and the corresponding gene expression score in breast cancer. *Modern Pathology* 2013;26(1):79-86. doi: https://dx.doi.org/10.1038/modpathol.2012.151

Comparator not relevant

68. Barton MK. Researchers find discordance between standard human epidermal growth factor receptor 2 (HER2) testing and HER2 status reported on Oncotype DX. *CA: A Cancer Journal for Clinicians* 2012;62(2):71-72. doi: https://dx.doi.org/10.3322/caac.21133
69. Baxter E, Gondara L, Lohrisch C, et al. Using proliferative markers and Oncotype DX in therapeutic decision-making for breast cancer: the B.C. experience. *Current Oncology* 2015;22(3):192-98. doi: https://dx.doi.org/10.3747/co.22.2284
70. Clark BZ, Dabbs DJ, Cooper KL, et al. Impact of progesterone receptor semiquantitative immunohistochemical result on Oncotype DX recurrence score: a quality assurance study of 1074 cases. *Applied Immunohistochemistry & Molecular Morphology* 2013;21(4):287-91. doi: https://dx.doi.org/10.1097/PAI.0b013e31826f80c9

Outcome not relevant

71. Acs G, Khkapour N, Kiluk J, et al. Prediction of the oncotype dx recurrence score: Prospective validation of two equations with clinicopathologic analysis and outcome. *Laboratory Investigation* 2015;95:31A. doi: http://dx.doi.org/10.1038/labinvest.2015.5
72. Acs G, Khkapour N, Kiluk J, et al. Can we predict the oncotype DX recurrence score? A clinicopathologic analysis of discrepant cases as determined by two equations with clinical outcome. *Laboratory Investigation* 2015;95:31A. doi: http://dx.doi.org/10.1038/labinvest.2015.5
73. Ahmad FK, Deris S, Abdullah MS. Synergy network based inference for breast cancer metastasis. *World Conference on Information Technology* 2011;3 doi: 10.1016/j.procs.2010.12.178
74. Alvarado MD, Prasad C, Rothney M, et al. A prospective comparison of the 21-gene recurrence score and the PAM50-based Prosigna in estrogen receptor-positive early-stage breast cancer. *Advances in Therapy* 2015;32(12):1237-47. doi: https://dx.doi.org/10.1007/s12325-015-0269-2
75. Atehortua NA, Issa AM. A method to measure clinical practice patterns of breast cancer genomic diagnostics in health systems. *Personalized Medicine* 2012;9(6):585-92. doi: http://dx.doi.org/10.2217/pme.12.72
76. Azim HA, Michels S, Bedard P, et al. Elucidating the biological basis of prognosis in young women with early breast cancer (BC) using gene expression profiling. *Journal of Clinical Oncology Conference: ASCO Annual Meeting* 2011;29(15 SUPPL. 1)
77. Azim HA, Michiels S, Bedard P, et al. The prognostic performance of first generation gene expression signatures (FGS) in young women with operable breast cancer (BC). *Annals of Oncology* 2011;22:ii61. doi: http://dx.doi.org/10.1093/annonc/mdr091
78. Baehner FL, Butler SM, Anderson JM, et al. Consistency and control in clinical assay technology over time: The oncotype DX recurrence score and assessment of single

gene expression levels. *Cancer Research Conference: 34th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2011;71(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS11-P1-07-11

79. Balassanian R, Engelberg JA, Bishop JW, et al. Harmonization of immunohistochemical stains for breast cancer biomarkers-an athena pathology collaboration. *Laboratory Investigation* 2013;93:29A. doi: http://dx.doi.org/10.1038/labinvest.2013.14
80. Barnes M, Srinivas C, Xu C, et al. Early stage breast cancer prognostication using whole tumor or Ki67 heterogeneity-based digital imaging. *Cancer Research Conference: 38th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2016;76(4 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS15-P1-01-04
81. Barry WT, Marcom PK, Geradts J, et al. Retrospective evaluation of precision of gene-expression-based signatures of prognosis and tumor biology in replicate surgical biospecimens from patients with breast cancer. *Cancer Research Conference: 35th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2012;72(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS12-P4-09-01
82. Bartlett JMS, Sabine VS, Haider S, et al. Theranostic multiparametric tests improve residual risk assessment in early luminal breast cancer. *Cancer Research Conference: 37th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2015;75(9 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS14-S2-01
83. Barton S, Zabaglo L, A'Hern R, et al. Assessment of the contribution of the IHC4+C score to decision making in clinical practice in early breast cancer. *British Journal of Cancer* 2012;106(11):1760-65. doi: https://dx.doi.org/10.1038/bjc.2012.166
84. Bastien RR, Rodriguez-Lescure A, Ebbert MT, et al. PAM50 breast cancer subtyping by RT-qPCR and concordance with standard clinical molecular markers. *BMC Medical Genomics* 2012;5:44. doi: https://dx.doi.org/10.1186/1755-8794-5-44
85. Bastien RRL, Vaughn C, Elsberry D, et al. Clinical validation of the Prosigna breast cancer prognostic gene signature assay on formalin-fixed paraffin embedded breast cancer tumors with comparison to standard molecular markers. *Journal of Clinical Oncology Conference* 2014;32(15 SUPPL. 1)
86. Beamish R, McCaffrey J, Smith M, et al. A comparative analysis of distant recurrence risk assessments by Oncotype DX recurrence score alone and integrated with clinicopathologic factors in early-stage breast cancer. *Journal of Clinical Oncology Conference* 2013;31(15 SUPPL. 1)
87. Beck AH, West RB, Van De Vijver M, et al. Computational image analysis identifies new morphologic features that predict breast cancer outcome. *Laboratory Investigation* 2011;91:28A.
88. Beitsch P, Jia A, De Snoo F, et al. Mammaprint 70-gene assay predicts risk of local-regional recurrence. *Annals of Surgical Oncology* 2012;1):25-26. doi: http://dx.doi.org/10.1245/s10434-012-2344-06p
89. Bernard PS, Davis C, Munarriz B, et al. Determining agreement between immunohistochemistry and RT-qPCR for standard biomarkers in breast cancer: Validation on GEICAM 9906 clinical trial. *Journal of Clinical Oncology Conference: ASCO Annual Meeting* 2011;29(15 SUPPL. 1)
90. Bhargava R, Brufsky AM, Lembersky BC, et al. Treatment decisions and five year outcomes of the oncotype Dx test: An independent assessment. *Cancer Research Conference: 36th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2013;73(24 SUPPL. 1) doi: http://dx.doi.org/10.1158/0008-5472.SABCS13-P6-06-46
91. Biermann J, Neusser S, Philipp L. Retrospective cost analysis of the EndoPredict test in patients with primary breast cancer in a German breast center. *Abstract Senology Meeting* 2016

92. Bishop JW, Engelberg J, Apple S, et al. Raising the bar: Breast cancer biomarkers IHC4 harmonization from University of California-Athena pathology collaboration. *Journal of Clinical Oncology Conference: ASCO's Quality Care Symposium* 2012;30(34 SUPPL. 1)
93. Blohmer JU, Rezai M, Kummel S, et al. Using the 21-gene assay to guide adjuvant chemotherapy decision-making in early-stage breast cancer: a cost-effectiveness evaluation in the German setting. *Journal of Medical Economics* 2013;16(1):30-40. doi: https://dx.doi.org/10.3111/13696998.2012.722572
94. Boer K, Rozsa P, Horvath Z, et al. Economic impact of oncotype DX results guided adjuvant treatments in Hungary. *Breast* 2013;22:S114.
95. Bonneterre J, Prat A, Galvan P, et al. Value of a gene signature assay in patients with early breast cancer and intermediate risk: a single institution retrospective study. *Current Medical Research & Opinion* 2016;32(5):835-39. doi: https://dx.doi.org/10.1185/03007995.2016.1146664
96. Borowsky A, Balassanian R, Yau C, et al. Interobserver Agreement of Breast Cancer IHC4 after "Score the Core" Training. *Modern Pathology* 2016;29:33A-33A.
97. Borowsky A, Balassanian R, Yau C, et al. Interobserver agreement of breast cancer IHC4 after "score the core" training. *Laboratory Investigation* 2016;96:33A. doi: http://dx.doi.org/10.1038/labinvest.2016.3
98. Borowsky A, Thompson CK, Balassanian R, et al. Phenotype and immunophenotype analysis of gene expression defined "indolent risk" breast cancers. *Laboratory Investigation* 2016;96:33A. doi: http://dx.doi.org/10.1038/labinvest.2016.3
99. Bosl A, Spitzmueller A, Haid A, et al. Comparison of EndoPredict and MammaPrint in hormone receptor positive, HER2 negative breast cancer. *Breast* 2015;24:S46-S47. doi: http://dx.doi.org/10.1016/S0960-9776%2815%2970108-6
100. Brase JC, Muller BM, Haufe F, et al. Comparison of an RNA-based multigene test between core biopsies and corresponding surgical breast cancer sections. *Annals of Oncology* 2012;23:ii28. doi: http://dx.doi.org/10.1093/annonc/mds041
101. Brownlie D, Chatterjee S, Saad Z, et al. Impact of Oncotype DX on the decision for adjuvant chemotherapy: Retrospective analysis of the Salford Royal Foundation Trust cohort. *European Journal of Surgical Oncology* 2015;41 (6):S74. doi: http://dx.doi.org/10.1016/j.ejso.2015.03.210
102. Buechler S, Badve SS, Gokmen-Polar Y. INDUCT: A risk score to predict relapse in estrogen-receptor-positive breast cancer. *Journal of Clinical Oncology Conference* 2014;32(15 SUPPL. 1)
103. Buscariollo D, Cronin A, Bleicher RJ, et al. Association Between the 21-Gene Recurrence Score and Isolated Local-Regional Recurrence in Hormone Receptor-Positive Breast Cancer. *International Journal of Radiation Oncology Biology Physics* 2016;96(2):E13-E13.
104. Cai L, Riojas J, McDowell G, et al. Evaluation and clinical specimen testing experiences on prosigna breast cancer prognostic gene signature assay. *Journal of Molecular Diagnostics* 2015;17 (6):826.
105. Chadha M, Stewart R, Wallach J. The association of clinical-pathologic factors and Oncotype Dx recurrence score (RS) in the outcome of early stage breast cancer. *Journal of Clinical Oncology Conference* 2015;33(28 SUPPL. 1)
106. Chen YY, Tseng LM, Yang CF, et al. Adjust cut-off values of immunohistochemistry models to predict risk of distant recurrence in invasive breast carcinoma patients. *Laboratory Investigation* 2016;96:36A. doi: http://dx.doi.org/10.1038/labinvest.2016.3
107. Clough KB, Poulet B, Jamshidian F, et al. Risk classification of Early Stage Breast Cancer as Assessed by MammaPrint and Oncotype DX Genomic Assays. *Cancer Research Conference: 35th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2012;72(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS12-P6-07-03

108. Conlon N, Ross DS, Howard J, et al. Is there a role for Oncotype Dx testing in invasive lobular carcinoma? *Breast Journal* 2015;21(5):514-19. doi: https://dx.doi.org/10.1111/tbj.12445
109. Connell LC, Teo M, O'Reilly S, et al. Does Oncotype DX recurrence score correlate with survival benefit in "good prognosis" patients as estimated by AdjuvantOnline!? *Journal of Clinical Oncology Conference* 2011;29(27 SUPPL. 1)
110. Cotzia P, Corney D, Mollaee M, et al. Impact of tumor microenvironment on prosigna breast cancer prognostic score. *American Journal of Clinical Pathology* 2015;144:A279.
111. Crager M, Tang G, Shak S. Using the 21-gene recurrence score and the recently developed Recurrence Score-Clinical-Pathologic to assess recurrence risk in patients with node-negative, ER-positive early-stage breast cancer receiving aromatase inhibitor treatment alone. *Journal of Clinical Oncology Conference: ASCO Annual Meeting* 2011;29(15 SUPPL. 1)
112. Cuallar MT, Telatar M, Louie C, et al. Validation of the prosigna breast cancer prognostic gene signature in the clinical lab. *Journal of Molecular Diagnostics* 2015;17 (6):831.
113. Dabbs DJ, Brufsky A, Jankowitz RC, et al. Comparison of test results and clinical outcomes of patients assessed with both MammaPrint and Oncotype DX with pathologic variables: An independent study. *Journal of Clinical Oncology* 2014;32(15)
114. Dakin A, Cotter M, Walshe J, et al. Oncotype DX testing reduces the number of patients receiving chemotherapy for invasive breast carcinoma. *Irish Journal of Medical Science* 2013;1):S495. doi: http://dx.doi.org/10.1007/s11845-013-1039-2
115. Deck KB, Sinha R, Kerlin D, et al. Comparison of MammaPrint and TargetPrint with clinical parameters in patients with breast cancer: Findings from a prospective U.S. cohort. *Journal of Clinical Oncology Conference: ASCO Annual Meeting* 2011;29(15 SUPPL. 1)
116. Denduluri N, Rugo HS, Davis SE, et al. Concordance between the 21-gene recurrence score (RS) and the 70-gene profile (MP) in breast cancer (BC) patients (pts). *Journal of Clinical Oncology Conference* 2011;29(27 SUPPL. 1)
117. Dodson A, Zabaglo L, Martins V, et al. Between-lab variability in Ki67 scoring by a standardised method in core-cuts has little impact on risk estimates by the IHC4+Clinical (IHC4+C) Score. A study presented on behalf of the International Ki67 in Breast Cancer Working Group of the Breast International Group. *European Journal of Cancer* 2016;57:S142-S43.
118. Dowsett M. Pathology challenges for biology-driven trials: The Ki67 experience. *European Journal of Cancer* 2013;49:S1-S2. doi: http://dx.doi.org/10.1016/S0959-8049%2813%2970081-9
119. Dowsett M, Leung SCY, Zabaglo L, et al. Analytical validation of a standardized scoring protocol for Ki67: Phase-3 of an international multicenter collaboration. *Cancer Research Conference: 38th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2016;76(4 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS15-P1-01-01
120. Drak Alsibai K, Azoulay S, Languille-Mimoune O, et al. KI67 evaluation in breast cancer: Molecular pathology using digital imaging. *Laboratory Investigation* 2013;93:492A. doi: http://dx.doi.org/10.1038/labinvest.2013.37
121. Drukker CA, Elias SG, Nijenhuis MV, et al. Gene expression profiling to predict the risk of locoregional recurrence in breast cancer: a pooled analysis. Erratum appears in Breast Cancer Res Treat. 2015 Jan;149(2):567. *Breast Cancer Research & Treatment* 2014;148(3):599-613. doi: https://dx.doi.org/10.1007/s10549-014-3188-z
122. Drukker CA, Nijenhuis MV, Elias SG, et al. Gene expression profiling to predict the risk of locoregional recurrence in breast cancer. *Cancer Research Conference: 35th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United*

States Conference Start 2012;72(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS12-P2-10-42
123. Ebbert MTW, Bastien RRL, Rowe LR, et al. PAM50 breast cancer intrinsic classifier: Clinical validation of a multianalyte laboratory developed test. *Journal of Clinical Oncology Conference: ASCO Annual Meeting* 2011;29(15 SUPPL. 1)
124. Ellis MJ, Chen B, Chapman JAW, et al. CADER prognostic gene signature for disease free survival in hormone receptor positive breast cancer: NCIC CTG MA.12 phase III placebo-controlled tamoxifen trial. *Cancer Research Conference: 37th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2015;75(9 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS14-P4-11-11
125. Espinel CF, Keating S, Hibshoosh H, et al. MammaPrint Feasibility in a Large Tertiary Urban Medical Center: An Initial Experience. *Scientifica* 2012;2012:942507. doi: https://dx.doi.org/10.6064/2012/942507
126. Ettl J, Grose Lackmann K, Klein E, et al. Feasibility and utility of the endopredict test in clinical routine. *Breast* 2013;22:S53.
127. Fitzal F, Filipits M, Fesl C, et al. Predicting local recurrence using PAM50 in postmenopausal endocrine responsive breast cancer patients. *Journal of Clinical Oncology Conference* 2014;32(15 SUPPL. 1)
128. Fitzal F, Filipits M, Rudas M, et al. The genomic expression test EndoPredict is a prognostic tool for identifying risk of local recurrence in postmenopausal endocrine receptor-positive, her2neu-negative breast cancer patients randomised within the prospective ABCSG 8 trial. *British Journal of Cancer* 2015;112(8):1405-10. doi: https://dx.doi.org/10.1038/bjc.2015.98
129. Fitzal F, Filipits MF, Rudas M, et al. Tailoring local therapy in post-menopausal endocrine responsive HER2neu negative breast cancer patients based on their genetic risk profile using Endopredict. *European Journal of Cancer* 2014;50:S97. doi: http://dx.doi.org/10.1016/S0959-8049%2814%2970079-6
130. Freitas MR, Simon SD. Comparison between Oncotype DX test and standard prognostic criteria in estrogen receptor positive early-stage breast cancer. *Einstein* 2011;9(3):354-58. doi: https://dx.doi.org/10.1590/S1679-45082011AO2039
131. Fresno C, Gonzalez GA, Merino GA, et al. A novel non-parametric method for uncertainty evaluation of correlation-based molecular signatures: its application on PAM50 algorithm. *Bioinformatics* 2017;33(5):693-700. doi: https://dx.doi.org/10.1093/bioinformatics/btw704
132. Galatenko VV, Lebedev AE, Nechaev IN, et al. On the construction of medical test systems using greedy algorithm and support vector machine. *Bulletin of Experimental Biology & Medicine* 2014;156(5):706-09. doi: https://dx.doi.org/10.1007/s10517-014-2430-3
133. Garcia-Estevez L, Hernandez E, Acosta D, et al. Effect of Oncotype DX Recurrence Score (RS) on chemotherapy (CT) decision-making by providing information beyond intrinsic subtypes in both luminal A and B breast cancer (BC) patients (pts): A retrospective study in the Spanish population. *Journal of Clinical Oncology Conference* 2016;34(no pagination)
134. Garcia-Estevez L, Hernandez E, Acosta D, et al. Use of Oncotype DX Recurrence Score (RS) reduces chemotherapy (CT) beyond treatment decisions using Ki67-based determinations of luminal A and B breast cancer subtypes: A retrospective study in the Spanish population. *Annals of Oncology Conference: 41st European Society for Medical Oncology Congress, ESMO* 2016;27(no pagination) doi: http://dx.doi.org/10.1093/annonc/mdw364.36
135. Gilman D, Del Rosario G, Moore TD, et al. Oncotype DX recurrence score correlation with histologic grade and older decision-making tools for predicting breast cancer recurrence risk. *Annals of Surgical Oncology* 2011;18:S64. doi: http://dx.doi.org/10.1245/s10434-011-1552-3

136. Gnant M, Sestak I, Filipits M, et al. Identifying clinically relevant prognostic subgroups of postmenopausal women with node-positive hormone receptor-positive early-stage breast cancer treated with endocrine therapy: a combined analysis of ABCSG-8 and ATAC using the PAM50 risk of recurrence score and intrinsic subtype. *Annals of Oncology* 2015;26(8):1685-91. doi: https://dx.doi.org/10.1093/annonc/mdv215
137. Goossens-Beumer I, Witteveen A, Delahaye L, et al. MammaPrint in large clinical studies (MINDACT) and in diagnostics; how similar are they? *European Journal of Cancer* 2016;57:S147.
138. Gurard-Levin ZA, Pancaldi V, Wilson LOW, et al. Epigenetic profiling of chemotherapy sensitivity. *Cancer Research Conference: 106th Annual Meeting of the American Association for Cancer Research, AACR* 2015;75(15 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.AM2015-LB-155
139. Hanna MG, Bleiweiss IJ, Nayak A, et al. Correlation of Oncotype DX Recurrence Score with Histomorphology and Immunohistochemistry in over 500 Patients. *International Journal of Breast Cancer* 2017;2017:1257078. doi: https://dx.doi.org/10.1155/2017/1257078
140. Hedjazi L, Le Lann MV, Kempowsky T, et al. Symbolic data analysis to defy low signal-to-noise ratio in microarray data for breast cancer prognosis. *Journal of Computational Biology* 2013;20(8):610-20. doi: https://dx.doi.org/10.1089/cmb.2012.0249
141. Hideo S, Horimoto Y, Sonoue H, et al. Application of a 70-gene expression profiling tool to Japanese breast cancer patients. *Breast* 2013;22:S58.
142. Hu Z, Li Y, Fan C, et al. Validation of PAM50 breast cancer intrinsic subtypes using nanostring ncounter gene expression assay. *Journal of Molecular Diagnostics* 2012;14 (6):647. doi: http://dx.doi.org/10.1016/S1525-1578%2812%2900211-5
143. Hu Z, Li Y, Treece A, et al. Verification of the nanostring prosigna breast cancer prognostic gene signature assay in a clinical laboratory. *Journal of Molecular Diagnostics* 2014;16 (6):748.
144. Jegadeesh NK, Kim S, Prabhu RS, et al. The 21-gene recurrence score and locoregional recurrence in breast cancer patients. *Annals of Surgical Oncology* 2015;22(4):1088-94. doi: https://dx.doi.org/10.1245/s10434-014-4252-y
145. Jiang H, Denduluri N, Majure M, et al. **Comparison** of risk prediction with the 21-gene recurrence score (**oncotype** DX) and the 70-gene signature (**MammaPrint**) in patients with estrogen receptor-positive early stage breast cancer. *Cancer Research Conference: 38th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2016;76(4 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS15-P5-08-06
146. Kao KJ, Chang KM, Hsu HC, et al. Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization. *BMC Cancer* 2011;11:143. doi: https://dx.doi.org/10.1186/1471-2407-11-143
147. Karagiannis GS, Goswami S, Jones JG, et al. Selective gene-expression profiling of metastasizing breast tumor cell subpopulations complements the predictive power of Mammaprint Dx and Oncotype Dx. *Cancer Research Conference: 107th Annual Meeting of the American Association for Cancer Research, AACR* 2016;76(14 Supplement) doi: http://dx.doi.org/10.1158/1538-7445.AM2016-1528
148. Kelly CM, Beamish R, McCaffrey J, et al. A comparative analysis of distant recurrence risk assessments by Oncotype DX recurrence score alone and integrated with clinicopathologie factors in early-stage breast cancer. *Journal of Clinical Oncology* 2013;31(15)
149. Khawaja S, Udayasankar S, Munir A, et al. The long term follow up of patients undergoing oncotype Dx testing in a multicenter study in Southwest Wales, UK. *European Journal of Cancer* 2016;57:S7.
150. Krishnamurti U, Wetherilt CS, Yang J, et al. Tumor-infiltrating lymphocytes are significantly associated with better overall survival and disease-free survival in triple

negative but not estrogen receptor positive breast cancers. *Human Pathology* 2017;30:30. doi: https://dx.doi.org/10.1016/j.humpath.2017.01.004

151. Kuderer NM, Barry WT, Geradts J, et al. A cross-platform comparison of genomic signatures and OncotypeDx score to discover potential prognostic/predictive genes and pathways. *Cancer Research Conference: 35th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2012;72(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS12-P2-10-03
152. Kunz G. Use of a genomic test (MammaPrintTM) in daily clinical practice to assist in risk stratification of young breast cancer patients. *Archives of Gynecology & Obstetrics* 2011;283(3):597-602. doi: https://dx.doi.org/10.1007/s00404-010-1454-9
153. Lakhanpal R, Sestak I, Shadbolt B, et al. IHC4 score plus clinical treatment score predicts locoregional recurrence in early breast cancer. *Breast* 2016;29:147-52. doi: https://dx.doi.org/10.1016/j.breast.2016.06.019
154. Lamond NW, Skedgel C, Rayson D, et al. Cost-utility of the 21-gene recurrence score assay in node-negative and node-positive breast cancer. *Breast Cancer Research & Treatment* 2012;133(3):1115-23. doi: https://dx.doi.org/10.1007/s10549-012-1989-5
155. Lee AV, Lin Y, Lin HM, et al. Intra-tumor heterogeneity affects multi-gene test prognostic risk stratification. *Cancer Research Conference: 36th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2013;73(24 SUPPL. 1) doi: http://dx.doi.org/10.1158/0008-5472.SABCS13-P2-11-10
156. Lewin R, Sulkes A, Shochat T, et al. Oncotype-DX recurrence score distribution in breast cancer patients with BRCA1/2 mutations. *Breast Cancer Research & Treatment* 2016;157(3):511-16. doi: https://dx.doi.org/10.1007/s10549-016-3836-6
157. Li H, Zhu Y, Burnside ES, et al. MR Imaging Radiomics Signatures for Predicting the Risk of Breast Cancer Recurrence as Given by Research Versions of MammaPrint, Oncotype DX, and PAM50 Gene Assays. *Radiology* 2016;281(2):382-91. doi: https://dx.doi.org/10.1148/radiol.2016152110
158. Liang Y, Chen XS, Wu JY, et al. Analysis of multidisciplinary team decision-making modification related factors in early breast cancer adjuvant chemotherapy. Chinese. *Chinese Journal of Cancer Prevention and Treatment* 2015;22(22):1769-73.
159. Lima Pereira AA, Santini FC, Shimada AK, et al. Discordance between Oncotype Dx and Saint Gallen criteria, Adjuvant!, NCCN 2011 guidelines, and initial physician treatment recommendation. *Journal of Clinical Oncology Conference* 2012;30(15 SUPPL. 1)
160. Lima Pereira AA, Shimada AK, Santini FC, et al. Oncotype DX-the sirio-libanes hospital cancer center experience. *Annals of Oncology* 2012;23:ix112. doi: http://dx.doi.org/10.1093/annonc/mds392
161. Lo SS, Mai H, McCroskey Z, et al. The 70-gene signature (70-GS) in a lymph node-negative patient series with intermediate risk 21-gene recurrence scores (RS) and known adjuvant treatment recommendations. *Cancer Research Conference: 36th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2013;73(24 SUPPL. 1) doi: http://dx.doi.org/10.1158/0008-5472.SABCS13-P2-11-13
162. Lund MJ, Mosunjac M, Davis KM, et al. 21-Gene recurrence scores: racial differences in testing, scores, treatment, and outcome. *Cancer* 2012;118(3):788-96. doi: https://dx.doi.org/10.1002/cncr.26180
163. Luo J, Chang LW, Tao Y, et al. A copy number aberration driven endocrine response gene signature stratifies risk in estrogen receptor positive breast cancer. *Cancer Research Conference: 37th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2015;75(9 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS14-P5-10-02
164. Luo J, Chang LW, Van Tine BA, et al. An amplicon-driven aromatase inhibitor response (ADAIR) signature provides an orthogonal risk classifier for ER+ breast cancer. *Cancer Research Conference: 34th Annual CTRC AACR San Antonio Breast Cancer*

Symposium San Antonio, TX United States Conference Start 2011;71(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS11-P1-06-13

165. Mai HP, Kliethermes S, McCroskey Z, et al. Comparison of the intermediate risk 21-gene recurrence score (RS) with the 70-gene signature (GS) as a continuous variable, Magee equations, and Adjuvant! Online (AOL). *Journal of Clinical Oncology Conference* 2014;32(26 SUPPL. 1)
166. Mamounas E, Tang G, Fisher B, 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. *Journal of Clinical Oncology* 2010;28(10):1677-83.
167. Mamounas EP, Liu Q, Paik S, et al. 21-gene recurrence score and locoregional recurrence in node-positive/ER-positive breast cancer treated with chemo-endocrine therapy. *Journal of the National Cancer Institute* 2017;109(4) doi: https://dx.doi.org/10.1093/jnci/djw259
168. Mamounas EP, Tang G, Paik S, et al. The 21-gene recurrence score (RS) predicts risk of loco-regional recurrence (LRR) in node (+), ER (+) breast cancer (BC) after adjuvant chemotherapy and tamoxifen: Results from NSABP B- 28. *Annals of Surgical Oncology* 2013;1):S6. doi: http://dx.doi.org/10.1245/s10434-013-2877-x
169. Marcinkowski EF, Kauffman R, Ottesen R, et al. Acceptance of adjuvant chemotherapy in patients with early stage breast cancer. *Annals of Surgical Oncology* 2015;1):S59. doi: http://dx.doi.org/10.1245/s10434-015-4372-z
170. Markopoulos C, Xepapadakis G, Venizelos V, et al. Clinical experience of using Oncotype DX as an additional treatment decision tool in early breast cancer - a retrospective analysis from 5 Greek institutions. *European Journal of Surgical Oncology* 2012;38(5):413-19. doi: https://dx.doi.org/10.1016/j.ejso.2012.02.183
171. Maroun R, Saleh R, Asselah J, et al. A head-to-head comparison of mammaprint and oncotype Dx: A mcgill university health center experience. *Journal of Clinical Oncology Conference* 2015;33(15 SUPPL. 1)
172. McVeigh TP, Hughes LM, Miller N, et al. The impact of Oncotype DX testing on breast cancer management and chemotherapy prescribing patterns in a tertiary referral centre. *European Journal of Cancer* 2014;50(16):2763-70. doi: https://dx.doi.org/10.1016/j.ejca.2014.08.002
173. Michaud P, Mouysset J, Dohollou N, et al. French prospective multi-center cohort on the decision impact assessment. *Value in Health* 2016;19 (3):A297.
174. Miller DV, Jensen TA, Covington AD, et al. Comparing recurrence risk scores derived from 4 immunohistochemistry stains (IHC4) versus oncoType DX. *Cancer Research Conference: 36th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2013;73(24 SUPPL. 1) doi: http://dx.doi.org/10.1158/0008-5472.SABCS13-P6-06-34
175. Moinfar F, Luschin-Ebengreuth G, Tsybrovskyy O, et al. Prediction of results of MammaPrint's 70 gene signatures by conventional histopathological and biological approaches in patients with breast cancer. *Cancer Research Conference: 34th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2011;71(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS11-P5-11-02
176. Morales-Estevez C, Moreno A, Porras I, et al. The genomic profile (GP) of early breast cancer (EBC): Daily practice analysis. *Journal of Clinical Oncology Conference* 2016;34(no pagination)
177. Mullinax J, Carr D, Vera N, et al. Distant recurrence risk with prospective use of the 21-gene assay at a single institution. *Journal of Clinical Oncology Conference* 2015;33(28 SUPPL. 1)
178. Muniz J, Kidwell KM, Henry NL. Metabolic syndrome and breast cancer recurrence within the 21-gene recurrence score assay. *Journal of Clinical Oncology Conference* 2015;33(28 SUPPL. 1)

179. Na KY, Kim KS, Lee JE, et al. The 70-gene prognostic signature for korean breast cancer patients. *Journal of Breast Cancer* 2011;14(1):33-38. doi: https://dx.doi.org/10.4048/jbc.2011.14.1.33
180. Nijenhuis MV, Drukker CD, Elias SG, et al. The 70-gene signature predicts the risk of locoregional recurrence after adequate breast surgery. *European Journal of Cancer* 2013;49:S400. doi: http://dx.doi.org/10.1016/S0959-8049%2813%2970062-5
181. Ole Eriksen J, Jensen MB, Laenkholm AV, et al. Validation of prediction of local recurrence (LR) by Prosigna (PAM50) in a Danish breast cancer cooperative group (DBCG) cohort of hormone receptor positive (HR+), postmenopausal early breast cancer (EBC) patients allocated to 5yr of endocrine therapy (ET). *Cancer Research Conference: 38th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2016;76(4 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS15-P2-08-10
182. Onoda T, Yamauchi H, Yagata H, et al. Evaluating conventional pathological factors against the standardized 21-gene signature for hormone receptor positive breast cancer patients. *Breast* 2011;20:S46-S47.
183. Osaki A, Takeuchi H, Nakamiya N, et al. Feasibility of Mammaprint risk assessment using vacuum-assisted breast biopsy (Mammotome) in early breast cancer. *Breast* 2011;20:S46.
184. Ozmen V, Atasoy A, Gokmen E, et al. Correlations between oncotype DX recurrence score and classic risk factors in early breast cancer: Results of A prospective multicenter study in Turkey. *Meme Sagligi Dergisi / Journal of Breast Health* 2016;12(3):107-11. doi: http://dx.doi.org/10.5152/tjbh.2016.2874
185. Pelaez Garcia A, Yebenes L, Angulo A, et al. Comparison of risk classification between EndoPredict and Mammaprint scores in ER+/HER2-invasive breast cancer. *Virchows Archiv* 2016;469:S22. doi: http://dx.doi.org/10.1007/s00428-016-1997-7
186. Prendergast A, Martin K, Doolan P, et al. Can the Ki67 proliferation index predict the Oncotype DX recurrence score in lymph node negative, ER positive breast cancer? *Cancer Research Conference: 34th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2011;71(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS11-P5-13-12
187. Quintyne KI, Woulfe B, Gupta RK. Correlation between 21-gene recurrence score assay with Nottingham prognostic index (NPI) and Adjuvant! Online (AO) prognostic tools in newly diagnosed patients with early-stage breast cancer in midwestern Ireland. *Journal of Clinical Oncology Conference* 2012;30(15 SUPPL. 1)
188. Retel VP, Joore MA, Drukker CA, et al. Prospective cost-effectiveness analysis of genomic profiling in breast cancer. *European Journal of Cancer* 2013;49(18):3773-79. doi: https://dx.doi.org/10.1016/j.ejca.2013.08.001
189. Rouzier R, Gentien D, Guinebretiere JM, et al. Prospective multicenter study of the impact of the Prosigna assay on adjuvant clinical decision-making in women with early stage breast cancer: Which patients are the best candidates? *Journal of Clinical Oncology Conference* 2016;34(no pagination)
190. Schlake G, Kronenwett R, Tiecke F, et al. EndoPredict-based treatment decision can reduce chemotherapy usage in ER+, HER2-breast cancer. *Breast* 2015;24:S107. doi: http://dx.doi.org/10.1016/S0960-9776%2815%2970270-5
191. Sestak I, Dowsett M, Ferree S, et al. Retrospective analysis of molecular scores for the prediction of distant recurrence according to baseline risk factors. *Breast Cancer Research and Treatment* 2016;159(1):71-78. doi: https://dx.doi.org/10.1007/s10549-016-3868-y
192. Shimizu H, Horimoto Y, Arakawa A, et al. Application of a 70-Gene Expression Profile to Japanese Breast Cancer Patients. *Breast Care* 2015;10(2):118-22. doi: https://dx.doi.org/10.1159/000376562
193. Shivers SC, Clark L, Esposito N, et al. Direct comparison of risk classification between MammaPrint, Oncotype DX and MammoStrat assays in patients with early stage breast cancer. *Cancer Research Conference: 36th Annual CTRC AACR San Antonio*

Breast Cancer Symposium San Antonio, TX United States Conference Start 2013;73(24 SUPPL. 1) doi: http://dx.doi.org/10.1158/0008-5472.SABCS13-P6-06-02

194. Smyth L, Watson G, Walsh EM, et al. Economic impact of 21-gene recurrence score testing on early-stage breast cancer in Ireland. *Breast Cancer Research & Treatment* 2015;153(3):573-82. doi: https://dx.doi.org/10.1007/s10549-015-3555-4
195. Solin LJ, Gray R, Goldstein LJ, et al. Prognostic value of biologic subtype and the 21-gene recurrence score relative to local recurrence after breast conservation treatment with radiation for early stage breast carcinoma: results from the Eastern Cooperative Oncology Group E2197 study. *Breast Cancer Research and Treatment* 2012;134(2):683-92. doi: https://dx.doi.org/10.1007/s10549-012-2072-y
196. Sparano JA, Goldstein LJ, Davidson NE, et al. TOP2A RNA expression and recurrence in estrogen receptor-positive breast cancer. *Breast Cancer Research & Treatment* 2012;134(2):751-57. doi: https://dx.doi.org/10.1007/s10549-012-2112-7
197. Stull TS, Goodwin MC, Anderson JM, et al. Comparison of Oncotype DX recurrence scores between surgical and core biopsy specimens in breast cancer patients. *Cancer Research Conference: 34th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2011;71(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS11-P3-06-05
198. Svedman C, Clough K, Poulet B, et al. A comparison of risk classification as assessed by the MammaPrint and oncotype DX assays. *Breast* 2013;22:S65.
199. Torrisi R, Garcia-Etienne CA, Losurdo A, et al. Potential impact of the 70-gene signature in the choice of adjuvant systemic treatment for ER positive, HER2 negative tumors: a single institution experience. *Breast* 2013;22(4):419-24. doi: https://dx.doi.org/10.1016/j.breast.2013.03.013
200. van den Broek AJ, Broeks A, Horlings HM, et al. Association of the germline TP53 R72P and MDM2 SNP309 variants with breast cancer survival in specific breast tumor subgroups. *Breast Cancer Research & Treatment* 2011;130(2):599-608. doi: https://dx.doi.org/10.1007/s10549-011-1615-y
201. Van Der Velden BHM, Loo CE, Gilhuijs KGA. Parenchymal enhancement of the contralateral breast in DCE-MRI and outcome of patients with early breast cancer: Complementary value of the 70-gene signature. *European Journal of Cancer* 2016;57:S17.
202. Varga Z, Sinn P, Fritzsche F, et al. Comparison of EndoPredict and Oncotype DX test results in hormone receptor positive invasive breast cancer. Erratum appears in PLoS One. 2013;8(10). doi:10.1371/annotation/f715f38e-7aee-4d2b-8bbf-da0411dc6ef3. *PLoS ONE [Electronic Resource]* 2013;8(3):e58483. doi: https://dx.doi.org/10.1371/journal.pone.0058483
203. Viale G, Snoo F, Van't Veer L, et al. Are all small tumors low risk? Characterization of small invasive node negative breast cancers (BC) enrolled in the EORTC 10041/BIG 3-04 (MINDACT) trial. *Cancer Research Conference: 37th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2015;75(9 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS14-P2-03-01
204. Vidal M, Peg V, Galvan P, et al. Gene expression-based classifications of fibroadenomas and phyllodes tumours of the breast. *Molecular Oncology* 2015;9(6):1081-90. doi: https://dx.doi.org/10.1016/j.molonc.2015.01.003
205. Wagner LI, Gray RJ, Sledge GW, et al. Patient-reported cognitive impairments among women with breast cancer randomly assigned to hormonal therapy (HT) alone versus chemotherapy followed by hormonal therapy (C+HT): Results from the Trial Assigning Individualized Options for Treatment (TAILORx). *Journal of Clinical Oncology Conference* 2012;30(15 SUPPL. 1)
206. Wang X, Ring B, Hicks D, et al. The expression of tocopherol associated protein (TAP) is associated with recurrence and survival rates in node positive breast cancer patients. *Laboratory Investigation* 2011;91:68A-69A.

207. Wang X, Ring BZ, Seitz RS, et al. Expression of a-Tocopherol-Associated protein (TAP) is associated with clinical outcome in breast cancer patients. *BMC Clinical Pathology* 2015;15:21. doi: https://dx.doi.org/10.1186/s12907-015-0021-5
208. Wexelman BA, Gallagher A, Smith BL, et al. The impact of the Oncotype DX recurrence score pathology-clinical (RSPC) on the predicted recurrence risk for node negative breast cancer patients: A cancer center experience. *Journal of Clinical Oncology Conference* 2014;32(15 SUPPL. 1)
209. Woltmann A, Chen B, Lascorz J, et al. Systematic pathway enrichment analysis of a genome-wide association study on breast cancer survival reveals an influence of genes involved in cell adhesion and calcium signaling on the patients' clinical outcome. *PLoS ONE* 2014;9(6):e98229. doi: https://dx.doi.org/10.1371/journal.pone.0098229
210. Woodward WA, Barlow WE, Jagsi R, et al. The 21-gene recurrence score and locoregional recurrence rates in patients with node-positive breast cancer treated on SWOG S8814. *International Journal of Radiation Oncology Biology Physics* 2016;96(2):S146-S46.
211. Yan L, Tian L, Liu S. Combining large number of weak biomarkers based on AUC. *Statistics in Medicine* 2015;34(29):3811-30. doi: https://dx.doi.org/10.1002/sim.6600
212. Yao K, Turk M, Goldschmidt R, et al. MammaPrint as a predictor of local-regional recurrence: Findings from a United States early-stage breast cancer patient cohort. *Annals of Surgical Oncology* 2014;21:128. doi: http://dx.doi.org/10.1245/s10434-014-3672-z
213. Yerushalmi R, Raiter A, Nalbandyan K, et al. Cell surface GRP78: A potential marker of good prognosis and response to chemotherapy in breast cancer. *Oncology Letters* 2015;10(4):2149-55. doi: https://dx.doi.org/10.3892/ol.2015.3579
214. Zanconati F, Cusumano P, Tinterri C, et al. The 70-gene expression profile, Mammaprint, for breast cancer patients in mainly European hospitals. *Breast* 2011;20:S45.
215. Zardavas D, Fumagalli D, Borwn DN, et al. Understanding the biology and prognosis of PIK3CA gene mutations in primary breast cancer using gene expression profiling: A pooled analysis. *Cancer Research Conference: 36th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2013;73(24 SUPPL. 1) doi: http://dx.doi.org/10.1158/0008-5472.SABCS13-P2-11-02
216. Zhao X, Rodland EA, Tibshirani R, et al. Molecular subtyping for clinically defined breast cancer subgroups. *Breast Cancer Research* 2015;17:29. doi: https://dx.doi.org/10.1186/s13058-015-0520-4

Study type not relevant

217. Andre F, Delaloge S, Guinebretiere JM, et al. Proliferation of breast cancer and decisional biomarkers in RPC practice (RPC 2013). *Oncologie* 2013;15(12):594-606. doi: 10.1007/s10269-013-2341-3
218. Anonymous. MammaPrint reduces breast cancer overtreatment. *Cancer Discovery* 2016;6(6):OF4. doi: https://dx.doi.org/10.1158/2159-8290.CD-NB2016-047
219. Augustovski F, Soto N, Caporale J, et al. Response to real-life decision-making impact of Oncotype DX. *Breast Cancer Research and Treatment* 2015;154(1):211. doi: https://dx.doi.org/10.1007/s10549-015-3602-1
220. Azim HA. Has the time come for genomic tests to guide the use of adjuvant chemotherapy in node-positive breast cancer? *Annals of Oncology* 2015;26(8):1521-23. doi: http://dx.doi.org/10.1093/annonc/mdv248
221. Baker H. 21-gene assay identifies patients who can avoid chemotherapy. *Lancet Oncology* 2015;16(15):e531. doi: https://dx.doi.org/10.1016/S1470-2045(15)00376-9
222. Bargallo-Rocha JE, Lara-Medina F, Perez-Sanchez V, et al. Cost-effectiveness of the 21-gene breast cancer assay in Mexico. *Advances in Therapy* 2015;32(3):239-53. doi: https://dx.doi.org/10.1007/s12325-015-0190-8

223. Bernhardt SM, Dasari P, Walsh D, et al. Hormonal modulation of breast cancer gene expression: Implications for intrinsic subtyping in premenopausal women. *Frontiers in Oncology* 2016;6:241. doi: https://dx.doi.org/10.3389/fonc.2016.00241
224. Bertelli G, Holt S. Real-life decision-making impact of Oncotype DX. *Breast Cancer Research & Treatment* 2015;154(1):209-10. doi: https://dx.doi.org/10.1007/s10549-015-3600-3
225. Blank PR, Filipits M, Dubsky P, et al. Cost-effectiveness analysis of prognostic gene expression signature-based stratification of early breast cancer patients. *Pharmacoeconomics* 2015;33(2):179-90. doi: https://dx.doi.org/10.1007/s40273-014-0227-x
226. Bonastre J, Marguet S, Lueza B, et al. Cost effectiveness of molecular profiling for adjuvant decision making in patients with node-negative breast cancer. *Journal of Clinical Oncology* 2014;32(31):3513-19. doi: https://dx.doi.org/10.1200/JCO.2013.54.9931
227. Christgen M, Harbeck N, Gluz O, et al. Recognition and handling of discordant negative human epidermal growth factor receptor 2 classification by Oncotype DX in patients with breast cancer. *Journal of Clinical Oncology* 2012;30(26):3313-14; author reply 14-15. doi: https://dx.doi.org/10.1200/JCO.2012.42.1990
228. Dowsett M, Salter J, Zabaglo L, et al. Predictive algorithms for adjuvant therapy: TransATAC. *Steroids* 2011;76(8):777-80. doi: https://dx.doi.org/10.1016/j.steroids.2011.02.032
229. Gerson Cwilich R, Alban de la Torre LF, Villalobos Prieto A, et al. Clinicopathological features, prognosis and influence in the adjuvant treatment of the risk recurrence groups determined by the 21 gene expression profile, Oncotype Dx, in early breast cancer. *Gaceta Medica de Mexico* 2012;148(2):117-24.
230. Hall PS, McCabe C, Stein RC, et al. Economic evaluation of genomic test-directed chemotherapy for early-stage lymph node-positive breast cancer. *Journal of the National Cancer Institute* 2012;104(1):56-66. doi: https://dx.doi.org/10.1093/jnci/djr484
231. Katz G, Romano O, Foa C, et al. Economic impact of gene expression profiling in patients with early-stage breast cancer in France. *PLoS ONE* 2015;10(6):e0128880. doi: https://dx.doi.org/10.1371/journal.pone.0128880
232. Kondo M, Hoshi SL, Yamanaka T, et al. Economic evaluation of the 21-gene signature (Oncotype DX) in lymph node-negative/positive, hormone receptor-positive early-stage breast cancer based on Japanese validation study (JBCRG-TR03). *Breast Cancer Research and Treatment* 2011;127(3):739-49. doi: https://dx.doi.org/10.1007/s10549-010-1243-y
233. Margenthaler JA, Cyr AE. Molecular profiling assays in breast cancer: Beyond prime time and into syndication. *Oncology* 2012;26(4)
234. Nerich V, Curtit E, Bazan F, et al. Economic assessment of the routine use of Oncotype DX assay for early breast cancer in Franche-Comte region. *Bulletin du Cancer* 2014;101(7-8):681-89. doi: https://dx.doi.org/10.1684/bdc.2014.1923
235. Nicolini A, Ferrari P, Fallahi P, et al. An iron regulatory gene signature in breast cancer: more than a prognostic genetic profile? *Future Oncology* 2012;8(2):131-34. doi: https://dx.doi.org/10.2217/fon.11.148
236. Printz C. New genetic test reduces chemotherapy use in patients with early-stage breast cancer. *Cancer* 2016;122(13):1964-65. doi: http://dx.doi.org/10.1002/cncr.30125
237. Rahilly-Tierney C, Walton SM. Cost-effectiveness of the 70-gene signature versus adjuvant! Online and systematic chemotherapy for risk stratification of patients with node-negative breast cancer: does accuracy matter? *Journal of Clinical Oncology* 2015;33(14):1628-29. doi: https://dx.doi.org/10.1200/JCO.2014.60.5006
238. Raison C. NanoString launches its first commercial diagnostic product. *Expert Review of Molecular Diagnostics* 2013;13(3):229. doi: http://dx.doi.org/10.1586/ERM.13.21

239. Ross JS. A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. *Breast Diseases* 2013;24(3):264-65. doi: http://dx.doi.org/10.1016/j.breastdis.2013.07.029
240. Sakata S, Cronk M. The financial burden of using Oncotype Dx for patients with lymph node-negative and estrogen receptor-positive breast cancer in Australia. *Asia-Pacific Journal of Clinical Oncology* 2014;10(1):94-95. doi: https://dx.doi.org/10.1111/ajco.12077
241. Schmidt C. Mammaprint reveals who can skip chemotherapy for breast cancer. *Journal of the National Cancer Institute* 2016;108(8) doi: https://dx.doi.org/10.1093/jnci/djw197
242. Segui MA, Crespo C, Cortes J, et al. In response: Genomic profile of breast cancer. *Expert Review of Pharmacoeconomics & Outcomes Research* 2015;15(3):395-97. doi: https://dx.doi.org/10.1586/14737167.2015.1025760
243. Sparano JA. A 21-gene expression assay in breast cancer. *New England Journal of Medicine* 2016;374(14):1387. doi: https://dx.doi.org/10.1056/NEJMc1515988
244. Storz-Pfennig P, Wolf K. Biomarkers in early breast cancer and beyond: Who needs all those tests? *The Lancet Oncology* 2014;15(9):919-20. doi: http://dx.doi.org/10.1016/S1470-2045%2814%2970270-0
245. Teig B. Genomic signature in breast cancer: Oncotype DX(). *Annales de Pathologie* 2013;33(3):225-28. doi: https://dx.doi.org/10.1016/j.annpat.2013.04.001
246. Thewes B, Prins J, Friedlander M. 70-gene signature in early-stage breast cancer. *New England Journal of Medicine* 2016;375(22):2199-200. doi: https://dx.doi.org/10.1056/NEJMc1612048
247. van't Veer LJ, Rutgers EJT, Piccart M, et al. 70-gene signature in early-stage breast cancer. *New England Journal of Medicine* 2016;375(22):2200-01. doi: https://dx.doi.org/10.1056/NEJMc1612048
248. Vignatelli L, Negro A, Giovannini T, et al. In vitro test for the evaluation of individual risk of metastasis in surgically treated women for breast cancer (Structured abstract). *Health Technology Assessment Database* 2011(4)
249. Webb E. Biomarker index that may accurately identify women at long-term risk of breast cancer relapse. *Women's Health* 2013;9(6):511. doi: http://dx.doi.org/10.2217/whe.13.64
250. Yamauchi H, Nakagawa C, Yamashige S, et al. Societal cost-effectiveness analysis of the 21-gene assay in estrogen-receptor-positive, lymph-node-negative early-stage breast cancer in Japan. *BMC Health Services Research* 2014;14:372. doi: https://dx.doi.org/10.1186/1472-6963-14-372

Could not obtain full text

251. Gene expression profiling of 21 genes in breast cancer to quantify the risk of disease recurrence and predict adjuvant chemotherapy benefit (Structured abstract). *Health Technology Assessment Database* 2013(4)
252. Oncotype DX® breast cancer assay to quantify the risk of disease recurrence and predict adjuvant hemotherapy benefit (Structured abstract). *Health Technology Assessment Database* 2014(4)

No novel data (secondary reference to other study)

253. Albanell J, Colomer R, Ruiz-Borrego 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.
254. Albanell J, Gligorov J, Holt SD, et al. Pooled analysis of 4 European studies assessing the impact of oncotype DX on treatment decisions. *Breast* 2013;22:S62.
255. Alvarado M, Prasad C, Rothney M, et al. A laboratory comparison of the 21-gene assay and PAM50-ROR. *Breast* 2015;24:S110-S11. doi: http://dx.doi.org/10.1016/S0960-9776%2815%2970279-1

256. Alvarado M, Prasad C, Rothney M, et al. A laboratory comparison of the 21-gene assay and PAM50-ROR. *Annals of Surgical Oncology* 2015;1):29. doi: http://dx.doi.org/10.1245/s10434-015-4561-9
257. Armstrong A, Howell S, Wardley A, et al. Experience with the Oncotype DX assay in a UK centre. *European Journal of Cancer* 2014;50:S54. doi: http://dx.doi.org/10.1016/S0959-8049%2814%2970075-9
258. Bailey H, Alvarado M, Baehner F, et al. A laboratory comparison of the 21-gene assay and PAM50-ror. *Laboratory Investigation* 2015;95:34A. doi: http://dx.doi.org/10.1038/labinvest.2015.5
259. Barcenas CH, Sinha AK, Raghavendra AS, et al. Outcomes after chemotherapy in early-stage breast cancer (EBC) patients who underwent a 21-gene expression assay. *Journal of Clinical Oncology Conference* 2016;34(no pagination)
260. Bargallo-Rocha JE, Lara F, Shaw Dulin RJ, et al. A study of the impact of the 21-gene breast cancer assay on the use of adjuvant chemotherapy in women with breast cancer in a Mexican public hospital. *Annals of Oncology* 2012;23:ix107-ix08. doi: http://dx.doi.org/10.1093/annonc/mds392
261. Bartlett JMS, Stein RC, Bayani J, et al. Comparison of multiparameter tests in the UK OPTIMA-Prelim trial. *Cancer Research Conference: 37th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2015;75(9 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS14-P4-11-07
262. Bender RA, Knauer M, Rutgers EJ, et al. The 70-gene profile and chemotherapy benefit in 1,600 breast cancer patients. *Journal of Clinical Oncology* 2009;27(15):512.
263. Blohmer JU, Kuhn T, Rezai M, et al. German multicentre decision impact study of Oncotype DX recurrence score (RS) on adjuvant treatment in estrogen receptor positive (ER+) node negative (N0) and node positive (N+) early breast cancer. *Breast* 2011;20:S46.
264. Blohmer JU, Rezai M, Kummel S, et al. Impact of the Oncotype DX Recurrence Score Assay ontherapy recommendations for ER-positive (ER+), node negative (N0) and node positive (N+) early breast cancer -Results of an interim analysis of the German decisionimpact study. *Onkologie* 2011;34:119-20. doi: http://dx.doi.org/10.1159/000333301
265. Bonneterre J, Prat A, Galvan P, et al. Is PROSIGNA useful to determine adjuvant treatment in intermediate prognosis early breast cancer (EBC)? *European Journal of Cancer* 2015;51:S312.
266. Braybrooke J, Kuchel A, Robinson T, et al. A prospective multi-centre study of the impact of Oncotype DX on adjuvant treatment decisions in patients in the UK with estrogen receptor positive early breast cancer. *European Journal of Cancer* 2013;49:S466-S67. doi: http://dx.doi.org/10.1016/S0959-8049%2813%2970063-7
267. Bundred N, Armstrong A, Howell S, et al. Experience with the Oncotype DX assay in a UK centre. *European Journal of Surgical Oncology* 2014;40 (5):606-07. doi: http://dx.doi.org/10.1016/j.ejso.2014.02.211
268. Cheng WY, Ou Yang TH, Maurer M, et al. BCAM (breast cancer attractor metagenes): A new tool for assessing breast cancer prognosis. *Cancer Research Conference: 105th Annual Meeting of the American Association for Cancer Research, AACR* 2014;74(19 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.AM2014-2878
269. Cusumano G, Generali D, Lifrange E, et al. Inter-institutional comparison of impact of Mammaprint on the routine treatment decision-making process in early breast cancer patient populations from three European hospitals. *Breast* 2011;20:S45.
270. Cusumano P, Generali DG, Ciruelos E, et al. European inter-institutional impact study of mammaprint. *Annals of Oncology* 2012;23:ix96. doi: http://dx.doi.org/10.1093/annonc/mds392
271. Cuzick J, Sestak I, Ferree S, et al. Prediction of late breast cancer recurrence by the ROR (PAM50) score in postmenopausal women in the TransATAC cohort. *Annals of Oncology* 2012;23:ix75-ix76. doi: http://dx.doi.org/10.1093/annonc/mds391

272. D'Alfonso TM, Van Laar RK, Flinchum R, et al. Breastprs effectively separates oncotypedx intermediate risk group to low and high risk groups. *Laboratory Investigation* 2013;93:35A. doi: http://dx.doi.org/10.1038/labinvest.2013.14
273. Dalton L. Invasive breast cancer: stratification of histological grade by gene-based assays: a still relevant example from an older data set. *Histopathology* 2014;65(3):429-33. doi: https://dx.doi.org/10.1111/his.12423
274. Davidson JA, Cromwell I, Ellard S, et al. A prospective clinical utility study of the impact of the 21-gene recurrence score assay (Oncotype DX) in estrogen receptor positive (ER+) node negative (pN0) breast cancer in academic Canadian centers. *Journal of Clinical Oncology Conference* 2012;30(15 SUPPL. 1)
275. De Boer RH, Baker C, Speakman D, et al. Australian decision impact study: The impact of oncotype DX recurrence score (RS) on adjuvant treatment decisions in hormone receptor positive (HR+), node negative (N0) and node positive (N+) early stage breast cancer (ESBC) in the multidisciplinary clinic (MDC). *Cancer Research Conference: 34th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2011;71(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS11-P4-09-18
276. Deck K, Kerlin D, Barone JL, et al. Comparison of the mammaprint 70-gene expression profile with clinical parameters in patients with breast cancer: Findings from a united states cohort. *Annals of Surgical Oncology* 2011;18:S166. doi: http://dx.doi.org/10.1245/s10434-011-1680-9
277. Dodson A, Zabaglo L, Yeo B, et al. Risk of recurrence estimates with IHC4 are tolerant of variations in staining and scoring. *Cancer Research Conference: 37th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2015;75(9 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS14-P5-10-06
278. Dowsett M, Lopez-Knowles E, Sidhu K, et al. Comparison of PAM50 Risk of Recurrence (ROR) score with oncotypedx and IHC4 for predicting residual risk of RFS and distant-(D)RFS after endocrine therapy: A TransATAC study. *Cancer Research Conference: 34th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2011;71(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS11-S4-5
279. Dowsett M, Sestak I, Buus R, et al. EndoPredict (EPClin) score for estimating residual distant recurrence (DR) risk in ER+/HER2- breast cancer (br ca) patients treated with 5 years adjuvant endocrine therapy alone: Validation and comparison with the oncotype DX recurrence score (RS). *Cancer Research Conference: 38th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2016;76(4 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS15-S3-01
280. Drukker C, Van Den Hout HC, Sonke GS, et al. Risk estimations and treatment decisions in early stage breast cancer; Agreement among oncologists and the impact of the 70-gene signature. *European Journal of Cancer* 2013;49:S469. doi: http://dx.doi.org/10.1016/S0959-8049%2813%2970063-7
281. Drukker CA, Bueno-De-Mesquita JM, Retel VP, et al. Comparing the 70-gene signature to the Dutch Breast Cancer guidelines in the prospective RASTER study. *European Journal of Cancer* 2013;49:S402-S03. doi: http://dx.doi.org/10.1016/S0959-8049%2813%2970062-5
282. Drukker CA, Nijenhuis MV, Bueno-De-Mesquita JM, et al. Optimized prediction of clinical outcome by the PREDICT plus tool and the 70-gene signature in early stage node-negative breast cancer. *Cancer Research Conference: 36th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2013;73(24 SUPPL. 1) doi: http://dx.doi.org/10.1158/0008-5472.SABCS13-P6-06-13

283. Drukker CA, Van Tinteren H, Schmidt MK, et al. Impact of the 70-gene signature on long term breast cancer outcome. *European Journal of Cancer* 2014;50:S203. doi: http://dx.doi.org/10.1016/S0959-8049%2814%2970103-0
284. Dubsky P, Brase JC, Fisch K, et al. The EndoPredict score identifies late distant metastases in ER+/HER2-breast cancer patients. *Cancer Research Conference: 35th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2012;72(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS12-S4-3
285. Eiermann W, Rezai M, Kummel S, et al. Using the 21-gene breast cancer assay in adjuvant decision-making in ER-Positive (ER+) early breast cancer (EBC) is cost-effective: Results of a large prospective german multicenter study. *European Journal of Cancer* 2012;48:S130.
286. Eiermann W, Rezai M, Kummel SK, et al. Using the 21-gene Breast Cancer Assay in Adjuvant Decision-making in ER-positive (ER plus) Early Breast Cancer (EBC) is Cost-effective: Results of a Large Prospective German Multicenter Study. *European Journal of Cancer* 2012;48:S130-S30.
287. Ettl J, Lackmann KG, Hapfelmeier A, et al. Prospective comparison of uPA/PAI-1 and EndoPredict-clin score in ER-positive, HER2-negative breast cancer: Impact on risk stratification and treatment decisions. *Journal of Clinical Oncology Conference* 2013;31(15 SUPPL. 1)
288. Filipits M, Dubsky PC, Rudas M, et al. Impact of the EndoPredict-clin score on risk stratification in ER-positive, HER2-negative breast cancer after considering clinical guidelines. *Journal of Clinical Oncology Conference* 2012;30(15 SUPPL. 1)
289. Fried G, Moskovitz M. Retrospective analysis of treatment decisions in patients with intermediate recurrence score results. *Breast* 2013;22:S84-S85.
290. Gligorov J, Pivot XB, Naman HL, et al. Prospective study of the impact of using the 21-gene recurrence score assay on clinical decision making in women with estrogen receptor-positive, HER2-negative, early-stage breast cancer in France. *Journal of Clinical Oncology Conference* 2012;30(15 SUPPL. 1)
291. Gluz O, Nitz U, Christgen M, et al. Prospective WSG phase III PlanB trial: Clinical outcome at 5 year follow up and impact of 21 Gene Recurrence Score result, central/local-pathological review of grade, ER, PR and Ki67 in HR+/HER2-high risk node-negative and-positive breast cancer. *European Journal of Cancer* 2016d;57:S6.
292. Gluz O, Nitz U, Kreipe HH, et al. Clinical impact of risk classification by central/local grade or luminal-like subtype vs. Oncotype DX: First prospective survival results from the WSG phase III planB trial. *European Journal of Cancer* 2015;51:S311.
293. Gnant M, Dowsett M, Filipits M, et al. Identifying clinically relevant prognostic subgroups in node-positive postmenopausal HR+ early breast cancer patients treated with endocrine therapy: A combined analysis of 2,485 patients from ABCSG-8 and ATAC using the PAM50 risk of recurrence (ROR) score and intrinsic subtype. *Journal of Clinical Oncology Conference* 2013;31(15 SUPPL. 1)
294. Gnant M, Filipits M, Dubsky P, et al. Pr predicting risk for late metastasis: The PAM50 risk of recurrence (ROR) score after 5 years of endocrine therapy in postmenopausal women with HR+ early breast cancer: A study on 1,478 patients from the ABCSG-8 trial. *Annals of Oncology* 2013;24:iii29. doi: http://dx.doi.org/10.1093/annonc/mdt078
295. Gnant M, Filipits M, Mlineritsch B, et al. Clinical validation of the PAM50 risk of recurrence (ROR) score for predicting residual risk of distant-recurrence (DR) after endocrine therapy in postmenopausal women with ER+ early breast cancer (EBC): An ABCSG study 54. *Cancer Research* 2012;72(24 Supplement):Abstract no P2-10-02.
296. Gnant M, Filipits M, Mlineritsch B, et al. Clinical validation of the PAM50 risk of recurrence (ROR) score for predicting residual risk of distant-recurrence (DR) after endocrine therapy in postmenopausal women with ER+ early breast cancer (EBC): An ABCSG study. *Cancer Research Conference: 35th Annual CTRC AACR San*

Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start 2012;72(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS12-P2-10-02

297. Gong C, Tan W, Chen K, et al. Prognostic value of a BCSC-associated microRNA signature in hormone receptor-positive her2-negative breast cancer. *Cancer Research Conference: 107th Annual Meeting of the American Association for Cancer Research, AACR* 2016;76(14 Supplement) doi: http://dx.doi.org/10.1158/1538-7445.AM2016-LB-139
298. Halpern N, Sonnenblick A, Uziely B, et al. Oncotype Dx recurrence score among BRCA1/2 germline mutation carriers with hormone receptors positive breast cancer. *International Journal of Cancer* 2017;24:24. doi: https://dx.doi.org/10.1002/ijc.30616
299. Hannouf MB, Xie B, Brackstone M, et al. Cost-effectiveness of a 21-gene recurrence score assay versus Canadian clinical practice in women with early-stage estrogen- or progesterone-receptor-positive, axillary lymph-node negative breast cancer. *BMC Cancer* 2012;12:447. doi: https://dx.doi.org/10.1186/1471-2407-12-447
300. Hannouf MB, Xie B, Brackstone M, et al. Cost effectiveness of a 21-gene recurrence score assay versus Canadian clinical practice in post-menopausal women with early-stage estrogen or progesterone-receptor-positive, axillary lymph-node positive breast cancer. *Pharmacoeconomics* 2014;32(2):135-47. doi: https://dx.doi.org/10.1007/s40273-013-0115-9
301. Hartmann S, Gerber B, Elling D, et al. The 70-gene Mammaprint signature is an independent prognostic factor for elderly women with hormone receptor-positive, HER2-negative breast cancer. *Breast* 2011;20:S43-S44.
302. Holt S, Bertelli G, Brinkworth E, et al. Results from a prospective clinical study on the impact of Oncotype DX on adjuvant treatment decision making in a cohort of 142 UK patients. *Cancer Research Conference: 34th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2011;71(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS11-P5-14-26
303. Holt SDH, Bertelli G, Pudney D, 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;20:S43.
304. Jaafar H, Taher A, Quasmeh K, et al. Impact of the oncotype DX assay on decision-making in estrogen receptor positive early breast cancer. *Breast* 2013;22:S69.
305. Jegadeesh N, Prabhu RS, Oprea G, et al. The 21-gene recurrence score predicts local recurrence in breast cancer patients treated with mastectomy alone but not in patients treated with radiation. *International Journal of Radiation Oncology Biology Physics* 2013;1):S101. doi: http://dx.doi.org/10.1016/j.ijrobp.2013.06.261
306. Jones J, Xue X, Lin H, et al. Tumor microenvironment of metastasis: An imaging based marker of risk for distant metastasis of breast cancer. *European Journal of Cancer* 2013;49:S16. doi: http://dx.doi.org/10.1016/S0959-8049%2813%2970134-5
307. Jones J, Xue X, Lin HM, et al. Tumor MicroEnvironment of Metastasis (TMEM) Is a Novel Approach to the Assessment of Metastatic Risk in Breast Cancer. *Laboratory Investigation* 2014;94:57A. doi: http://dx.doi.org/10.1038/labinvest.2014.16
308. Knauer M. Genomic profiling in breast cancer. *European Journal of Surgical Oncology* 2012;38 (9):756. doi: http://dx.doi.org/10.1016/j.ejso.2012.06.078
309. Krystel-Whittemore M, Brogi E, Bowser ZL, et al. Distant metastases in breast cancer patients with oncotype Dx recurrence score lower than 18. *Laboratory Investigation* 2016;96:52A. doi: http://dx.doi.org/10.1038/labinvest.2016.3
310. Krystel-Whittemore M, Brogi E, Bowser ZL, et al. Distant Metastases in Breast Cancer Patients with Oncotype Dx Recurrence Score Lower Than 18. *Modern Pathology* 2016;29:52A-52A.
311. Kuijer A, Straver M, Elias SG, et al. Impact of the 70-gene signature on adjuvant systemic therapy decisions in Dutch early breast cancer patients: Preliminary results of a prospective multicentre observational study. *European Journal of Cancer* 2016;57:S90-S91.

312. Kuijer A, Straver M, Elias SG, et al. Impact of the 70-gene signature on adjuvant systemic therapy decisions in early breast cancer patients: Preliminary results of a prospective multicenter observational study. *Annals of Surgical Oncology* 2016;1):S50-S51. doi: http://dx.doi.org/10.1245/s10434-015-5010-5
313. Kummel S, Eiermann W, Rezai M, et al. The Oncotype DX Recurrence Score Assay impacts adjuvant therapy recommendations for ER-positive (ER+), node negative (N0) and node positive (N+) early breast cancer-final results of the German decision impact study. *Journal of Cancer Research and Clinical Oncology* 2012;138:24-25. doi: http://dx.doi.org/10.1007/s00432-011-1144-4
314. Le Du F, Theriault RL, Willey JS, et al. Impact of 21-gene RT-PCR assay on adjuvant therapy for stage I breast cancer. *Journal of Clinical Oncology Conference* 2014;32(15 SUPPL. 1)
315. Levine MN, Cochrane BL, Julian JA, et al. Population-based evaluation of 21-gene assay in treatment decision making for early breast cancer in Ontario. *Journal of Clinical Oncology Conference* 2014;32(15 SUPPL. 1)
316. Linn SC, Drukker CA, Retel VP, et al. 70-gene signature prospectively predicts prognosis of patients with node-negative breast cancer: 5 year follow-up of the raster study. *European Journal of Cancer* 2012;48:S104.
317. Liu MC, Pitcher BN, Mardis ER, et al. PAM50 gene signature is prognostic for breast cancer patients treated with adjuvant anthracycline and taxane based chemotherapy. *Cancer Research Conference: 35th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2012;72(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS12-P2-10-01
318. Losk K, Vaz Duarte Luis I, Camuso K, et al. Factors associated with delays in chemotherapy initiation among patients with breast cancer. *Cancer Research Conference: 38th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2016;76(4 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS15-P1-12-08
319. Mamounas E, Wolmark N, Baehner FL, et al. Predicting late distant recurrence risk in ER+ breast cancer after five years of tamoxifen. *Breast* 2015;24:S117. doi: http://dx.doi.org/10.1016/S0960-9776%2815%2970296-1
320. Mamounas EP, Tang G, Paik S, et al. Association between the 21-gene recurrence score (RS) and benefit from adjuvant paclitaxel (Pac) in node-positive (N+), ER-positive breast cancer patients (pts): Results from NSABP B-28. *Cancer Research Conference: 35th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2012;72(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS12-S1-10
321. Marchionni L, Afsari B, Geman D, et al. A simple and reproducible breast cancer prognostic test. *BMC Genomics* 2013;14:336. doi: https://dx.doi.org/10.1186/1471-2164-14-336
322. Martin M, Brase JC, Perou CM, et al. Comparison of PAM50 risk of recurrence (ROR) scores with EndoPredict for predicting risk of distant metastasis in ER+/HER2-, early node-positive breast cancer patients treated with adjuvant chemotherapy-A GEICAM/ 9906 sub-study. *Cancer Research Conference: 36th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2013;73(24 SUPPL. 1) doi: http://dx.doi.org/10.1158/0008-5472.SABCS13-P2-11-06
323. Martin M, Brase JC, Ruiz-Borrego M, et al. Prognostic performance of the EndoPredict score in node-positive chemotherapy-treated ER+/HER2-breast cancer patients: Results from the GEICAM/9906 trial. *Cancer Research Conference: 35th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2012;72(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS12-P2-10-11
324. Martin M, Gonzalez-Rivera M, Morales S, et al. Prospective study of the impact of the ProsignaTM assay on adjuvant clinical decision-making in women with estrogen receptor-positive, HER2-negative, node-negative breast cancer: A GEICAM study.

Cancer Research Conference: 37th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start 2015;75(9 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS14-P6-08-10

325. Miller DP, Roberts M, Petkov VI, et al. Breast cancer-specific survival in >4,600 patients with lymph node-positive (LN+) hormone receptor-positive (HR+) invasive breast cancer (BC) and 21-gene recurrence score (RS) results in the SEER registries. *Annals of Oncology Conference: 41st European Society for Medical Oncology Congress, ESMO* 2016;27(no pagination) doi: http://dx.doi.org/10.1093/annonc/mdw364.7
326. Mittempergher L, De Ronde JJ, Nieuwland M, et al. Microarray gene expression analysis on formalin-fixed, paraffin embedded material is feasible and the resulting profiles are comparable to profiles from fresh frozen matched material. *Cancer Research Conference: 102nd Annual Meeting of the American Association for Cancer Research, AACR* 2011;71(8 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.AM2011-4834
327. Nagarajan R, Upreti M. An approach for deciphering patient-specific variations with application to breast cancer molecular expression profiles. *Journal of Biomedical Informatics* 2016;63:120-30. doi: https://dx.doi.org/10.1016/j.jbi.2016.07.022
328. Nielsen T, McDonald S, Kulkarni S, et al. Analytical reproducibility of the breast cancer intrinsic subtyping test and ncounter analysis system using formalin-fixed paraffin-embedded (FFPE) breast tumor specimens. *Laboratory Investigation* 2013;93:498A. doi: http://dx.doi.org/10.1038/labinvest.2013.37
329. Nitz U, Gluz O, Kates RE, et al. Prognostic impact of discordance between different risk assessment tools in early breast cancer (recurrence score, central grade, Ki67): Early outcome analysis from the prospective phase III WSG-PlanB trial. *Cancer Research Conference: 37th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2015;75(9 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS14-P4-11-01
330. Nitz U, Gluz O, Kreipe H, et al. Risk assessment by St.Gallen 2013 recommendation and Oncotype DX: Results from the WSG PlanB trial. *Breast* 2015;24:S110. doi: http://dx.doi.org/10.1016/S0960-9776%2815%2970278-X
331. Ozmen V, Atasoy A, Gokmen E, et al. Results of the Turkish prospective multi-center study utilizing the 21-gene Oncotype DX assay: Decision impact analysis. *Journal of Clinical Oncology Conference* 2015;33(15 SUPPL. 1)
332. Park MM, Ebel JJ, Zynger DL. ER and PR immunohistochemistry and HER2 FISH vs. Oncotype DX: Implications for breast cancer treatment. *Laboratory Investigation* 2013;93:62A. doi: http://dx.doi.org/10.1038/labinvest.2013.14
333. Perez E, Butler S, Dueck A, et al. The relationship between quantitative HER2 gene expression by the 21-gene RT-PCR assay and adjuvant trastuzumab (H) benefit in NCCTG (alliance) N9831. *Laboratory Investigation* 2015;95:61A. doi: http://dx.doi.org/10.1038/labinvest.2015.5
334. Perez EA, Butler SM, Dueck AC, et al. The relationship between quantitative HER2 gene expression by the 21-gene RT-PCR assay and adjuvant trastuzumab (H) benefit in NCCTG (Alliance) N9831. *Journal of Clinical Oncology Conference* 2013;31(15 SUPPL. 1)
335. Petkov VI, Miller DP, Howlader N, et al. Outcome disparities by age and 21-gene recurrence score (RS) in hormone receptor positive (HR+) breast cancer (BC). *Journal of Clinical Oncology Conference* 2016;34(no pagination)
336. Piccart M, Rutgers E, Van't Veer L, et al. Primary analysis of the EORTC 10041/ BIG 3-04 MINDACT study: A prospective, randomized study evaluating the clinical utility of the 70-gene signature (MammaPrint) combined with common clinical-pathological criteria for selection of patients for adjuvant chemotherapy in breast cancer with 0 to 3 positive nodes. *Cancer Research Conference: 107th Annual Meeting of the American Association for Cancer Research, AACR* 2016;76(14 Supplement) doi: http://dx.doi.org/10.1158/1538-7445.AM2016-CT039

337. Prasad C, Rothney M, Cherbavaz DB, et al. A pilot laboratory study comparing the 21-gene assay and PAM50-ROR. *Journal of Clinical Oncology Conference* 2014;32(15 SUPPL. 1)
338. Prat A, Cheang MC, Martin M, et al. Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically-defined luminal a breast cancer. *Annals of Oncology* 2012;23:ii17-ii18. doi: http://dx.doi.org/10.1093/annonc/mds039
339. Prat A, Galvan P, Buckingham W, et al. Feasibility of the PROSIGNA multigene test in core biopsies and comparison to corresponding surgical breast cancer sections. *Cancer Research Conference: 37th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2015;75(9 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS14-P6-01-06
340. Prat A, Parker JS, Fan C, et al. Concordance among gene-expression-based predictors for ER-positive breast cancer treated with adjuvant tamoxifen. *Journal of Clinical Oncology Conference: ASCO Annual Meeting* 2011;29(15 SUPPL. 1)
341. Rezai M, Eiermann W, Kummel S, et al. Impact of the recurrence score on adjuvant decision-making in ER-positive early breast cancer - Results of a large prospective multicentre decision impact study in node negative and node positive disease. *Cancer Research Conference: 34th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2011;71(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS11-P2-12-26
342. Saghatchian M, Mittempergher L, Michiels S, et al. Characterization of breast cancer distant metastasis based on outcome over time using a gene expression profiling approach and identification of pathway activities of late relapse. *Cancer Research Conference: 34th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2011;71(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS11-S1-6
343. Saghatchian M, Mittempergher L, Michiels S, et al. Microarray anlyses of breast cancers identify CH25H, a cholesterol gene, as a potential marker and target for late metastatic reccurences. *Cancer Research Conference: 35th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2012;72(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS12-P4-09-05
344. Saghatchian M, Mook S, Pruneri G, et al. 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. *Cancer Research* 2009;69(24):506S-06S.
345. Saito M, Shimizu H, Miura H, et al. Discordance of prognostic risk between histopathology and gene signature in Japanese early breast cancer. *Breast* 2011;20:S47.
346. Sestak I, Cuzick J, Dowsett M, et al. Prediction of late distant recurrence after 5 years of endocrine treatment: A combined analysis of 2485 patients from the ABCSG-8 and transATAC studies using the PAM50 risk of recurrence (ROR) score. *Cancer Research Conference: 36th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2013;73(24 SUPPL. 1) doi: http://dx.doi.org/10.1158/0008-5472.SABCS13-S6-04
347. Sestak I, Cuzick J, Dowsett M, et al. Prediction of residual risk of recurrence after 5 years of follow-up by clinicopathologic variables and 4 IHC markers: A TransATAC study. *Cancer Research Conference: 34th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2011;71(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS11-P2-12-09
348. Sestak I, Dowsett M, Ferree S, et al. Analysis of molecular scores for the prediction of distant recurrence according to body mass index and age at baseline. *Annals of Oncology* 2014;25:i8. doi: http://dx.doi.org/10.1093/annonc/mdu066.1

349. Sestak I, Dowsett M, Sgroi D, et al. Pr comparison of five different scores for the prediction of late recurrence for oestrogen receptor-positive breast cancer. *Annals of Oncology* 2013;24:iii29. doi: http://dx.doi.org/10.1093/annonc/mdt078
350. Sestak I, Zhang Y, Schnabel CA, et al. Clinical impact of differential risk stratification by breast cancer index (BCI) versus recurrence score (RS) in HR+ early-stage breast cancer: A TransATAC study. *Journal of Clinical Oncology Conference* 2014;32(15 SUPPL. 1)
351. Sgroi DC, Sestak I, Cuzick J, et al. Comparative performance of breast cancer index (BCI) vs. Oncotype Dx and IHC4 in the prediction of late recurrence in hormonal receptor-positive lymph node-negative breast cancer patients: A TransATAC Study. *Cancer Research Conference: 35th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2012;72(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS12-S1-9
352. Shak S, Miller DP, Howlader N, et al. Outcome disparities by age and 21-gene recurrence score (RS) result in hormone receptor positive (HR+) breast cancer (BC). *Annals of Oncology Conference: 41st European Society for Medical Oncology Congress, ESMO* 2016;27(no pagination) doi: http://dx.doi.org/10.1093/annonc/mdw364.3
353. Shak S, Petkov V, Miller DP, et al. Breast cancer specific mortality in patients with early-stage hormone receptor-positive invasive breast cancer and oncotype DX recurrence score results in the SEER database. *Journal of Clinical Oncology Conference: ASCO's Quality Care Symposium* 2016;34(7 SUPPL. 1)
354. Shak S, Petkov VI, Miller DP, et al. Breast cancer specific survival in 38,568 patients with node negative hormone receptor positive invasive breast cancer and oncotype DX recurrence score results in the SEER database. *Cancer Research Conference: 38th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2016;76(4 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS15-P5-15-01
355. Sinn BV, Muller BM, Keil E, et al. A novel multigene assay in clinical practice-performance and impact on clinical decisions. *Annals of Oncology* 2013;24:iii33-iii34. doi: http://dx.doi.org/10.1093/annonc/mdt078
356. Smyth L, Watson G, Kelly CM, et al. Economic impact of 21-gene recurrence score testing on early stage breast cancer in Ireland. *Breast* 2015;24:S113. doi: http://dx.doi.org/10.1016/S0960-9776%2815%2970285-7
357. Sparano J, Gray R, Zujewski JA, et al. Prospective trial of endocrine therapy alone in patients with estrogen-receptor positive, HER2-negative, node-negative breast cancer: Results of the TAILORx low risk registry. *European Journal of Cancer* 2015;51:S714.
358. Sparano JA, Goldstein LJ, Davidson Jr NE, et al. Topoisomerase 2 alpha (TOP2A) RNA expression provides prognostic information in hormone receptor positive breast cancer that is complementary to a simulated algorithm for recurrence score. *Cancer Research Conference: 34th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2011;71(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS11-P2-12-13
359. Sparano JA, Gray RJ, Makower DF, et al. Prospective trial of endocrine therapy alone in patients with estrogen receptor positive, HER2-negative, node-negative breast cancer: Results of the TAILORx low risk registry. *Cancer Research Conference: 38th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2016;76(4 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS15-P2-08-01
360. Sparano JA, Rohan TE, Xue X, et al. A tumor microenvironment of metastasis (TMEM) biomarker in early breast cancer provides prognostic information that is complementary to IHC4: A prospective validation study. *Cancer Research Conference: 36th Annual CTRC AACR San Antonio Breast Cancer Symposium San*

Antonio, TX United States Conference Start 2013;73(24 SUPPL. 1) doi: http://dx.doi.org/10.1158/0008-5472.SABCS13-P2-11-03

361. Stalhammar G, Rosin G, Kis L, et al. Digital image analysis outperforms manual scoring for breast cancer subclassification and prognostication. *Cancer Research Conference: 38th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2016;76(4 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS15-P1-01-06
362. Stein RC, Dunn JA, Bartlett JM, et al. OPTIMA prelim: a randomised feasibility study of personalised care in the treatment of women with early breast cancer. *Health Technology Assessment* 2016;20(10):xxiii-xxix, 1-201. doi: https://dx.doi.org/10.3310/hta20100
363. Stein RC, Makris A, Hughes-Davies L, et al. Results of the OPTIMA (Optimal Personalized Treatment of early breast cancer usIng Multi-parameter Analysis) prelim study. *European Journal of Cancer* 2015;51:S268.
364. Stemmer S, Steiner M, Rizel S, et al. First prospective outcome data in 930 patients with more than 5 year median follow up in whom treatment decisions in clinical practice have been made incorporating the 21-Gene Recurrence Score. *European Journal of Cancer* 2015;51:S321.
365. Stephen J, Murray G, Cameron D, et al. Time dependence of biomarkers: Non-proportional effects of immunohistochemical panels predicting relapse risk in early breast cancer. *Cancer Research Conference: 37th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2015;75(9 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS14-P4-11-10
366. Stephen J, Murray G, Cameron D, et al. Comparison of immunohistochemical residual risk panels to predict risk in early breast cancers treated with endocrine therapy. *Cancer Research Conference: 37th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2015;75(9 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS14-P4-11-09
367. Stork-Sloots L, Generali D, Ciruelos E, et al. European inter-institutional impact study of MammaPrint. *Breast* 2013;22:S62-S63.
368. Tobin NP, Lindstrom LS, Carlson JW, et al. Strong prognostic concordance between Ki67 and binary, but not multi-level gene expression signatures. *Cancer Research Conference: 35th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2012;72(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS12-PD06-08
369. Torres S, Trudeau ME, Gandhi S, et al. Impact of the 21-gene Recurrence Score assay on the adjuvant treatment of breast cancer patients with 1-3 positive lymph nodes in an academic centre in Ontario. *Journal of Clinical Oncology Conference* 2016;34(no pagination)
370. Torrisi R, Garcia-Etienne CA, Losurdo A, et al. Can we predict the benefit of the 70-gene signature in the choice of adjuvant systemic treatment for ER positive, HER2 negative tumors in daily practice? A single institution experience. *Cancer Research Conference: 34th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2011;71(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS11-P4-09-27
371. Vollan HKM, Rueda OM, Borresen-Dale AL, et al. A tumor DNA complexity index is an independent predictor of survival in a dataset of 1950 breast cancers; a METABRIC group study. *Cancer Research Conference: 35th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2012;72(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS12-P2-10-20
372. Wolmark N, Mamounas EP, Baehner FL, et al. Recurrence score and quantitative ER expression to predict in late distant recurrence risk in ER+ BC after 5 years of tamoxifen. *Journal of Clinical Oncology Conference* 2014;32(15 SUPPL. 1)

373. Wuerstlein R, Sotlar K, Gluz O, et al. Significance of prospective multicenter decision impact WSG-BCIST Study in post-menopausal ER+ HER2-N0 early breast cancer (EBC) for molecular testing for intrinsic subtype definition. *Journal of Clinical Oncology Conference* 2015;33(15 SUPPL. 1)
374. Wuerstlein R, Sotlar K, Gluz O, et al. Prosigna results impact on adjuvant decision making in early breast cancer (EBC): Final analysis of the prospective WSG study. *Cancer Research Conference: 38th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2016;76(4 SUPPL. 1) doi: http://dx.doi.org/10.1158/1538-7445.SABCS15-P5-07-03
375. Wurstlein R, Sotlar K, Gluz O, et al. WSG BCIST study: Prosigna results impact on adjuvant decision making in early breast cancer (EBC). *Oncology Research and Treatment* 2016;39:49. doi: http://dx.doi.org/10.1159/000444354
376. Yamauchi H, Nakagawa C, Yamashige S, et al. Decision impact and economic evaluation of the 21-gene recurrence score (RS) assay for physicians and patients in Japan. *European Journal of Cancer* 2011;47:S378. doi: http://dx.doi.org/10.1016/S0959-8049%2811%2971603-3
377. Yamauchi H, Nakagawa C, Yamashige S, et al. Societal economics of the 21-gene Recurrence Score in estrogen-receptor-positive early-stage breast cancer in Japan. *Cancer Research Conference: 35th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2012;72(24 SUPPL. 3) doi: http://dx.doi.org/10.1158/0008-5472.SABCS12-P5-15-06
378. Yao K, Turk M, Kaul K, et al. Molecular subtyping using MammaPrint and BluePrint as an outcome predictor in U.S. breast cancer (BC) patients. *Journal of Clinical Oncology Conference* 2012;30(27 SUPPL. 1)
379. Yao K, Turk M, Kaul K, et al. MammaPrint and BluePrint in early breast cancer: Clinical implications of prognostic stratification and molecular subtyping. *Cancer Research Conference: 36th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start* 2013;73(24 SUPPL. 1) doi: http://dx.doi.org/10.1158/0008-5472.SABCS13-P2-11-23
380. Yeo BJ, Dowsett M, Smith IE, et al. The clinical impact of using the IHC4 score: Our MDT experience in a prospective series of postmenopausal women with er positive early breast cancer. *Annals of Oncology* 2014;25:i2. doi: http://dx.doi.org/10.1093/annonc/mdu062.1

Appendix 3: IHC4 methodolgies

Author, year	Lab methods	Algorithm	Advice from IHC4 team

Bartlett 2016	**DAB (conventional 3,3'-diaminobezidine) method:** Formalin-fixed paraffin-embedded tissue blocks were received at a central laboratory and replicate tissue microarrays constructed. Tissue microarrays were analysed by conventional IHC (DAB)… using the Ariol SL50 image analysis platform previously validated for generation of quantitative H-scores[1] Staining with DAB was performed centrally as previously described.[2] Antibodies used were a single batch of antibody (1:50; ER clone 6F11, Novocastra, Newcastle, United Kingdom; 1:50, PgR clone PgR636; HER2 HerceptTest; and 1:50, Ki-67 clone MIB1; all from Dako, Cambridge, United Kingdom) and reagents were used to perform all assays; incubations were temperature controlled. Replicate tissue microarrays were analyzed for ER (n= 6), PgR (n = 6), HER2/neu (n = 3), and Ki-67 (n = 3) staining by using the average score for HER2/neu across all cores analysed and the summed value for both percentages of positive cells and staining intensity (1þ, 2þ, 3þ) based on individual cell counts for ER/PgR and Ki-67 in the final analysis, as previously described.[1]	The IHC4 model[3]1 used a linear combination of multiple markers (ER, PR, HER2/neu, and Ki-67). For DAB scores, ER histoscores were divided by 30; PgR percentage positive cells were divided by 10; and Ki-67, represented as percentage positive cells, was included in the model without modification. HER2/neu was treated as a dichotomous variable on the basis of current guidelines.[4 5]	DAB: Compatible QIF: incompatible

Cuzick 2011[3]	Tissue microarrays (TMAs) were constructed by using a manual tissue arrayer (MTA-1; Beecher Instruments, Sun Prairie, WI) with 600-m tissue cores. Hematoxylin and eosin–stained slides were reviewed by a pathologist and/or an experienced technician, and three representative areas that contained invasive tumor cells were selected. Areas of invasive tumor away from in situ or benign tissue components were marked on both the slides and corresponding paraffin blocks for TMA construction. Three cores were extracted from each donor block and were assembled into three recipient blocks. ER and Ki-67 analyses were performed on 4-μm sections from the triplicate TMA blocks, and PgR and HER2 analyses were performed on single 4-μm whole sections from the donor blocks used in the TMA construction. Sections were picked up on charged slides, dewaxed in xylene, and rehydrated in decreasing grades of industrial methylated spirits. Antigen retrieval was performed for all markers: ER, PgR, and Ki-67 were microwaved for 10 minutes in citrate buffer pH 6.0, and HER2 was heated for 40 minutes in HercepTest antigen retrieval buffer (Dako, Copenhagen, Denmark) at 97°C in a waterbath. All slides were stained on the Dako autostainer by using either the REAL detection kit protocol or HercepTest. ER, PgR, and Ki-67 were demonstrated by using the 6F11 antibody (Vector Laboratories, Burlingame, CA) diluted 1:40, clone 16 (Vector Laboratories) diluted 1:100, or SP6 antibody (Abcam, Cambridge, MA) diluted 1:100, respectively. All dilutions and washes were performed with Dako antibody diluent and Dako wash buffer, respectively. Sections were then counterstained with Mayer's hematoxylin. HER2 was demonstrated by using the HercepTest kit per manufacturer's instructions followed by Vysis PathVysion (fluorescent in situ hybridization [FISH]) in those samples scored at 2+ by immunohistochemisty (IHC).		Compatible

	ER was quantified by using the H-score, which is defined as the percentage of cells staining weakly plus two times the percentage of cells staining moderately plus three times the percentage of cells staining strongly. ER was considered positive if the H-score was greater than 1. The variable ER10 was obtained by dividing the H-score by 30 to obtain a variable with a range of 0 to 10. PgR was scored as a percentage of cells staining positive with a positive cutoff of 10%. PgR10 was obtained by dividing this percentage by 10 to obtain a variable with a range of 0 to 10. HER2 was scored according to the manufacturer's recommendation: 3+ was positive, and equivocal 2+ cases underwent FISH analysis to determine the level of HER2 amplification. Tumors that were 3+ positive or 2+ positive with a FISH ratio of more than 2.0 were regarded as HER2 positive. Ki-67–stained slides were scanned with the Applied Imaging Ariol image analysis system (Genetix, San Jose, CA) by using the TMAsight assay with a ×20 objective. Images acquired through three filters (red, green, and blue) were converted by Ariol software into color reconstructions. MultiStainHighRes script was used to analyze images by using classifiers established during training. The analysis was performed only on invasive tumor areas in the individual cores. Ki-67 scores were recorded as the percentage of positively staining malignant cells. Further validation of the immunohistochemistry score for four markers (IHC4; ER, PgR, HER2, and Ki-67) was performed by using a cohort of 786 women treated for primary operable invasive breast cancer in Nottingham from 1990 to 1998. All of these patients were ER positive (H-score > 10) and received either adjuvant tamoxifen or no endocrine treatment. Information on local,		

	regional, and distant recurrence and survival is maintained on a prospective basis. Similar methods and scoring algorithms were used for the Nottingham cohort, except that the MiB1 antibody was used on whole sections for Ki-67, and TMAs were used for ER, PgR, and HER2.		
Stephen, 2014[6]	Immunohistochemical staining for a panel of biomarkers including ER, PgR, HER2, Ki67, HTF9C, CEACAM5,NDRG1, p53 and SLC7A5 and FISH (fluorescence in situ hybridisation) for HER2 was performed using either sextuplet(ER and PgR) or triplicate (all other markers) 0.6mm2 TMA cores. Results were derived from dual scoring by expert observers(as described by Kirkegaard et al (2006)) for the Edinburgh BCScohort for all markers. For TEAM patients, ER, PgR and Ki67scores were derived by quantitative image analysis using the Ariolsystem with algorithms validated against both whole sections andmanual assessment (Faratian et al, 2009; Bartlett et al, 2011a). Data for ER were recorded as a histoscore (Kirkegaard et al, 2006) andfor Ki67 and PgR as a percentage of positive cells (ATAC and Ki67guidelines; Dowsett et al, 2011). Results for HER2 were scoredaccording to the UK guidelines (Walker et al, 2008; Bartlett et al,2011b), with cases regarded as HER2-amplified if any core showed amplification/overexpression. Positivity for p53, HTF9C (recentlyre-named TRIMT2A), CEACAM5, NDRG1 and SLC7A5 wasrecorded as previously described.[7-9]	The IHC4 model (Cuzick et al, 2011[10]) utilised a linear combination of multiple markers: ER, PgR,HER2 and Ki67. Continuous marker scores were normalised prior to inclusion in the IHC4 model. ER histoscores were divided by 30, and PgR scores as a percentage of cells staining positive were divided by 10 to obtain continuous values between 0 and 10. Ki67 scores were represented as percentage positive cells and HER2 was treated as a dichotomous variable. The IHC4 risk score was generated according to the previously specified algorithm (Cuzick et al, 2011).[10] The IHC4 score is analysed as a continuous risk score, except for Kaplan–Meier analyses, in which the IHC4 score is categorised into three groups using two cutoff points that correspond to a 10-year distant recurrence rate of 10% and 20% from the original study; however, these cutoffs have not been previously validated (Cuzick et al, 2011).[10]	Similar
Gluz, 2016c[11] WSG-AGO-Doc[12]	Primary tumor blocks were requested for all patients who had completed 5 years of follow-up or who had experienced relapse or death (see supplementary Data, available at Annals of Oncology online; Figure. 1). Intrinsic subtypes were assigned as follows: luminal-A-like (HR-positive and HER2-negative and Ki-67 <20%);	IHC4 was computed according to the established formulas.[3 13] However, instead of using the H-score reported in Cuzick et al for estimating the semiquantitative	Broadly compatible, but less granularity

	luminal-B-like [HR-positive and (Ki-67 ≥20% or HER2-positive)]; HER2 subtype (HR-negative and HER2-positive); triple negative (TN: ER/PR/HER2)-basal-like [TN and (EGFR-positive or CK5/6-positive)]; TN-non-basal-like (TN and EGFR-negative and CK5/6-negative); here TN denotes HR-negative and HER2-negative.	expression of ER, we determined a general intensity score value of 0 to 3 and multiplied this value by the percentage of ER-positive tumor cells for a final ER score of 0 to 300.	
Nitz 2017[14-16] WSG-Plan B	Slide review, IHC, and fluorescence in situ hybridization analysis were performed in an independent central laboratory (Institute of Pathology, Hannover Medical School, Hannover, Germany). One experienced breast pathologist (M.C.) assessed histology and central grade using hematoxylin and eosin–stained slides, and a second pathologist (H.H.K.) reviewed them; both were blinded to the clinical data and to Ki-67 expression. Tissue microarrays (diameter, 1.4 mm) were constructed during the first slide review by choosing one morphologically representative region from each tumor sample. Slides were stained for ER (rabbit [SP1]; Neomarkers, Fremont, CA), PR (mouse monoclonal PgR636; DAKO, Glostrup, Denmark), and Ki-67 (clone 30-9 rabbit monoclonal; Ventana, Tucson, AZ) using standard protocols. Tumors were classified as ER or PR positive if immunostaining was present in ≥ 1% of tumor nuclei. Ki-67 was evaluated by one experienced breast pathologist, specialized in proliferation measurement (H.H.K.) in at least 100 tumor cells within the highest-density area; the measurement was performed semiquantitatively (in 5% increments) and quantitatively (in 1% increments).	IHC4 was computed as previously described.[3 13] However, instead of using the H-score reported in Cuzick et al for estimating the semiquantitative expression of ER, we determined a general intensity score value of 0 to 3 and multiplied this value by the percentage of ER-positive tumor cells for a final ER score of 0 to 300.	Incompatible: Ki67 assessed in 5% increments, which will alter IHC4 score
Gong 2016[17] N=611	ER was quantified by using the H-score and was considered positive if greater than 1%. The variable ER10 was obtained by dividing the H-score by 30 to obtain a variable with a range of 0 to 10. PgR was scored as the percentage of cells staining positive with a positive cutoff of 10%. PgR10 was obtained by dividing this percentage by 10 to obtain a variable with a range of 0 to 10. HER2 was scored	As per Cuzik 2011[3]	Unclear

	according to the manufacturer's recommendation: 3+ was positive and equivocal 2+ samples underwent fluorescent in situ hybridization analysis and were considered positive only if the ratio was more than 2. Ki-67 scores were recorded as the percentage of positively staining malignant cells. A histogram of the IHC4 score for all the patients is shown in Fig. S4. The median is 5.86 and the interquartile range (IQR, Q2) is 20.97 to 12.25. The hazard ratio (HR) for a change from the 25th (quartile 1, Q1) to 75th (quartile 3, Q3) percentile of the IHC3 score for all patients was 2.58(95% CI, 1.73 to 3.83) in a univariate analysis in 611 patients. Thus, we stratified the patients into low (Q1)-, intermediate (Q2) - or high (Q3) - risk group for convenient description.		
Lin, 2015[18]	Tumours were stained for ER, PgR, and HER2 by using IHC. The ER and PgR statuses were determined using the Ventana Benchmark system (Ventana Medical Systems Inc., Tucson, AZ, USA) and prediluted antibodies (anti-ER clone 6F11 and anti-PgR clone 16). ER and PgR were scored as percentage of tumor cells positively staining nuclei, and tumors with ≥ 10% positively stained cells were considered positive. The HER2 status was determined according to the American Society of Clinical Oncology/College of American Pathologists updated guideline[19]. Briefly, scores of 0 and 1+ by IHC were considered negative and 3 + was considered positive. Cases with a score of 2+ were tested for gene amplification by dual probe fluorescence in situ hybridization. HER2/CEP17 ratio ≥ 2.0 and/ or an average HER2 copy number ≥ 6.0 signals/cell were considered positive. The primary antibody for staining Ki67 was anti-Ki67 (1:200 dilution, clone MIB-1, DakoCytomation, Denmark)[20 21], and tumors with ≥ 13.25% positively stained nuclei were considered as highly expressed.[22]	According to the study by Cuzick et al.[3] the IHC4 score of each tumor was computed as IHC4 = 94.7 × (− 0.100 · ER10 − 0.079 · PgR10 + 0.586 · HER2 + 0.240 ln [1 + 10 · Ki67]). To avoid the bias caused by the differences in methodology and the antibodies between the present study and the study by Cuzick et al.,[3] we categorized our study participants into low, intermediate, and high risk groups according to the IHC4 scores of < 25th, 25th–75th, and > 75th percentiles, respectively.	Unlikely to be compatible – used image analysis for ER+ and PgR, Ki67method unclear
Rohan,	Staining for ER, PR, and HER2/neu was performed and interpreted	Cuzik et al. 2011[3]	Unlikely to

2014[23]	as per standard surgical pathology practice in accordance with American Society of Clinical Oncology/College of American Pathologists (ASCO-CAP) guidelines (25,26); Ki67 staining was performed as described elsewhere (27). ER/PR positivity was defined as 1% of cells or more staining positive (25), and HER2 positivity was defined as a score of three or greater. More detail provided in supplement to Rohan et al. 2014[23], not extracted here.		be compatible – applied re-fitted IHC4+C algorithm to the population
Viale 2013[24]	Biomarker expression was measured by IHC. HER2 was confirmed by FISH if ≥IHC2+. Tumours were deemed positive for ER/PR if IHC ≥1% or Allred ≥3 & for HER2 if IHC 3+ or if FISH amplified. Ki67 was high if > 11% LI (median).	NR	Unclear
Vincente-Salomon 2013[25]	Immunostaining was done according to previously published protocols[26]. The expression of ER (clone 6F11; 1/200; Novocastra), progesterone receptor (PR; clone 1A6; 1/200; Novocastra), ERBB2 (clone CB11; 1/1,000; Novocastra), epidermal growth factor receptor (HER1; clone 31G7; 1/40; Zymed; Clinisciences), cytokeratin 5/6 (clone D5/16B4; 1/50; Dako), and cytokeratin 8/18 (clone DC10; 1/100; Zymed; Clinisciences) were evaluated. For each antibody, internal and external controls were included in the experiments. ER, progesterone receptor, HER2 receptor and KI67 status were assessed by immunohistochemistry on representative formalin-fixed tumor blocks, according to previously published protocols[27]. The semiquantitative KI67 assessment was performed as previously published[28] and as recommended[29]. A cut-off of 14% was used to define tumors with a high KI67 score (according to St Gallen recommendations[30] and cut-off for molecular classification.[13] Internal (normal glands surrounding the carcinoma) and external controls (for ER, PR and HER2: tissue-microarrays composed of	Cuzik et al. 2011[3] Used IHC3 algorithm as patients HER2-	Compatible

	tumors with known ER, PR status, and known numbers of HER2 gene copiestogether with normal mammary tissue; for KI67: normal lymph node with germinal centers as positive controls) were included in all immunostaining experiments.		
Prat 2012	Sections were air-dried overnight before storage at 4°C (unless IHC was ran the next day). The oldest samples were 7-10 days approximately before staining was performed. The normal breast tissue adjacent to carcinoma was used as internal positive control as well as an external positive control (ie, a well-characterized sample with a weak expression of the biomarker assessed). Two pathologists assessed. Ki-67 Mouse MIB-1; PR Mouse PgR636; ER, Rabbit SP1, all supplied by DAKO, stained using DAKO autostainer, detected using Dako EnVision+.	IHC4 was computed according to the established formulas.[3 13] However, instead of using the H-score reported in Cuzick et al for estimating the semiquantitative expression of ER, we determined a general intensity score value of 0 to 3 and multiplied this value by the percentage of ER-positive tumour cells for a final ER score of 0 to 300.	Compatible

References for Appendix 3

1. Faratian D, Kay C, Campbell F, et al. Automated image analysis for high-throughput quantitative detection of ER and PgR expression levels in large-scale clinical studies: The TEAM trial experience. *Breast Cancer Research and Treatment* 2007;106:S215.
2. Bartlett JM, Brookes CL, Robson T, et al. Estrogen receptor and progesterone receptor as predictive biomarkers of response to endocrine therapy: a prospectively powered pathology study in the Tamoxifen and Exemestane Adjuvant Multinational trial. *Journal of Clinical Oncology* 2011;29(12):1531-38.
3. Cuzick J, Dowsett M, Pineda S, 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. *Journal of Clinical Oncology* 2011;29(32):4273-78. doi: https://dx.doi.org/10.1200/JCO.2010.31.2835
4. Welsh AW, Moeder CB, Kumar S, et al. Standardization of estrogen receptor measurement in breast cancer suggests false-negative results are a function of threshold intensity rather than percentage of positive cells. *Journal of Clinical Oncology* 2011;29(22):2978-84.
5. Bartlett J, Going JJ, Mallon EA, et al. Evaluating HER2 amplification and overexpression in breast cancer. *The Journal of pathology* 2001;195(4):422-28.
6. Stephen J, Murray G, Cameron DA, et al. Time dependence of biomarkers: non-proportional effects of immunohistochemical panels predicting relapse risk in early breast cancer. *British Journal of Cancer* 2014;111(12):2242-47. doi: https://dx.doi.org/10.1038/bjc.2014.530
7. Bartlett JM, Bloom KJ, Piper T, et al. Mammostrat as an immunohistochemical multigene assay for prediction of early relapse risk in the tamoxifen versus exemestane adjuvant multicenter trial pathology study. *Journal of Clinical Oncology* 2012;30(36):4477-84.
8. Ring BZ, Seitz RS, Beck R, et al. Novel prognostic immunohistochemical biomarker panel for estrogen receptor–positive breast cancer. *Journal of Clinical Oncology* 2006;24(19):3039-47.
9. Ross DT, Kim C-y, Tang G, et al. Chemosensitivity and stratification by a five monoclonal antibody immunohistochemistry test in the NSABP B14 and B20 trials. *Clinical cancer research* 2008;14(20):6602-09.
10. Cuzick J, Dowsett M, Wale C, et al. Prognostic value of combined ER, PgR, Ki67, ER2 immunohistochemical score (IHC4) and comparison with the GHI recurrence score in early breast cancer. *Journal of Clinical Oncology* 2011
11. Gluz O, Liedtke C, Huober J, et al. Comparison of prognostic and predictive impact of genomic or central grade and immunohistochemical subtypes or IHC4 in HR+/HER2- early breast cancer: WSG-AGO EC-Doc Trial. *Annals of Oncology* 2016;27(6):1035-40. doi: https://dx.doi.org/10.1093/annonc/mdw070
12. Nitz U, Gluz O, Huober J, et al. Final analysis of the prospective WSG-AGO EC-Doc versus FEC phase III trial in intermediate-risk (pN1) early breast cancer: efficacy and predictive value of Ki67 expression. *Annals of Oncology* 2014;25(8):1551-57.
13. Prat A, Cheang MCU, Martín M, et al. Prognostic significance of progesterone receptor–positive tumor cells within immunohistochemically defined luminal A breast cancer. *Journal of clinical oncology* 2012;31(2):203-09.
14. Gluz O, Nitz U, Chrlstgen M, et al. Prognostic impact of 21 gene recurrence score, IHC4, and central grade in high-risk HR+/HER2-early breast cancer (EBC): 5-year results of the prospective Phase III WSG PlanB trial. *Journal of Clinical Oncology Conference* 2016a;34(no pagination)
15. Gluz O, Nitz UA, Christgen M, et al. West German Study Group Phase III PlanB trial: First prospective outcome data for the 21-gene recurrence score assay and concordance of prognostic markers by central and local pathology assessment. *Journal of Clinical Oncology* 2016b;34(20):2341-49. doi: https://dx.doi.org/10.1200/JCO.2015.63.5383
16. Nitz U, Gluz O, Christgen M, et al. Reducing chemotherapy use in clinically high-risk, genomically low-risk pN0 and pN1 early breast cancer patients: five-year data from the prospective, randomised phase 3 West German Study Group (WSG) PlanB trial. *Breast Cancer Research and Treatment* 2017 doi: 10.1007/s10549-017-4358-6

17. Gong C, Tan W, Chen K, et al. Prognostic value of a BCSC-associated microRNA signature in hormone receptor-positive HER2-negative breast cancer. *EBioMedicine* 2016;11:199-209. doi: https://dx.doi.org/10.1016/j.ebiom.2016.08.016
18. Lin CH, Chen IC, Huang CS, et al. TP53 mutational analysis enhances the prognostic accuracy of IHC4 and PAM50 assays. *Scientific Reports* 2015;5:17879. doi: https://dx.doi.org/10.1038/srep17879
19. Wolff AC, Hammond MEH, Hicks DG, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *Journal of clinical oncology* 2013;31(31):3997-4013.
20. De Azambuja E, Cardoso F, de Castro Jr G, et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12 155 patients. *British journal of cancer* 2007;96(10):1504.
21. Yerushalmi R, Woods R, Ravdin PM, et al. Ki67 in breast cancer: prognostic and predictive potential. *The lancet oncology* 2010;11(2):174-83.
22. Cheang MC, Chia SK, Voduc D, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *JNCI: Journal of the National Cancer Institute* 2009;101(10):736-50.
23. Rohan TE, Xue X, Lin HM, et al. Tumor microenvironment of metastasis and risk of distant metastasis of breast cancer. *Journal of the National Cancer Institute* 2014;106(8) doi: https://dx.doi.org/10.1093/jnci/dju136
24. Viale G, Speirs V, Bartlett JM, et al. Pr prognostic and predictive value of IHC4 and erb1 in the intergroup exemestane study (IES)-on behalf of the pathies investigators. *Annals of Oncology* 2013;24:iii29-iii30. doi: http://dx.doi.org/10.1093/annonc/mdt078
25. Vincent-Salomon A, Benhamo V, Gravier E, et al. Genomic instability: a stronger prognostic marker than proliferation for early stage luminal breast carcinomas. *PLoS ONE [Electronic Resource]* 2013;8(10):e76496. doi: https://dx.doi.org/10.1371/journal.pone.0076496
26. Azoulay S, Laé M, Fréneaux P, et al. KIT is highly expressed in adenoid cystic carcinoma of the breast, a basal-like carcinoma associated with a favorable outcome. *Modern pathology* 2005;18(12):1623.
27. Vincent-Salomon A, Lucchesi C, Gruel N, et al. Integrated genomic and transcriptomic analysis of ductal carcinoma in situ of the breast. *Clinical cancer research* 2008;14(7):1956-65.
28. Reyal F, Bollet MA, Caly M, et al. Respective prognostic value of genomic grade and histological proliferation markers in early stage (pN0) breast carcinoma. *Plos one* 2012;7(4):e35184.
29. Dowsett M, Nielsen TO, A'Hern R, et al. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. *Journal of the National Cancer Institute* 2011;103(22):1656-64.
30. Goldhirsch A, Wood W, Coates A, et al. Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Annals of oncology* 2011;22(8):1736-47.
31. Camp RL, Chung GG, Rimm DL. Automated subcellular localization and quantification of protein expression in tissue microarrays. *Nature medicine* 2002;8(11):1323.
32. Giltnane JM, Rimm DL. Technology insight: Identification of biomarkers with tissue microarray technology. *Nature Reviews Clinical Oncology* 2004;1(2):104.

Appendix 4: Microarray data relating to one test only

Note: Ahn 2013 also reported other data relating to Oncotype-DX versus Mammaprint and appears in the main report.

Table 1 Study characteristics

Author, year, Number patients	Cohorts	Country	O-DX	EP	MMP	PRO	Other tests	Population	Nodal Status	Endo/Chemo
Ahn 2013[1] a)N=186 b)N=82	Gananam Severance Hospital (1997-2007)	Korea	O-DX		MMP			100% ER+ 12% HER2+ a) all patients b) subset with RS 19-30	a)47.8% LN+ (% LN>3 NR) b)43.9% LN+ (LN>3 NR)	a)84% ET 13% CT b) 94% ET 82% CT
O-DX only										
Cockburn 2016[2] a) N=230 b) N=132	NCBI Gene Expression Omnibus: a) GSE17705 (MD Anderson) b) GSE6532	a)USA b) UK, Sweden	O-DX			Excluded*		100% ER+ 100% HER2-	a) 39.6% LN+ (LN>3 NR) b) 67.4% LN+ (LN>3 NR)	100% ET CT NR
Loi 2007[3] N=249	John Radcliffe Hospital, UK; Guys Hospital, UK; Uppsala University Hospital, Sweden (GSE6532)	UK, Sweden	O-DX					100% HR+ HER2- NR	LN0 47% (% LN>3 NR) **SG:** a) LN0 100% b) LN+ 100%	ET 100% CT 0%
Naoi, 2013[4] N=459	Osaka University Hospital; public databases (GSE17705, GSE12093)	Japan, NR	O-DX					100%ER+ HER2- NR	LN0 100% (% LN>3 NR)	100% ET 0% CT
MMP only										
Bianchini, 2013[5] N= 683	GSE6532, GSE9195, GSE17705, GSE12093	NR			MMP			100% ER+ 95% HER2-	38% LN+ (LN>3 NR)	100% ET CT NR

Zemmour, 2015[6] N=197	TRANSBIG**	France, Sweden UK			MMP			ER+ 69% HER2- 94%	LN0 100%	ET 0% CT 0%

Table 2 Data from microarray studies for one test only

Reference; N	**Cohorts**	**Population**	**Nodal status**	**Endo / chemo**	**% pts per group**			**Outcome**	**Test**	**Outcomes HR (95% CI) unless stated otherwise**		
					Low	**Inter**	**High**			**0-5 yr**	**0-10 yr**	**5-10yr**
O-DX only												
Ahn 2013[1] a)N=186	Gananam Severance Hospital	100% ER+ 12% HER2+ a) all patients	a)47.8% LN+ (% LN>3 NR)	a)84% ET 13% CT	27	82	77	OS	**O-DX**		HR NR, P=0.361	
Cockburn 2016[2] a) N=230 b) N=132	NCBI Gene Expression Omnibus: a) GSE17705 - training	100% ER+ 100% HER2-	a) 39.6% LN+ (LN>3 NR)	100% ET CT NR	-	-	-	DRFS	**O-DX**		HR (O-DX continuous): 1.74 (0.99 to 3.07, p=0.055)*	
			a) LN0 (n=139)		-	-	-				HR (O-DX continuous): 3.58)1.38 to 9.27, p=0.012)*	
			a) LN+ (n=91)		-	-	-				HR (O-DX continuous): 1.16 (0.57 to 2.34, p=0.68)*	

	b) GSE6532		b) LN0 (n=43)			-	-	-				HR (O-DX continuous): 0.36 (95% CI NR) p=0.0001*	
			b) LN+)N=89)			-	-	-				HR (O-DX continuous): HR 0.82 (95% CI NR)P=0.306*	
Loi 2007[3] N=249 a) 118 b) 131	John Radcliffe Hospital, UK; Guys Hospital, UK; Uppsala University Hospital, Sweden	HR+ 100% HER2- NR	LN0 47%		ET 100% CT 0%	30		70	TDM	**O-DX**		**Rates Inter/Low vs High:** 81% vs 60% **AUC:** 0.69	
			a) LN0 100%			34		66				**Rates Inter/Low vs High:** 84% vs 64%	
			b) LN+ 100%			27		73				**Rates Inter/Low vs High:** 78% vs 57%	
Naoi, 2013[4] N=459	Osaka University Hospital; public databases (GSE17705, GSE12093)	100%ER+ HER2- NR	LN0 100%		100% ET 0% CT	62	18	20	RFS	**O-DX**		**Low vs Intermediate:** HR NR, p=0.0014 **Low vs High:** HR NR, p<0.01	
Jonsdottir, 2014[7] N=94	NR - Norway		a-i) 100% ER+, HER2- NR		a-i) NR	-	-	-		**O-DX**	**HR NR** **14 year Rates:** low: 83%; Inter: 81%; High: 61%, p=0.293		
Gyorffy 2015[8] b-i) N=113	b) University Hospitals (Frankfurt & Hamburg)		SG b-i): 100% ER+; HER2- NR	**SG b-i): ER+, LN0**	NR	-	-	-		**O-DX**	2.21 (0.80 to 6.11, p=0.116)		
MMP only													

Bianchini, 2013 [5] N= 683	GSE6532, GSE9195, GSE17705, GSE12093	100% ER+ 95% HER2-	38% LN+	100% ET	NR	NR	NR	DRFS	**MMP**	HR: 3.59 (2.02 to 6.30) p<0.0001	HR: 2.93 (1.91 to 4.49) p<0.0001	HR: 2.30 (1.16 to 4.56) p=0.017
Zemmour, 2015[6] N=197	TRANSBIG**	ER+ 69% HER2- 94%	LN0 100%	ET 0% CT 0%	-	-	-		**MMP**	**Year (5 or 10) NR** 15.19 (2.08 to 110.88, p<0.001) **Sens:** 97% **Spec:** 34% **Accuracy:** 45% **5 year multivariate HR:**[a] 17.03 (95% CI 2.31 to 125.55, p=0.005)		

[a] multivariate analysis adjusted for age, tumour size, tumour grade, ER status, HER2 status.

1. Ahn SG, Lee HM, Lee HW, et al. Prognostic discrimination using a 70-gene signature among patients with estrogen receptor-positive breast cancer and an intermediate 21-gene recurrence score. *International Journal of Molecular Sciences* 2013;14(12):23685-99. doi: https://dx.doi.org/10.3390/ijms141223685
2. Cockburn JG, Hallett RM, Gillgrass AE, et al. The effects of lymph node status on predicting outcome in ER+ HER2- tamoxifen treated breast cancer patients using gene signatures. *BMC Cancer* 2016;16:555. doi: https://dx.doi.org/10.1186/s12885-016-2501-0
3. Loi S, Haibe-Kains B, Desmedt C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor–positive breast carcinomas through genomic grade. *Journal of Clinical Oncology* 2007;25(10):1239-46.
4. Naoi Y, Kishi K, Tsunashima R, et al. Comparison of efficacy of 95-gene and 21-gene classifier (Oncotype DX) for prediction of recurrence in ER-positive and node-negative breast cancer patients. *Breast Cancer Research and Treatment* 2013;140(2):299-306. doi: https://dx.doi.org/10.1007/s10549-013-2640-9
5. Bianchini G, Pusztai L, Karn T, et al. Proliferation and estrogen signaling can distinguish patients at risk for early versus late relapse among estrogen receptor positive breast cancers. *Breast Cancer Research* 2013;15(5):R86. doi: https://dx.doi.org/10.1186/bcr3481
6. Zemmour C, Bertucci F, Finetti P, et al. Prediction of early breast cancer metastasis from DNA microarray data using high-dimensional cox regression models. *Cancer Informatics* 2015;14(Suppl 2):129-38. doi: https://dx.doi.org/10.4137/CIN.S17284
7. Jonsdottir K, Assmus J, Slewa A, et al. Prognostic value of gene signatures and proliferation in lymph-node-negative breast cancer. *PLoS ONE* 2014;9(3):e90642. doi: https://dx.doi.org/10.1371/journal.pone.0090642
8. Gyorffy B, Karn T, Sztupinszki Z, et al. Dynamic classification using case-specific training cohorts outperforms static gene expression signatures in breast cancer. *International Journal of Cancer* 2015;136(9):2091-98. doi: https://dx.doi.org/10.1002/ijc.29247

Appendix 5: EAG post-test chemotherapy use survey disseminated to UKBCG members

The following questionnaire was circulated via email to members of the UK Breast Cancer Group.

Questionnaire: Use of adjuvant chemotherapy for breast cancer based on the results of genomic/immunohistochemical tests

A team of researchers at the University of Sheffield is undertaking an assessment of the clinical and cost-effectiveness of alternative risk stratification tests for ER-positive, HER2-negative women with early breast cancer. The cost-effectiveness analysis element of this work requires estimates of the proportion of patients who go on to receive adjuvant chemotherapy based on the results of these tests.

Please consider the following three populations of women with ER-positive, HER2-negative with early breast cancer:

(1) Node-negative NPI<3.4
(2) Node-negative NPI>3.4
(3) Node-positive (1-3 nodes)

Based on your own subjective opinion, please estimate the probability that a woman in each of these subgroups and with each genomic/immunohistochemical test result would go on to receive adjuvant chemotherapy. Please complete both Tables 1 and 2.

Table 1: Chemotherapy decisions based on risk score for tests which give 3 classifications (e.g. Oncotype DX, Prosigna)

Risk score	**Probability patient with test result would receive chemotherapy**		
	(1) Node-negative NPI<3.4	**(2) Node-negative NPI>3.4**	**(3) Node-positive (1-3 nodes)**
Low-risk	PLEASE COMPLETE	PLEASE COMPLETE	PLEASE COMPLETE
Intermediate-risk	PLEASE COMPLETE	PLEASE COMPLETE	PLEASE COMPLETE
High-risk	PLEASE COMPLETE	PLEASE COMPLETE	PLEASE COMPLETE

Table 2: Chemotherapy decisions based on risk score for tests which give 2 classifications (e.g. MammaPrint and EndoPredict)

Risk score	Probability patient with test result would receive chemotherapy		
	(1) Node-negative NPI<3.4	(2) Node-negative NPI>3.4	(3) Node-positive (1-3 nodes)
Low-risk	PLEASE COMPLETE	PLEASE COMPLETE	PLEASE COMPLETE
High-risk	PLEASE COMPLETE	PLEASE COMPLETE	PLEASE COMPLETE

Survey results

Eleven oncologists completed the questionnaire. The mean probabilities obtained from the survey are presented in Tables 3 and 4.

Table 3: Chemotherapy decisions based on risk score for tests which give 3 classifications (e.g. Oncotype DX, Prosigna)

Risk score	Probability patient with test result would receive chemotherapy		
	(1) Node-negative NPI<3.4	(2) Node-negative NPI>3.4	(3) Node-positive (1-3 nodes)
Low-risk	0%	4%	41%
Intermediate-risk	20%	41%	72%
High-risk	77%	91%	95%

Table 4: Chemotherapy decisions based on risk score for tests which give 2 classifications (e.g. MammaPrint and EndoPredict)

Risk score	Probability patient with test result would receive chemotherapy		
	(1) Node-negative NPI<3.4	(2) Node-negative NPI>3.4	(3) Node-positive (1-3 nodes)
Low-risk	1%	14%	36%
High-risk	74%	91%	96%

Appendix 6: Additional inputs used in EAG sensitivity analyses

Table 1: Cusumano *et al*[216] post-test chemotherapy probabilities (node-negative and node-positive)

Test risk classification	Post-test chemotherapy probability	
	Node-negative	Node-positive
Low-risk	0.05	0.36
High-risk	0.92	0.99

Table 2: Penault-Llorca *et al*[213] post-test chemotherapy probabilities (node-negative)

Test risk classification	Post-test chemotherapy probability
Low-risk	0.01
High-risk	0.87

Table 3: Baseline probability of chemotherapy adjusted by Oncotype RS score[255] (node-negative, intermediate clinical risk)

Oncotype DX risk classification	Probability (no test)
Low-risk	[redacted]
Intermediate-risk	
High-risk	

Table 4: Risk classification probabilities and 10-year DMFI probabilities for RSPC (from TransATAC analysis[43])

Test risk classification	Classification probability	10-year DMFI
LN0, NPI≤3.4		
Low-risk	[redacted]	[redacted]
Intermediate-risk		
High-risk		
LN0, NPI>3.4		
Low-risk	[redacted]	[redacted]
Intermediate-risk		
High-risk		

Table 5: Prosigna risk classification and distant metastases probabilities derived from Gnant and Filipits[54]

Test risk classification	Classification probability	10-year DMFS
LN+		
Low-risk	0.04	1.00
Intermediate-risk	0.34	0.94
High-risk	0.62	0.76

Table 6: EPClin risk classification and distant metastases probabilities derived from Dubsky *et al*[57]

Test risk classification	Classification probability	10-year DMFS
LN+		
Low-risk	0.24	0.95
High-risk	0.76	0.72

Table 7: MammaPrint risk classification and distant metastases probabilities derived from van t' Veer *et al*[292]

Test risk classification	Classification probability	10-year DMFS
LN0		
Low-risk	0.71	0.93
High-risk	0.29	0.85

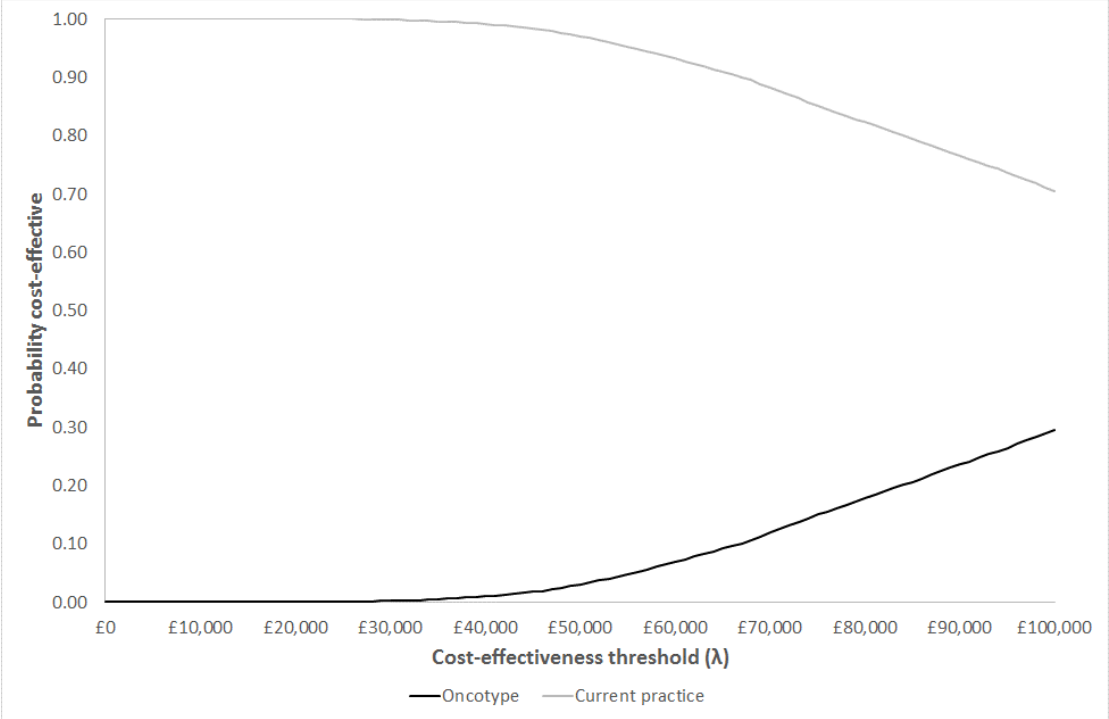
Appendix 7: Modelled chemotherapy use with and without tumour profiling tests (EAG model)

Oncotype DX versus current practice			
Subgroup	**Test**	**No test**	**Net change**
LN0 NPI≤3.4	0.076	0.072	0.004
LN0 NPI>3.4	0.273	0.430	-0.157
LN+ (1-3 nodes)	0.337	0.627	-0.290
IHC4+C versus current practice			
Subgroup	**Test**	**No test**	**Net change**
LN0 NPI≤3.4	0.030	0.072	-0.042
LN0 NPI>3.4	0.355	0.430	-0.075
LN+ (1-3 nodes)	0.554	0.627	-0.073
ProSigna versus current practice			
Subgroup	**Test**	**No test**	**Net change**
LN0 NPI≤3.4	0.075	0.072	0.003
LN0 NPI>3.4	0.435	0.430	0.005
LN+ (1-3 nodes)	0.709	0.627	0.082
EPClin versus current practice			
Subgroup	**Test**	**No test**	**Net change**
LN0 NPI≤3.4	0.140	0.072	0.068
LN0 NPI>3.4	0.438	0.430	0.008
LN+ (1-3 nodes)	0.603	0.627	-0.024
MammaPrint versus current practice			
Subgroup	**Test**	**No test**	**Net change**
MINDACT overall population	0.319	0.466	-0.148
mAOL high-risk	0.445	0.772	-0.327
mAOL low-risk	0.191	0.159	0.033

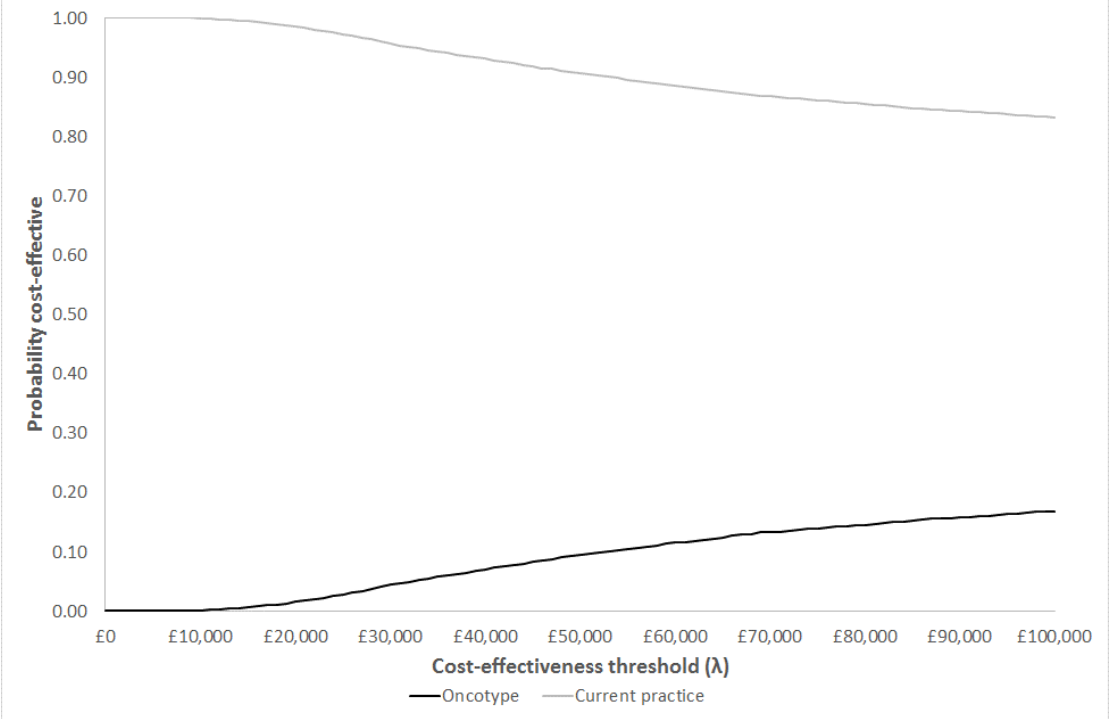
Appendix 8: EAG cost-effectiveness acceptability curves

Cost-effectiveness acceptability curves – Oncotype DX versus current practice

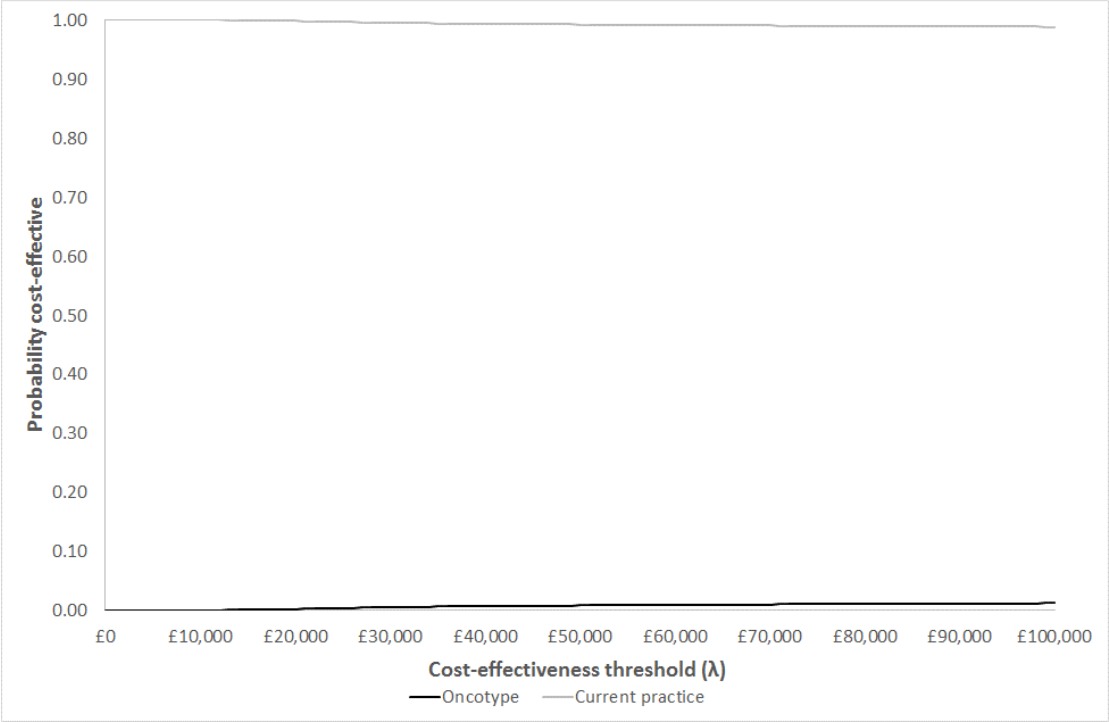
(1) Node-negative NPI≤3.4



(2) Node-negative NPI>3.4

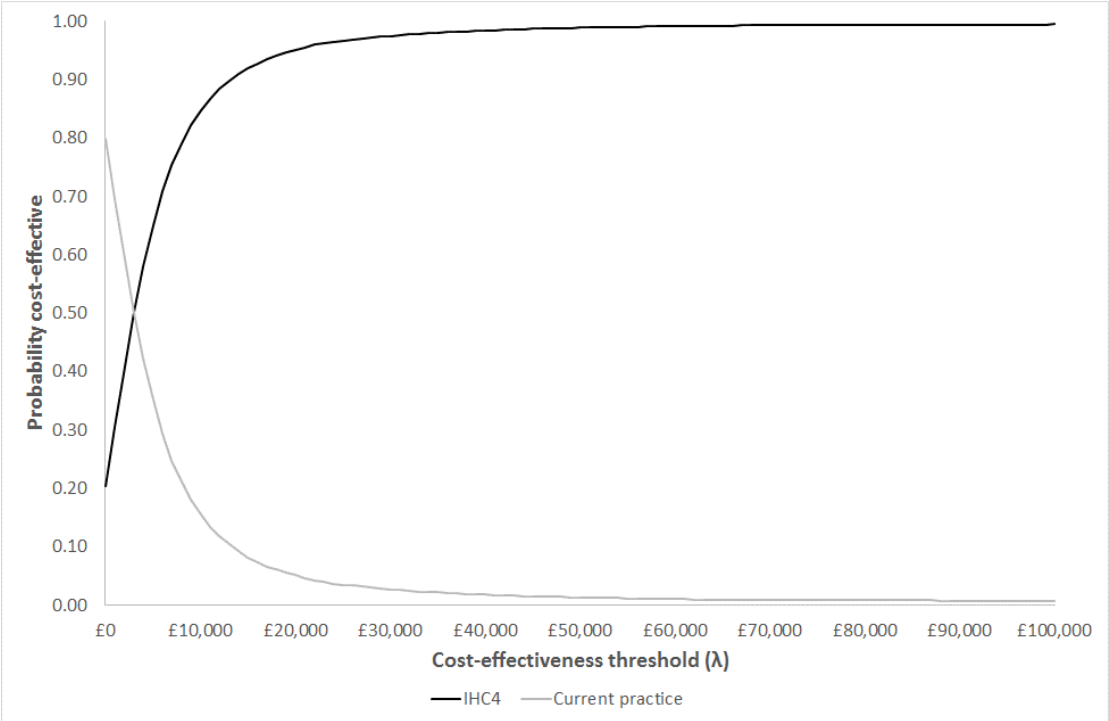


(3) Node-positive (1-3 nodes)

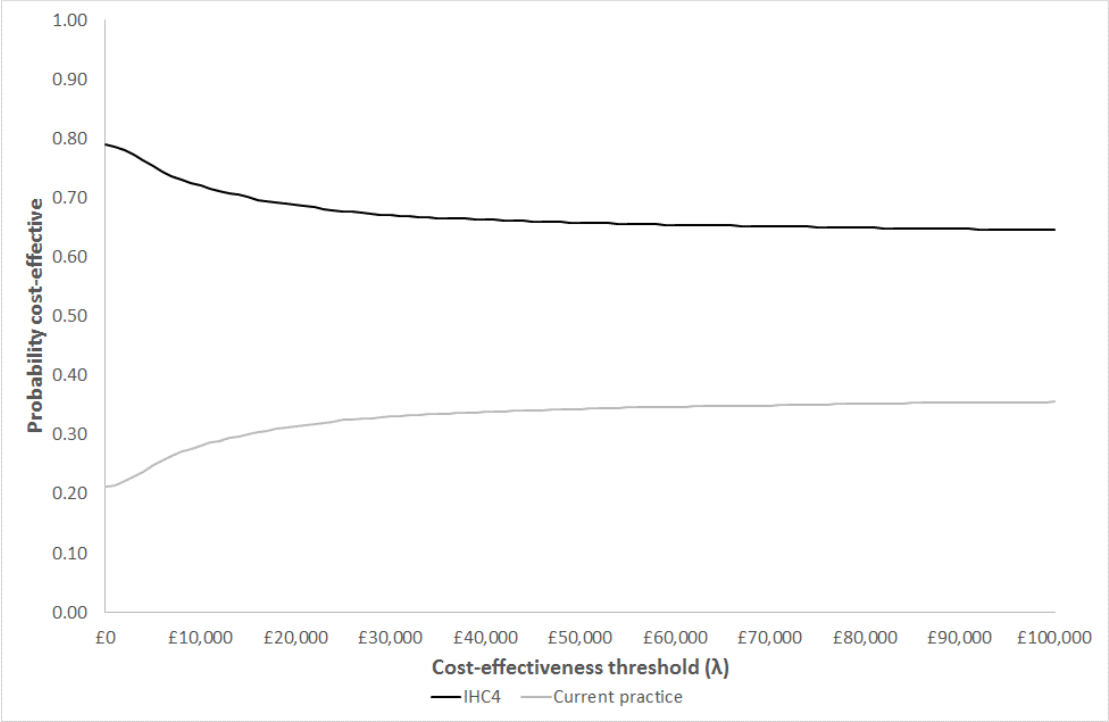


Cost-effectiveness acceptability curves – IHC4+Clin versus current practice

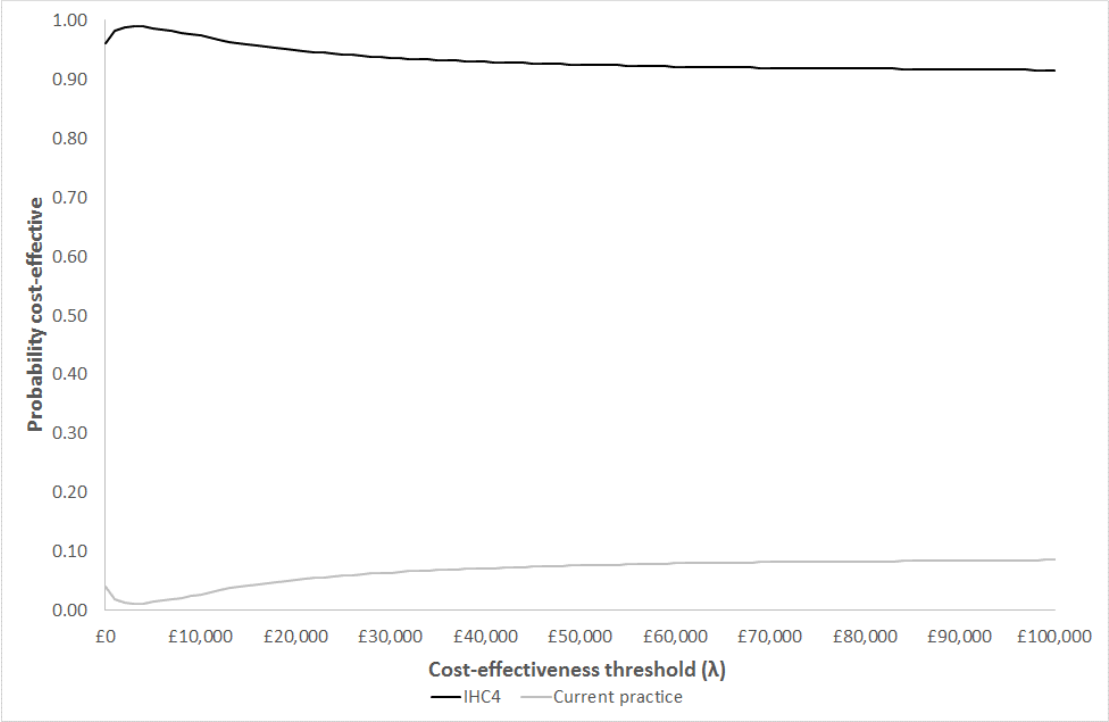
(1) Node-negative NPI≤3.4



(2) Node-negative NPI>3.4

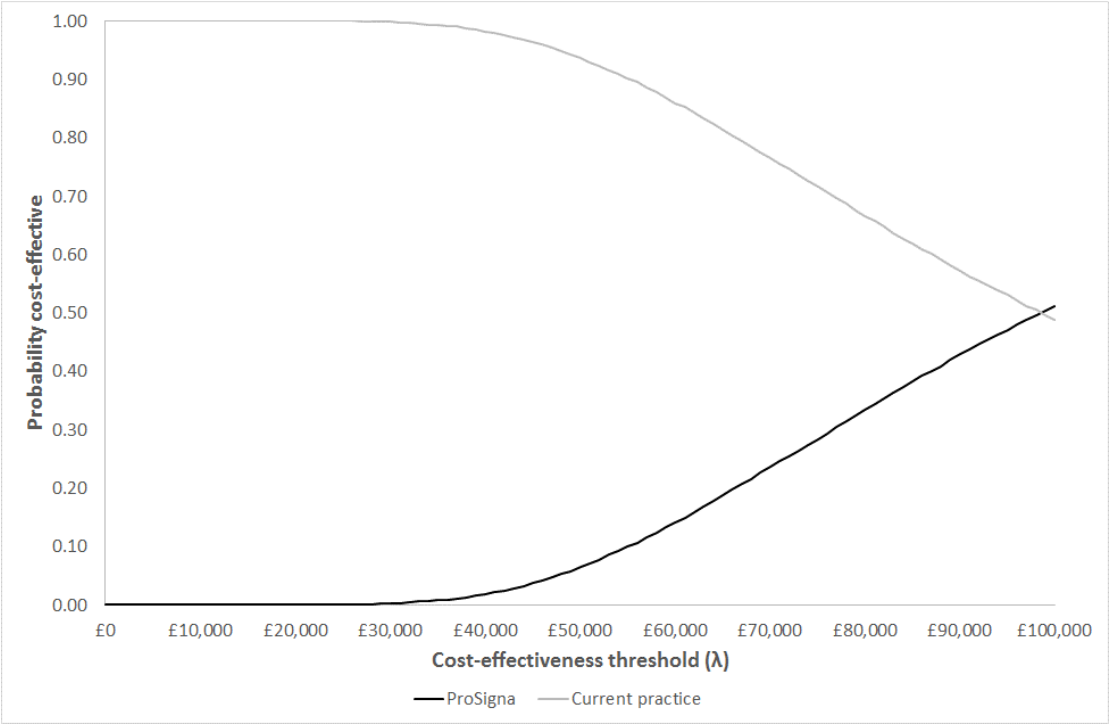


(3) Node-positive (1-3 nodes)

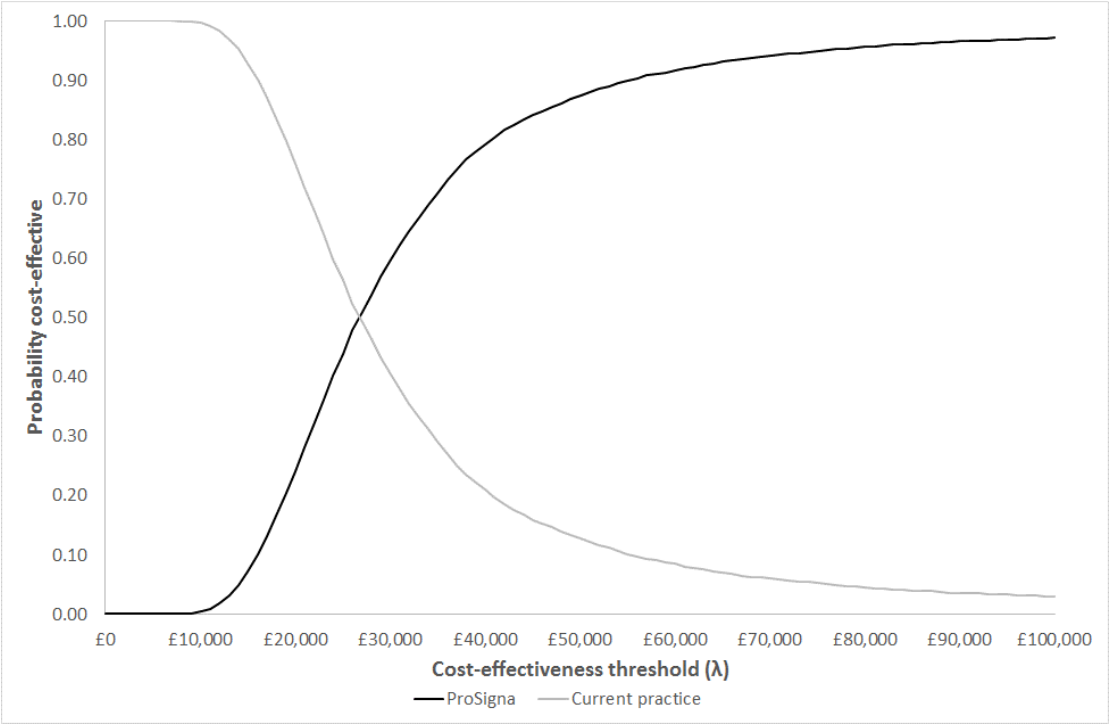


Cost-effectiveness acceptability curves – Prosigna versus current practice

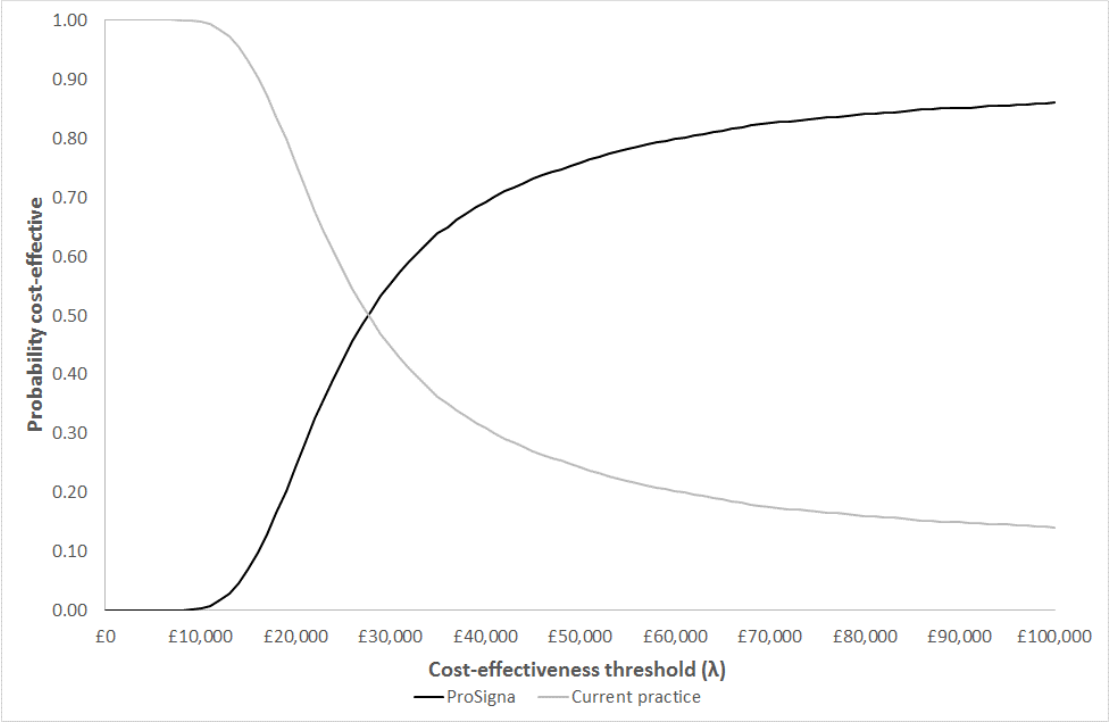
(1) Node-negative NPI≤3.4



(2) Node-negative NPI>3.4

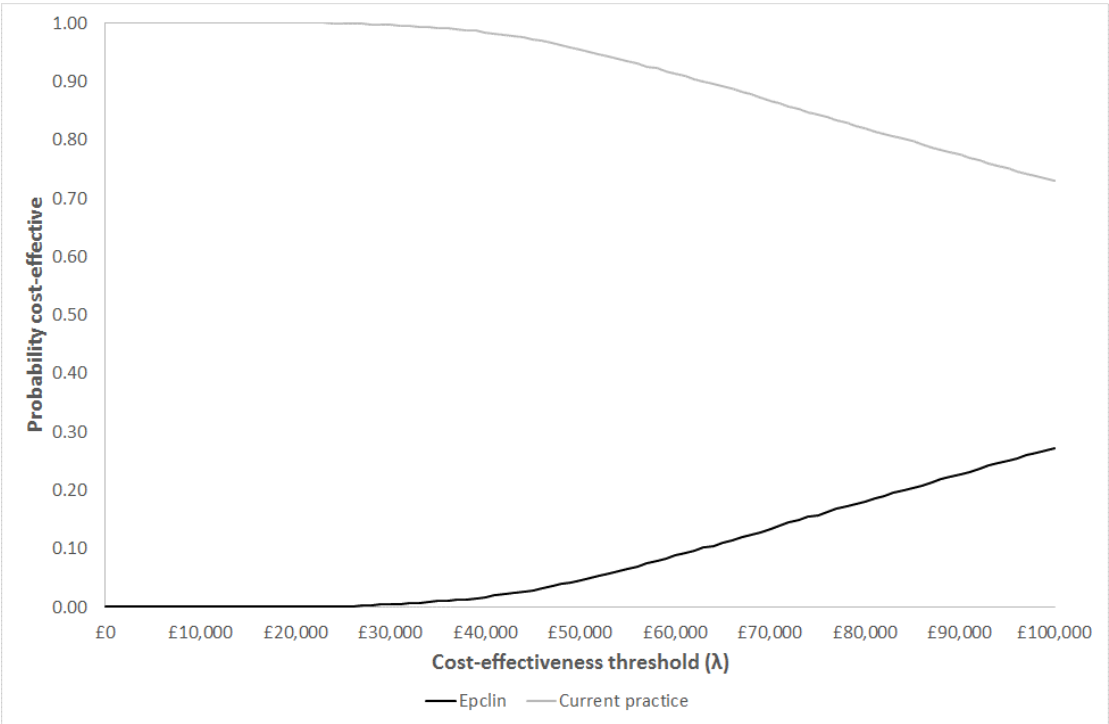


(3) Node-positive (1-3 nodes)

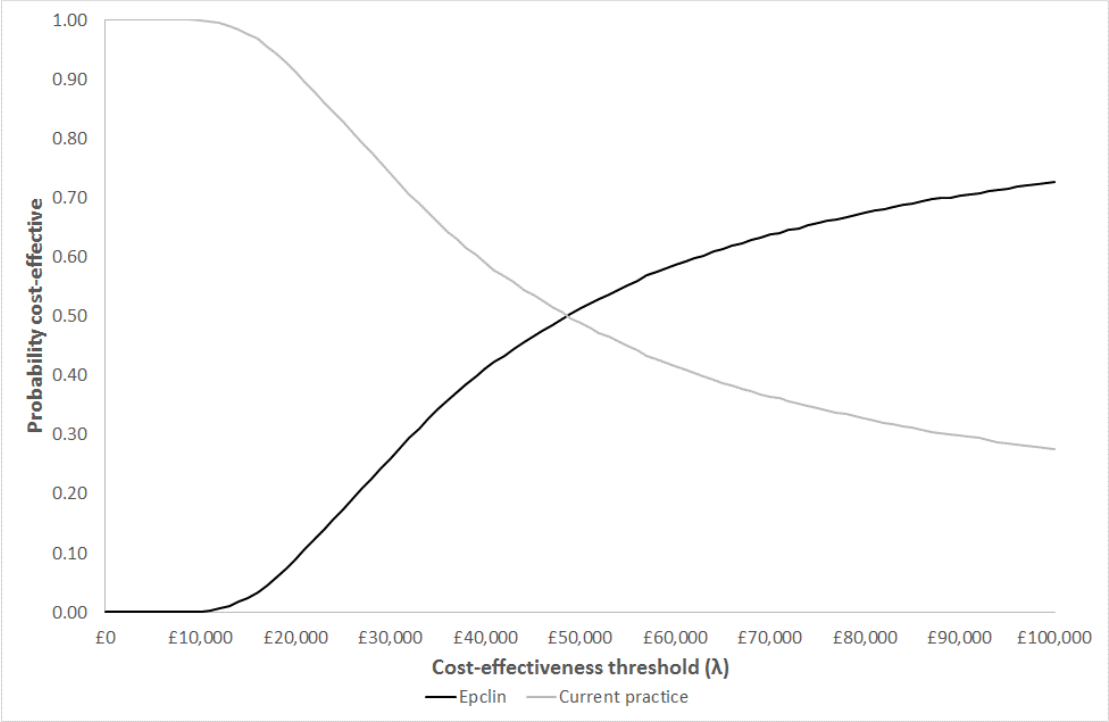


Cost-effectiveness acceptability curves – EPClin versus current practice

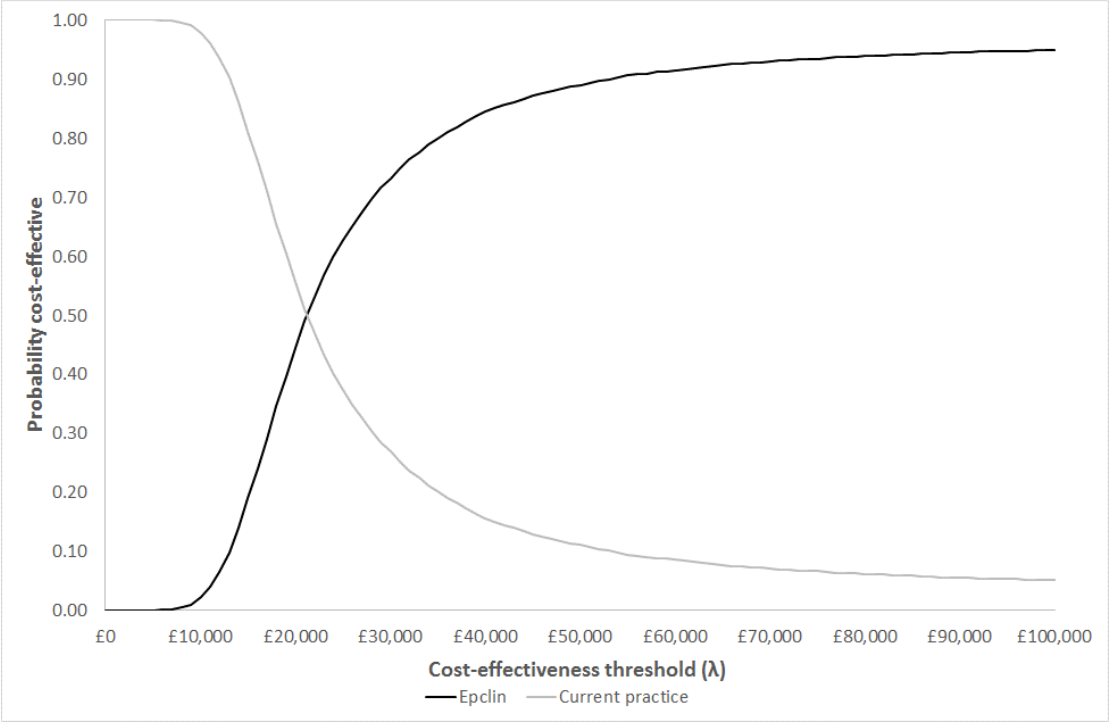
(1) Node-negative NPI≤3.4



(2) Node-negative NPI>3.4

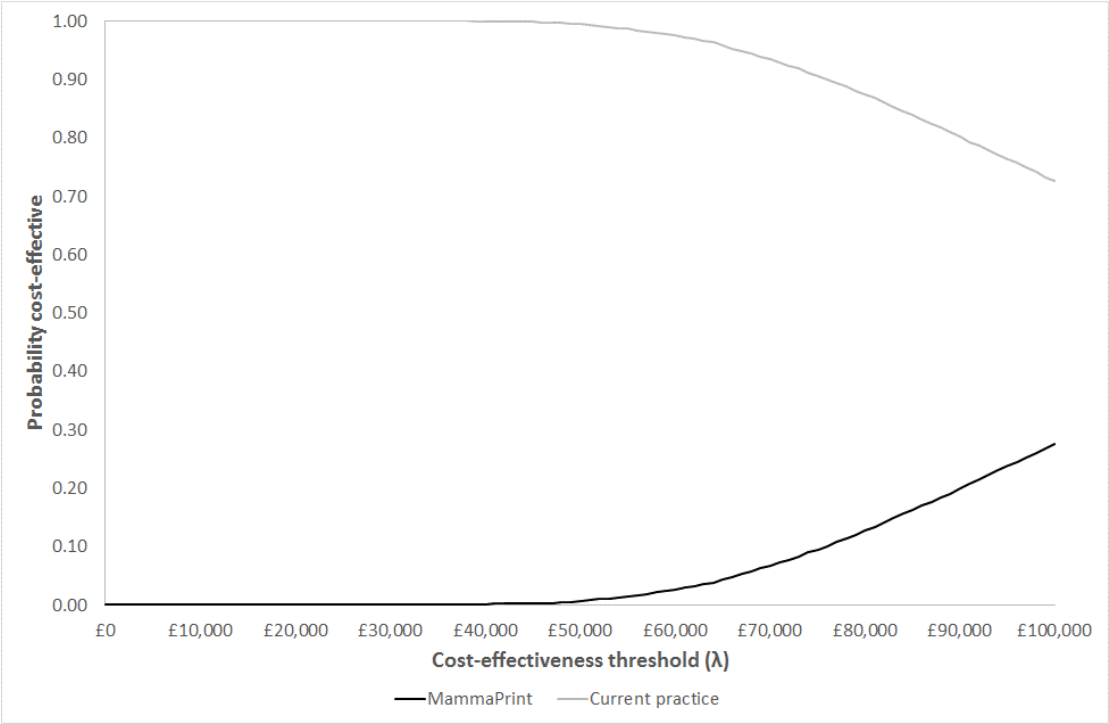


(3) Node-positive (1-3 nodes)

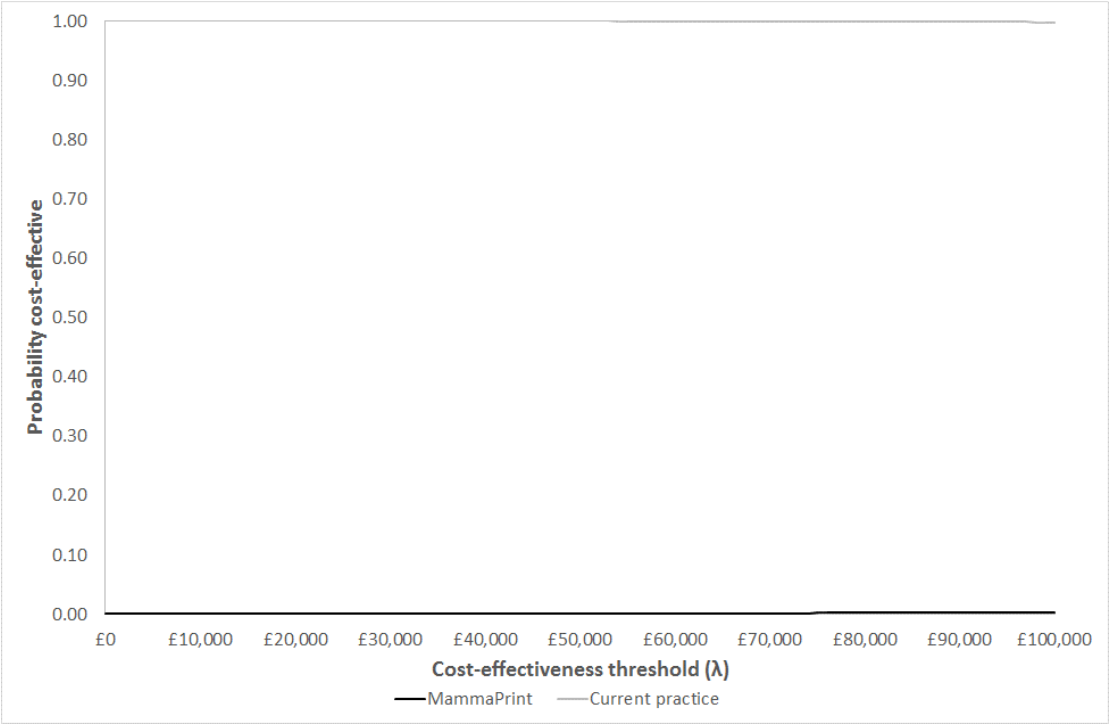


Cost-effectiveness acceptability curves – MammaPrint versus current practice

(1) Overall MINDACT population



(2) mAOL high-risk subgroup



(3) mAOL low-risk subgroup