CONFIDENTIAL UNTIL PUBLISHED

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Diagnostics Assessment Programme

Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer

Contents:

National Institute for Health and Care Excellence

DIAGNOSTICS ASSESSMENT PROGRAMME

Evidence overview

Tumour profiling tests to guide adjuvant chemotherapy decisions in people with breast cancer

(update of DG10)

This overview summarises the key issues for the diagnostics advisory committee's consideration. This document is intended to be read in conjunction with the final scope issued by NICE for the assessment and the diagnostics assessment report. A glossary of terms can be found in Appendix B. Academic in confidence information is marked [redacted]. Commercial-in-confidence information is marked [redacted].

1 Background

1.1 Introduction

The purpose of this assessment is to update NICE diagnostics guidance 10 on gene expression profiling and expanded immunohistochemistry tests for guiding adjuvant chemotherapy decisions in early breast cancer management. The clinical and cost effectiveness of 5 tumour profiling tests used to guide adjuvant chemotherapy decisions in people with breast cancer have been evaluated. Provisional recommendations on the use of these technologies will be formulated by the diagnostics advisory committee at the committee meeting on 30 November 2017.

This guidance will update the existing guidance, which included the following recommendations:

- Oncotype DX is recommended as an option for guiding adjuvant chemotherapy decisions for people with oestrogen receptor positive (ER+), lymph node negative (LN−) and human epidermal growth factor receptor 2 negative (HER2−) early breast cancer if:
  - the person is assessed as being at intermediate risk and
  - information on the biological features of the cancer provided by Oncotype DX is likely to help in predicting the course of the disease and would therefore help when making the decision about prescribing chemotherapy and
  - the manufacturer provides Oncotype DX to NHS organisations according to the confidential arrangement agreed with NICE.
- NICE encourages further data collection on the use of Oncotype DX in the NHS.
- MammaPrint, IHC4 and Mammostrat are only recommended for use in research in people with ER+, LN− and HER2− early breast cancer, to collect evidence about potentially important clinical outcomes and to determine the ability of the tests to predict the benefit of chemotherapy (see section 7). The tests are not recommended for general use in these people because of uncertainty about their overall clinical benefit and consequently their cost effectiveness.

This update has been done according to the standard update process.

Tumour profiling tests are designed to provide information on the activity of genes within tumour samples from people with early breast cancer. The results of the tests provide a risk profile of an individual's breast cancer which can be combined with other clinical risk factors that are routinely assessed, such as nodal status and tumour size. It is claimed that the risk profile can be used to better predict the risk of disease recurrence in the future. Some tests also claim to predict the benefit a patient may receive from chemotherapy.

This information is intended to help treatment decision-making with regard to adjuvant chemotherapy use.

Tumour profiling tests may improve the identification of people with early breast cancer who may not benefit from having adjuvant chemotherapy because they have a low risk of disease recurrence. These people could avoid unnecessary treatment, and would therefore not be exposed to the comorbidities and negative impacts on quality of life that are associated with chemotherapy. Additionally, people with early breast cancer who have been identified as at low risk of disease recurrence could be reclassified as being at high risk of recurrence, and therefore may benefit from chemotherapy. People with breast cancer and clinicians may also benefit from improved confidence in the appropriateness of the treatment they are having or recommending.

1.2 Scope of the assessment

Table 1 Scope of the assessment

Decision question	Do tumour profiling tests used for guiding adjuvant chemotherapy decisions for people with early stage breast cancer (described in section 1.6 of NICE CG80) represent a clinically and cost-effective use of NHS resources?
Populations	People with oestrogen receptor positive (ER+) (and/or progesterone receptor positive [PR+]), human epidermal growth factor receptor 2 negative (HER2−), early stage breast cancer (stages I or II) with 0 to 3 positive lymph nodes. Where data permits, the following subgroups may be considered: • people with lymph node negative cancer; people with micrometastases in the lymph nodes; and people with 1 to 3 positive lymph nodes • premenopausal women and postmenopausal women • people predicted to be at low, intermediate or high risk using a risk assessment tool, or using clinical and pathological features • men and women • people of different ethnicities.
Interventions	• EndoPredict

	• MammaPrint • Oncotype DX Breast Recurrence Score • Prosigna • IHC4 In combination with current decision-making.
Comparators	• Current decision-making, which may include any tool, or clinical and pathological features, used to assess risk.
Healthcare setting	Secondary and tertiary care.
Outcomes	Intermediate measures for consideration may include: • time to test results • analytical validity • prognostic ability • ability to predict benefit from chemotherapy • impact of test results on decision-making.
	Clinical outcomes for consideration may include: • disease-free survival • overall survival • distant recurrence • disease-related morbidity and mortality • chemotherapy-related morbidity and mortality.
	Patient-reported outcomes for consideration may include: • health-related quality of life • anxiety.
	Costs will be considered from an NHS and Personal Social Services perspective. Costs for consideration may include: • costs of treating breast cancer, including: drug cost, administration cost, outpatient appointments, and treatment of adverse events • costs of the tests, including equipment costs and reagents when relevant • costs of staff and associated training.
	The cost effectiveness of interventions should be expressed in terms of incremental cost per quality-adjusted life year.
Time horizon	The time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared.

Further details including descriptions of the interventions, comparators, care pathway and outcomes can be found in the final scope.

2 The evidence

This section summarises data from the diagnostics assessment report compiled by the External Assessment Group (EAG).

2.1 Clinical effectiveness

The EAG did a systematic review of the evidence on the effectiveness of 5 tumour profiling tests: EndoPredict (EPClin score unless EP score is specifically mentioned), IHC4/IHC4+C, MammaPrint, Oncotype DX Breast Recurrence Score (hereafter referred to as Oncotype DX) and Prosigna. A summary of the test characteristics is presented in table 2.

Table 2 Characteristics of tests included in the scope

Test	EndoPredict	MammaPrint	Oncotype DX	Prosigna	IHC4/IHC4+C
Purpose	Predict recurrence risk	Predict recurrence risk and chemotherapy benefit	Predict recurrence risk and chemotherapy benefit	Predict recurrence risk and intrinsic subtype[a]	Predict recurrence risk
Description	EP score = 12 gene assay (RT-qPCR) EPClin score = EP score + clinical factors	Microarray 70 gene array	RT-qPCR 21 gene assay	Direct mRNA counting + clinical factors 50 gene assay	IHC4 = 4 IHC tests IHC4+C = 4 IHC tests + 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
Menopausal status	Pre- and postmenopausal	Pre- and postmenopausal	Pre- and postmenopausal	Postmenopausal	Postmenopausal
Test result	Low risk, high risk	Low risk, high risk	Low risk, intermediate risk, high risk	Low risk, intermediate risk, high risk	Low risk, intermediate risk, high risk

[a] Evidence relating to intrinsic subtype was not reviewed in this assessment.

Abbreviations: EP, EndoPredict; IHC, immunohistochemistry; RT-qPCR, reverse transcriptase quantitative polymerase chain reaction.

The methods of the systematic review can be found starting on page 45 of the diagnostics assessment report. Evidence on the following outcomes was of interest in the clinical-effectiveness review:

- Prognostic ability – the degree to which the test can accurately predict the risk of an outcome such as disease recurrence.
- Prediction of chemotherapy benefit - 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.
- Clinical utility – the ability of the prospective use of the test to affect patient outcomes such as recurrence and survival compared with current practice.
- Decision impact – how the test influences decision-making in terms of which patients will be offered chemotherapy.

A total of 153 references were included in the review. Studies assessing prognostic ability and prediction of chemotherapy benefit were quality assessed using relevant criteria selected from the draft prediction model study risk of bias assessment tool (PROBAST). Clinical utility studies were quality assessed using the Cochrane risk of bias tool for randomised controlled trials (RCTs). Studies assessing decision impact were not quality assessed because of time constraints.

Evidence on prognostic ability

Study designs and patient characteristics

Studies with evidence on prognostic ability are summarised by test, starting on page 66 of the diagnostics assessment report. The EAG judged that studies done in East Asia may be less generalisable to England because usual clinical practice may differ between countries enough to affect prognostic outcomes. Also, it is possible that people of different ethnicities have different underlying risk profiles and natural history of disease.

Many studies treated some or all patients with chemotherapy. The EAG stated that results from these studies should be viewed with caution because this could reduce the apparent prognostic performance of a test as chemotherapy could affect event rates. As such, validation cohorts (a population studied to confirm the prognostic ability of a test) should ideally treat patients with endocrine monotherapy, but not chemotherapy.

Results are generally presented as unadjusted or adjusted analyses. Unadjusted analyses do not assess the question of whether a test has additional value over clinicopathological factors. Adjusted analyses show whether the test has prognostic value over clinicopathological variables. Studies with evidence of prognostic ability for the tumour profiling tests are summarised in table 3.

Table 3: Study designs and patient characteristics for studies providing evidence on prognostic ability

	Study designs	Study locations	Cohort characteristics	Treatments	Quality assessment
Oncotype DX (from page 66 of DAR)	11 data sets: 7 re-analyses of RCT data; 4 retrospective studies of routinely collected data/archived samples	RCTs: 5 USA, 1 UK, 1 France Retrospective studies: 1 USA, 2 China, 1 Japan	LN status: 3 studies - mixed; 4 studies - LN0; 3 studies - LN+; 1 study NR HR status: all studies 100% HR+ HER2 status: only 2 studies 100% HER2-	All ET no CT: 4 studies All ET all CT: 3 studies All ET and mixed/ unclear CT: 2 studies Mixed ET and CT: 1 study Unclear: 1 study	All studies were validation studies Only 4 studies with no CT treatment Concerns due to exclusion of tumour samples with insufficient tissue
MammaPrint (from page 128 of the DAR)	10 data sets: 1 re-analysis of RCT data; 9 retrospective studies of routinely collected data 4 studies pooled data on specific patients from studies above	RCT: Sweden Retrospective studies: 4 Netherlands; 2 USA; 2 Europe; 1 Japan	LN status: 3 cohorts – mixed; 6 cohorts – LN0; 1 cohort – LN+ HR status: 8 cohorts – mixed; 2 cohorts (and 1 subgroup) – 100% ER+ HER2 status: 6 cohorts – not reported; 4 cohorts – mixed	All ET no CT: 2 analyses No ET, no CT: 2 analyses Mixed ET, no CT: 1 cohorts Mixed ET, mixed CT: 6 cohorts	All studies were validation studies Only 5 analyses with no CT treatment Concerns due to exclusion of some patients recruited to the original trial/cohort and inclusion of patients out of scope
Prosigna (from page 182 of the DAR)	8 data sets: 6 re-analyses of RCT data; 3 retrospective analyses of 2 prospective cohorts	RCTs: 1 UK, 1 Austria, 1 Spain, 3 USA/Canada Retrospective studies: 1 Denmark, 1 Canada	LN status: Mixed – 6 studies; LN0 – 1 study; LN+ - 2 studies HR status: 6 studies 100% ER+ or HR+; 3 studies mixed HER2 status: 3 studies HER2-; 3 studies HER2 NR; 3 studies mixed	All ET no CT: 5 cohorts All ET and all CT: 1 cohort Some ET (or NR) and all CT: 3 cohorts	All studies were validation studies 5 cohorts had no CT treatment Concerns due to exclusion of some patients recruited to the original trial/cohort
EndoPredict (from page 198 of the DAR)	3 data sets: all re-analyses of RCT data	1 UK, 1 Austria, 1 Spain	LN status: Mixed – 2 studies; LN+ 1 study HR status: all 100% ER+ HER2 status: all 100% HER2-	All ET (5 years) no CT: 2 cohorts All ET (5 years) all CT: 1 cohort	All studies were validation studies 2 studies with no CT Concerns due to exclusion of some patients recruited to the original trial (or unclear)

IHC4 and IHC4+C (from page 212 of the DAR)	12 data sets: 6 re-analyses of RCT data (5 validation, 1 derivation); 6 analyses of routinely collected patient data	RCTs: 1 UK, 2 Germany, 1 Spain, 1 Europe, 1 various Retrospective studies: 2 UK, 1 USA, 1 France, 1 China, 1 Taiwan	LN status: Mixed – 9 studies; LN0 – 1 study; LN+ - 2 studies HR status: 10 studies 100% HR+ or ER+; 2 studies HR+ NR HER2 status: 7 studies 100% HER2-; 3 studies HER2 NR; 2 studies mixed	All ET, no CT: 2 studies All ET and mixed/unclear CT: 4 studies Some ET all CT: 2 studies Some ET no CT: 1 study Mixed/unclear ET and CT: 3	One study was the derivation cohort Only 2 studies with no CT Concerns due to exclusion of tumour samples with insufficient tissue
Abbreviations: CT, chemotherapy; DAR, diagnostics assessment report; ER, oestrogen receptor; ET, endocrine therapy; HER2, human epidermal growth factor receptor 2; HR, hormone receptor; LN, lymph node; NR, not reported; RCT, randomised controlled trial;					

Distribution of patients across risk categories

Among studies of lymph node negative patients treated with endocrine monotherapy, around 70 to 80% of patients were categorised as low or low/intermediate risk across all tests (table 4). Some studies did not report the human epidermal growth factor receptor 2 (HER2) status of patients, which, if mixed, could affect the proportions of patients in each category. There was only 1 MammaPrint study which treated patients with endocrine monotherapy. In MammaPrint studies with mixed endocrine and chemotherapy use, cohorts generally had mixed hormone receptor status, and/or mixed HER2 status, so results may not be comparable with other tests (low risk 20 to 61%; 6 studies; not tabulated). Most IHC4/IHC4+C studies used quartiles or tertiles to define risk groups. These studies will have been specific to each cohort and do not provide useful information on the distribution of patients across risk categories (not tabulated).

Table 4: Risk categories for lymph node negative cohorts not treated with chemotherapy

	Low risk category	Intermediate risk category	High risk category	Number of studies	Patients
Oncotype DX	48 to [redacted]%	20 to [redacted]%	[redacted] to 33%	4 studies	ER+, HER2+/-
MammaPrint	71%	-	29%	1 study	ER+, HER2 NR
Prosigna	48 to [redacted]%	[redacted] to 32%	[redacted] to 20%	3 studies	Most ER+, HER2-
EndoPredict	[redacted] to [redacted]%	-	[redacted] to [redacted]%	2 studies	ER+, HER2-
IHC4+C	[redacted]%	[redacted]%	[redacted]%	1 study	ER+, 95% HER2-
Abbreviations: ER, oestrogen receptor; HER2, human epidermal growth factor receptor 2; NR – not reported					

The proportion of low/intermediate risk patients was generally much lower in lymph node positive than in lymph node negative cohorts. When Oncotype DX was used, however, the proportion of patients with low/intermediate risk was only slightly lower in the lymph node negative group than in the lymph node positive group (table 5). Studies of MammaPrint in lymph node positive patients were all done in cohorts with mixed hormone receptor status and mixed or unknown HER2 status, so results may not be comparable with other tests (low risk 38 to 41%; 2 studies). Most IHC4/IHC4+C studies used quartiles or tertiles to define risk groups, and these will have been specific to each cohort (not tabulated).

Table 5: Risk categories for lymph node positive cohort not treated with chemotherapy

	Low risk category	Intermediate risk category	High risk category	Number of studies	Patients
Oncotype DX	[redacted]%	[redacted]%	[redacted]%	1 study	ER+, HER2-
MammaPrint	-	-	-	No studies limited to HR+ patients	
Prosigna	4 to 25%	27 to 34%	48 to 62%	3 studies	Most ER+, HER2-
EndoPredict	[redacted] to [redacted]%	-	[redacted] to [redacted]%	2 studies	ER+, HER2-
IHC4+C	[redacted]%	[redacted]%	[redacted]%	1 study	ER+, HER2-
Abbreviations: ER, oestrogen receptor; HER2, human epidermal growth factor receptor 2; HR, hormone receptor; NR – not reported					

Prognostic performance: Oncotype DX

The 10 year distant recurrence-free interval rates (table 6) suggest that:

- the lymph node negative, low-risk group is at very low risk of recurrence in the absence of chemotherapy
- the lymph node negative intermediate risk group maybe at slightly higher risk of recurrence
- The lymph node positive group was generally at higher risk of recurrence than the lymph node negative group in both low and intermediate categories.

Table 6 Percentage DRFI risk (0 to 10 years) by lymph node status and Oncotype DX test result

LN status	Oncotype DX risk score	10 year DRFI rates [a]	Notes
LN negative	Low risk	93 to 97%	4 studies; patients had endocrine monotherapy
		96%	1 study; patients had endocrine and chemotherapy
LN positive	Low risk	82%	1 study; patients had endocrine monotherapy
		81%	1 study; patients had endocrine and chemotherapy
LN negative	Intermediate risk	86 to 100%	4 studies; patients had endocrine monotherapy
		89%	1 study; patients had endocrine and chemotherapy
LN positive	Intermediate risk	[redacted]	1 study; patients had endocrine monotherapy
		65%	1 study; patients had endocrine and chemotherapy
LN negative	High risk	61 to 77%	4 studies; patients had endocrine monotherapy
		88%	1 study; patients had endocrine and chemotherapy
LN positive	High risk	[redacted]	1 study; patients had endocrine monotherapy
		59%	1 study; patients had endocrine and chemotherapy

Abbreviations: DRFI, distant recurrence-free interval; LN, lymph node

[a] Note than Sun et al. was excluded from the ranges because it appeared to be an outlier with very low DRFI rates (18 to 63%)

Unadjusted analyses indicated that Oncotype DX had prognostic power (there were statistically significant differences between low-risk and high-risk groups) across various recurrence outcomes, regardless of lymph node status. However, hazard ratios between the intermediate-risk group and the high- or low-risk groups were not always statistically significant, particularly in the lymph node positive group.

In adjusted analyses, Oncotype DX provided additional prognostic information over most commonly used clinicopathological variables (age, grade, size, nodal status) regardless of lymph node status. Analyses also showed that Oncotype DX provided additional prognostic information (statistically significant) [redacted].

Prognostic performance: MammaPrint

MammaPrint had prognostic power (there were statistically significant differences between low-risk and high-risk groups) for 10 year distant recurrence-free survival in almost all unadjusted analyses of lymph node negative and lymph node positive patients. The 10-year distant recurrence-free survival and distant recurrence-free interval rates for low-risk patients are shown in table 7.

Table 7 Percentage DRFS/DRFI risk (0 to 10 years) by lymph node status and MammaPrint risk category

LN status	MammaPrint risk category	10-year DRFS/DRFI rates	Notes
Pooled LN negative / LN positive	Low risk group	87%	1 analysis; 33% had endocrine and 25% chemotherapy
LN negative	Low risk group	93%	1 analysis; endocrine monotherapy
		83%	1 analysis; no endocrine or chemotherapy
LN negative	Low risk group	80% to 90%	3 analyses; varying rates of endocrine and chemotherapy use
LN positive	Low risk group	79% to 91%	2 analyses; varying rates of endocrine and chemotherapy use
Abbreviations: DRFS, distant recurrence-free survival; DRFI, distant recurrence-free interval; LN, lymph node			

In lymph node negative patients, 4 of 5 unadjusted analyses showed statistically significant differences in hazard ratios between risk groups for 10-year distant recurrence-free survival or distant recurrence-free interval rates. Among lymph node positive patients, 2 unadjusted analyses reported statistically significant prognostic performance of MammaPrint based on hazard ratios for 10-year distant recurrence-free survival between risk groups.

In adjusted analyses, a pooled analysis of lymph node negative and positive patients showed that MammaPrint had prognostic power (there were statistically significant differences between low-risk and high-risk groups) for 10-year distant recurrence-free survival in a multivariable analysis adjusting for clinicopathological variables. Among lymph node negative patients, MammaPrint had prognostic power for distant recurrence-free interval when adjusted for Adjuvant! Online or NPI in 3 cohorts. In lymph node positive patients, MammaPrint had prognostic power (statistically significant

differences or borderline statistically significant differences between low-risk and high-risk groups) for 10-year distant recurrence-free survival and interval in adjusted analyses.

Prognostic performance: Prosigna

Prosigna had prognostic power (there were statistically significant differences between low-risk and high-risk groups) for 10-year distant recurrence-free survival and 10-year distant recurrence-free interval in all unadjusted analyses of lymph node negative and lymph node positive patients. The 10-year distant recurrence-free survival and distant recurrence-free interval rates for low- and intermediate-risk patients are shown in table 8. In analyses adjusted for clinicopathological variables, Prosigna was found to be prognostic for 10-year distant metastasis-free survival and distant recurrence-free survival. In lymph node negative patients the results were statistically significant and in lymph node positive patients the results were statistically or borderline significant.

Table 8 Percentage DRFS/DRFI risk (0 to 10 years) by lymph node status and Prosigna category

LN status	Prosigna risk category	10 year DRFS/DRFI rates	Notes
LN negative	Low risk	95% to [redacted]	3 studies; patients had endocrine monotherapy
LN positive	Low risk	[redacted]	2 studies; patients had endocrine monotherapy
		92%	1 study; all patients had endocrine and chemotherapy
LN negative	Intermediate risk	[redacted] to 93%	2 studies; patients had endocrine monotherapy
LN positive	Intermediate risk	[redacted] to 94%	2 studies; patients had endocrine monotherapy
		74%	1 study; all patients had endocrine and chemotherapy
Abbreviations: DRFS, distant recurrence-free survival; DRFI, distant recurrence-free interval; LN, lymph node			

Prognostic performance: EndoPredict

EndoPredict had prognostic power (there were statistically significant differences between low-risk and high-risk groups) for unadjusted analyses of 10-year distant recurrence-free survival and distant recurrence-free interval in lymph node negative and lymph node positive patients. The 10-year distant recurrence-free survival and distant recurrence-free interval rates for low-risk

patients are shown in table 9. In adjusted analyses of one study, EndoPredict [redacted].

Table 9 Percentage DRFS/DRFI risk (0 to 10 years) by lymph node status and EndoPredict risk category

LN status	EndoPredict risk category	10 year DRFS/DRFI rates	Notes
LN negative	Low risk	[redacted] to [redacted]	2 studies, patients had endocrine monotherapy
LN positive	Low risk	[redacted]	2 studies, patients had endocrine monotherapy
LN positive	Low risk	100%	1 study, patients had endocrine and chemotherapy
Abbreviations: DRFS, distant recurrence-free survival; DRFI, distant recurrence-free interval; LN, lymph node			

Prognostic performance: IHC4 and IHC4+C

Studies have shown that IHC4 provides statistically significant prognostic information in unadjusted analyses in both lymph node negative, lymph node positive and mixed cohorts. However, none of these studies reported survival or recurrence outcomes by risk group, but instead presented hazard ratios for high risk groups compared with low risk groups. In addition, most studies used quartiles or tertiles to define risk groups, and these were specific to each cohort and did not use pre-defined cut-off values. Also, many used laboratory methods that differed from the derivation study methodology (the study population in which the test was established). In adjusted analyses, IHC4 was shown to have additional prognostic value over clinicopathological factors in some studies.

Data on IHC4+C came from the derivation cohort and 1 validation cohort. These studies showed that IHC4+C had prognostic value in unadjusted analyses. In adjusted analyses, IHC4+C provided statistically significantly more information than NPI in lymph node negative patients but not lymph node positive patients.

Evidence on prediction of chemotherapy benefit

Chemotherapy benefit: Oncotype DX

Five data sets reported across 11 published references and 1 confidential manuscript conducted analyses that assessed the ability of Oncotype DX to predict benefit of chemotherapy (table 10).

Table 10 Studies reporting the ability of Oncotype DX to predict benefit of chemotherapy

Data set	Reference(s)	Study design	Patient population	Study quality
SWOG-8814 study	Albain et al. 2010	Phase 3, open-label, parallel-group RCT	All postmenopausal, HR+, LN+ patients (38.1% with ≥4 positive lymph nodes) and 12% HER2+.	Some risk of bias, mainly because of patient spectrum bias
NSABP B-20 trial	Paik et al. 2006, Tang et al. 2011a and Tang et al. 2011b	RCT	ER+, LN0 patients, with an unreported percentage being HER2	Some risk of bias, mainly because of patient spectrum bias
MD Anderson Center		Retrospective observational	HR+, HER2-, LN0 patients	High risk of confounding and unclear generalisability to decision problem
Clalit Health Services		Retrospective observational	ER+, HER2- patients, [redacted] and with LN1-3	High risk of confounding and unclear generalisability to decision problem
SEER registry		Retrospective observational	HR+, HER2-, LN0 patients	High risk of confounding and unclear generalisability to decision problem

Abbreviations: ER, oestrogen receptor; HER2, human epidermal growth factor receptor 2; HR, hormone receptor; LN, lymph node; NSABP, national surgical adjuvant breast and bowel project; RCT, randomised controlled trial; SEER, surveillance, epidemiology and end results; SWOG, southwest oncology group

There is some evidence from 2 re-analyses of RCTs to suggest that Oncotype DX may predict benefit from chemotherapy. Based on hazard ratios for disease-free survival for patients having chemotherapy versus those having no chemotherapy, the greatest benefit appeared to be for patients in the Oncotype DX high risk recurrence score category. Unadjusted interaction tests between Oncotype DX risk group and chemotherapy benefit were mainly statistically significant, but adjusted interaction tests were not always statistically significant. Therefore the significant results could be a consequence of omitting potentially important covariates from the statistical model.

From the 3 observational studies evidence was mixed and at high risk from confounding; 1 reported a statistically significant interaction test but this was only adjusted for a limited number of factors; 2 reported hazard ratios for various outcomes at 5 years which were statistically non-significant.

The recurrence score pathology-clinical (RSPC) algorithm incorporates Oncotype DX plus age, tumour size and grade. The RSPC algorithm was derived in TransATAC and NSABP B-14, and validated in NSABP B-20. There was a non-significant interaction test result between chemotherapy benefit and RSPC risk group. This suggests that the interaction between treatment effect and recurrence score risk group may be confounded by clinicopathological variables.

Chemotherapy benefit: MammaPrint

Two studies reported the ability of MammaPrint to predict the benefit of chemotherapy (table 11). Knauer et al. (2010) reported a pooled analysis of 541 patients, from 6 consecutive patient series, and Mook et al. (2009) reported a pooled analysis of 2 of the 6 patient series from Knauer et al. with an extended follow-up (10 years).

Table 11 Studies reporting the ability of MammaPrint to predict chemotherapy benefit

Reference(s)	Study design	Patient population	Study quality
Knauer et al. 2010	Retrospective analysis of 6 patient series	90% ER+, 89% HER2-, 50% LN0, 50% 1-3 LN+ 100% had endocrine therapy, 42% had chemotherapy	High risk of confounding and included a proportion of patients who were ER- and HER2+
Mook et al. 2009	Retrospective analysis of 2 of the 6 patient series from Knauer et al.	Restricted to LN1-3 patients (including micrometastases)	High risk of confounding and included a proportion of patients who were ER- and HER2+
Abbreviations: ER, oestrogen receptor; HER2, human epidermal growth factor receptor 2; LN, lymph node			

The evidence for the ability of MammaPrint to predict chemotherapy benefit is very limited. 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. In analyses adjusted for clinicopathological

variables results were not statistically significant. Further, the interaction test for chemotherapy treatment and risk group was non-significant. In the analysis restricted to LN1-3 patients, a statistically non-significant interaction between chemotherapy treatment and risk group was reported.

Evidence on clinical utility

Clinical utility: Oncotype DX

Five data sets reported across 9 published references and 1 confidential manuscript reported evidence relating to the clinical utility of Oncotype DX (table 12). Another study, based on the SEER database (National Cancer Institute database, US) and Genomic Health's clinical laboratory database, did not meet the inclusion criteria (because of insufficient follow-up length), but is discussed because it presents subgroup data according to age, lymph node status and race. Studies generally reported different outcomes, making comparisons across studies difficult.

Table 12 Studies providing evidence on the clinical utility of Oncotype DX

Study	Study design	Patients
TAILORx	Women with RS<11 were assigned to endocrine therapy alone. Women with RS 11 to 25 were randomised to either endocrine therapy plus chemotherapy or endocrine therapy alone. As of July 2017, only results for the low-risk (RS<11) group (n=1,626) were available. Data for this group are effectively prospective observational data.	HR+, HER2-, LN0 with tumours sized 1.1 to 5 cm (or 0.6 to 1.0 cm in intermediate or high-risk)
WSG Plan B trial	Patients with RS≥12 were randomised to 2 different sorts of chemotherapy. Another aim was to assess the risk of recurrence in patients with RS<12 who were not treated with adjuvant chemotherapy. Data for this group are effectively prospective observational data.	HR+, HER2-, clinically high-risk patients with 0-3 positive LN
MD Anderson Cancer Center (USA)	Retrospective analyses of routinely collected data. 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.	ER+, HER2-, LN0-LNmic patients who had had an Oncotype DX test
Clalit Health Services (Israel)	Retrospective analyses of routinely collected data. 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.	ER+, HER2-, LN0-LNmic or LNmic – LN3 patients who had had an Oncotype DX test.
Memorial Sloan Kettering Center (USA)	Retrospective analyses of routinely collected data. 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.	ER+, HER2-, LN0-LNmic patients who had had an Oncotype DX test.
Abbreviations: ER, oestrogen receptor; HER2, human epidermal growth factor receptor 2; HR, hormone receptor; LN, lymph node; LNmic, lymph node micrometastases; RS, risk score; TAILORx, the trial assigning individualised options for treatment WSG, West German study group		

Based on the evidence available, the EAG stated that it is difficult to conclude whether patient outcomes would be affected by using the test in a clinical setting. In lymph node negative patients, using the test in clinical practice appeared to result in low rates of chemotherapy use in low-risk patients (2% to 12%), with acceptable outcomes (distant recurrence-free survival, distant recurrence-free interval or invasive disease-free survival 96% to 99.6%). Rates of chemotherapy use increased with increasing risk category, and were generally higher in lymph node positive patients.

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

Clinical utility: MammaPrint

Two studies reported evidence relating to clinical utility of MammaPrint (table 13).

Table 13 Studies providing evidence on the clinical utility of MammaPrint

Study	Study design	Patients
MINDACT	A partially randomised trial of MammaPrint versus clinical practice. Patients with discordant risk scores (high/low or low/high) via MammaPrint and modified AO were randomised to chemotherapy or no chemotherapy; patients with concordant risk were followed as prospective cohorts (high-risk patients were all recommended to receive chemotherapy and low-risk patients were all recommended no chemotherapy).	Overall, 88% were HR+; 90% HER2-; 79% were LN0 and 21% LN1-3. However, this varied by group.
RASTER	A prospective observational study. Chemotherapy use was guided by MammaPrint in combination with the Dutch Institute of Healthcare Improvement guidelines of 2004 and clinician and patient preference. As such, estimates of prognostic performance are confounded by the differing rates of chemotherapy in different risk groups. [redacted]	LN0 patients, age <61 years, 80% ER+ and 84% HER2-
Abbreviations: AO, Adjuvant! Online; ER, oestrogen receptor; HER2, human epidermal growth factor receptor 2; HR, hormone receptor; LN, lymph node		

In the MINDACT study, for the high-clinical, low-MammaPrint risk group, 5 year distant metastasis-free survival was 95.9% with chemotherapy and 94.4% without chemotherapy, an absolute difference of 1.5%. The authors suggested that chemotherapy could possibly be avoided in these patients. For

the low-clinical, high-MammaPrint risk group, 5-year distant metastasis-free survival was 95.8% with chemotherapy and 95.0% without chemotherapy, an absolute difference of 0.8%. The EAG suggested that this result shows that low-clinical risk patients with a high-risk MammaPrint result 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. However, the comparator was modified Adjuvant! Online, and it is unclear whether the same would be true for other clinical risk scores.

Results from the RASTER study suggest that distant recurrence-free interval rates were sufficiently low in the MammaPrint low-risk group for these patients to avoid chemotherapy. The 5-year distant recurrence free interval rate for lymph node negative patients was 97.0% for low-risk patients (15% had chemotherapy) and 91.7% for high-risk patients (81% received chemotherapy). [redacted] Further, MammaPrint provided additional prognostic information over Adjuvant! Online and NPI, but not over another NHS risk scoring tool, PREDICT Plus.

Evidence comparing tests with each other

Studies comparing more than one test

Data were reported for 6 cohorts: 4 re-analyses of RCTs and 2 observational studies. The design and results from these studies are described in more detail in the diagnostics assessment report starting on page 241. The most comprehensive analysis in terms of the number of tests assessed was from TransATAC, which assessed 4 tests: EndoPredict, Prosigna, Oncotype DX and IHC4+C. A bespoke analysis of the TransATAC data which focused on the population in the scope for this assessment was provided by the trial investigators.

As the data comparing the tests with each other are limited, only broad conclusions can be drawn. Evidence shows that generally when a test placed

more patients in a low-risk category compared with another test, the event-free survival was reduced. Also, the tests generally performed differently in lymph node negative and lymph node positive patients. In TransATAC, [redacted] In terms of additional prognostic information the tests provide over clinicopathological variables or algorithms, [redacted] provided statistically significantly more prognostic information [redacted]

Microarray studies

Thirteen studies reported data from microarray analyses on more than one of the tests. These studies had methodological limitations, but they have value because they provide comparative prognostic data. All the studies reported data on Oncotype DX and MammaPrint, and 2 also report data on EndoPredict. The results of these studies are described in the diagnostics assessment repot starting on page 263.

The 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 lymph node status. In terms of additional prognostic performance of the tests over clinicopathological variables, EndoPredict appeared to have the greatest benefit, followed by Oncotype DX and then MammaPrint, though the evidence base was limited.

OPTIMA prelim

The OPTIMA prelim study analysed concordance between different tests, that is, the degree to which the tests assign the same patients to the same risk groups. It is described starting on page 282 of the diagnostics assessment

report. The study included Oncotype DX, MammaPrint, Prosigna and IHC4 plus 2 other tests not included in this assessment. 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 of 30 mm or greater if node negative.

Out of the 4 in-scope tests, MammaPrint assigned the most patients to the low-risk category (table 14), but unlike the other 3 tests it does not have an intermediate category. When low and intermediate categories were treated as 1 category for the 3 tests that have 3 risk groups, Oncotype DX assigned the most to the low/intermediate category, and MammaPrint the least. Kappa statistics indicated modest agreement between tests ranging from 0.33 to 0.53. Further, across 5 tests in the study that reported risk groups, only 39% of tumours were uniformly classified as either low/intermediate or high by all 5 tests. Of these, 31% were low/intermediate by all tests and 8% were high-risk by all tests.

The authors of the study concluded that although the 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.

Table 14 Proportion of patients assigned to each risk category in OPTIMA prelim

Test	% Low risk	% Intermediate risk	% High risk
Oncotype DX	54	28	18
MammaPrint	61	-	39
Prosigna	36	29	35
IHC4	24	48	28

Evidence on decision impact

The review of decision impact focused on studies done in the UK or the rest of Europe. The studies are described starting on page 284 of the diagnostics assessment report. The following studies were identified:

- Oncotype DX: 6 UK studies and 12 other European studies

- EndoPredict: 1 UK study and 3 other European studies
- IHC4+C: 1 UK study and 0 other European studies
- Prosigna: 0 UK studies and 3 other European studies
- MammaPrint: 0 UK studies and 8 other European studies

The percentage of patients with any change in treatment recommendation or decision (either to or from chemotherapy) in UK studies was 29% to 49% across 4 Oncotype DX studies, 37% in 1 EndoPredict study and 27% in 1 IHC4+C study. Ranges across European (non-UK) studies were 5% to 70% for Oncotype DX, 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 (pre-test to post-test) among UK studies was a reduction of 8% to 23% across 4 Oncotype DX studies, an increase of 1% in 1 EndoPredict study, and a reduction of between 2% and 26% in 1 IHC4+C study (depending on whether the decision was defined as 'recommend chemotherapy' or 'discuss chemotherapy'). Net changes across European (non-UK) studies were a reduction of 0% to 64% for Oncotype DX, a reduction of 13% to 26% for EndoPredict, a reduction of 2% to an increase of 9% for Prosigna, and reduction of 31% to an increase of 8% for MammaPrint.

Evidence on anxiety and health-related quality of life

The EAG identified 6 studies that reported outcomes relating to anxiety (including worry and distress) and health-related quality of life, which are discussed starting on page 298 of the diagnostics assessment report. The lack of use of a comparator in the studies made it difficult to tell whether changes in anxiety experienced with the use of tumour profiling tests would also have occurred if patients received a definitive decision based on clinical risk factors alone. Overall, evidence suggests that tumour profile testing may reduce state anxiety in some patients in some contexts, but generally there was little impact on health-related quality of life.

2.2 Costs and cost effectiveness

The EAG conducted a review of existing studies investigating the cost effectiveness of tumour profiling tests to guide treatment decisions in people with early breast cancer, and critiqued economic analyses provided by Agendia (MammaPrint), Genomic Health (Oncotype DX) and, and the chief investigator of a UK decision impact study (EndoPredict). The EAG also constructed a de novo economic model to assess the cost effectiveness of Oncotype DX, MammaPrint, Prosigna, IHC4+C, and EndoPredict compared with current practice.

Systematic review of cost-effectiveness evidence

A total of 26 studies were identified that had been published since the original assessment for diagnostics guidance 10. The models reported in the 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 (18 studies) or MammaPrint (8 studies) with comparators such as Adjuvant! Online, the St Gallen guidelines, standard practice or other conventional diagnostic tools. Only 1 study compared EndoPredict against a comparator, which was comprised of a combination of 3 different guidelines. There was variation between the analyses in the patient populations evaluated, their disease type and other patient characteristics.

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 models used a Markov or hybrid decision tree–Markov approach, 2 studies used a partitioned survival approach, and 1 study used a discrete event simulation approach. The time horizons ranged from 10 years to the patient's remaining lifetime, with cycle lengths ranging from 1 month to 1 year when reported. Most of the models that evaluated Oncotype DX assumed that the test could predict the benefit of chemotherapy.

None of the identified models included all of the intervention tests included in the scope of the assessment. Further details of the published models can be found starting on page 305 of the diagnostics assessment report.

Review and critique of economic analyses provided by companies

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. The EAG review and critique of these models is presented in the diagnostics assessment report starting on page 317.

Economic analysis

The EAG developed a de novo economic model designed to assess the cost effectiveness of Oncotype DX, MammaPrint, Prosigna, IHC4+C and EndoPredict compared with current practice. It is described in the diagnostics assessment report starting on page 346. The model assessed the health outcomes and costs associated with each test over a lifetime horizon (42 years) from the perspective of the UK NHS and Personal Social Services. All costs and health outcomes were discounted at a rate of 3.5% per annum. Unit costs were valued at 2015/16 prices. The principal source of evidence used to inform the analyses of Oncotype DX, Prosigna, IHC4+C and EndoPredict was the TransATAC study. ATAC was an international trial with a translational research continuation (TransATAC) that investigated the prognosis of breast cancer. Only UK data were included in the bespoke analysis provided to the EAG, which was also restricted to hormone receptor positive, HER2 negative patients with 0 to 3 positive lymph nodes. A comparison of the TransATAC data with other study data is provided in the diagnostics assessment report on page 303. As this study excluded MammaPrint, the MINDACT study was used as the basis for evaluating the cost effectiveness of MammaPrint. PREDICT scores were not available in either dataset, and so this tool could not be considered as a comparator or used to determine different risk subgroups. Therefore, the comparator for the analyses of Oncotype DX, Prosigna, IHC4+C and EndoPredict was current practice (various tools and algorithms),

and the comparator for the analysis of MammaPrint was a modified version of Adjuvant! Online.

Model structure

The de novo model was a hybrid decision tree–Markov model, and was based on the model previously developed by Ward et al. to inform diagnostics guidance 10. The decision tree component of the model classified patients in the current practice (no test) group and the tumour profiling test group into high, intermediate and low risk categories. For EndoPredict and MammaPrint, the intermediate risk category was excluded as the test provides results in terms of high and low risk only. Within both the test group and the current practice group, the decision tree determined the probability that a patient would be in 1 of 6 groups: low-risk, chemotherapy; low risk, no chemotherapy; intermediate risk, chemotherapy; intermediate risk, no chemotherapy; high risk, chemotherapy, and high risk, no chemotherapy (for the analyses of EndoPredict and MammaPrint, 4 branches were used due to the absence of an intermediate risk category). Each of the branches was then linked to a Markov model which predicted lifetime quality-adjusted life-years (QALYs) and costs according to the patient's risk of distant recurrence and whether or not they received chemotherapy. The decision tree structure is shown in Figure 1.

Figure 1: Decision tree

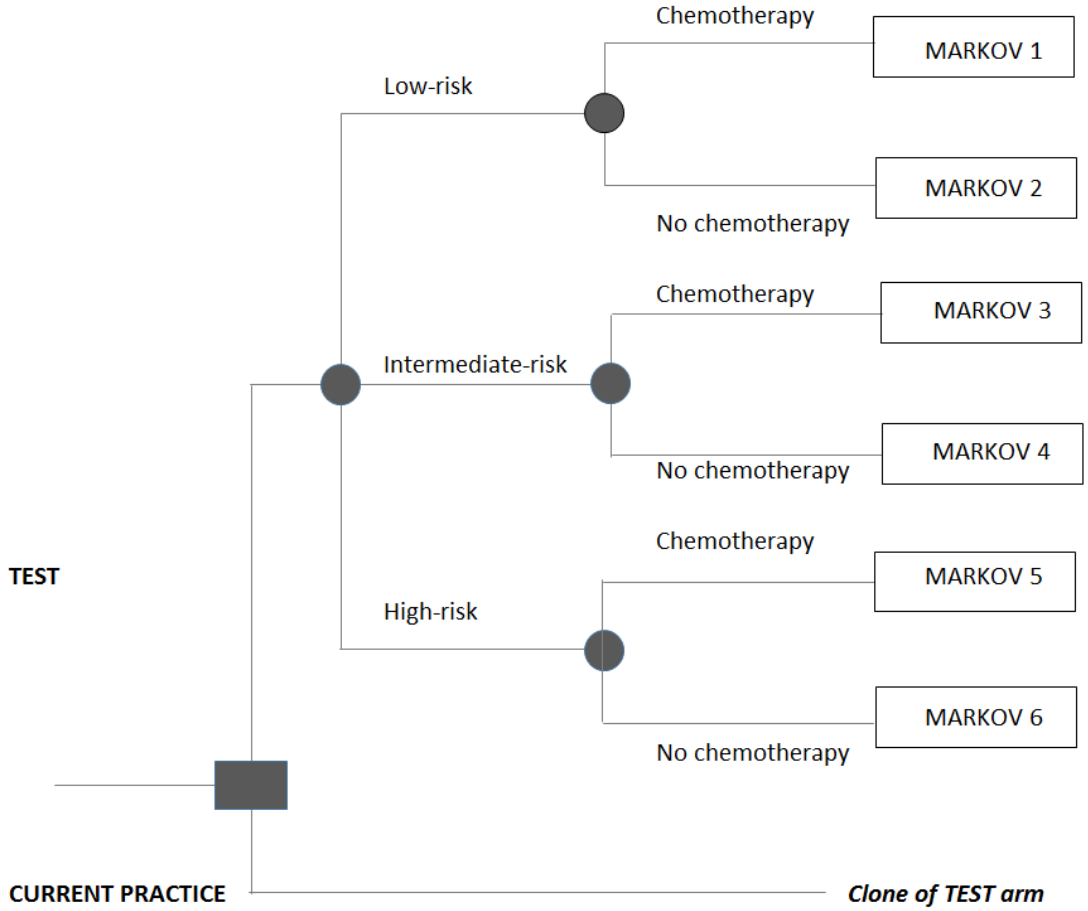
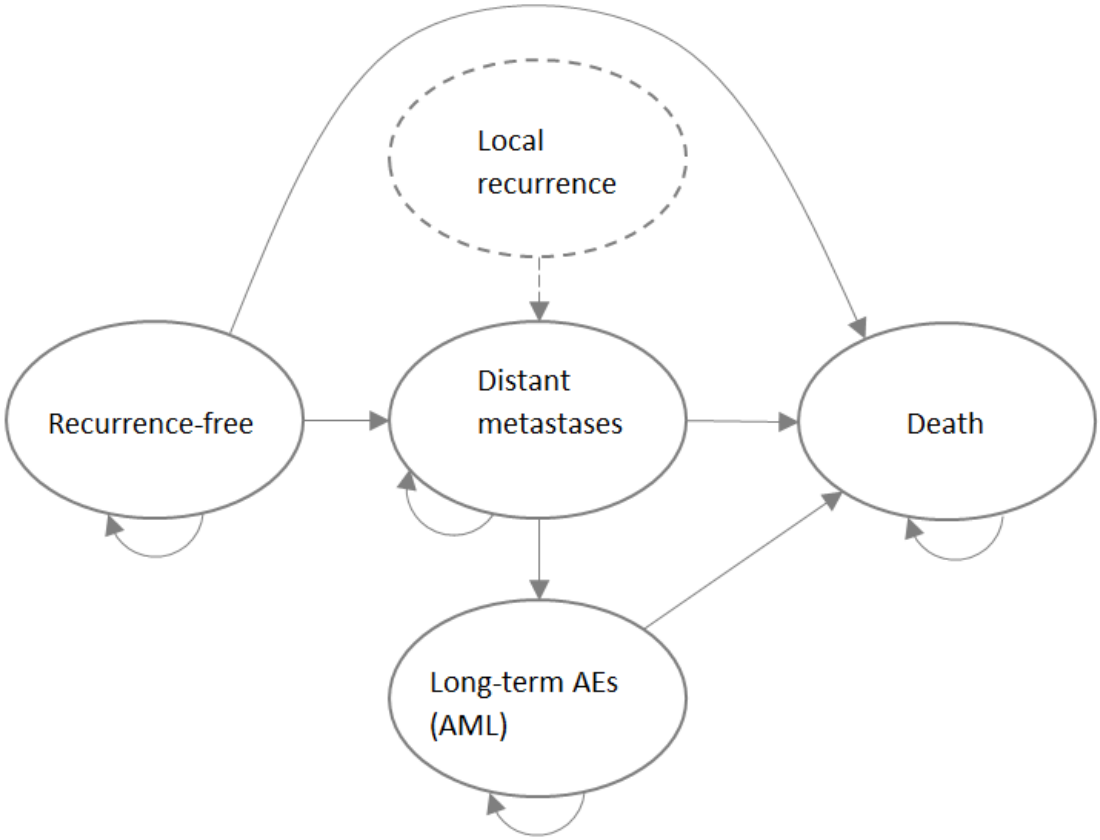


Figure 2 shows the Markov nodes of the model. Each Markov node was evaluated over 84 cycles of 6 months (42 years). Each node included 4 health states: recurrence-free; distant recurrence; long-term adverse events (acute myeloid leukaemia); and dead. Patients entered the model in the recurrence-free health state. A health-related quality of life decrement was applied during the first model cycle to account for health losses associated with short-term adverse events for patients receiving adjuvant chemotherapy. The benefit of adjuvant chemotherapy was modelled using a relative risk reduction for distant recurrence within each risk classification group. The benefit of the test was therefore captured in the model by changing the probability that patients with each test risk classification received adjuvant chemotherapy. A full

description of the model structure can be found starting on page 348 of the diagnostics assessment report.

Figure 2: Markov nodes



Abbreviations: AEs, adverse events; AML, Acute myeloid leukaemia

Model inputs

The model was populated with data from the clinical-effectiveness review, the NHS England access scheme dataset (provided as commercial-in-confidence by Genomic Health), the National Cancer Registration and Analysis Service (NCRAS), published literature and expert opinion. Full details of the model inputs can be found starting on page 352 of the diagnostics assessment report.

Risk classification probabilities

Risk classification probabilities for Oncotype DX, Prosigna, IHC4+C and EndoPredict were obtained from a bespoke data analysis of the TransATAC

trial (table 15). The analysis provided data on hormone receptor positive, HER2 negative patients for the 3 modelled subgroups (node-negative NPI≤3.4, node-negative NPI>3.4, and 1 to 3 positive nodes). Risk classification probabilities for MammaPrint were obtained from the MINDACT trial (table 16).

Table 15 Risk classification probabilities from TransATAC

Test (number samples)	Proportion of patients with risk classification		
	Low-risk	Intermediate-risk	High-risk
Node-negative, 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
EndoPredict (423)	0.90	-	0.10
Node-negative, 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
EndoPredict (254)	0.47	-	0.53
Node-positive (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
EndoPredict (198)	0.24	-	0.76
Abbreviations: NPI, Nottingham prognostic index			

Table 16 Risk classification probabilities from MINDACT

Population	Proportion of patients with risk classification	
	MammaPrint low-risk	MammaPrint high-risk
MINDACT overall population (n=6,693)	0.64	0.36
MINDACT Adjuvant! Online clinical high-risk subgroup (n=3,370)	0.46	0.54
MINDACT Adjuvant! Online clinical high-risk subgroup (n=3,324)	0.82	0.18

Probability of developing distant metastases

For Oncotype DX, Prosigna, IHC4+C and EndoPredict the probability of developing distant metastases in each cohort and risk category was based on 10-year recurrence-free interval data taken from the bespoke data analysis of the TransATAC trial (table 17). For MammaPrint the probability of developing distant metastases was based on an adjusted analysis of 5-year distant metastases-free survival data from Cardoso et al. 2016 (table 18). The model

assumed that the risk of distant metastases between 10 and 15 years was halved, and after 15 years was zero. This assumption was made because there is uncertainty about the sustained benefit of adjuvant chemotherapy.

Table 17 Ten year distant recurrence rates by risk classification

Population	Proportion distant recurrence-free at 10 years (95% CI)			
	Oncotype DX	Prosigna	IHC4+C	EndoPredict
LN negative, 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)
LN negative, 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
LN negative, 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)
LN negative, 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)
LN negative, 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
LN negative, 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 (0.688-0.838)
LN positive, low risk	0.818 (0.727-0.880)	1 (n/a)	0.961 (0.851-0.990)	0.95 (0.811-0.988)
LN positive, intermediate risk	0.754 (0.630-0.842)	0.807 (0.679-0.889)	0.758 (0.635-0.845)	n/a
LN positive, 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)
Abbreviations: LN, lymph node; NPI, Nottingham prognostic index				

Table 18 Calculated 10-year distant metastases-free survival by group for MammaPrint

Risk group	Treatment	10-year distant metastases-free survival probability for group[a]
Modified Adjuvant! Online low, MammaPrint low	Chemotherapy	0.953
	No chemotherapy	0.953
Modified Adjuvant! Online high, MammaPrint low	Chemotherapy	0.920
	No chemotherapy	0.891
Modified Adjuvant! Online low, MammaPrint high	Chemotherapy	0.918
	No chemotherapy	0.903
Modified Adjuvant! Online high, MammaPrint high	Chemotherapy	0.821
	No chemotherapy	0.766[b]
[a] Extrapolated from 5-year data assuming a constant event rate [b] Estimated by adjusting to remove the effect of chemotherapy		

Probability of having chemotherapy

The probability of having chemotherapy in the current practice group was taken either from data resulting from a bespoke request placed with NCRAS to obtain aggregate data on the use of adjuvant chemotherapy in women with early breast cancer in England (described on page 359 of the diagnostics assessment report), or from the NHS England Access dataset (relating only to node-negative disease and NPI>3.4; described on page 360 of the diagnostics assessment report). The NCRAS dataset reflects unselected patients who were not necessarily eligible for tumour profile testing; therefore the proportion of women who are eligible for testing who go on to receive chemotherapy may be greater than the estimates generated using this dataset.

In the groups for which tumour profiling tests were used, the probability of having chemotherapy was taken from either:

- The NHS England access scheme dataset: [redacted]
- Bloomfield et al. (2017): A UK study on the impact of EndoPredict results on adjuvant treatment decisions (149 patients). This is the only decision impact study on a 2-level tumour profiling test. It is unlikely to accurately represent the use of chemotherapy in node-positive disease.
- Loncaster et al. (2017): A prospective study to evaluate the clinical value of Oncotype DX testing in 201 patients who had been recommended chemotherapy. It provides separate estimates for node-negative and node-positive disease.
- Holt et al. (2011): A UK study on the impact of Oncotype DX on adjuvant treatment decisions with results available for 74 patients.
- UK breast cancer group (UKBCG) survey: a bespoke survey designed by the EAG, with 11 usable responses from oncologists. The results indicate considerable variation in practice.
- Expert opinion.

The probability of receiving adjuvant chemotherapy in the current practice group by test risk classification is presented in table 19. Where appropriate, the source not selected for inclusion in the base case was tested in the sensitivity analyses.

Table 19 Probability of receiving chemotherapy in the base case

Population	Source	Proportion of patients receiving chemotherapy		
		Low risk	Intermediate risk	High risk
Current practice group				
Node-negative, NPI≤3.4	NCRAS dataset	0.07		
Node-negative, NPI>3.4	NHS England access scheme dataset	0.43		
Node-positive (1-3 nodes)	NCRAS dataset	0.63		
Overall population (MammaPrint)	Expert opinion	0.46		
3-level tests (Oncotype DX, Prosigna and IHC4+C)				
Node negative, NPI≤3.4	UKBCG survey data	0.00	0.17	0.74
Node negative, NPI>3.4	NHS England access scheme dataset	0.01	0.33	0.89
Node-positive (1-3 nodes)	Loncaster et al. node-positive estimates	0.08	0.63	0.83
2-level tests (EndoPredict and MammaPrint)				
EndoPredict: All 3 subgroups	Bloomfield et al. (2017) study	0.07	-	0.77
MammaPrint: all subgroups	Bloomfield et al. (2017) study	0.07	-	0.77
Abbreviations: NCRAS, national cancer registration and analysis service; NPI, Nottingham prognostic index; UKBCG, UK breast cancer group				

Adjuvant chemotherapy treatment effect on distant recurrence

In the base-case analysis, the benefit of chemotherapy was assumed to be the same across all test risk groups, that is, all tests were assumed to be associated with prognostic benefit only. For Oncotype DX, Prosigna, IHC4+C and EndoPredict the relative risk of recurrence for chemotherapy versus no chemotherapy was based on a meta-analysis reported by the early breast cancer trialists' collaborative group (EBCTCG; 2012). A 10-year relative risk of distant recurrence of 0.76 was estimated, and was assumed to apply to the lymph node negative and positive groups. For MammaPrint the relative risk of distant recurrence for chemotherapy versus no chemotherapy was based on

data from the MINDACT trial on the discordant clinical and genomic risk groups. A 10-year relative risk of distant recurrence for the discordant populations was estimated to be 0.77. Sensitivity analyses explored the relative risks of distant recurrence in the modified Adjuvant! Online low- and high-risk subgroups, which were estimated to be 0.84 and 0.74, respectively.

In sensitivity analyses, the impact of assuming that Oncotype DX could predict the benefit of chemotherapy was explored, based on the studies reported by Paik et al. (2006) and Albain et al. (2010). In the node-negative group, the 10 year relative risks of relapse with chemotherapy versus no chemotherapy were 1.31, 0.61 and 0.26 for the low-, intermediate- and high-risk categories respectively. For the node-positive group, the 10-year relative risks of relapse with chemotherapy versus no chemotherapy were 1.02, 0.72 and 0.59 respectively.

Survival following onset of distant metastases

The survival prognosis of patients with distant metastases was based on an analysis of 77 women randomly selected from 232 women who had relapsed breast cancer between 2000 and 2005 (Thomas et al. 2009). Median survival was 40.1 months following distant recurrence. From this, the 6-month probability of death following distant recurrence was estimated to be 0.098, assuming a constant rate. The rate of death due to distant metastases was assumed to be the same across the different subgroups and across each test risk group.

The probability of local recurrence, developing acute myeloid leukaemia and survival thereafter

The model assumes that 10.5% of patients entering the distant recurrence health state had previously had a local recurrence, based on de Bock et al. (2009). The 6-month probability of developing acute myeloid leukaemia (AML) was estimated to be 0.00025, based on Wolff et al. (2015). Survival following the onset of AML was estimated to be approximately 8 months; assuming a constant event rate, this gave a 6-month probability of death following AML of 0.53. This was taken from the NICE technology appraisal guidance on

azacitidine for the treatment of myelodysplastic syndromes, chronic myelomonocytic leukaemia and acute myeloid leukaemia.

Costs

Costs and resource use inputs are described starting on page 371 of the diagnostics assessment report.

Test costs

The costs of the tumour profiling tests were based on prices submitted by companies, as shown in table 20.

Table 20 Test costs used in the model

Test	Cost	Comments
Oncotype DX	£2,580	Tests carried out in Genomic Health laboratory in US. Cost includes sample handling and customer service. A commercial-in-confidence discounted test cost was used in the model.
Prosigna	£1,970	Based on conducting the test in an NHS laboratory, which includes the laboratory costs (£240), the Prosigna kit (£1,650) and the nCounter System (£194,600)
EndoPredict	£1,500	Tests carried out in Myriad's laboratory in Munich
IHC4	£203	The cost was submitted using 2014 prices. The total cost of the test (£198) was uplifted using the HCHS indices to current prices.
MammaPrint	£2,326	Converted from Euros to UK Pounds Sterling assuming exchange rate of 1 British Pound to 1.15 Euro.
Abbreviations: HCHS, hospital and community health services		

Costs of adjuvant chemotherapy acquisition and administration

The costs associated with adjuvant chemotherapy were obtained from a previous costing analysis of the OPTIMA Prelim trial (Hall et al. 2017). The EAG model assumed that women with ER+, HER2-, early breast cancer with 0 to 3 nodes typically receive 1 of 4 adjuvant chemotherapy regimens:

- FEC100-T (fluorouracil, epirubicin, cyclophosphamide and docetaxel; 3+3 cycles; assumed to be given to 25% of patients)
- TC (docetaxel and cyclophosphamide; 4 cycles; assumed to be given to 20% of patients)
- FEC75 (fluorouracil, epirubicin and cyclophosphamide; 6 cycles; assumed to be given to 45% of patients)

- FEC100-Pw (fluorouracil, epirubicin, cyclophosphamide and weekly paclitaxel; 3+3 cycles; assumed to be given to 10% of patients)

The weighted mean cost of adjuvant chemotherapy acquisition, delivery and toxicity was estimated to be £3,145 per course.

Costs of endocrine therapy

The model assumed that all surviving patients received endocrine therapy for a period of between 5 and 8 years and may have received 1 of 4 endocrine therapy regimens:

- tamoxifen for 5 years (40% of patients, annual cost £35.06)
- anastrozole for 5 years (20% of patients, annual cost £14.09)
- letrozole for 5 years (20% of patients, 10% of patients were assumed to receive extended letrozole for 3 further years, annual cost £32.87)
- tamoxifen for 2 years then exemestane for 3 years (20% of patients, annual cost of exemestane £69.52).

Costs of additional treatments

The model assumed that 30% of women with early breast cancer would receive 4 milligrams of bisphosphonates (zoledronic acid) every 6 months by intravenous infusion for up to 3 years (cost per 36-month course = £58.50).

Follow-up costs

The model assumed that all patients received 2 routine follow-up visits during the first year after surgery, with annual visits thereafter for 5 years. Patients were also assumed to have a routine annual mammogram for up to 5 years. The cost of a routine follow-up visit was estimated to be £162.84, and the cost of a mammogram was estimated to be £46.37.

Costs of treatments for local and distant recurrence

Costs associated with treating local recurrence were taken from a UK-based costing analysis (Karnon et al. 2007) and uplifted to current prices using the HCHS (Hospital and Community Health Service) index (£13,913). This was applied as a once-only cost upon the incidence of distant recurrence.

Costs associated with treating distant metastases were derived from Thomas et al. (2009), and included 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 £4,541.

Health-related quality of life and QALY decrements

Health utilities were taken from various published studies as shown in table 21. The studies are described in detail starting on page 367 of the diagnostics assessment report.

Table 21 Health utilities applied in the model

Health state / event	Duration applied in model	Mean	Standard error	Source
Recurrence-free	Indefinite	0.824	0.002	Lidgren et al.
Distant metastases	Indefinite	0.685	0.004	
Disutility distant metastases	Indefinite	-0.14	0.11	Calculated using difference method
Local recurrence	Once-only QALY loss applied on transition to distant recurrence state	-0.108	0.04 (assumed)	Campbell et al.
Chemotherapy AEs	6 months	-0.038	0.004	Campbell et al.
AML	Indefinite	0.26	0.04 (assumed)	Younis et al.
Abbreviations: AEs, adverse events; AML, acute myeloid leukaemia; QALY, quality-adjusted life year				

Base-case results

For the purposes of decision-making, the incremental cost-effectiveness ratios (ICERs) per QALY gained or lost are considered. The following assumptions were applied in the base-case analysis:

- The proportion of patients who received chemotherapy under current practice (no test) was assumed to be the same for each test risk classification (low, intermediate, and high risk). This proportion was however assumed to differ between subgroups defined according to clinical risk (LN0 NPI≤3.4, LN0 NPI>3.4, LN1-3, MINDACT overall population, MINDACT modified Adjuvant! Online low risk, and MINDACT modified Adjuvant! Online high risk).

- Clinicians interpret each of the 3-level tests in the same way (for example, an Oncotype DX high-risk score would lead to the same chemotherapy decision as a Prosigna high-risk score).
- Clinicians interpret each of the 2-level tests in the same way (for example, a MammaPrint high-risk score would lead to the same chemotherapy decision as an EndoPredict high-risk score).
- The benefit of adjuvant chemotherapy was the same across all risk score categories for all tests.
- A proportion of patients (10.5%) who developed distant metastases had previously developed local recurrence (QALY losses and costs associated with local recurrence were applied once only).
- The prognosis of patients with AML and the costs and QALYs accrued within the AML state were independent of whether the patient had previously developed distant metastases.
- A disutility associated with adjuvant chemotherapy was applied once during the first model cycle only (while the patient is receiving the regimen).
- Costs associated with endocrine therapy, bisphosphonates, follow-up appointments and mammograms were assumed to differ according to time since model entry.
- Across all 3 analysis subgroups, patients entered the model aged 58 years, based on the mean age of patients in the NHS England Access dataset.
- The model included both pre- and postmenopausal women; however, the TransATAC study related only to postmenopausal women.

The results of the model are presented in the diagnostics assessment report starting on page 379, and are summarised below. In addition, the modelled chemotherapy use with and without the tumour profiling tests is presented in appendix 7 in the diagnostics assessment report (page 502). All estimates presented here are based on the probabilistic version of the model.

In the node-negative population, in the subgroup with an NPI of 3.4 or less, for tumour profiling tests compared with current practice the model gave ICERs of £147,419 per QALY gained (EndoPredict), £122,725 per QALY gained

(Oncotype DX), £91,028 per QALY gained (Prosigna) and £2,654 per QALY gained (IHC4+C).

In the node-negative population, in the subgroup with an NPI greater than 3.4, for tumour profiling tests compared with current practice the model gave ICERs of £46,788 per QALY gained (EndoPredict) and £26,058 per QALY gained (Prosigna). Oncotype DX was dominated by current practice (that is, Oncotype DX was more expensive and less effective) and ICH4+C was dominant over current practice (that is, ICH4+C was less expensive and more effective).

In the node-positive population, the tumour profiling tests compared with current practice (NPI) had ICERs of £28,731 per QALY gained (Prosigna) and £21,458 per QALY gained (EndoPredict). Oncotype DX was dominated by current practice and ICH4+C was dominant over current practice.

In the overall MINDACT population, MammaPrint compared with current practice (modified Adjuvant! Online) had an ICER of £131,482 per QALY gained. In the modified Adjuvant! Online high-risk subgroup, MammaPrint was dominated by current practice, and in the modified Adjuvant! Online low-risk subgroup, MammaPrint compared with current practice had an ICER of £414,202 per QALY gained.

QALYs, costs and ICERs for each test compared with current practice are presented in tables 22 to 26.

Table 22 Probabilistic ICERs for Oncotype DX compared with current practice

Option	QALYs	Costs	Inc. QALYs	Inc. costs	ICER (per QALY gained)
Node-negative NPI≤3.4					
Oncotype DX	13.89	£5,474	0.01	£1,313	£122,725
No test	13.88	£4,161	-	-	-
Node-negative NPI>3.4					
Oncotype DX	12.73	£11,806	-0.01	£881	Dominated
No test	12.74	£10,925	-	-	-
Node-positive (1-3 nodes)					
Oncotype DX	12.48	£13,212	-0.07	£687	Dominated

No test	12.55	£12,525	-	-	-
Abbreviations: ICER, incremental cost-effectiveness ratio; NPI, Nottingham prognostic index; QALY, quality adjusted life year					

Table 23 Probabilistic ICERs for IHC4+C compared with current practice

Option	QALYs	Costs	Inc. QALYs	Inc. costs	ICER (per QALY gained)
Node-negative NPI≤3.4					
IHC4+C	13.86	£4,291	0.01	£22	£2,654
No test	13.86	£4,269	-	-	-
Node-negative NPI>3.4					
IHC4+C	12.73	£10,941	0.01	-£90	Dominating
No test	12.72	£11,031	-	-	-
Node-positive (1-3 nodes)					
IHC4+C	12.59	£12,268	0.05	-£287	Dominating
No test	12.54	£12,554	-	-	-
Abbreviations: ICER, incremental cost-effectiveness ratio; NPI, Nottingham prognostic index; QALY, quality adjusted life year					

Table 24 Probabilistic ICERs for Prosigna compared with current practice

Option	QALYs	Costs	Inc. QALYs	Inc. costs	ICER (per QALY gained)
Node-negative NPI≤3.4					
Prosigna	13.87	£6,201	0.02	£1,884	£91,028
No test	13.84	£4,318	-	-	-
Node-negative NPI>3.4					
Prosigna	12.65	£13,330	0.06	£1,686	£26,058
No test	12.59	£11,644	-	-	-
Node positive (1-3 nodes)					
Prosigna	12.47	£15,172	0.07	£1,936	£28,731
No test	12.40	£13,236	-	-	-
Abbreviations: ICER, incremental cost-effectiveness ratio; NPI, Nottingham prognostic index; QALY, quality adjusted life year					

Table 25 Probabilistic ICERs for EndoPredict compared with current practice

Option	QALYs	Costs	Inc. QALYs	Inc. costs	ICER (per QALY gained)
Node-negative NPI≤3.4					
EndoPredict	13.85	£6,034	0.01	£1,679	£147,419
No test	13.84	£4,355	-	-	-
Node-negative NPI>3.4					
EndoPredict	12.71	£12,612	0.03	£1,388	£46,788
No test	12.68	£11,224	-	-	-
Node-positive (1-3 nodes)					
EndoPredict	12.52	£14,080	0.05	£1,164	£21,458

No test	12.46	£12,916	-	-	-
Abbreviations: ICER, incremental cost-effectiveness ratio; NPI, Nottingham prognostic index; QALY, quality adjusted life year					

Table 26 Probabilistic ICERs for MammaPrint compared with current practice

Option	QALYs	Costs	Inc. QALYs	Inc. costs	ICER (per QALY gained)
Overall MINDACT population					
MammaPrint	13.51	£9,151	0.01	£1,760	£131,482
No test	13.49	£7,391	-	-	-
MINDACT modified Adjuvant! Online high-risk group					
MammaPrint	12.86	£12,727	-0.04	£1,413	Dominated
No test	12.90	£11,313	-	-	-
MINDACT modified Adjuvant! Online low-risk group					
MammaPrint	13.70	£7,777	0.01	£2,410	£414,202
No test	13.69	£5,366	-	-	-
Abbreviations: ICER, incremental cost-effectiveness ratio; NPI, Nottingham prognostic index; QALY, quality adjusted life year					

Probabilistic sensitivity analysis

Probabilistic sensitivity analyses of pairwise (test compared with current practice) results (table 27) indicated that:

- In the lymph node negative, NPI of 3.4 or less subgroup, the only test with a non-zero probability of producing more net benefit compared with current practice at maximum acceptable ICERs of £20,000 and £30,000 per QALY gained was IHC4+C.
- In the lymph node negative, NPI of greater than 3.4 subgroup, at a maximum acceptable ICER of £20,000 per QALY gained, IHC4+C had a probability of 0.69 of being the most cost-effective option. All other tests had less than 0.24 probability of bring more cost effective than current practice. In the same subgroup, at a maximum acceptable ICER of £30,000 per QALY gained, IHC4+C had a probability of 0.67 of being the most cost-effective option and Prosigna had a probability of 0.60 of being the most cost-effective option. Oncotype DX had a probability of 0.04 and EndoPredict had a probability of 0.26 of producing more net benefit compared with current practice.

- In the lymph node positive subgroup, IHC4+C had probabilities of 0.95 and 0.94 of producing more net benefit compared with current practice at maximum acceptable ICERs of £20,000 and £30,000 respectively. In the same subgroup at the same maximum acceptable ICERs, the probability of EndoPredict producing more net benefit than current practice ranged from 0.44 to 0.73, and for Prosigna the range was 0.24 to 0.55. In this subgroup Oncotype DX had very low probabilities of producing more net benefit than current practice at the same maximum acceptable ICERs (0.01 or lower).
- In the overall MINDACT population and in the subgroups, the probability that MammaPrint would produce more net benefit than current practice at maximum acceptable ICERs of £20,000 and £30,000 per QALY gained was approximately zero.

Table 27: Probabilities of tests being cost effective

Test	Subgroup	Probability of being cost effective compared with current practice	
		At maximum acceptable ICER of £20,000	At maximum acceptable ICER of £30,000
Oncotype DX	LN0 NPI≤3.4	0.00	0.00
	LN0 NPI>3.4	0.01	0.04
	LN+ (1-3 nodes)	0.00	0.01
IHC4+C	LN0 NPI≤3.4	0.95	0.97
	LN0 NPI>3.4	0.69	0.67
	LN+ (1-3 nodes)	0.95	0.94
Prosigna	LN0 NPI≤3.4	0.00	0.00
	LN0 NPI>3.4	0.24	0.60
	LN+ (1-3 nodes)	0.24	0.55
EndoPredict	LN0 NPI≤3.4	0.00	0.00
	LN0 NPI>3.4	0.09	0.26
	LN+ (1-3 nodes)	0.44	0.73
MammaPrint	MINACT overall population	0.00	0.00
	Modified AO high risk	0.00	0.00
	Modified AO low risk	0.00	0.00

Abbreviations: AO, Adjuvant! Online; ICER, incremental cost-effectiveness ratio; LN, lymph node; NPI, Nottingham prognostic index

Deterministic sensitivity analyses

The EAG did several deterministic sensitivity analyses for each test. Full details are on page 377 of the diagnostics assessment report. Results for Oncotype DX were:

- Node-negative with NPI≤3.4: ICERs for Oncotype DX compared with current practice remained over £34,000 per QALY gained across all analyses.
- Node-negative with NPI>3.4: Oncotype DX was either dominated or had an ICER of more than £35,000 per QALY gained across almost all analyses. The only exception was when Oncotype DX was assumed to predict chemotherapy benefit. Within this analysis, Oncotype DX dominated current practice.
- Node-positive (1 to 3 nodes): Oncotype DX remained dominated across the majority of analyses. The exceptions were: when Oncotype DX was assumed to predict chemotherapy benefit (was dominant), and when the cost of chemotherapy was doubled (£3,700 saved per QALY lost).

Deterministic sensitivity analysis results for IHC4+C:

- Node-negative with NPI≤3.4: ICERs for IHC4+C compared with current practice remained below £16,000 per QALY gained across all analyses, except when post-test chemotherapy probabilities were derived from Holt et al. 2011 (£36,259 per QALY gained); in addition, IHC4+C dominated current practice when the cost of chemotherapy was doubled.
- Node-negative with NPI>3.4: IHC4+C dominated current practice or had an ICER below £6,000 per QALY gained across all scenarios.
- Node-positive (1 to 3 nodes): IHC4+C dominated current practice across all but 1 scenario. When the probability of receiving chemotherapy was based on the UKBCG survey, the ICER was estimated to be £1,929 per QALY gained.

Deterministic sensitivity analysis results for Prosigna:

- Node-negative with NPI≤3.4: ICERs for Prosigna compared with current practice were greater than £71,000 per QALY gained across all analyses.
- Node-negative with NPI>3.4: ICERs for Prosigna compared with current practice were below £34,000 per QALY gained across all analyses.
- Node-positive (1 to 3 nodes): ICERs for Prosigna compared with current practice were below £38,000 per QALY gained across all analyses.

Deterministic sensitivity analysis results for EndoPredict:

- Node-negative with NPI≤3.4: ICERs for EndoPredict compared with current practice remained greater than £91,000 per QALY gained across all analyses.
- Node-negative with NPI>3.4: ICERs for EndoPredict compared with current practice remained greater than £30,000 per QALY gained across all but 2 of the analyses. Exceptions were: when the UKBCG survey was used to inform the probability of receiving chemotherapy (£25,250 per QALY gained), and when Cusumano et al. (2014) was used to inform the probability of receiving chemotherapy conditional on the EndoPredict test result (£26,689 per QALY gained).
- Node-positive (1 to 3 nodes): ICERs for EndoPredict compared with current practice remained below £30,000 per QALY gained across all scenarios.

Deterministic sensitivity analysis results for MammaPrint:

- In the overall MINDACT population, the ICER for MammaPrint compared with current practice was estimated to be greater than £76,000 per QALY gained across all scenarios.
- In the modified Adjuvant! Online high-risk subgroup, MammaPrint was dominated by current practice across almost all scenarios.
- In the modified Adjuvant! Online low-risk subgroup, the ICER for MammaPrint compared with current practice was greater than £161,000 per QALY gained across all analyses.

Comparison between the new EAG model and the EAG model used for diagnostics guidance 10

The differences between the models are described in the diagnostics assessment report starting on page 398. The new EAG model suggested that in the lymph node negative with NPI>3.4 subgroup, Oncotype DX was dominated by current practice. In the same subgroup, the previous EAG model produced a base-case ICER for Oncotype DX compared with current practice of £22,600 per QALY gained. This ICER was also based on Oncotype DX offered at a confidential price through a patient access scheme.

The models had a similar general modelling approach. In both models, data on risk reclassification and risk of distant recurrence in the absence of chemotherapy were taken from analyses of the ATAC trial, although different datasets were used. The proportions of women who were assumed to receive chemotherapy conditional on the Oncotype DX risk score were taken from the NHS England access scheme dataset in the current EAG model, but the previous EAG model used unpublished data (Holt et al. 2013) to estimate this. In addition, the proportion of patients receiving chemotherapy in the standard care arm was taken from the NHS England access scheme dataset in the current EAG model, but was taken from English cancer registry datasets in the previous model.

When both models used pre- and post-test chemotherapy probabilities from the NHS England access scheme dataset and no predictive benefit is assumed, both models produce the same economic conclusion: Oncotype DX is dominated by current practice.

3 Summary

Clinical effectiveness

Among studies of lymph node negative patients receiving endocrine monotherapy, percentages of patients categorised as high risk ranged from 9 to 33% across all 5 tests. In studies of patients receiving endocrine monotherapy, 3 tests (Prosigna, EndoPredict and IHC4+C) categorised more

lymph node positive patients as high risk than lymph node negative patients. Oncotype DX, however, categorised similar numbers of patients as high risk in lymph node negative and positive groups. Oncotype DX also categorised more lymph node positive patients as low risk than other tests, but led to worse 10-year distant recurrence free survival/interval outcomes in this group compared with other tests.

All tests had statistically significant prognostic power in unadjusted analyses in lymph node negative and lymph node positive populations. All tests provided additional prognostic information over most commonly used clinicopathological factors and over clinical treatment score and Nottingham Prognostic Index (NPI) in lymph node negative patients. Results were more varied in lymph node positive patients.

There was some evidence of differential chemotherapy benefit between risk groups assessed by Oncotype DX, shown by significant interaction tests between risk group and chemotherapy treatment in unadjusted analyses. However, the interaction test results sometimes became non-significant when clinicopathological factors were adjusted for. Evidence on the ability of MammaPrint to predict benefit from chemotherapy was extremely limited, but suggested no statistically significant difference in effect of chemotherapy between risk groups. Evidence of differential chemotherapy benefit was not available for the other 3 tests.

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.

Decision impact studies 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 (Oncotype DX, EndoPredict and IHC4+C) and from 5% to 70% across European studies (all tests except

IHC4). Across all tests, 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% in European studies

Cost effectiveness

The EAG developed a de novo health economic model to assess the cost effectiveness of Oncotype DX, MammaPrint, Prosigna, EndoPredict and IHC4+C, each versus current practice. The base-case model suggested the following results:

Oncotype DX: In the lymph node negative NPI≤3.4 group, the incremental cost-effectiveness ratio (ICER) for Oncotype DX compared with current practice was estimated to be £122,725 per quality-adjusted life year (QALY) gained (£34,245 per QALY gained assuming prediction of chemotherapy benefit). In the lymph node negative NPI>3.4 and lymph node positive groups, Oncotype DX was dominated by current practice (but Oncotype DX dominated current practice if prediction of chemotherapy benefit was assumed). The results were primarily driven by the modelled reduction in the use of adjuvant chemotherapy using the Oncotype DX test because it categorised more lymph node positive patients as low risk than other tests, but this led to worse 10-year distant recurrence free survival/interval outcomes.

IHC4+C: In the lymph node negative NPI≤3.4 group, the ICER for IHC4+C compared with current practice was estimated at £2,654 per QALY gained. In the lymph node negative NPI>3.4 and lymph node positive groups, IHC4+C dominated current practice.

Prosigna: In the lymph node negative NPI≤3.4 group, the ICER for Prosigna compared with current practice was estimated to be £91,028 per QALY gained. In the lymph node negative NPI>3.4 and lymph node positive groups, the ICERs for Prosigna compared with current practice were estimated to be £26,058 and £28,731 per QALY gained, respectively.

EndoPredict: In the lymph node negative NPI≤3.4 group, the ICER for EndoPredict compared with current practice was estimated to be £147,419 per QALY gained. In the lymph node negative NPI>3.4 and lymph node positive groups, the ICERs for EndoPredict compared with current practice were estimated to be £46,788 and £21,458 per QALY gained, respectively.

MammaPrint: In the overall MINDACT population, the ICER for MammaPrint compared with current practice was estimated to be £131,482 per QALY gained. In the modified Adjuvant! Online high-risk group, MammaPrint was expected to be dominated by current practice. In the modified Adjuvant! Online low-risk group, the ICER for MammaPrint compared with current practice was estimated to be £414,202 per QALY gained.

4 Issues for consideration

Clinical effectiveness

Many of the included studies were retrospective analyses of randomised controlled trials (RCTs) or observational data sets which used stored tumour samples. Nearly all of these studies excluded patients who did not have a large enough tissue sample for testing, 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) are likely to be under-represented. However, this issue is unavoidable in retrospective analyses.

The IHC4/IHC4+C evidence base was 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 where fixed cut-offs are likely to be more practicable. In addition, there are known problems with conducting the analyses required for IHC4, in particular the reliability and reproducibility of the Ki-67 marker measurement.

Prosigna, EndoPredict and IHC4+C categorise more lymph node positive patients than lymph node negative patients as high risk. However, Oncotype

DX categorised similar percentages of lymph node positive and lymph node negative patients as high risk.

In terms of the prognostic ability of the tumour profiling tests, much of the evidence base was results from unadjusted analyses, which did not assess whether a test had additional value over clinicopathological factors. In adjusted analyses, the clinicopathological variables included were not consistent. Further, the retrospective observational studies reporting evidence on prognostic ability were at risk of confounding and spectrum bias, which can affect estimates of prognostic performance. This is because chemotherapy rates may differ by risk group, and if patients who received chemotherapy were excluded, these patients would be likely to be systematically different to those who did not. These problems were particularly relevant to the MammaPrint evidence base, as most studies were observational in nature rather than re-analyses of RCTs.

There were relatively limited data relating to the ability of Oncotype DX and MammaPrint to predict benefit from chemotherapy and on the ability of the tests to affect patient outcomes. These types of evidence were not available for the other 3 tests.

Concordance between tests was not fully reviewed, but 1 UK study (OPTIMA prelim) which compared Oncotype DX, MammaPrint, Prosigna and IHC4 concluded that although the 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.

No data were available for men, who do account for a proportion of breast cancer cases seen in practice. It is not certain whether the prognostic and clinical-effectiveness data are applicable to men.

Cost effectiveness

The EAG model is subject to a number of uncertainties and limitations.

With the exception of Oncotype DX in the lymph node negative NPI>3.4 group, the evidence surrounding the pre- and post-test chemotherapy probabilities is subject to considerable uncertainty. The model results are sensitive to the assumptions made about pre- and post-test chemotherapy use. The inclusion of data collected through the NHS England access scheme dataset has a significant impact on the model results for Oncotype DX compared with the model results from the original assessment for diagnostics guidance 10 (in the lymph node negative NPI>3.4 subgroup, Oncotype DX was dominated by current practice in the current model, but had an incremental cost-effectiveness ration (ICER) of £22,600 per QALY gained in the previous EAG model). Further, there is only 1 UK-based decision impact study relating to a 2-level tumour profiling test (Bloomfield et al. 2017). Sensitivity analyses showed that alternative estimates of post-test chemotherapy use generally led to more favourable cost-effectiveness estimates for EndoPredict and MammaPrint.

The comparator in the model is defined as a modified version of Adjuvant! Online for the MammaPrint analyses, and as current practice for the other 4 tests. In clinical practice in England other tools may be used to define risk, such as the PREDICT algorithm. It was not possible to do a comparison with PREDICT, or to define clinical risk groups by PREDICT because data were not available from the TransATAC trial, the NCRAS data set or the MINDACT trial. The cost effectiveness of the tumour tests compared with current NHS practice is therefore highly uncertain.

There is uncertainty about whether Oncotype DX can predict chemotherapy benefit. The inclusion of this potential test characteristic in the model has a substantial impact on the results. When a predictive benefit was included, Oncotype DX dominated current practice in both the lymph node negative NPI>3.4 and lymph node positive (1 to 3 nodes) groups, and had an ICER of £34,245 per QALY gained in the lymph node negative NPI≤3.4 group.

The analysis of MammaPrint was based on a different data source from the other 4 tests. In addition, the MINDACT trial, which was used to inform the analysis of MammaPrint, had a follow-up period limited to 5 years.

The test cost for Prosigna was based on an efficient level of throughput. This may not hold if centres do not undertake the anticipated number of tests.

There is the potential for the prognostic performance of IHC4+C to have been overestimated. This is because the TransATAC trial was the derivation study for IHC4 and it is not certain how generalisable the prognostic model fitted from this dataset is.

The test risk classification probabilities and distant metastases-free survival probabilities for Oncotype DX, Prosigna, IHC4+C and EndoPredict were based on a postmenopausal population only (TransATAC). It is expected that the tumour profiling tests will also be used in premenopausal women.

5 Equality considerations

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others.

Breast cancer can occur in men, and it is often underdiagnosed and undertreated in this group. No data were identified for male only cohorts.

Women of African family origin are more likely to develop breast cancer at an earlier age and to have a more aggressive form of the disease compared with other women. Data relating to people of different ethnicities were difficult to interpret because of differences in treatment practices in different countries.

6 Implementation

NanoString does not offer a centralised testing service for Prosigna, so a local testing service would need to be established.

Standardisation and quality assurance programmes would be required before IHC4 could be used routinely in the NHS.

7 Authors

Frances Nixon

Topic Lead

Rebecca Albrow

Technical Adviser

November 2017

Appendix A: Sources of evidence considered in the preparation of the overview

A. The diagnostics assessment report for this assessment was prepared by the School of Health and Related Research (ScHARR), The University of Sheffield:

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

B. The following organisations accepted the invitation to participate in this assessment as stakeholders. They were invited to attend the scoping workshop and to comment on the diagnostics assessment report.

Manufacturers of technologies included in the final scope:

- Agendia NV
- Genomic Health UK
- Myriad Genetics
- Nanostring Technologies
- Royal Marsden Hospital Trust

Other commercial organisations:

- Decision Resources Group, Abacus
- Oncomark
- Roche Diagnostics

Professional groups and patient/carer groups:

- Association of Breast Surgery
- Breast Cancer Now
- The Royal College of Physicians
- The Royal College of Radiologists

Research groups:

- Cancer Research UK

Associated guideline groups:

None

Others:

- Colchester Hospital NHS Foundation Trust
- Department of Health
- Greater Manchester Cancer / NHS Trafford clinical commissioning group
- Healthcare Improvement Scotland
- Medicines and Healthcare products Regulatory Agency
- NHS England
- Peony Breast Care Unit
- The London Breast Clinic
- Welsh Government

Appendix B: Glossary of terms

Adjuvant therapy
Additional cancer treatment given after primary treatment to lower the risk that cancer will come back. Adjuvant therapy may include chemotherapy, radiotherapy, hormone therapy or biological therapy.

Distant recurrence
Cancer that comes back in a different area to the original cancer after initial treatment.

Hormone (endocrine) therapy
Hormones such as oestrogen and progesterone can fuel the growth of breast cancer. Hormone therapies, such as tamoxifen and aromatase inhibitors, aim to block the availability of hormones such as oestrogen and progesterone and prevent the cancer growing.

Local recurrence
Cancer that comes back in the same place as the original cancer after initial treatment.

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

A rapid evidence review of the analytical validity of IHC4: ADDENDUM to "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 27th October

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

Background

IHC4 relies on the quantification of the immunohistochemistry (IHC) markers oestrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and Ki67 for each patient. Whilst a widely adopted technique, IHC can be criticised for a lack of stringency,[1 2] which in turn can lead to problems with reproducibility between laboratories. Problems with IHC that can lead to variations in quantitative values produced include:

- Pre-analytical methods (e.g. sample type, fixation, storage)
- Analytical methods (e.g. antibodies, staining techniques and reagents) and
- Interpretation (e.g. manual versus automated scoring, using whole slides versus using hot spots or heterogeneous areas, edge areas versus central areas).

The authors of the IHC4 derivation study[3] note that the use of the IHC4 score in laboratories beyond their own (Royal Marsden Hospital) would raise concerns relating to the reproducibility of the component IHC assays.[3] This summary aims to highlight the main issues relating to the use of IHC4 in laboratories other than the Royal Marsden Hospital laboratory (where the score originated) and the recent work that attempts to address some of these concerns.

Methods

It was not possible, within the time-frame of the review, to conduct a full systematic review of the analytical validity of all components of the IHC4 (namely ER, PR, HER2 and Ki67). Instead, we have conducted a rapid review, using systematic search and snowballing search techniques, to identify the most recent and most relevant literature. We have focussed on studies which consider the analytical validity of the IHC4 test, and on studies which consider the analytical validity of Ki67, as this is the most problematic of the four components.[4]

In order to select the most relevant and recent literature we created a long list of potentially relevant studies and then selected the most relevant literature from this, in three stages:

1) Studies from the following sources:

- The main search (primary or secondary studies, including expert reviews). The search was designed to identify studies relating to the analytical validity of IHC4, but not to the component elements (ER, PR, HER2 and Ki67)
- The reference lists of studies included in the prognostic review of IHC4[3 5-17]
- The reference lists of studies included or cited in existing systematic or expert reviews[18-21]
- Suggestions from clinical experts

2) Identified key studies and conducted citation searches of these within Google Scholar, and added relevant citations to the long list created in step 1. Where the number of citations for a single study was in excess of 100 studies, these were limited (using the Google Scholar "search within citing articles" facility) to those containing the words "analytical validity". The key studies selected for citation searching were:

- Dowsett 2011[22]: International Ki67 in Breast Cancer Working Group recommendations
- Dodson 2016[4]: IHC4 analytical validity study.
- Engelberger 2015[23]: "Score the Core" development study. This was chosen as it relates directly to attempts to improve IHC4 analytical validity
- Polley 2013; Polley 2015; Leung 2016:[24-26] Ki67 analytical validity studies resulting from the International Ki67 in Breast Cancer Working Group[22]. These were chosen as they are recent developmental studies relating to Ki67.

3) Selected the most relevant studies to include in this summary. These were chosen considering the following factors:

- Inter-laboratory reproducibility of IHC4 or Ki67 compared to the Royal Marsden, as this is the centre where the IHC4 score was generated
- Inter-rater reliability of IHC4 or Ki67

As there were no systematic reviews on the analytical validity of IHC4, recent expert reviews and the discussion points raised in the IHC4 prognostic literature[3 5-16 27] were consulted to ensure all points of interest were covered.

Summary of findings

A total of 308 titles were screened for relevance. No systematic review relating to the analytical validity of IHC4 or its components was identified. Eight studies (one Working Group report[22] and 7 primary studies[4 23-26 28-31]) were included (Table 1). These are broadly split into:

1. **Analytical validity of IHC4 between Royal Marsden and external centres**
2. **Analytical validity of IHC4 within other centres**
3. **Analytical validity of Ki67: Studies related to Ki67 Working Group and Royal Marsden**

1. Analytical validity of IHC4 between Royal Marsden and external centres

Dodson et al. 2016[4]

Methods: This study[4] (N=28) originated from the Royal Marsden Hospital (London, UK) and conducted two main assessments (Table XX). In the first assessment, sections from ER+, HER2- breast cancer tissue micro-arrays were distributed to three centres, where ER, PR, HER2 and Ki67 were stained according to each centre's own standard procedures, and scored at the Royal Marsden Hospital. Individual IHC scores (ER and PR only) and IHC4+C scores were then compared with those produced from slides stained by the Royal Marsden Hospital. This essentially compares different staining techniques, as all other variables are constant. In the second assessment, tissue microarray sections that had been stained at the Royal Marsden were scored by simplified non-counting methods and compared to results obtained through counting. This essentially compares different scoring methods as all other variables are constant. For ER, two different methods of scoring were used: a "simplified H-Score" where each of the four categories were "eye-balled" (instead of counted) and scored as per the usual protocol where the H-Score = (% cells weakly stained x 1) + (% cells moderately stained x 2) + (% cells strongly stained x 3); and an "estimated H-Score" where the proportion of stained cells was eye-balled and multiplied by the modal intensity score (estimated on a scale of 1-3). For PR and Ki67, the simplified method was an "eye-balled" estimate of the proportion stained cells, regardless of intensity of staining.

Results: Correlations between the external centres and the Royal Marsden were high for ER (r=0.93-0.96) and PR (r=0.91-0.98) but moderate for Ki67 (r=0.80-0.89). Upon calculation of the IHC4 scores, these translated to high correlation for IHC4 (r=0.90-0.93) and IHC4+C (0.98-0.99). For risk of distant recurrence at 10 years the correlation was also high (r=0.97-0.98).

The different scoring methods were also highly correlated for ER (r=0.92-0.93) and PR (r=0.98) but correlations were poorer for Ki67 (r=0.86). Again, correlations for IHC4 (r=0.90 to 0.97) and IHC4+C (r=0.97 to 1.00) were high, as were those for distant recurrence (0.97 to 1.00).

Conclusions: The authors conclude that IHC4+C is tolerant of variation in staining and scoring methods, and that additional confirmatory, comparative studies are required.

Critique: The EAG note that only one variable was altered at a time, namely staining technique and counting technique, and that it is unclear whether similar correlations would be achieved in routine clinical practice, where multiple and potentially different variations could occur. The authors themselves acknowledge this limitation and refer to an ongoing study involving 20 centres which may address some of these concerns. In addition, the authors note that HER2 assessment was not included in this analysis (as all patients were HER2-), and cite the high levels of proficiency in this assay in UK centres reported by UK NEQAS.[32]

The authors also have concerns relating to the Ki67 component, and advise the use of formal counting rather than simplified eye-balling methods. The logarithmic transformation of Ki67 data in the IHC4

algorithm is likely to accentuate differences at the lower end of the scoring scale (ie. 0-20% stained cells), where most patients score, and in could lead to a change in risk category for individual patients.

Engelberg 2015[23]

This study aimed to improve the precision and accuracy of assessing ER, PR, Ki-67, and HER2 (IHC4) through use of the online training tool developed and used in Balassanian 2013[28] & Bishop 2012[29] (see below), now termed "Score the Core" (STC). In Engelberg 2015[23], slides were stained at the Royal Marsden Hospital and scored by two pathologists. The *H* scores had a concordance of 0.90 between the first and second pathologist. Slides were then scanned as whole slide images (WSI) and uploaded to the software and distributed to nine pathologists in the Athena Breast Health Network (University of California), and was opened to pathology residents at the University of California Davis as well. Quantitative image analysis (QIA, an overlay of software-generated image analysis) was not available until after the user had submitted their score. HER2 data were excluded from the analysis as only one tumour was HER2+. As slides were stained at one laboratory, this study tests inter-observer reproducibility in scoring after training.

The training programme resulted in a decrease in error in relation to the reference slides for the Athena pathologists for ER and Ki-67 (ER: from 11.4 to 8.6 on a 100-point scale, p=0.03; Ki-67: from 7.8 to 5.7 percentage points, p=0.03), but not for PR which had reasonable agreement to begin with (6.8 to 4.8 on a 100-point scale, p=0.08). When the residents were included, all improvements were statistically significant.

Kappa scores between the reference slides (Royal Marsden Hospital) and the pathologists (Athena network) after training were ER: 0.73; PR: 0.96; Ki67: 0.87. Kappa scores between pathologists (Athena network) after training were ER: 0.77; PR: 0.87; Ki67: 0.62.

Critique: HER2 was not assessed. These results indicate that training improved scoring agreement, but Kappa values (between Royal Marsden pathologists and Athena pathologists, and between Athena pathologists compared to each other) were not always excellent even after training (range 0.62 to 0.96). Kappas for ER were surprisingly lower than might be expected for an established assay (0.73 and 0.77 respectively). Because slides were pre-stained, this study only provides information about inter-rater reliability and it is unclear whether similar Kappa scores would be achieved in routine clinical practice, where multiple and potentially different variations in pre-analytical, analytical and post-analytical factors could occur.

2. Analytical validity of IHC4 within other centres

Evidence from the main review

None of the prognostic studies identified by the main review[3 5-16 27] reported data relating to analytical validity. If the score had demonstrated prognostic value in multiple analyses, it could be argued that the analytical validity was sufficient for the purpose of prognosis. However, the evidence was somewhat mixed (see section XXX of main report), with some studies reporting statistically significant prognostic value and some not, though this did not seem to be associated with the assay methodologies which sometimes differed from those reported in the derivation study.[3] .

Balassanian 2013[28] & Bishop 2012[29]

Two abstracts reported on work conducted by the University of California Athena pathology collaboration, to investigate variance in, and harmonise IHC4 staining and scoring across labs. They report some analytical validity results, but also some attempts to improve standardisation of IHC4 methods. Both are reported here.

The first abstract[28] states that five slides from phenotypically different tumours were sent to 5 University of California laboratories, where IHC4 and HER2 FISH tests were conducted according to the prevailing methodology at each lab. Digital whole slide images (DWSI) were also captured, and analysed using quantitative image analysis (QIA). This study therefore tests staining and scoring variance. The abstracts report that there was variance between technical procedures, and between pathologist's scores, but this was not sufficient to affect the clinical score, and that technical staining variance by different laboratories was observed significantly more often for Ki-67than other IHC tests. Antibody vendor or clone did not explain the variance. Parallel analyses using DWSI with QIA suggests that the main source of variance was technical differences, and that WSI with QIA is a robust method to aid harmonisation of IHC4 scoring.

In a second abstract[29] (assumed to be part of, or an extension of, the same study), a similar (or the same) experiment as reported in Balassanian et al.[28] was described, along with two attempts to improve harmonisation . "Technical variance reduction" was attempted, using a Delphi voting process to identify an "ideal slide". Labs then made technical adjustments to their processes to match the appearance (depth of colour, contrast etc) of the ideal slide, and these slides were then scored by pathologists and by quantitative image analysis. "Scoring variance reduction" was attempted through creation of a digital pathology training tool, later to become "Score the Core".

In addition to some of the results reported by Balassanian et al.[28], mean values and variance were similar between WSI and traditional glass slides, except for HER2. Only early results from the quantitative image analysis relating to the "technical variance reduction" efforts were reported, which suggested that there was reduced variance. No results were reported for the "Scoring variance reduction" efforts.

Critique: the analytical validity data from these abstracts suggest that IHC4 scores conducted according to somewhat heterogeneous technical methods do not vary enough to affect clinical practice. There are more problems with Ki67 than ER, PR and HER2. The study further suggests novel concepts to improve harmonisation across labs, including reference slides to harmonise technical differences, use of WSI with QIA to improve scoring differences, and training through a digital tool.

Borowsky 2016[30]

This study used the "Score the Core" training, as developed and used in Balassanian 2013[28] & Bishop 2012[29] and Engelberg 2015[23] and measured inter-observer variance across four sites and nine pathologists after web-based training. 727 tumour samples were sectioned and stained in one laboratory (not reported which), and scored in a random order by two pathologists, hence testing scoring reproducibility. Kappa values were ER: 0.94; PR: 0.84; Her2: 0.91.

Critique: Excellent agreement was reported after training for ER, PR and HER2. Ki67 was not reported. Because slides were pre-stained, this study only provides information about scoring and it is unclear whether similar Kappa scores would be achieved in routine clinical practice, where multiple and potentially different variations in pre-analytical, analytical and post-analytical factors could occur.

3. Analytical validity of Ki67: Studies related to Ki67 Working Group and Royal Marsden

Because Ki67 is more problematic than the other components of IHC4 (see Dodson 2016[4] above), we have included some additional literature on this topic. However, the search strategy for the assessment report included search terms for IHC4, but not for Ki67 as this was not included in the scope of the assessment. Therefore, a systematic identification of all studies reporting data relating to Ki67 analytical validity has not been conducted. Instead, we focus on studies stemming from the "International Ki67 in Breast Cancer Working Group" (IKBCWG) and/or studies relating to the Royal Marsden hospital where the IHC4 score was generated, as these have highest relevance to the decision problem. However, it should be noted that there is a much larger body of literature on Ki67 which may address some of the issues not addressed by the selected studies.

The IKBCWG produced a set of recommendations in 2011[22] relating to the pre-analytical and analytical assessment, and interpretation and scoring of Ki67, in an attempt to aid harmonization of methodology. They concluded that, at the time, heterogeneity in pre-analytic and analytical methods were not the major source of variation in Ki67 measurements, and that a lack of standardization in scoring procedures (eg, core-cuts vs whole-tumor sections vs tissue microarrays) was problematic. They also stated that

the lack of quality assurance schemes made values produced in different labs non-comparable (though an individual lab may have high reproducibility), making use of the score in clinical decision-making (either on its own or in an algorithm such as IHC4) problematic without labs having their own reference data upon which to standardize values.

From this working group stemmed a series of three studies,[24-26] reported below.

Polley et al. (2013)[26]

This study assessed three questions assessing reproducibility between and within laboratories. The first question was reproducibility for Ki67 between laboratories due to differences in scoring. For this, 100 samples were stained centrally (at the Royal Marsden), then sent to eight laboratories (all having published papers on Ki67 i.e. with expertise in this field) where Ki67 was assessed using local methods of scoring. Reproducibility between local and central laboratories was moderate (intraclass correlation (ICC) 0.71, 95% CI: 0.47 to 0.78), implying that differences in scoring have an impact on Ki67. The second was reproducibility between laboratories due to both staining and scoring; this time, 100 samples were both stained and scored locally. Reproducibility between local and central laboratories was lower than above (ICC 0.59, 95% CI: 0.37 to 0.68), implying that differences in staining also impact on Ki67. The third was within-laboratory reproducibility for Ki67, in which 6 labs locally stained 50 samples each and repeated the scoring on three separate days; reproducibility within laboratories was high (ICC 0.94, 95% CI: 0.93 to 0.97). Factors contributing to between-laboratory discordance included tumour region selection, counting method, and subjective assessment of staining positivity. Formal counting methods gave more consistent results than visual estimation (eye-balling).

Polley et al. (2015)[25]

This study assessed reproducibility for Ki67 between laboratories following web-based training in scoring. For this, 50 samples were stained centrally (at the Royal Marsden) and sent to 16 laboratories in 8 countries. Participants scored Ki67 according to a specific protocol after undertaking training. Reproducibility between laboratories was high (ICC 0.94, 95% credible interval (CrI): 0.90, 0.97) when using central staining and web-based training in scoring.

Leung et al. (2016)[24]

This study compared three methods of Ki67 scoring: global method (assessing four fields of 100 cells each); weighted global method (as global but weighted by estimated percentage of total area); and hot-spot method (assessing a single field of 500 cells). For this, 30 samples were stained centrally (at the Royal Marsden) and sent to 22 laboratories in 11 countries. There was moderate inter-laboratory reproducibility for all three methods: unweighted global (ICC 0.87, 95% CrI 0.81, 0.93); weighted global (ICC 0.87, 95% CrI 80, 0.93) and hot-spot (ICC 0.84, 95% CrI 0.77, 0.92). A few cases still

showed large scoring discrepancies. Interestingly, a conference abstract for the same study (Dodson et al., 2016) reported that when these Ki67 assessments were integrated into the IHC4+C score, the correlation for risk of recurrence was very high (ICC 0.99, 95% CI: 0.99 to 1.00), implying that variability in Ki67 had little impact on the combined IHC4+C score.

Discussion

Only two studies reported data relating to the analytical validity of IHC4 in centres external to the Royal Marsden and reported good to moderate correlations for ER, PR and Ki67 when comparing different staining techniques, different scoring methods and different observers. Both studies isolated one analytical or counting variable to alter at a time, and one included additional training and standardisation practices, making it unclear if the same favourable correlations would be achievable when comparing samples prepared in totality at different sites or in isolation of the training programme (Score the Core).

Interestingly, despite moderate Ki67 correlations in Dodson 2016a, the IHC4+C correlations were very high (0.98 to 0.99), suggesting the algorithm is robust to a degree of variation in the scoring of component parts. Similar results were reported in a conference abstract (Dodson 2016b[31]) for the Leung 2016[24] study of Ki67, where incorporation of Ki67 values (by any of three methods of counting) into the IHC4+C score resulted in risk category agreement of 98.6%, and in Balassanian 2013[28] where several labs stained and scored 5 slides, but IHC4 scores were not affected by variance in component scores. Whilst these results are reassuring, they represent only a small number of laboratories, and it seems likely that whilst problems with variance in IHC results persist, clinician confidence in using the score may be affected.

Data relating to the analytical validity of IHC4 within other centres was scarce, though our searches are not comprehensive. One study showed that despite considerable heterogeneity between methods of preparation and interpretation the IHC4 scores did not differ enough to change clinical decisions. Excellent agreement between scoring of ER, PR and Ki67 was achieved after training using "Score the Core" on slides stained at one site.

Notably, across these four studies, only one reported correlation data for HER2 (0.91),[30] meaning this is poorly evidenced. Ki67 was not reported in one study, and identified as more problematic than the other factors in three studies; Dodson 2016,[4] Engleberg 2015[23] (though the kappa for Ki67 was 0.87 between more experienced pathologists, and ER also reported Kappas <0.8, for both experienced and resident pathologists), Balassanian 2013[28]& Bishop 2012.[29]

Attempts to standardise Ki67 appear promising as a result of the IKBCWG programme of work, with high levels of correlation within labs, or when using centrally-stained slides. Web-based training for scoring appears to improve agreement, but has not been used on whole sections and biopsy samples. Problems with variations in staining that were evident in Polley 2013[26] do not appear to have been addressed in the selected literature, probably as the original Working Group[22] findings pointed to problems with scoring being the main source of variance.

It should be noted that there are many examples of attempts to improve IHC measurement in the literature that have not been reviewed here due to time and scope limitations. These include digital imaging (which was used as a reference method in some of the studies included here), double staining, variance in antibodies, use of quantum dots, and even novel ways of measuring the markers themselves, such as use of mRNA, chromogenic in situ hybridization and quantitative immunofluorescence (QIF, e.g AQUA which has been used to validate the IHC4 algorithm).[17]

Conclusions

Excellent levels of agreement appear achievable (with web-based training) when slides are prepared centrally. Standardisation of staining may be achievable with training, but has not yet been fully reported or robustly tested (N=5 tumours). Variance in IHC or Ki67 assays may not affect the IHC4 risk scores in clinically meaningful way, but evidence is extremely limited. Efforts to improve Ki67 appear promising but have not yet addressed all variance issues. External quality assessment schemes may improve inter-laboratory agreement.

Table 1Study characteristics and results

Reference	Targets	Topic	Samples/setting	Experimental variable	Findings	Conclusions
1. Analytical validity of IHC4 between Royal Marsden and external centres						
Dodson 2016a (full paper)[4]	IHC4+C Ki67 ER PR	1) Inter-laboratory reproducibility for ER, PR & Ki67: slides stained at 3 external centres compared with staining at RMH; RMH scoring of all samples by single assessor (i.e. assessing effect of staining method) 2) Scoring via counting methods vs. simplified non-counting-based methods (all stained & scored at RMH)	N=28 tumour samples, ER+, HER2- 4 centres (all UK)	1)Staining 2) Scoring method	1) External vs RMH staining: High correlation for ER (r=0.93-0.96) and PR (r=0.91-0.98) but moderate for Ki67 (r=0.80-0.89). Translated to high correlation for IHC4 (r=0.90-0.93), IHC4+C (0.98-0.99) and risk of distant recurrence (r=0.97-0.98) 2) Non-counting methods vs counting: high correlation for ER (r=0.92-0.93) and PR (r=0.98) but poorer correlation for Ki67 (r=0.86)	1) External vs RMH staining: high reproducibility for ER and PR, moderate for Ki67. Translated to high correlation for IHC4 and IHC4+C scores and distant recurrence 2) Non-counting vs. counting methods of scoring (same lab): high reproducibility for ER and PR, moderate for Ki67. Recommend formal counting for ki67

Reference	Targets	Topic	Samples/setting	Experimental variable	Findings	Conclusions
Engelberg 2015 (full paper)[23]	IHC4 Ki67 ER PR HER2	Development of "score the core" web-based training 1) 1 RMH pathologist stained and scored reference slides, 2nd pathologist re-scored 2)Athena pathologists scored the RMH reference slides after training 3) Athena pathologists scoring RMH slides after training, compared to each other 4) Pathology Residents scored the RMH reference slides after training	N=32 samples from RMH, 9 pathologists at international centres	1-4) Inter-observer reproducibility in scoring after training	1) Scoring agreement between two RMH pathologists for *H* scores on slide stained at RMH, r=0.90 2) Agreement (kappa) between RMH and Athena pathologists after training on scanned slide stained at RMH: ER: 0.73; PR: 0.96; Ki67: 0.87 3) Agreement (kappa) between Athena pathologists after training on scanned slide stained at RMH: ER: 0.77; PR: 0.87; Ki67: 0.62 4) Agreement between reference slides (RMH) and pathology residents after training: lower correlation for PR (P = .03, pooled 2-sample t test) and no significant difference for ER or Ki-67.	"Score the core" web-based training can improve agreement to reference score and between pathologists. Agreement on IHC4 elements scored by different pathologists were not always good.
2. Analytical validity of IHC4 within other centres						

Reference	Targets	Topic	Samples/setting	Experimental variable	Findings	Conclusions
Balassanian 2013 (CA)[28] Bishop 2012 (CA)[29]	IHC4 ER PR HER2 Ki67	1) IHC4 scoring via traditional techniques versus quantitative image analysis (QIA) with whole slide imaging (WSI); stained and scored at local labs within University of California-Athena pathology collaboration 2) Technical variance reduction through use of "ideal slide" 3) Scoring variance reduction through use of web-based training (Score the Core)	N=5 tumour samples, 5 labs,10 pathologists at University of California	1) Inter-lab variance in staining and scoring 2) intervention to reduce technical (staining) variance 3) intervention to reduce scoring variance	1) Considerable and significant technical and interpretational variances exist between laboratories but IHC4 scores do not differ to a clinically meaningful extent. There are more problems with Ki67 than ER, PR and HER2. 2) Early results suggest reduction in staining variance after intervention 3) Results not reported	See findings
Borowsky 2016 (CA)[30]	IHC4 Ki67 ER PR HER2	Interobserver agreement of IHC4 components after "score the core" web-based training (using tissue microarrays to visually score ER, PR and Ki-67). Sections stained at one lab (not named)	N=727 samples, 4 sites, 9 pathologists (Conf abs)	Inter-observer reproducibility after training	"Experts at multiple sites trained with the Score the Core tool can provide high precision IHC quantitation suitable for clinical decision making." Kappa scores: ER: 0.94; PR: 0.84; HER2: 0.91; Ki67: assessed but no correlation reported	After "score the core" web-based training, agreement between pathologists was good for ER, PR, HER2 (assessed but not reported for Ki67)

Reference	Targets	Topic	Samples/setting	Experimental variable	Findings	Conclusions
3. Analytical valdidty of Ki67: Studies related to Ki67 Working Group and RMH						
Dowsett 2011 (recommendations from Ki67 working group)[22]	Ki67	Summary of issues affecting Ki67 reproducibility and recommendations to mitigate these		NA	Issues include: • Preanalytical (type of biopsy, fixative, storage) • Analytic (antibodies, staining etc) • Interpretation and scoring: determination of percentage positive cells; differences between areas of slide (edge vs central, hot spots), visual vs automated • Data analysis: issues with cutpoints Most problematic is methods of counting and a lack of quality assurance schemes.	
Polley 2013[26] (full paper)	Ki67	1&2) Inter-laboratory reproducibility for Ki67, using central or local staining and own method of scoring 3) Intra-laboratory reproducibility for Ki67, local staining, scored on 3 separate days All used MIB-1 antibody	1&2) 8 labs scored n=100 samples, local and central staining (RMH) 3) 6 labs repeated n=50 slides on 3 days Labs USA & Europe, all had papers on Ki67 i.e. experts	1) Scoring 2) Staining and scoring 3) Intra-lab reproducibility of counting	1&2) Interlab reproducibility was only moderate (central staining: ICC = 0.71, 95% CI = 0.47 to 0.78; local staining: ICC = 0.59, 95% CI = 0.37 to 0.68) "Factors contributing to interlaboratory discordance included tumor region selection, counting method, and subjective assessment of staining positivity. Formal counting methods gave more consistent results than visual estimation." 3) Intralab reproducibility was high (ICC=0.94, 95% CI;0.93, 0.97)	Reproducibility for Ki67 scoring was high within laboratories but only moderate between laboratories (using central or local staining, and local scoring methods)

Reference	Targets	Topic	Samples/setting	Experimental variable	Findings	Conclusions
Polley 2015[25] (full paper)	Ki67	Inter-laboratory reproducibility for Ki67 after web-based training in scoring. Centrally-stained slides (RMH) sent to external labs for scoring according to specific protocol.	N=50 samples 16 labs, 8 countries	1) inter-Laboratory after training	High inter-laboratory reproducibility following web-based training in scoring (ICC 0.94, 95% CrI 0.90, 0.97) "Although these data are potentially encouraging, suggesting that it may be possible to standardize scoring of Ki67 among pathology laboratories, clinically important discrepancies persist. Before this biomarker could be recommended for clinical use, future research will need to extend this approach to biopsies and whole sections, account for staining variability, and link to outcomes."	Reproducibility for Ki67 scoring was high between laboratories when using central staining AND web-based training in scoring
Leung 2016[24] (full paper) Dodson 2016b (CA)[31]	Ki67	Compares three methods of Ki67 counting: global (4 fields of 100 cells) vs. weighted global (as global but weighted by estimated % of total area) vs. hot-spot method (single field of 500 cells). Centrally-stained slides (RMH)	N=30 samples 22 labs in 11 countries	Counting method	Moderate inter-laboratory reproducibility for all methods: unweighted global (ICC 0.87, 95% CrI 0.81, 0.93); weighted global (ICC 0.87, 95% CrI 80, 0.93) and hot-spot (ICC 0.84, 95% CrI 0.77, 0.92). A few cases still showed large scoring discrepancies. When integrated into IHC4+C, ICC for risk of recurrence was 0.99 (95% CI 0.99, 1.00) and risk category agreement (low/intermediate/high) was 98.6% (Dodson 2016 CA) [31] "Establishment of external quality assessment schemes is likely to improve the agreement between laboratories further."	Moderate reproducibility for Ki67 between laboratories for each of three pre-specified scoring methods (using central staining). Translated to very high correlation for IHC4+C recurrence risk (i.e. variability in Ki67 had little impact on IHC4+C)

Reference	Targets	Topic	Samples/setting	Experimental variable	Findings	Conclusions
RMH, Royal Marsden Hosptial; ER, oestrogen receptor; PR, Progesterone receptor; HER2, human epidermal growth factor receptor 2; IHC, immunohistochemistry; CA conference abstract						

1. Elliott K, McQuaid S, Salto-Tellez M, et al. Immunohistochemistry should undergo robust validation equivalent to that of molecular diagnostics. *Journal of clinical pathology* 2015:jclinpath-2015-203178.
2. Goldstein NS, Hewitt SM, Taylor CR, et al. Recommendations for improved standardization of immunohistochemistry. *Applied Immunohistochemistry & Molecular Morphology* 2007;15(2):124-33.
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. Dodson A, Zabaglo L, Yeo B, et al. Risk of recurrence estimates with IHC4+C are tolerant of variations in staining and scoring: an analytical validity study. *Journal of Clinical Pathology* 2016;69(2):128-35. doi: https://dx.doi.org/10.1136/jclinpath-2015-203212
5. Christiansen J, Bartlett JMS, Gustavson M, 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(15 SUPPL. 1)
6. 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
7. 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)
8. 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
9. 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
10. 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
11. 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
12. Prat A, Cheang MC, Martin M, et al. Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically defined luminal A breast cancer. *Journal of Clinical Oncology* 2013;31(2):203-09. doi: https://dx.doi.org/10.1200/JCO.2012.43.4134
13. 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
14. 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
15. 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
16. 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* 2013;8(10):e76496. doi: https://dx.doi.org/10.1371/journal.pone.0076496

17. Bartlett JM, Christiansen J, Gustavson M, 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(1):66-74. doi: 10.5858/arpa.2014-0599-OA [published Online First: 2015/12/31]
18. Ward S, Scope A, Rafia R, 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(44):1-302. doi: 10.3310/hta17440 [published Online First: 2013/10/04]
19. Harbeck N, Sotlar K, Wuerstlein R, et al. Molecular and protein markers for clinical decision making in breast cancer: today and tomorrow. *Cancer Treatment Reviews* 2014;40(3):434-44. doi: https://dx.doi.org/10.1016/j.ctrv.2013.09.014
20. Hayes DF. Clinical utility of genetic signatures in selecting adjuvant treatment: Risk stratification for early vs. late recurrences. *Breast* 2015;24 Suppl 2:S6-S10. doi: https://dx.doi.org/10.1016/j.breast.2015.07.002
21. Institut für Qualität und Wirtschaftlichkeit im G. Biomarker-based tests for the decision for or against adjuvant systemic chemotherapy in primary breast cancer. 2016
22. 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.
23. Engelberg JA, Retallack H, Balassanian R, et al. "Score the Core" Web-based pathologist training tool improves the accuracy of breast cancer IHC4 scoring. *Human Pathology* 2015;46(11):1694-704. doi: https://dx.doi.org/10.1016/j.humpath.2015.07.008
24. Leung SC, Nielsen TO, Zabaglo L, et al. Analytical validation of a standardized scoring protocol for Ki67: phase 3 of an international multicenter collaboration. *NPJ Breast Cancer* 2016;2:16014.
25. Polley M-YC, Leung SC, Gao D, et al. An international study to increase concordance in Ki67 scoring. *Modern Pathology* 2015;28(6):778-86.
26. Polley M-YC, Leung SC, McShane LM, et al. An international Ki67 reproducibility study. *Journal of the National Cancer Institute* 2013;105(24):1897-906.
27. 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.
28. 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
29. 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)
30. 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
31. 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.
32. Bartlett JM, Ibrahim M, Jasani B, et al. External quality assurance of HER2 FISH and ISH testing: three years of the UK national external quality assurance scheme. *American Journal of Clinical Pathology* 2009;131(1):106-11.

Addendum to: 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.*

28th November 2017

In response to the DAR consultation responses collated by NICE and sent to the EAG on 13th November 2017, the EAG provide the following addenda to the report. These are grouped by test, and include addenda on Oncotype DX, MammaPrint and Endopredict.

1. Oncotype DX

1.1 Inclusion of NSABP B-20 endocrine monotherapy arm in the prognostic data set

In response to Agendia's comment (#2), the EAG agree that the NSABP B-20 patients who were used as part of the derivation set (the monotherapy arm of the trail, N=233) should not have been included in the set of studies reporting prognostic performance in patients with LN0 disease and treated with endocrine monotherapy (Paik et al 2006)[1-3]. This reduces the number of studies reporting this subgroup to three for DRFI, but does not alter the conclusions drawn, as the three remaining studies all demonstrated prognostic performance in this group. The inclusion of NSABP B-20[1-3] patients in the set of studies reporting patients with LN0 disease treated with endocrine therapy and chemotherapy was not problematic, as the chemotherapy patients did not form part of the derivation set.

1.2 Inclusion of NSABP B-20 endocrine monotherapy arm in the chemotherapy benefit data set

The EAG also agree with Agendia's comment (#2) on the inclusion of the NSABP B-20 analysis[1-3] in the studies reporting chemotherapy benefit, in that 233 (the endocrine monotherapy patients) of the 651 patients were from the derivation cohort. The remaining patients were from the two endocrine therapy plus chemotherapy arms of the same trial. It is unclear whether inclusion of the derivation set patients would augment or reduce any apparent interaction between chemotherapy and RS, but it does put the study at high risk of bias. However, because there is no other analysis in LN0 patients, the data is still of relevance, as the next level of evidence.

Agendia also note that the interaction test p value relates to a 50 point difference in recurrence score in the Albain et al 2010[4] analysis of SWOG-8814 patients. There is a lack of consistency between the methods section and results section of the journal article, and the EAG had interpreted this as relating to an analysis using the continuous score.[4] However, upon closer inspection, the EAG agree that the analysis stated in the methods section using the continuous score has not been reported, and instead an analysis using the 50-point difference has been reported. This means the study can be considered at high risk of reporting bias, and the analysis has very low clinical relevance.

In the report, the EAG concluded that there is weak evidence for the prediction of chemotherapy benefit by Oncotype DX. Now, with the high risk of bias associated with the B-20 cohort analysis[1-3] and the Albain et al 2010[4] analysis, the evidence base could be judged to be very weak with a very high risk of bias.

1.3 Inclusion of NSABP B-20 endocrine monotherapy arm in the validation set for RSPC

Whilst not mentioned by Agendia, the B-20 cohort was also used to validate the RSPC algorithm.[2] This validation data should therefore be interpreted with caution, as some patients were included in the derivation of Oncotpye DX, which is one component of the RSPC algorithm.

1.4 Use of 50-point difference in analysis of chemotherapy benefit (Albain 2010).[4]

The use of the 50-point difference in the analysis of an interaction between RS and chemotherapy benefit does not indicate the clinical significance of the 18 -30 RS cut points. However, the study does conclude that there is little benefit from chemotherapy at RS<20.

1.5 Use of 50-point difference in adjusted analyses of prognostic performance.

The use of the 50-point difference in the adjusted analyses of prognostic performance indicate that RS is prognostic after adjusting for clinicopathological factors, but does not provide information about the clinical significance of the 18 -30 RS cut points.

2. EndoPredict

2.1 EP score adds information to clinicopathological factors in years 0-10 as well as years 5-10 (section 4.6.2)

The EAG report already notes that the EP score adds significant information to clinicopathological factors or Adjuvant! Online in ABCSG6 and ABCSG8, shown via c-index analyses, in years 5-10. We agree with comment #8 from Myriad that this also applies in years 0-10 (reported in Filipits 2011).[5]

2.2 Time to test results for EndoPredict (section 4.11)

The EAG agrees with the comment (#18) from Myriad that the publication by Müller et al. (2013)[6] reports the time to test result for EndoPredict. In this study the median handling time was three working days (range 0 to 11 days), while 59% of tests were performed within 3 days or less.

3. MammaPrint

3.1 MINDACT trial provides randomised controlled trial evidence of treatment guided by test versus usual practice, in patients who are high-risk via either mAOL or MammaPrint (section 4.4.4)

We agree with Agendia (comment #1a) that MammaPrint is the only one of the five tests to have reported evidence of a RCT (MINDACT) where patients were randomised to treatment guided by the test or by usual clinical practice. These patients were high-risk via either mAOL or MammaPrint. Patients with high-clinical but low-MammaPrint risk showed a non-significant effect of chemotherapy.

3.2 MINDACT data for prognostic performance of MammaPrint (section 4.4.2 and 4.4.4)

Agendia (comment #1e) note that it may be possible to generate prognostic performance data from MINDACT by comparing outcomes for low-MMP vs high-MMP patients using the concordant-risk groups plus the discordant-risk groups in which treatment was determined by mAOL rather than MammaPrint. However, we were not able to locate these data in the time available to respond to these comments. The EAG report does note that, 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). This analysis does not omit the patients treated according to MammaPrint, but the adjustment for other factors may mitigate this. These data could potentially be considered prognostic data. This is consistent with the findings of other MammaPrint prognostic studies which showed that MammaPrint was statistically significantly prognostic in multivariable analyses.

3.3 Difficulties of obtaining trial data/samples to assess chemotherapy benefit

The EAG agrees with Agendia (comment #17) that it is difficult to undertake further assessments of predictive ability for chemotherapy benefit, since there are few trials in which patients were randomised to chemotherapy versus no chemotherapy, and the few trials of this type that are available have insufficient tumour samples left on which to undertake tumour profiling tests.

3.4 Correction for inclusion of derivation patients in Van de Vijver 2002[7] study

Re Agendia comment #54, the Van de Vijver 2002[7] study included a correction for the fact that a small proportion of patients derived from the derivation set were included in the validation study. The small proportion (n=61) were included to avoid selection bias, since the previous study included a disproportionately large number of patients in whom distant metastases developed within five years. The correction in analysis was made using the “leave-one-out” cross-validated classification to predict the outcomes among these patients. This approach minimizes to some extent the possibility of overestimating the value of the prognosis profile while it keeps the consecutive series complete. The study also provides validation results taking only the new samples into account.

3.5 Reference 292

The EAG agree with Agendia that reference 292 is incorrect. However, the authors are correct, and the title and bibliographic information was incorrect, rather than the other way araound. All “Author et al. year” citations relating to 292 should read “van’t Veer *et al.* 2017”, and the reference should read:

292. van‘t Veer, L.J., Yau, C., Nancy, Y.Y., Benz, C.C., Nordenskjöld, B., Fornander, T., Stål, O., Esserman, L.J. and Lindström, L.S. Tamoxifen therapy benefit for patients with 70-gene signature high and low risk. Breast Cancer Research and Treatment (2017): 1-9.

1. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *Journal of Clinical Oncology* 2006;24(23):3726-34.
2. Tang G, Cuzick J, Costantino JP, 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(33):4365-72. doi: https://dx.doi.org/10.1200/JCO.2011.35.3714
3. Tang G, Shak S, Paik S, 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 Research & Treatment* 2011a;127(1):133-42. doi: https://dx.doi.org/10.1007/s10549-010-1331-z
4. Albain K, Barlow W, Shak S, 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(1):55-65.
5. Filipits M, Rudas M, Jakesz R, 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(18):6012-20. doi: https://dx.doi.org/10.1158/1078-0432.CCR-11-0926
6. Muller BM, Keil E, Lehmann A, et al. The EndoPredict gene expression assay in clinical practice - performance and impact on clinical decisions. *PLoS ONE* 2013;8(6):e68252. doi: https://dx.doi.org/10.1371/journal.pone.0068252
7. van de Vijver MJ, He Y, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *New England Journal of Medicine* 2002;347(25):1999-2009.

Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer Addendum: EAG responses to key themes within the Comments on the Diagnostics Consultation Document

As part of the Diagnostic Assessment Programme topic "Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer", following the 1st Diagnostics Appraisal Committee meeting on 30 November 2017, NICE produced a Diagnostics Consultation Document (DCD, dated 10 January 2018).[1] Commentators provided comments on the DCD, and the EAG has responded to these comments in a separate document. This addendum provides responses to key themes within the comments document.

1. Use of TransATAC data in the economic model

1.1. Rationale for using TransATAC data in the EAG health economic model

All studies reporting prognostic ability or prediction of chemotherapy benefit and meeting the inclusion criteria were included in the clinical review. The rationale for using the TransATAC data in the EAG model was that it could be restricted to the population in the NICE scope (ER+ HER2- 0-3 positive nodes) and it was possible to split the node-negative patients into clinically low-risk and clinically intermediate-risk (according to NPI score above or below 3.4).

1.2. The TransATAC analysis is unreported and has not been subjected to scientific peer review

Several analyses of TransATAC focussing on different tumour profiling tests have been published in peer-reviewed journals. On behalf of the EAG, the TransATAC authors produced a bespoke analysis[2] which covered four of the five tests included in the DAR (Oncotype DX, EndoPredict, Prosigna and IHC4+C) and which was restricted to the relevant population as above.

Subsequent to the publication of the EAG report, the TransATAC authors have published a pre-planned analysis of these data in a peer-reviewed journal (Sestak *et al.,* 2018[3]).

Table **1** presents some key data from the bespoke analysis for the EAG[2] alongside the data from Sestak *et al*., 2018.[3] Whilst there are some small differences, these data are largely consistent. It is not possible to use the newly-published data[3] in our model since LN0 patients are not stratified into clinically low-risk and clinically intermediate-risk, and hazard ratios (HRs) are reported for a 1 standard deviation (1SD) change rather than between risk groups.

Table 1: A comparison of key analyses reported in the data request analysis[2] and in Sestak 2018[3]

Test	LN0 HR (95% CI) for 1SD 10 year		LN1-3 HR (95% CI) for 1SD 10 year		ΔLR-χ2 to CTS 10 year		
	Data request[2]	**Sestak 2018[3]**	**Data request[2]**	**Sestak 2018[3]**	**Data request[2] LN0**	**Data request LN1-3**	**Sestak 2018[3] LN0-3**
Oncotype DX	1.67 (1.39-2.01)	1.69 (1.40-2.03)	1.42 (1.05-1.91)	1.39 (1.05-1.85)	22.78 *p*<0.0001	4.75 *p*=0.023	15.2
IHC4+C	2.56 (1.98-3.33)	NR	1.83 (1.31-2.56)	NR	48.55 *p*<0.0001	12.60 *p*<0.001	NR
IHC4	NR	1.95 (1.55-2.45)	NR	1.33 (0.99-1.78)	NR	NR	20.1
Prosigna	2.58 (1.97-3.38)	2.56 (1.96-3.35)	1.59 (1.16-2.17)	1.58 (1.16-2.15)	50.77 *p*<0.0001	8.51 *p*=0.004	26.3
EPClin	2.34 (1.82-3.02)	2.14 (1.71-2.68)	1.84 (1.34-2.53)	1.69 (1.29-2.22)	40.60 *p*<0.0001	12.91 *p*<0.001	24.4

1.3 Patient numbers per subgroup are small

The number of patients per subgroup were: at least 410 for LN0 NPI<3.4 (more for some tests), at least 253 for LN0 NPI>3.4, and at least 192 for LN1-3. The EAG do not consider the subgroups to be unreliably small.

1.4. Overlapping confidence intervals for recurrence rates between risk groups and between tests

The EAG agrees that there is some overlap between confidence intervals. However, this does not prevent the data from being useable. The point estimates for recurrence per test risk group (for LN0 and LN+ patients) are consistent with estimates from other studies (see point 2 of this addendum, distant recurrence rates by risk classification). The EAG's probabilistic sensitivity analysis fully characterises the uncertainty surrounding these estimates.

1.5. Bias in the patient spectrum due to exclusion of small tumours with insufficient tissue

The EAG report noted this limitation. This is a limitation of most analyses using stored tumour samples and is not limited to TransATAC. A comparison of some basic population-level statistics between the MINDACT trial and the TransATAC data population was provided for the previous round of comments on the DAR, and no major differences were observed.

1.6. TransATAC includes postmenopausal women who were not suitable for chemotherapy

TransATAC selected patients who had not received chemotherapy in order to assess prognostic ability of tumour profiling tests, which required calculation of distant recurrence rates in the absence of

chemotherapy. The EAG report noted this limitation. Many other prognostic studies included in the systematic review also included patients receiving no chemotherapy, to allow a consistent assessment of prognostic ability. TransATAC does appear to include some patients who would be currently indicated for chemotherapy in the UK (e.g. LN>3).

2. Distant recurrence rates by risk classification

2.4. Consistency of Oncotype 10-yr outcomes across re-analyses of RCTs included in the review

Table 2 shows distant recurrence-free rates at 10 years across re-analyses of RCTs with endocrine monotherapy. Distant recurrence-free rates at 10 years in LN0 Oncotype DX low-risk patients (not subgrouped by clinical risk) are consistent across TransATAC publications (94.9% in the bespoke analysis;[2] 94.1% in the Sestak 2016 SABCS presentation;[4] 96% in Dowsett et al. 2010,[5] the latter being measured at 9 years rather than 10 years). These rates are also consistent with those from other studies: B14[6] (93.2%) and B20[7] (96.8%), for patients in the no-chemotherapy arms. Outcomes for other risk groups were also consistent across studies (Table 2).

Table 2: 10-year distant recurrence for Oncotype DX (RCT re-analyses; endocrine monotherapy)

Nodal status	**Oncotype DX risk group**	**Percent of patients distant recurrence-free at 10 years (95% CI)**				
		TransATAC data request[2]	**TransATAC (Sestak 2016 SABCS[4])**	**TransATAC (Dowsett 2010;[5] 9yr recurrence)**	**B14 (Paik 2004,[6] Tang 2011a[8])**	**B20 (Paik 2006[7])**
LN0	ODX low	94.9	94.1	96 (93 to 97)	93.2 (90.4, 96.0)	96.8 (93.7, 99.9)
LN0	ODX int	87.7	83.3	88 (82 to 92)	85.7 (79.7, 91.7)	90.9 (82.5, 99.4)
LN0	ODX high	77.2	72.8	75 (66 to 83)	69.5 (62.6, 76.4)	60.5 (46.2, 74.8)
		LN1-3 only	**Incl LN4+**	**Incl LN4+**		
LN+	ODX low	81.8 (72.7-88.0)	73.8	83 (76 to 88)		
LN+	ODX int	75.4 (63.0-84.2)	65.3	72 (61 to 80)		
LN+	ODX high	68.6 (44.7-83.9)	51.2	51 (36 to 65)		

Data from Table 12 in EAG report. No additional RCTs of endocrine monotherapy reported distant recurrence in LN+ patients.

2.5. Consistency of outcomes across studies: Oncotype low-risk patients subgrouped by clinical risk

There are several comments referring to the 10-year distant recurrence rate of 15% in the LN0 Oncotype DX low-risk group in the TransATAC analysis (i.e. 85.4% distant recurrence-free). It is vital to point out that this does not represent the Oncotype DX low-risk group as a whole (see response 2.1 and Table 2 for the whole Oncotype DX low risk group). Instead, it represents the LN0 NPI>3.4 subgroup (i.e. LN0 and clinically intermediate-risk).

Table 3 shows distant recurrence-free rates at 10 years for LN0 patients, subgrouped by clinical risk. For TransATAC, these were subgrouped according to NPI score (which includes nodal status, tumour grade and tumour size). For the Oncotype DX low-risk, clinically intermediate subgroup (NPI>3.4), the distant recurrence-free rate at 10 years was 85.4%. We could not identify any other studies subgrouping by NPI score. However, the B14 analysis subgrouped by various other measures of clinical risk: tumour size, grade and Adjuvant! Online (AOL).[6, 8] B14 results appeared consistent with TransATAC, with similar 10-year distant recurrence-free rates for Oncotype DX low-risk, clinically intermediate-risk patients (tumour >4cm, 87%; grade poor-differentiated, 86%; AOL intermediate-risk, 86.6%, AOL high-risk, 95.0%). Outcomes for other Oncotype DX risk groups sub-grouped by clinical status were also consistent across studies (Table 3).

Table 3: 10-year distant recurrence for Oncotype DX by clinical risk group (RCT re-analyses)

Oncotype DX risk group	**Clinical risk**	**TransATAC data request[2] LN0**		**B14 (Paik 2004,[6] Tang 2011a[8]) LN0**	
		Definition of clinical risk	**% DRF at 10yr (95% CI)**	**Definition of clinical risk**	**% DRF at 10yr**
ODX low	Clinical low	NPI≤3.4	98.3 (96.3-99.2)	Tumour <1cm	100
				Grade well-diff	96
				AOL low-risk	94.4
	Clinical intermediate	NPI>3.4	85.4 (77.6-90.7)	Tumour >4cm	87
				Grade poor-diff	86
				AOL int-risk	86.6
				AOL high-risk	95.0
ODX int	Clinical low	NPI≤3.4	93.1 (86.7-96.5)	Tumour <1cm	87
				Grade well-diff	91
				AOL low-risk	90.0
	Clinical intermediate	NPI>3.4	79.8 (69.4-86.9)	Tumour >4cm	88
				Grade poor-diff	76
				AOL int-risk	86.1
				AOL high-risk	76.6
ODX high	Clinical low	NPI≤3.4	83.8 (57.7-94.5)	Tumour <1cm	83
				Grade well-diff	69
				AOL low-risk	81.8
	Clinical intermediate	NPI>3.4	74.9 (59.8-85.1)	Tumour >4cm	47
				Grade poor-diff	60
				AOL int-risk	56.8
				AOL high-risk	68.5

TransATAC data from Table 124 in EAG report. B14 data by size/grade estimated from graphs in Paik 2004.[6] DRF, distant recurrence-free

2.6. Consistency of Oncotype 5yr outcomes between TransATAC and observational studies

There were several comments suggesting that the TransATAC recurrence rates used in the EAG model were less favourable than the recurrence rates from observational studies of Oncotype DX. Table 4 shows outcomes at 5 years for TransATAC and for observational studies of Oncotype DX (no 5-year data were available for other reanalyses of RCTs). Outcomes at 5 years were similar between

TransATAC and observational studies of Oncotype DX. It should be noted that some patients in the observational studies received chemotherapy; this may have improved observed outcomes.

The differences between the TransATAC recurrence rates used in the EAG model and the recurrence rates reported in observational studies appear to be due to: (a) the model data being stratified by clinical risk (those with NPI >3.4 had less favourable outcomes), and (b) the observational data being reported at a 5-year rather than 10-year follow-up.

Table 4: 5-year outcomes for Oncotype DX (RCTs and observational studies; some chemotherapy use)

Oncotype DX risk group			**LN0-mic**						**LN0-3, clin high risk**
	TransATAC data request[2] (LN0) N=829		**CT use in obs. studies**	TAILORx (Sparano 2015[9]) N=1626	MD Anderson (Le Du 2015[10]) N=1030	Clalit (Stemmer 2016[11]) N=1594	Memorial Sloan Kettering (Wen 2017[12]) N=1406	SEER (Petkov 2016,[13] Roberts 2016[14]) N=38,568	WSG PlanB (Nitz 2017 [15-17]) N=2646
	CT use	**DRFI 5yr**		**DRFS 5yr**	**DRFS 5yr**	**DRFI 5yr**	**DRFI 5yr**	**BCSS 5yr**	**IDFS 5yr**
ODX very low (<11/12)	None		0%	99.3 (98.7, 99.6)			99.9%	99.6 (99.4, 99.8)	94.2 (91.2, 97.3)
ODX low (RS<18)	None	99.1	1-12%	-	95.9 (93.0, 97.6)	99.5 (98.4, 99.8)	99.6%	99.6 (99.4, 99.7)	
ODX int (RS 18-30)	None	94.0	26-43%		-	98.8 (97.2, 99.4)		98.6 (98.3, 98.9)	94.3 (92.8, 95.8) **(RS 12-25)**
ODX high (RS >30)	None	88.9	89-90%		76.4 (59.2, 87.1)	93.1 (87.1, 96.3)		95.6 (94.4, 96.6)	84.2 (80.6, 87.8) **(RS ≥25)**

Data from Table 26 in EAG report. CT, chemotherapy; DRFS, distant recurrence-free survival; DRFI, distant recurrence-free interval; IDFS, invasive disease-free survival; BCSS, breast cancer-specific survival

3. Ability of Oncotype DX to predict differential relative benefit from adjuvant chemotherapy

3.1. Clarification on the difference between absolute and relative benefit

A key issue for clinical and cost-effectiveness of tumour profiling tests is whether the **relative** benefit from chemotherapy differs between test risk groups. It is important to note that this relates to relative rather than absolute benefit. We concluded in our EAG report that all the tests have additional prognostic ability over clinicopathological factors, at least in LN0 patients, i.e. that recurrence rates are higher in higher-risk groups. This means that the **absolute** benefit of chemotherapy is also higher in higher-risk groups. However, this does not necessarily mean that the relative benefit differs between groups.

As an example, if distant recurrence rates in the test high-risk group were 30% without chemotherapy and 20% with chemotherapy, the absolute benefit of chemotherapy would be 10%. Likewise, if distant recurrence rates in the test low-risk group were 3% without chemotherapy and 2% with chemotherapy, the absolute benefit of chemotherapy would be 1% (i.e. much smaller). However, the relative benefit would be the same in both groups (relative risk of 0.67, i.e. chemotherapy reduces recurrence by one-third).

3.2. Summary of data on the ability of Oncotype DX to predict benefit from chemotherapy

Data on ability of Oncotype DX to predict differential relative chemotherapy benefit is summarised in this section. Limitations of the chemotherapy benefit studies are summarised in Section 3.3. The EAG's overall view on chemotherapy benefit data is provided in Section 3.4.

Data on the ability of Oncotype DX to predict chemotherapy benefit comes mainly from two re-analyses of RCTs: one in LN0 patients (NSABP-B20; Paik 2006,[7] Tang 2011a[8]) and one in LN+ (SWOG-8814, Albain 2010[7, 8, 18]). In both, patients were randomised to endocrine monotherapy or endocrine plus chemotherapy. Summary results are provided in Table 5.

Relative and absolute benefit per risk group (adjusted and unadjusted): Both studies showed that unadjusted HRs for the effect of chemotherapy vs. no chemotherapy on survival and recurrence outcomes were most favourable in the higher-risk groups. HRs were generally statistically significant in high-risk groups but not in low- or intermediate-risk (). In the B20[7, 8] study (LN0), unadjusted HRs for 10-year distant recurrence-free interval (DRFI) in the low, intermediate and high-risk groups were 1.31, 0.61 and 0.26. HRs restricted to HER2- patients (adjusted and unadjusted) showed the same pattern (Table 5; not reported in journal article - provided via personal communication with Dr Tang via NICE). However, it is interesting to note that absolute differences (for chemotherapy vs. no chemotherapy) were very small in the low and intermediate-risk groups (1.1% and 1.8%, both favouring no chemotherapy), though greater in the high-risk group (27.6% favouring chemotherapy).

In SWOG-8814 (LN+),[18] DRFI was not reported. HRs for 10-year disease-free survival (DFS) for low, intermediate and high-risk groups, adjusted for number of positive nodes, were 1.02, 0.72 and 0.59.

Interaction tests (adjusted and unadjusted): Interaction tests indicate whether the difference in chemotherapy effect for a change in RS score is statistically significant. In B20 (LN0), the unadjusted interaction test for 10-year DRFI (for continuous RS score by chemotherapy) was reported as p=0.031[8] or p=0.038,[7] indicating a statistically significant difference in chemotherapy benefit as RS changes (Table 5). Interaction tests adjusted for clinicopathological factors were borderline significant for the full cohort (p=0.035, p=0.039 and p=0.068; difference due to method of assessing grade), while for the HER2- subgroup they were statistically significant (p=0.007, p=0.018 and p=0.022). The EAG report stated that it was unclear whether all factors were adjusted for simultaneously in B20; however, personal communication with the biostatistician (via NICE) confirms that this was the case.

In SWOG-8814 (LN+), the interaction test for 10-year DFS (for continuous RS score by chemotherapy; adjusted for number of nodes) was p=0.053 for all years and p=0.029 for years 0-5. Interaction tests adjusted individually for each of age, ethnicity, tumour size, grade, PR, P53 and HER2 were also statistically significant (p=not reported). Initially, the EAG interpreted this as a model including all clinicopathological variables; however, clarification from the authors in a personal communication to the EAG stated that each variable was included in a separate model. However, an interaction test adjusted for Allred-scored ER status was not significant (p=0.15). No interaction test was available that included all clinicopathological variables together.

Observational studies: Three observational studies had some data on chemotherapy benefit: two studies in patients with LN0 disease (MD Anderson[10, 19] and SEER[14, 20]) and one study in patients with LN+ disease (Clalit Health[21, 22]). Evidence was mixed and at high risk from confounding, since receipt of chemotherapy was influenced by Oncotype DX score, and patients receiving chemotherapy were likely to be at higher risk. Only one study (SEER) reported an interaction test; this was statistically significant (p=0.03), but only adjusted for grade, tumour size, age and race (omitting ER and PR).[13, 14] The other two studies only reported HRs for chemotherapy versus no chemotherapy in intermediate (MD Anderson and Clalit Health)[10, 11, 19, 21, 22] and high-risk patients (MD Anderson),[10, 19] and these were statistically non-significant, even after adjustment for confounders in one study.[10, 19]

Table 5: Prediction of chemotherapy benefit by Oncotype DX – Reanalyses of RCT data

Study	Outcome	% recurrence-free; absolute benefit			Hazard ratio for CT vs no CT (95% CI)			Interaction tests	Adjusted interaction tests
		Low	Intermediate	High	Low	Intermediate	High		
NSABP-B20 **LN0** ER+ N=651 Paik 2006[7] Tang 2011a[8] Personal comm.	**DRFI 10yr Unadjusted** **HER2- Unadjusted** **HER2- Adjusted[a]**	CT: 95.6% No CT: 96.8% Abs diff -1.1%	CT: 89.1% No CT: 90.9% Abs diff -1.8%	CT: 88.1% No CT: 60.5% Abs diff 27.6%	1.31 (0.46, 3.78), *p*=0.61 1.21 (0.41, 3.55), *p*=0.73 1.18 (0.40, 3.53), *p*=0.76[a]	0.61 (0.24, 1.59), *p*=0.39 0.78 (0.29, 2.11), *p*=0.62 0.67 (0.24, 1.87), *p*=0.44[a]	**0.26 (0.13, 0.53), p<0.001** **0.21 (0.08, 0.53), p<0.001** **0.20 (0.07, 0.52), p=0.001[a]**	Interaction (continuous RS) ***p*=0.031** or ***p*=0.038** (Tang 2011a[8] and Paik 2006[7])	Interaction[a] (continuous RS) adjusted for age, tumour size, grade, ER and PR: - All pts: ***p*=0.035, 0.039**, 0.068[b] - HER2-: ***p*=0.007, 0.018, 0.022** [b]
	DFS 10yr				0.91 (0.57, 1.45)	0.79 (0.43, 1.47)	**0.41 (0.23, 0.71)**	p=0.082	
	OS 10yr				1.37 (0.63, 3.01)	0.94 (0.4, 2.25)	**0.31 (0.16, 0.60)**	**p=0.011**	
SWOG-8814 **LN+** HR+ HER2+/- N=367 Albain 2010[18]	**DFS 10yr**	CT: 64% No CT: 60% Abs diff 4%		CT: 55% No CT: 43% Abs diff 12%	1·02 (0·54, 1·93); *p*=0·97 [c]	0·72 (0·39, 1·31); *p*=0·48[c]	0·59 (0·35, 1·01); ***p*=0·033**[c]		- Interaction (continuous RS) adjusted for positive nodes: All years: *p*=0.053 [c] 0-5 years: ***p*=0.029** [c] 5-10 years: *p*=0.58 [c] - Interaction (continuous RS) adjusted for each of age, ethnicity, size, grade, PR, P53, HER2: significant (p=NR). - Interaction adjusted for Allred-scored ER: p=0·15
	BCSS 10yr			CT: 73% No CT: 54% Abs diff 19%	*p*=0.56	*p*=0.89	***p*=0.033** [c]		
	OS 10yr			CT: 68% No CT: 51% Abs diff 17%	1·18 (0·55, 2·54, p=0·68)[c] p=0.63 log-rank	0·84 (0·40, 1·78, p=0·65) [c] p=0.85 log-rank	0·56 (0·31, 1·02, *p*=0·057) [c] ***p*=0.027** log-rank		Interaction (continuous RS) [c] All yrs: ***p*=0.026** 0-5 yrs: ***p*=0.016** 5-10 yrs: *p*=0.87

Data from Table 22 in EAG report. [a]Adjusted for age, tumour size, grade, ER and PR.[b]p-values correspond to analyses using different assessments of tumour grade. [C]Adjusted for number of positive nodes (1 to 3 vs. 4 or more)

3.3. Key limitations of studies assessing chemotherapy benefit

a) *Lack of data on chemotherapy benefit for the clinically intermediate-risk group:* NICE currently recommends Oncotype DX only for patients who are clinically intermediate-risk, for whom the chemotherapy decision is uncertain. This is a key subgroup for the economic modelling (defined as NPI>3.4). There are no data on the chemotherapy effect in patients who are Oncotype DX low-risk but clinically intermediate-risk. It is plausible that even if there is no chemotherapy benefit for clinically-low Oncotype DX-low patients, there could be benefit for clinically-intermediate (NPI>3.4) Oncotype DX-low patients.

b) *Statistical significance of interaction tests:* Most unadjusted interaction tests were statistically significant (Table 5). In terms of adjusted interaction tests, these were significant or borderline significant in B20 (LN0); and more clearly significant for the new HER2- subgroup (personal communication via NICE). One of the key concerns in the EAG report was that it was unclear whether all factors were adjusted for simultaneously in B20; however, personal communication with the biostatistician confirms that this was the case. This, along with the new HER2- subgroup analysis, provides stronger evidence for an interaction than presented in the EAG report.

However, in SWOG-8814 (LN+), it is now apparent after clarification from the lead biostatistician that interaction tests were adjusted for each clinicopathological factor individually (not all together, as initially thought by the EAG). All were individually significant except for the interaction test adjusted for Allred-scored ER status (p=0.15). As such, it remains unclear whether the interaction test would remain significant after adjustment for all relevant clinicopathological variables.

This also raises an interesting point as to whether results should be adjusted for ER status. On the one hand, test results should be adjusted to account for the effect of clinicopathological factors for which data are available in routine practice. On the other hand, it is not clear to what extent quantitative ER results are routinely available in UK practice, or their level of analytic validity; the SWOG-8814 author noted in his personal communication that performance of the Allred score is subject to some variability between pathologists. The author further stated that *"It is certainly possible that by including other measures of HER2, ER degree, Ki-67, grade, nodal size etc that one could make the interaction nonsignificant. However … you do get the benefit of most of those in a single well controlled measure (RS) rather than relying on separate assays for each with high known variability."* In other words, the benefit of Oncotype DX could be more accurate prognosis, rather than the prediction of chemotherapy benefit.

c) *Possible overestimation of chemotherapy benefit due to B20 being derivation study:* Patients from the no-chemotherapy arm of B20 were used to derive the Oncotype DX score. Therefore, Oncotype DX

may be overfitted in this study arm (i.e. recurrence rates may be artificially low in Oncotype low-risk patients and artificially high in Oncotype DX high-risk patients). This could lead to an overestimate of chemotherapy benefit since the chemotherapy arm was not used in derivation, therefore recurrence rates in this arm may show less separation between the low and high risk groups.

B14 (Paik 2004)[6] is a validation study of Oncotype DX (tamoxifen only; no chemotherapy arm). Comment 162 notes that the prognostic effect of Oncotype DX in the no-chemotherapy arm of B20 is greater than that in B14. As shown in Table 6, in the absence of chemotherapy, there is greater separation in B20 than B14; in other words, low-risk patients have a better 10-year recurrence-free rate in B20 (96.8%) than B14 (93.2%), while high-risk patients have a worse recurrence-free rate in B20 (60.5%) than B14 (69.5%).

In terms of prediction of chemotherapy benefit, B20 has a worse recurrence-free rate in the chemotherapy arm in low-risk patients (95.6% with chemotherapy vs. 96.8% without). This is counter-intuitive, and gives a corresponding HR greater than 1 (HR=1.31). However, comparing the chemotherapy arm of B20 (95.6% recurrence-free) with the no-chemotherapy arm of B14 (93.2% recurrence-free) indicates a small benefit in low-risk patients, though this breaks randomisation and may be affected by population differences between trials.

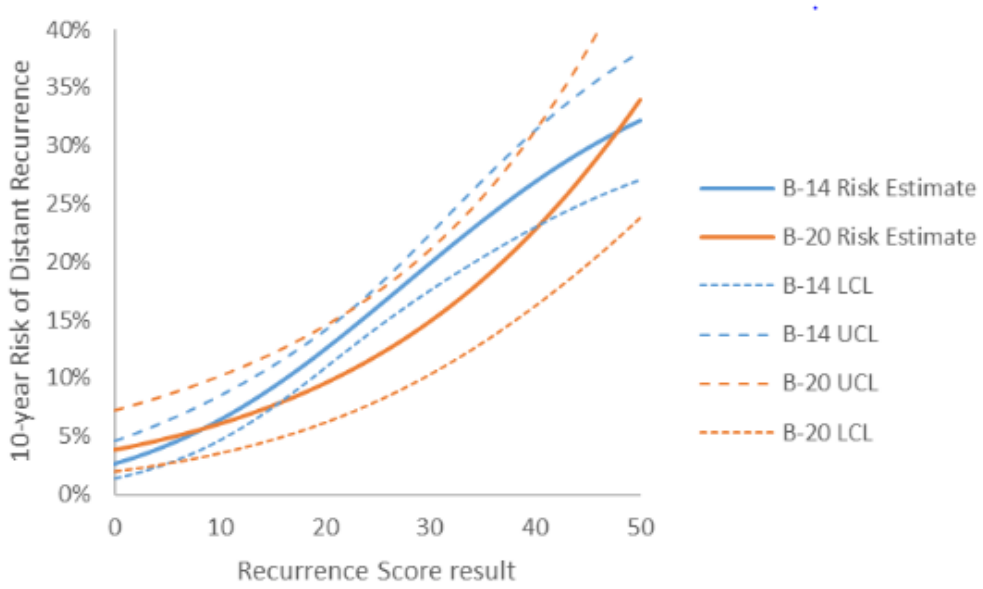
Additional data (personal communication with Dr Tang) compares the recurrence rates for a range of Oncotype DX scores in B14 and B20 (Figure 1). This analysis (which uses continuous Oncotype DX scores) is interpreted by Dr Tang as suggesting that the range of distant recurrence risk estimates, and slopes, are very similar between B20 and B14. However, the EAG still note that recurrence rates per risk group do appear to show greater separation in B20 than B14 (Table 6).

Table 6: Comparison of Oncotype prognostic ability in B14 and B20

Oncotype risk group	NSABP-B14 (Paik 2004)[6]		NSABP-B20 (Paik 2006[7])			
	Tamoxifen		Tamoxifen		Tamoxifen + chemotherapy	
	% patients per risk group (n)	% recurrence-free 10yr	% patients per risk group (n)	% recurrence-free 10yr	% patients per risk group (n)	% recurrence-free 10yr
Low	51% (388)	93.2%	60% (135)	96.8%	51% (218)	95.6%
Intermediate	22% (149)	85.7%	20% (45)	90.9%	21% (89)	89.1%
High	27% (181)	69.5%	21% (47)	60.5%	28% (117)	88.1%

Data from Table 12 in EAG report (also comment 161a in Comments on Diagnostics Consultation Document)

Figure 1: 10yr risk of distant recurrence in tamoxifen-alone groups: B20 and B14 (personal communication with Dr Tang, B20 study)



d) *Clinical relevance of chemotherapy benefit is unclear for the Oncotype DX intermediate-risk group:* Hazard ratios for chemotherapy benefit are available for this group, but it is unclear how they should be interpreted in clinical practice, i.e., would patients be treated, not treated, or would other clinicopathological variables be taken into consideration when making a decision?

e) *The number of events per subgroup is relatively low*, particularly for the B20 study (Table 7). Confidence intervals for the hazard ratios in low-risk and intermediate-risk groups are very wide in both B20 and SWOG-8814 (Table 5).

Table 7: Event rates for B14, B20 and SWOG-8814

Oncotype risk group	**Treatment**	**N events / N patients**		
		B14 (Paik 2004)[6] LN0	**B20 (Paik 2006)[7] LN0**	**SWOG-8814 (Albain 2010),[18] LN+**
Low	Chemo	-	10 / 218	26 / 91
Low	No chemo	28 / 338	5 / 135	15 / 55
Intermediate	Chemo	-	9 / 89	20 / 57
Intermediate	No chemo	25 / 149	7 / 45	22 / 46
High	Chemo	-	13 / 117	28 / 71
High	No chemo	56 / 181	18 / 47	26 / 47

3.4. EAG summary of evidence and limitations for prediction of chemotherapy benefit by Oncotype

Both B20 (LN0) and SWOG-8814 (LN+) showed that hazard ratios for chemotherapy vs. no chemotherapy were most favourable in the higher-risk groups, and were generally statistically significant in high-risk groups but not in low- or intermediate-risk groups. Unadjusted interaction tests

were statistically significant. Adjusted interaction tests were borderline significant in B20 (significant in HER2- patients), while in SWOG-8814 they were significant when adjusted for some clinicopathological variables individually, but not when adjusting for ER determined by Allred status.

Considering the limitations discussed above, the EAG considers that there remains uncertainty surrounding whether Oncotype DX is associated with a predictive benefit of chemotherapy (i.e. a difference in relative effect by genomic risk group), and if so, that there is uncertainty in the likely magnitude of this predictive effect within the clinical subgroups considered in this appraisal.

3.5. Observational studies showing low recurrence rates in test low-risk groups: to what extent does this bypass the issue of whether tests are predictive for chemotherapy benefit?

Some comments have noted the low recurrence rates within Oncotype low-risk groups in large observational studies. These are summarised in Table 4. LN0 patients with RS<18 have been reported as having a 5-year DRFS of 95.9%[10] and a 5-year DRFI of 99.5-99.6%.[10-14] For LN0-mic patients with RS<11/12, reported rates of 5-year DRFS, DRFI and BCSS range from 99.3-99.9%.[9, 12-14] The fact that TAILORx has not yet reported final results also indicates that recurrence rates are likely to be low.

Some commentators question whether these low recurrence rates in low-risk patients bypass the issue of whether tests are predictive for chemotherapy benefit. This is an important consideration. However, the EAG consider the following points to be important here:

a) The low-risk RS cut-off is currently 18 rather than 11 or 12, according to the NICE scope, the manufacturers, UK clinical practice, and NHS England Access Scheme data. Despite this, data using the RS<11/12 cut-point were included in the EAG clinical review for completeness.

b) NICE currently recommends Oncotype DX only for patients who are clinically intermediate-risk, for whom the chemotherapy decision is uncertain. This clinically-intermediate subgroup is a key subgroup for the economic modelling (defined as NPI>3.4). Conversely, the observational studies (as well as the reanalyses of RCTs) include a range of clinically low- and intermediate-risk patients. Patients who are RS low-risk but clinically intermediate-risk have a higher recurrence rate than the wider RS low-risk group, as shown in both TransATAC and B14 (see Table 3). The observational evidence may include patients who would not require an Oncotype DX test in UK clinical practice due to their low clinical risk, and may mask a subgroup of clinically-intermediate risk patients with higher recurrence rates.

c) The issue of predictive performance remains important for the modelling, because whether to accept the very different **relative** chemotherapy benefits between high-risk and low-risk patients (e.g. from the B20 study, with its limitations as discussed above) has a large impact on cost-effectiveness.

4. Risk of recurrence after 5 years

As noted in the EAG report, the assumptions employed in the model regarding the long-term risk of distant recurrence and the impact of chemotherapy are based on the earlier model reported by Ward *et al*[23] used in NICE DG10.[24] These assumptions are also applied in the Genomic Health model. As noted in the EAG's response to consultation on the assessment report, 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, 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. We undertook sensitivity analyses in which the risk tapering assumption is removed (see EAG report, Tables 139, 142, 145, 148 and 151); these sensitivity analyses indicate that removing the assumption of capped recurrence risk does not significantly impact upon the conclusions drawn from the analysis.

5. Adverse effects of chemotherapy

5.1 Additional EAG sensitivity analysis - Inclusion of additional adverse events

In response to the DCD, several commentators have criticised the EAG model for excluding long-term adverse events (AEs) associated with chemotherapy, for example, chronic heart failure (CHF), permanent alopecia and peripheral neuropathy. As noted in the original EAG report, CHF was excluded from the EAG model due to a lack of evidence on the joint survival impact of CHF and metastatic breast cancer.

Within this addendum, the EAG has undertaken exploratory analyses to assess the potential impact of including these potential late effects of chemotherapy on the cost-effectiveness of the tumour profiling tests.

Estimated lifetime QALY losses and costs associated with CHF were obtained from a re-analysis of the model previously developed as part of the OPTIMA-Prelim study (Hall *et al*[25]); this was one of a minority of studies identified within the EAG's review which included this late effect of chemotherapy. The lifetime impact of CHF was estimated using the Hall *et al* model by comparing two scenarios: (i) all patients receive adjuvant chemotherapy (including excess CHF risk), and; (ii) the excess CHF risk is set equal to zero (although background levels of CHF are still included).

In addition, the EAG has included additional disutilities associated with permanent alopecia and peripheral neuropathy, based on studies identified within a systematic review of studies reporting utility values associated with AEs of chemotherapy (Shabaruddin *et al*[26]). Of the range of potentially relevant disutilities reported in the review, studies were considered potentially relevant for inclusion in the exploratory analysis if they: (a) included a counterfactual state for comparison (i.e. the same state

without the AE), and (b) if the valuations were elicited from the general public (rather than from patients experiencing the AE or from health care practitioners acting as proxy for patients). The selected disutility for alopecia was based on a general population time trade-off (TTO) study of lung cancer states reported by Nafees *et al.*[27] The disutility for peripheral neuropathy was based on a general population TTO study of colorectal cancer states reported by Shiroiwa *et al.*[28]

These additional HRQoL and cost impacts were included in the EAG's model, based on the assumptions set out in Table 8. The results of the analysis are shown in Table 9.

Table 8: Additional assumptions included in EAG's sensitivity analysis

Adverse event	Incidence	Health loss	Cost
Acute myeloid leukaemia (AML)	0.49% at 10-years (Wolff *et al*[29])	Health state utility = 0.26	Lifetime cost £10,400
CHF	Based on excess CHF risk relative to that of the general population	Net lifetime QALY loss -0.0385 QALYs (Hall *et al*[25])	Net lifetime cost -£2 (Hall *et al*[25])
Alopecia	15% of all patients receiving chemotherapy (commentator opinion)	Disutility = -0.04495 (Nafees *et al*[27])	Cost not included in analysis
Peripheral neuropathy	12% of all patients receiving chemotherapy (commentator opinion)	Disutility = -0.02 (Shiroiwa *et al*[28])	Cost not included in analysis

Table 9: Central estimates of cost-effectiveness

Test	Scenario	NPI≤3.4	NPI>3.4	LN+ (1-3 nodes)
Oncotype DX	EAG base case	£120,144	Dominated	Dominated
	Additional AEs included	£121,270	£548,524	Dominated
IHC4+C	EAG base case	£2,752	Dominating	Dominating
	Additional AEs included	£1,735	Dominating	Dominating
Prosigna	EAG base case	£89,693	£25,857	£28,666
	Additional AEs included	£88,114	£25,277	£31,807
EPClin	EAG base case	£141,848	£46,482	£21,489
	Additional AEs included	£350,042	£46,310	£19,911
Test	**Scenario**	**MINDACT ITT**	**MINDACT high-risk**	**MINDACT low-risk**
MammaPrint	EAG base case	£134,059	Dominated	£399,182
	Additional AEs included	£59,193	Dominated	£848,869

As shown in Table 9, the economic conclusions drawn from the analyses are largely unchanged by the inclusion of these additional AEs, although the inclusion of alternative disutilities may lead to different results. The EAG has a number of concerns regarding the reliability of this additional exploratory analysis:

- The QALY losses and costs associated with CHF have been derived from a separate model (Hall *et al*[25]).
- The baseline health state utilities for the relapse-free and post-relapse states included in the EAG model (taken from Lidgren *et al*[30]) may already include a proportion of patients who are experiencing AEs at the time of HRQoL assessment.
- The Lidgren *et al* study[30] and the AE utility studies identified from the Shabaruddin *et al* review[26] relate to different hypothetical populations; the selected utility estimates for peripheral neuropathy and alopecia do not relate to breast cancer states.
- The available AE utility studies[26] typically use stated preference elicitation techniques (standard gamble or time trade-off), hence both the measurement and valuation of AEs within these studies are from individuals who do not have breast cancer and who have not experienced the AE under consideration. This is not ideal.
- As they are based on comparisons of hypothetical health state scenarios, it is unlikely that the disutilities from the AE utility studies include the possibility of amelioration or resolution of the AE under consideration. It is also unclear how to quantify the distribution of severity of the AEs resulting from chemotherapy within the analysis.

5.2 QALY shortfall analysis

In light of the uncertainties associated with the analysis presented in Section 5.1, the EAG undertook a further analysis which presents the QALY shortfall associated with each test achieving an ICER of £20,000 and £30,000 per QALY gained, based on the deterministic version of the EAG model (see Table 10, Table 11, Table 12, Table 13, Table 14 and Table 15). Other things being equal, this additional analysis may further inform the Appraisal Committee's deliberations around whether other factors which cannot be reliably quantified might have a sufficient impact on the ICERs of the tumour profiling tests to change the interpretation of the model results.

Within each analysis, the QALY shortfall represents the additional number of incremental QALYs that would need to be accrued, given the currently quantified estimates of the incremental QALYs gained for the test and its incremental cost, in order for each test to achieve an ICER at a particular threshold (λ=£20,000 per QALY gained or λ=£30,000 per QALY gained). In health economic terms, this QALY shortfall is equivalent to net clinical benefit. The Committee may find it useful to consider whether the expected magnitude of the health losses avoided by reducing chemotherapy use via tumour profiling tests which are not captured in the EAG model is likely to be equal to or greater than this estimated QALY shortfall. It should be noted that this analysis is predicated on the commentators' assumption that the adverse effects of chemotherapy have been underestimated in the EAG's model. However, the EAG model suggests that with the exception of IHC4+C, all tests increase chemotherapy use at least in

some subgroups (see EAG report, Appendix 7); where this is the case, changing the balance of the net health gains and losses of chemotherapy will produce less favourable ICERs for the tumour profiling tests. It should also be noted that any potential underestimation of QALY losses only apply to those patients who would have received chemotherapy and who would have experienced associated late effects who now do not receive chemotherapy due to the tumour profiling test result and thus avoid these late effects.

The QALY shortfall analysis operates as follows. As shown in Table 10, within the LN0 NPI>3.4 group, Oncotype DX (assuming prognostic benefit only) is estimated to lead to -0.02 QALYs and additional costs of £869 compared with no testing, hence it is expected to be dominated by no testing. In this subgroup, Oncotype DX would need to make up a further 0.06 QALYs in order to achieve an ICER of £20,000 per QALY gained given its incremental cost (£869 / [0.06+-0.02] = £20,000). Within this subgroup, the EAG model suggests that the probability of receiving chemotherapy is reduced by 16% due to the use of Oncotype DX. Assuming that 25% of these patients experience late effects of chemotherapy which are not accounted for within the EAG model, this means that 4% (0.16 x 0.25) of those forgoing chemotherapy will avoid late effects. Given the overall QALY shortfall of 0.06 QALYs and the probability of avoiding late effects of 0.04, this means that each patient who would have experienced a late effect of chemotherapy would have had to have lost 1.49 QALYs (0.06/0.04) due to that AE in order for Oncotype DX to be cost-effective at a threshold of £20,000 per QALY gained.

The results for this analysis are summarised below.

Oncotype DX (prognostic benefit assumed) – refer to Table 10

LN0, NPI≤3.4 – Analysis not relevant as more patients receive chemotherapy in the test group.

LN0, NPI>3.4 – Each patient who avoids chemotherapy and avoids experiencing a late AE not quantified in the EAG model would have to save 1.49 QALYs due to the unquantified AE in order for Oncotype DX to have an ICER of £20,000 per QALY gained. Assuming a threshold of £30,000 per QALY gained, the equivalent value is 1.12 QALYs per patient.

LN+ (1-3 nodes) – Each patient who avoids chemotherapy and avoids experiencing a late AE not quantified in the EAG model would have to save 1.44 QALYs due to the unquantified AE in order for Oncotype DX to have an ICER of £20,000 per QALY gained. Assuming a threshold of £30,000 per QALY gained, the equivalent value is 1.29 QALYs per patient.

Oncotype DX (predictive benefit assumed) – refer to Table 11

LN0, NPI≤3.4 – Analysis not relevant as more patients receive chemotherapy in the test group.

LN0, NPI>3.4 – Analysis not relevant as test dominates.

LN+ (1-3 nodes) – Analysis not relevant as test dominates.

IHC4+C – refer to Table 12

LN0, NPI≤3.4 – Analysis not relevant as ICER already below £20,000 per QALY gained.

LN0, NPI>3.4 – Analysis not relevant as test dominates.

LN+ (1-3 nodes) – Analysis not relevant as test dominates.

Prosigna – refer to Table 13

LN0, NPI≤3.4 – Analysis not relevant test increases chemotherapy use.

LN0, NPI>3.4 – Analysis not relevant test increases chemotherapy use.

LN+ (1-3 nodes) – Analysis not relevant test increases chemotherapy use.

EPClin – refer to Table 14

LN0, NPI≤3.4 – Analysis not relevant test increases chemotherapy use.

LN0, NPI>3.4 – Analysis not relevant test increases chemotherapy use.

LN+ (1-3 nodes) – Analysis not relevant at threshold of £30,000 per QALY gained as ICER is below this. Each patient who avoids chemotherapy and avoids experiencing a late AE not quantified in the EAG model would have to save 0.69 due to the unquantified AE in order for EPClin to have an ICER of £20,000 per QALY gained.

MammaPrint – refer to Table 15

MINDACT ITT - Each patient who avoids chemotherapy and avoids experiencing a late AE not quantified in the EAG model would have to save 2.03 QALYs due to the unquantified AE in order for MammaPrint DX to have an ICER of £20,000 per QALY gained. Assuming a threshold of £30,000 per QALY gained, the equivalent value is 1.23 QALYs per patient.

MINDACT high-risk - Each patient who avoids chemotherapy and avoids experiencing a late AE not quantified in the EAG model would have to save 1.39 QALYs due to the unquantified AE in order for MammaPrint to have an ICER of £20,000 per QALY gained. Assuming a threshold of £30,000 per QALY gained, the equivalent value is 1.11 QALYs per patient.

MINDACT low-risk - Analysis not relevant test increases chemotherapy use.

Table 10: QALY shortfall analysis - Oncotype DX (prognostic benefit only)

Oncotype DX (prognostic)	**LN0, NPI<3.4**	**LN0, NPI>3.4**	**LN+ (1-3 nodes)**
Inc. QALYs	0.01	-0.02	-0.07
Inc. costs	£1,317	£869	£647
ICER	£120,144	Dominated	Dominated
QALY shortfall to achieve ICER=£20,000/QALY gained	**0.05**	**0.06**	**0.10**
QALY shortfall to achieve ICER=£30,000/QALY gained	**0.03**	**0.04**	**0.09**
Proportion patients avoiding chemo due to testing	0.00	0.16	0.29
Proportion patients unaccounted AEs (assumption based on consultation responses)	0.25	0.25	0.25
Proportion patients tested avoiding chemo with unaccounted AEs	n/a - more get chemo in test group	0.04	0.07
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£20,000/QALY	**n/a - more get chemo in test group**	**1.49**	**1.44**
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£30,000/QALY	**n/a - more get chemo in test group**	**1.12**	**1.29**

Table 11: QALY shortfall analysis - Oncotype DX (predictive benefit)

Oncotype DX (predictive)	**LN0, NPI<3.4**	**LN0, NPI>3.4**	**LN+ (1-3 nodes)**
Inc. QALYs	0.04	0.27	0.09
Inc. costs	£1,211	-£364	-£68
ICER	£34,245	Dominating	Dominating
QALY shortfall to achieve ICER=£20,000/QALY gained	**0.03**	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**
QALY shortfall to achieve ICER=£30,000/QALY gained	**0.01**	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**
Proportion patients avoiding chemo due to testing	0.00	0.16	0.29
Proportion patients unaccounted AEs (assumption based on consultation responses)	0.25	0.25	0.25
Proportion patients tested avoiding chemo with unaccounted AEs	n/a - more get chemo in test group	0.04	0.07
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£20,000/QALY	**n/a - more get chemo in test group**	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£30,000/QALY	**n/a - more get chemo in test group**	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**

Table 12: QALY shortfall analysis - IHC4+C

IHC4+C	**LN0, NPI<3.4**	**LN0, NPI>3.4**	**LN+ (1-3 nodes)**
Inc. QALYs	0.01	0.01	0.05
Inc. costs	£22	-£89	-£269
ICER	£2,752	Dominating	Dominating
QALY shortfall to achieve ICER=£20,000/QALY gained	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**
QALY shortfall to achieve ICER=£30,000/QALY gained	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**
Proportion patients avoiding chemo due to testing	0.04	0.08	0.07
Proportion patients unaccounted AEs (assumption based on consultation responses)	0.25	0.25	0.25
Proportion patients tested avoiding chemo with unaccounted AEs	0.01	0.02	0.02
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£20,000/QALY	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£30,000/QALY	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**

Table 13: QALY shortfall analysis - Prosigna

Prosigna	**LN0, NPI<3.4**	**LN0, NPI>3.4**	**LN+ (1-3 nodes)**
Inc. QALYs	0.02	0.07	0.07
Inc. costs	£1,891	£1,713	£1,967
ICER	£89,693	£25,857	£28,666
QALY shortfall to achieve ICER=£20,000/QALY gained	**0.07**	**0.02**	**0.03**
QALY shortfall to achieve ICER=£30,000/QALY gained	**0.04**	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**
Proportion patients avoiding chemo due to testing	0.00	-0.01	-0.08
Proportion patients unaccounted AEs (assumption based on consultation responses)	0.25	0.25	0.25
Proportion patients tested avoiding chemo with unaccounted AEs	n/a - more get chemo in test group	n/a - more get chemo in test group	n/a - more get chemo in test group
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£20,000/QALY	**n/a - more get chemo in test group**	**n/a - more get chemo in test group**	**n/a - more get chemo in test group**
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£30,000/QALY	**n/a - more get chemo in test group**	**n/a - more get chemo in test group**	**n/a - more get chemo in test group**

Table 14: QALY shortfall analysis - EPClin

EPClin	**LN0, NPI<3.4**	**LN0, NPI>3.4**	**LN+ (1-3 nodes)**
Inc. QALYs	0.01	0.03	0.06
Inc. costs	£1,686	£1,401	£1,185
ICER	£141,848	£46,482	£21,489
QALY shortfall to achieve ICER=£20,000/QALY gained	**0.07**	**0.04**	**0.00**
QALY shortfall to achieve ICER=£30,000/QALY gained	**0.04**	**0.02**	**n/a - ICER already below threshold**
Proportion patients avoiding chemo due to testing	-0.07	-0.01	0.02
Proportion patients unaccounted AEs (assumption based on consultation responses)	0.25	0.25	0.25
Proportion patients tested avoiding chemo with unaccounted AEs	n/a - more get chemo in test group	n/a - more get chemo in test group	0.01
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£20,000/QALY	**n/a - more get chemo in test group**	**n/a - more get chemo in test group**	**0.69**
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£30,000/QALY	**n/a - more get chemo in test group**	**n/a - more get chemo in test group**	**n/a - ICER already below threshold**

Table 15: QALY shortfall analysis - MammaPrint

MammaPrint	**MINDACT ITT**	**MINDACT high-risk**	**MINDACT low-risk**
Inc. QALYs	0.01	-0.04	0.01
Inc. costs	£1,757	£1,380	£2,415
ICER	£134,059	Dominated	£399,182
QALY shortfall to achieve ICER=£20,000/QALY gained	**0.07**	**0.11**	**0.11**
QALY shortfall to achieve ICER=£30,000/QALY gained	**0.05**	**0.09**	**0.07**
Proportion patients avoiding chemo due to testing	0.15	0.33	-0.03
Proportion patients unaccounted AEs (assumption based on consultation responses)	0.25	0.25	0.25
Proportion patients tested avoiding chemo with unaccounted AEs	0.04	0.08	n/a - more get chemo in test group
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£20,000/QALY	**2.03**	**1.39**	**n/a - more get chemo in test group**
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£30,000/QALY	**1.23**	**1.11**	**n/a - more get chemo in test group**

6. Probability of having chemotherapy

Several commentators have suggested other potentially relevant decision impact studies could or should have been included in the EAG report. However, the studies suggested are either already included in the EAG report, or were excluded from the report with justification. The only exception to this is a study reported by Rodriguez *et al*; this study was not identified by the EAG searches, however, the results appear to be consistent with other Prosigna decision impact studies already included in the EAG review.

7. EAG systematic review and meta-analysis

All major comments relating to this theme are discussed in the EAG's table of responses.

8. EAG economic model

8.1. Re-analysis of MammaPrint by Agendia within the EAG model

Agendia have undertaken a re-analysis of the cost-effectiveness of MammaPrint using the EAG model *"with corrected usage of available MammaPrint data in those instances where we [Agendia] strongly disagree with the chosen inputs in the current model."* With respect to this analysis, the company claims that on the basis of altered model inputs, the ICER for MammaPrint is now less than £30,000 per QALY gained. However, the EAG notes that within the company's re-analysis, chemotherapy is assumed to be associated with no additional benefit in terms of DRFS for any patient population (including those with clinical-high MammaPrint-high risk). If this was the case, genomic testing would have no value as clinicians would never give chemotherapy to any patient. The EAG considers Agendia's re-analysis of the EAG model to be inappropriate and believes that the results are not meaningful.

8.2. Additional EAG sensitivity analysis - Cost-effectiveness of adjuvant chemotherapy by subgroup

During the consultation on the EAG report and the DCD, it has been suggested that the EAG model is predisposed to find giving chemotherapy to all patients a clinically effective and cost-effective use of resources. This interpretation of the model is inaccurate. In the interests of clarity, Table 16 presents the results of an analysis comparing 100% chemotherapy versus 0% chemotherapy using the EAG model. As shown in the table, the strategy involving the indiscriminate use of chemotherapy is dominated by the no chemotherapy option for patients with NPI≤3.4 (i.e. chemotherapy generates fewer QALYs at a greater cost). Chemotherapy appears to have a favourable clinical and cost-effectiveness profile within the LN0, NPI>3.4 and LN+ subgroups.

Table 16: Cost-effectiveness of chemotherapy versus no chemotherapy

Subgroup	Option	QALYs	Costs	Inc. QALYs	Inc. costs	ICER
LN0, NPI≤3.4	100% chemotherapy	13.83	£7,454	-0.04	£3,670	**Dominated**
	No chemotherapy	13.87	£3,784	-	-	-
LN0, NPI>3.4	100% chemotherapy	12.85	£11,700	0.27	£2,316	**£8,449**
	No chemotherapy	12.58	£9,384	-	-	-
LN+	100% chemotherapy	12.63	£12,668	0.35	£2,011	**£5,787**
	No chemotherapy	12.28	£10,658	-	-	-

8.3. Additional EAG sensitivity analysis - Alternative estimates of chemotherapy benefit within clinical risk subgroups

Several commentators have raised issues regarding the estimated relative risk of distant recurrence associated with chemotherapy. The original EAG report acknowledged that there is uncertainty around this estimate and notes that the estimated relative risk of 0.76 was calculated using the most relevant data reported within the EBCTCG 2011 meta-analysis paper[31] (data specifically relating to distant recurrence). The EAG notes that it is possible that the relative benefit of chemotherapy could be different between clinical risk groups, although the EBCTCG meta-analysis does not provide sufficient information to determine the relative risk of distant recurrence within each of the three model subgroups (LN-, NPI≤3.4; LN- NPI>3.4, and LN+[1-3 nodes]). Tables 139, 142, 145, 148 and 151 of the EAG report presented sensitivity analyses using values of 0.70 and 0.80 to explore the impact of this uncertainty on the cost-effectiveness of the tests; these limits are similar to reported rate ratios for any recurrence (including local and regional) for ER+ patients with N0/N- and N1-3 within the EBCTCG meta-analysis paper.

Within this addendum, the EAG has expanded this existing sensitivity analysis to reflect a broader range of relative risk estimates. As shown in Table 17, the economic conclusions drawn from the model for Oncotype DX, IHC4+C and MammaPrint are unaffected by these alternative values. Conversely, within the scenarios in which chemotherapy is assumed to be less favourable, the ICERs for Prosigna and EPClin are markedly less favourable in the LN0 NPI>3.4 and LN+ subgroups.

Table 17: Additional EAG sensitivity analysis - Alternative estimates of chemotherapy benefit within subgroups

Test	Scenario	ICER (per QALY gained)		
		LN0 NPI≤3.4	**LN0 NPI>3.4**	**LN+**
Oncotype	**Chemotherapy RR = 0.76 (EAG base case)**	**£120,144**	**Dominated**	**Dominated**
	Chemotherapy RR = 0.6	£69,967	Dominated	Dominated
	Chemotherapy RR = 0.7	£94,920	Dominated	Dominated
	Chemotherapy RR = 0.8	£145,102	Dominated	Dominated
	Chemotherapy RR = 0.9	£297,925	£201,602	Dominated
IHC4+C	**Chemotherapy RR = 0.76 (EAG base case)**	**£2,752**	**Dominating**	**Dominating**
	Chemotherapy RR = 0.6	£1,326	Dominating	Dominating
	Chemotherapy RR = 0.7	£2,138	Dominating	Dominating
	Chemotherapy RR = 0.8	£3,223	Dominating	Dominating
	Chemotherapy RR = 0.9	£4,745	Dominating	Dominating
Prosigna	**Chemotherapy RR = 0.76 (EAG base case)**	**£89,693**	**£25,857**	**£28,666**
	Chemotherapy RR = 0.6	£52,504	£13,975	£14,678
	Chemotherapy RR = 0.7	£71,107	£19,926	£21,508
	Chemotherapy RR = 0.8	£107,875	£31,645	£36,018
	Chemotherapy RR = 0.9	£214,907	£65,467	£87,917
EPClin	**Chemotherapy RR = 0.76 (EAG base case)**	**£141,848**	**£46,482**	**£21,489**
	Chemotherapy RR = 0.6	£65,750	£26,202	£11,702
	Chemotherapy RR = 0.7	£99,445	£36,317	£16,663
	Chemotherapy RR = 0.8	£195,508	£56,485	£26,089
	Chemotherapy RR = 0.9	£2,680,967	£116,586	£50,984
MammaPrint	**Scenario**	**MINDACT ITT**	**mAOL High risk**	**mAOL Low risk**
	Chemotherapy RR = 0.76 (EAG base case)	**£134,059**	**Dominated**	**£399,182**
	Chemotherapy RR = 0.6	£176,352	Dominated	£113,124
	Chemotherapy RR = 0.7	£148,424	Dominated	£161,338
	Chemotherapy RR = 0.8	£127,971	Dominated	£276,670
	Chemotherapy RR = 0.9	£112,346	£216,964	£920,361

RR – relative risk

9. Company economic models - new model submitted by Agendia

In response to the diagnostic consultation document, Agendia submitted a revised version of their model based on the MINDACT trial. The EAG has scrutinised this new analysis. The EAG notes that the model trace shows that the proportion of patients remaining alive and recurrence-free increases over time, whilst the proportion of the modelled cohort who are dead is allowed to decrease over time (see Figure 2 and Figure 3); this is clearly incorrect and as such the model lacks any face validity. In addition, whilst the company states that extrapolation has now been included in the model in order to account for longer-term costs and health impacts (assuming a constant event rate), the model trace indicates that no additional events occur between years 7 and 10. This also indicates major programming errors. On the basis of these errors, the EAG does not consider the company's new analyses to be reliable.

Figure 2: Probability of being alive, genomic test group, new Agendia model

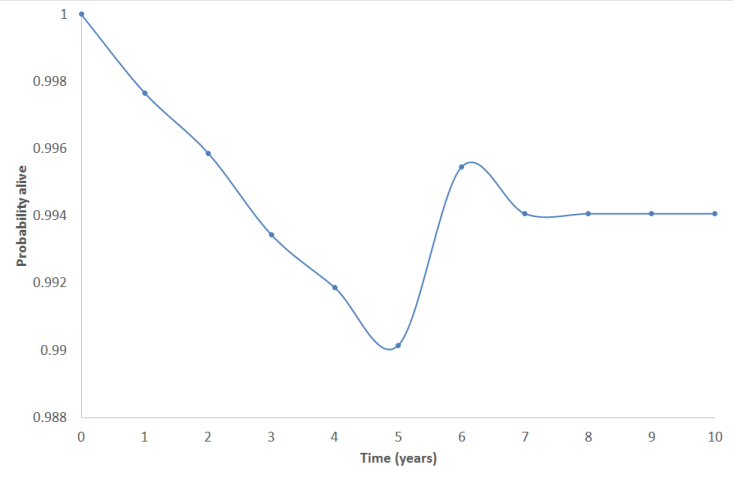
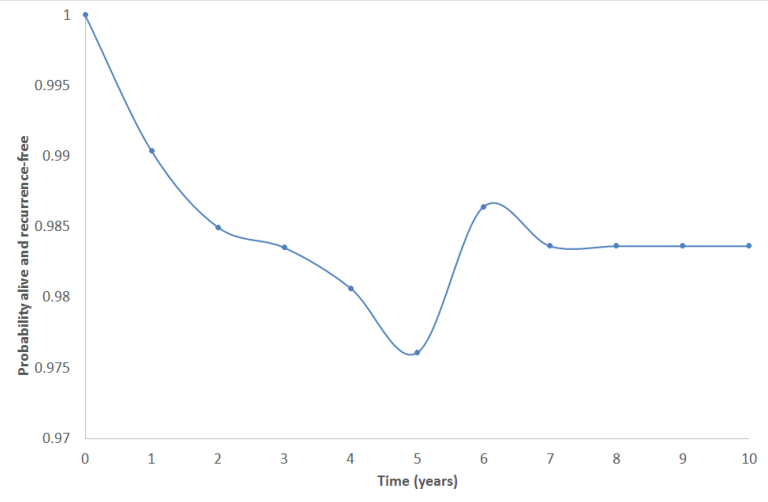


Figure 3: Probability of being alive and recurrence-free, genomic test group, new Agendia model



10. New commercial access schemes

Analyses based on company access proposals are included in a confidential addendum.

REFERENCES

1. National Institute for Health and Care Excellence. Diagnostics Consultation Document: Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer.; 2018.
2. Sestak I, Dowsett M, Cuzick J. NICE request - TransATAC data analysis. In; 2017.
3. Sestak I, Buus R, Cuzick J, et al. Comparison of the performance of 6 prognostic signatures for estrogen receptor–positive breast cancer: A secondary analysis of a randomized clinical trial. *JAMA Oncology* 2018; 10.1001/jamaoncol.2017.5524.

4. 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. 5480.

5. 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.

6. 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.

7. 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.

8. 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.

9. 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.

10. 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.

11. 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.

12. 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.

13. 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.

14. 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.

15. 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.

16. 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.

17. 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.

18. 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.

19. 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.

20. 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.

21. 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.

22. 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.

23. 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.

24. 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.

25. 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.

26. Shabaruddin FH, Chen LC, Elliott RA, Payne K. A systematic review of utility values for chemotherapy-related adverse events. *PharmacoEconomics* 2013;31:277-88.

27. Nafees B, Stafford M, Gavriel S, Bhalla S, Watkins J. Health state utilities for non small cell lung cancer. *Health & Quality of Life Outcomes* 2008;6.

28. Shiroiwa T, Fukuda T, Tsutani K. Health utility scores of colorectal cancer based on societal preference in Japan. *Quality of Life Research* 2009;18:1095-103.

29. 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.

30. 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.

31. 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.

Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer Addendum: Additional EAG analyses undertaken following the second appraisal committee meeting

Additional analyses of Oncotype DX versus usual practice including chemotherapy benefit based on naïve indirect comparisons of Study B20, Study B14 and TransATAC – LN0, NPI>3.4 subgroup

Table 1 presents estimated hazard ratios for chemotherapy versus no chemotherapy based on naïve indirect comparisons of Study B20, Study B14 and TransATAC.

Table 1: Hazard ratios for chemotherapy benefit by Oncotype DX risk score based on naïve indirect comparisons of Study B20, Study B14 and TransATAC

B20 versus B14			
Oncotype DX risk group	**No chemotherapy 10-yr DMFS**	**Chemotherapy 10-yr DMFS**	**Estimated HR**
Low	93.20%	95.60%	0.64
Intermediate	85.70%	89.10%	0.75
High	69.50%	88.10%	0.35
B20 versus TransATAC			
Oncotype DX risk group	**No chemotherapy 10-yr DMFS**	**Chemotherapy 10-yr DMFS**	**Estimated HR**
Low	94.90%	95.60%	0.86
Intermediate	87.70%	89.10%	0.88
High	77.20%	88.10%	0.49

Table 2 presents additional economic comparisons of Oncotype DX versus usual practice including chemotherapy benefit based on naïve indirect comparisons of Study B20, B14 and TransATAC. In each analysis, the modelled hazard ratio was calibrated against the estimates presented in Table 1. All analyses are based on the deterministic version of the EAG model.

Table 2: Additional analyses of Oncotype DX versus usual practice including chemotherapy benefit based on naïve indirect comparisons of Study B20, B14 and TransATAC – LN0, NPI>3.4 subgroup

Option	**QALYs**	**Costs**	**Inc. QALYs**	**Inc. costs**	**ICER (per QALY gained)**
Chemotherapy benefit based on indirect comparison of B20 and B14					
Oncotype DX	12.82	£10,664	0.03	£682	**£24,334**
No test	12.79	£9,981	-	-	-
Chemotherapy benefit based on indirect comparison of B20 and TransATAC					
Oncotype DX	12.74	£10,989	0.06	£525	**£8,150**
No test	12.68	£10,465	-	-	-

Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer:

EAG Addendum to responses to Comments on DCD2 (5 June 2018)

This document provides the EAG's responses to key themes in the Consultation Comments for the Diagnostics Consultation Document 2 (DCD2, April 2018) produced by NICE following the Diagnostics Appraisal Committee meeting on 14 March 2018.

1. Micrometastases

Some commentators questioned whether the tests might be useful in patients with micrometastases. Data relating to patients with micrometastases were very limited within the evidence base. Two studies relating to Oncotype DX(1-4) were identified that reported micrometastases subgroups, whilst the EAG identified no studies relating to the other tests that reported micrometastases data on its own.

Oncotype DX

The EAG identified two observational studies where patients were treated according to Oncotype DX score and local clinical practice.

1) Clalit Health Services study(1, 2) – reported 5 year DRFI and BCSS, for different cut points (11-25 and 18-30). [redacted]

At the cut point 18-30, Clalit Health reported DRFI for the subgroups LN0-1mic, LN1mic and LN1mic-LN3. Analyses were unadjusted and confounded by treatment with chemotherapy, which was lesser in the LN0-1mic group than in the LN1mic-LN3 group in RS<18 and RS18-30 risk groups, and similar in the RS>30 risk group (see column 5 in Summary

The data is uncertain due to high risk of confounding. LNmic patient outcomes were more similar to LN0 or LN0-1mic patients than to LN1mic-LN3 patients in RS<18 groups, more similar to LN1mic-LN3 patients in RS>30 groups, and variable in RS18-30.

Table 1). It is unclear how much chemotherapy LNmic patients received. DRFI in LN1mic-LN3 patients is likely to be improved by the greater use of chemotherapy, narrowing the difference between LN0-1mic and LNmic-LN3 groups.

However, it can be seen that LNmic low risk patients have DRFI similar to low risk LN0-mic patients, whilst LNmic intermediate and high risk patients have DRFI similar to LN1-3 high risk patients. Surprisingly, LNmic patients had worse DRFI than LN1mic-LN3 at intermediate and high RS scores, perhaps suggesting under-treatment of these patients.

2) SEER registry(3, 4) only reported BCSS using cut points 18-30, and had less than 5 years follow-up.

In the SEER registry analysis, which considers BCSS, the same problems are evident in terms of a lack of adjustment and differential chemotherapy use in LN0 versus LNmic-LN3 patients which mean the data is at high risk of confounding. In addition, the LN0 group is limited to ages 40-84 whereas the LN1mic-LN3 patients are not limited by age.

BCSS was similar in RS<18 and RS 18-30 groups for LN0, LN1mic and LNmic-LN3 groups, but LN1mic was more similar to LN1mic-LN3 in the RS>30 groups than to LN0.

Summary

The data is uncertain due to high risk of confounding. LNmic patient outcomes were more similar to LN0 or LN0-1mic patients than to LN1mic-LN3 patients in RS<18 groups, more similar to LN1mic-LN3 patients in RS>30 groups, and variable in RS18-30.

Table 1 Oncotype DX data on micrometastases

Study	Study design	Patients	Subgroup, N	Chemo per group	Cut off	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)
DRFI– 5 year										
Cut off 18-30										
Clalit Health Services Stemmer 2016(1) Stemmer 2016(2)	R	ER+, HER2-, had O-DX test	LN0-1mic N= 1,594(2)	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	NR
			LN1mic N =270(1)	RS<18: 7% RS18-30: 40% RS>30: 90%	18-30	**99.3** (NR)	**89.2** (NR)	**80.6** (NR)	NR	NR
			LN1mic – LN3 N=627(1)			96.8 (NR)	93.4 (NR)	83.6 (NR)		
Cut off 11-25										
Clalit Health Services Stemmer 2016(1)	R	ER+, HER2-, had O-DX test	**LN1mic** N=270(1)	RS<11: 7% RS11-25: 18% RS >25: 81%	11-25	**97.8** (NR)	**95.9** (NR)	**83.9** (NR)	NR	NR
			LN1mic – LN3 N =627(1)			95.1 (NR)	96.1 (NR)	86.8 (NR)		
BCSS – actuarial 5 year										
Cut off 18-30										
SEER registry Petkov 2016(3) Roberts 2016(4)	R	HR+, HER2-[m]	LN0 40-84 years of age, N =38,568	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(6)	NR		**98.9** (97.4, 99.6)	**99.1** (97.9, 99.6)	**84** (74.1, 90.4)	NR	NR
			LNmic-LN3 All ages, N =4691	RS <18: 23% RS 18-30: 47% RS >25: 75%		99.0 (98.0, 99.5)[n]	97.7 (95.9, 98.7)	85.7 (76.2, 91.6)	p<0.001	NR

References

1. 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(no pagination).
2. 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(4 SUPPL. 1).
3. 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.
4. 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(no pagination).
5. 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. 2017.
6. 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 Res Treat. 2017;163(2):303-10.

2. Critique of the new Agendia model (MammaPrint versus mAOL)

The EAG has scrutinised the new Agendia model of MammaPrint versus mAOL. This process has led to the identification of a number of important errors in the new model which render the results presented by Agendia invalid. These issues are discussed in the subsequent section.

(a) Errors and other issues identified within the new Agendia model

(1) Unconventional approach to half-cycle correction

The company's approach to half-cycle correction is unconventional and appears to assume that patients can only die if they have previously developed distant metastases. This assumption is not correct. The EAG notes that the company's overall survival estimates are unaffected by this assumption and the predicted survival estimates within each clinical-genomic risk group appear to be correct. However, the QALY gains are impacted. The impact of removing this assumption has not been tested by the EAG.

(2) Patients who have already died in previous cycles contribute to total QALYs in the current cycle

In the clinical low, genomic low group (worksheet "MODEL", cells AL32:AL40), the QALY calculation adds a QALY contribution from patients who have already died during previous cycles, rather than just those new patients who die within the current cycle. This is an error. The other three risk groups (clinical low, genomic high; clinical high, genomic low and clinical high, genomic high) are not affected, but are subject to a further error (see issue #3 below).

(3) The QALY calculations are different between the clinical-genomic risk subgroups

The approach used to calculate per-cycle QALY contributions is not the same between risk groups (e.g. calculations in worksheet "MODEL" column AZ). The calculations in cells AZ40:AZ49 (the mAOL group) draw on the number of new deaths, whilst those in cells AZ21:AZ30 (the MammaPrint group) do not. This appears to be an error. It is unclear what the company intended, although the EAG considers it likely that this approach reflects an inappropriate method for including a half-cycle correction.

(4) The QALY gains for patients in some clinical-genomic risk groups are counted 1.5 times each cycle

For some, but not all, of the clinical-genomic risk subgroups, the QALY gains are counted 1.5 times during each cycle. For example, the formula in cell AZ21 is *"=(((AR21*(1-AE_sec_prim))*u_CL_GH_ACT)-(du_CT*AR21)+(AR21*AE_sec_prim*u_AML)+(AS21*u_DM)+(AS21*u_DM*0.5))/(1+oDR)^cycle"*. This formula includes the per-cycle QALY contribution for those with progressed disease (cell AS21) 1.5 times. This is an error which applies to the discordant clinical-genomic risk groups in which chemotherapy is given. It is unclear what the company intended, although the EAG considers it likely that this approach reflects an inappropriate method for including a half-cycle correction.

(5) *The costs for patients in some clinical-genomic risk groups are counted 1.5 times each cycle*

For some, but not all, of the clinical-genomic risk subgroups, the costs are counted 1.5 times during each cycle. For example, the formula in cell AX21 is *"=((AR21*c_monitoring2)+(AR21*((p_ET_CL_GH)*c_ET))+(AR21*AE_CHF*c_CHF)+(AR21*AE_sec_prim*c_AML)+(AS21*c_DM)+(AS21*c_DM*0.5))/(1+cDR)^cycle".* This formula includes the per-cycle cost contribution for those with progressed disease (cell AS21) 1.5 times. This is an error which applies to the discordant clinical-genomic risk groups in which chemotherapy is given. Again, it is unclear what the company intended, although the EAG considers it likely that this approach reflects an inappropriate method for including a half-cycle correction.

(6) Very high per-cycle probability of AML and application of lifetime AML cost each cycle

The new Agendia model includes a per-cycle probability of AML following chemotherapy of 0.012 and applies the lifetime cost of treating AML to these patients during each cycle. According to the new Agendia model, the source used to inform this parameter is the MINDACT trial, although the EAG was unable to locate this value from the Cardoso trial paper or the accompanying supplementary material.[1] Based on the company's new model, this per-cycle AML probability results in around 13% of all patients developing AML by 10-years. The EAG considers this to be very high and notes that it is approximately 25 times higher than the probability applied in the EAG model, based on Wolff *et al*[2] (0.49% develop AML at 10-years). The EAG considers it likely that this value has been miscalculated and that it represents an error. The EAG notes that this value was applied in the company's original model, but was not identified as a major issue by the EAG due to the presence of other serious programming errors which invalidated the model results.

(7) Double counting of the HRQoL impact of chemotherapy-related adverse events

The company's new model includes clinical-genomic subgroup-specific health state utility values for the DMFS state during the first model cycle. The EAG believes that these differences between risk subgroups should account for disutilities associated with chemotherapy-related adverse events. The apparent source of these estimates is the MINDACT trial.[1] During all subsequent cycles, the utility for the DMFS state is based on Lidgren *et al*[3] (the source and value used in the EAG model). As the company also include separate parameters relating to the disutility associated with adjuvant chemotherapy, the EAG considers that simultaneous use of the MINDACT utilities does not make sense and leads to double-counting.

(8) Inconsistent assumptions regarding the magnitude and duration of chemotherapy-related adverse events

The EAG model applies a QALY loss of 0.038 for all patients receiving adjuvant chemotherapy during the first cycle. This estimate was taken from Campbell *et al*[4] and is assumed to relate to the first year

after starting chemotherapy, based on text reported in the paper. The company's new model applies a QALY loss of 0.076 (double the value reported by Campbell *et al*) during the first year and applies a second disutility of 0.038 during the second year. Noting the company's response to the second DCD, which states that the disutility is applied only in the first year (see table of responses to DCD2, comment 47), the EAG considers the disutilities applied in the company's new model to be incorrect.

(9) The cost of the MammaPrint test is partially included in the usual care (no testing) group and only partially accounted for in the MammaPrint group

The company's new model adopts a hybrid decision-tree Markov approach, based on the concordant and discordant clinical-genomic risk subgroups. This approach is different to the original model critiqued within the EAG report. The company's new model now includes parameters relating to the probability that clinical/genomic high-risk patients do/do not receive chemotherapy and that clinical/genomic low-risk patients do/do not receive chemotherapy. For those discordant patients who are assumed not to follow the test in the MammaPrint group, some costs and outcomes from the Markov sub-models of the usual care (no test) group are used. As a consequence of the way the model is programmed, this means that the cost of the MammaPrint test is not included in the MammaPrint group for those patients who are clinical low, genomic high risk but do not get chemotherapy or for those patients who are clinical high, genomic low risk and do get chemotherapy. These reflect errors.

For discordant patients in the usual care group who do not follow the chemotherapy decision indicated by their clinical risk level, the model now uses costs and outcomes from the Markov sub-models of the MammaPrint group. As a consequence of the way the model is programmed, this means that the cost of the MammaPrint test is partially included in the usual care group. This can be seen by changing the cost of MammaPrint to any alternative value – erroneously, this changes the total cost for the usual care group. This is an error.

(b) Additional EAG analysis exploring the impact of correcting errors in the new Agendia model

The EAG has attempted to rectify as many errors as possible in order to generate more reliable estimates of the cost-effectiveness of MammaPrint using the company's new model. The following analyses were undertaken:

(1) The QALY contribution of previously dead patients was removed and changed to reflect QALY contributions of those patients dying in the current cycle.
(2) The cost of the test was applied fully to the MammaPrint group and was removed from the usual care group.
(3) The probability of developing AML was divided by 25, thereby approximately reflecting the estimate applied in the EAG model.

(4) The half-cycle correction attempted by the company was removed. The EAG notes that it is better to exclude the half-cycle correction altogether than to include an adjustment which is known to be incorrect.
(5) The clinical-genomic risk group-specific utility values for the DMFS state in the first cycle were replaced with the utility value for DMFS (0.824).
(6) The disutility associated with chemotherapy in year 1 was set equal to 0.038. The disutility associated with chemotherapy in year 2 was set equal to zero.
(7) Analyses (1)-(6) were combined.

The results of the EAG's corrections are shown in Table 2. Each row of the table shows each individual correction; the final row shows the impact of all corrections combined. As shown in the table, the EAG's corrections to the new Agendia model suggest that MammaPrint is dominated by usual care in both the overall MINDACT population and the clinical high-risk subgroup.

Table 2: Results of the EAG's corrections to the new Agendia model

Scenario number	Scenario description	MINDACT ITT population			Clinical high-risk subgroup (based on mAOL)		
		Incremental QALYs	**Incremental costs**	**Incremental cost per QALY gained**	**Incremental QALYs**	**Incremental costs**	**Incremental cost per QALY gained**
-	Agendia new base case	0.0054	£687	**£126,104**	0.0198	-£928	**Dominating**
1	Remove QALY contribution of previously dead	0.0054	£687	£126,104	0.0198	-£928	Dominating
2	Apply MammaPrint cost fully to MammaPrint group, remove test cost from usual care	0.0054	£1,018	£186,893	0.0198	-£618	Dominating
3	Divide per-cycle AML probably by 25	-0.0014	£812	Dominated	0.0033	-£625	Dominating
4	Remove half-cycle correction	0.0121	£953	£78,805	0.0340	-£373	Dominating
5	All DMFS utilities=0.824 for all risk subgroups	0.0040	£687	£170,646	0.0169	-£928	Dominating
6	Apply chemotherapy-related AE QALY loss of 0.038 in first year only	-0.0038	£687	Dominated	-0.0025	-£928	Dominated
7	All EAG corrections combined	-0.0054	£1,410	**Dominated**	-0.0076	£240	**Dominated**

References

1. 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.
2. 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.
3. 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.
4. 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.

TAILORx (Sparano 2018): EAG summary of key points

This document provides the EAG's initial summary of key findings and implications of TAILORx, reported by Sparano et al. 2018.[1] This is based on a brief assessment of the article which has only just reported and may be subject to revision.

Trial design

- Patients with RS 0-10 received endocrine therapy, those with RS 26+ received chemo-endocrine therapy, while those with RS 11-25 were randomised to endocrine or chemo-endocrine therapy
- TAILORx used different RS cut-offs to those currently used (current cut-offs: 0-17, 18-30, 31+)
- Patients were HR+ HER2- LN0, and "met NCCN guidelines for recommendation or consideration of chemotherapy". In the randomised group, 73-74% were clinically low-risk according to modified AOL, so it is likely a similar proportion may have not been eligible for chemotherapy in UK
- Genomic Health funded part of the study and contributed to the manuscript

Key findings

- Across all patients, there was no clinically relevant (or statistically significant) difference between endocrine and chemo-endocrine therapy in patients with RS 11-25
    - Primary endpoint of 9-year invasive disease-free survival (IDFS): 84.3% with chemo, 83.3% without chemo, absolute difference 1.0%, ITT HR 1.08 (0.94 to 1.24), p=0.26 (similar for as-treated analysis). Upper CI was within non-inferiority margin (pre-specified: HR=1.322)
    - Freedom from distant recurrence (DRFI) at 9yr: 95% with chemo, 94.5% without chemo, absolute difference 0.5%, ITT HR 1.10 (0.85 to 1.41), p=0.48 (as-treated analysis similar)
- Prognostic ability: Continuous RS was associated with DRFI (i.e. DRFI increased as RS increased)
- **Exploratory** subgroup analyses were conducted; these did not appear to account for stratification factors and should be treated with caution. These include the following:
- Chemotherapy effect in different **RS subgroups**:
    - No significant interaction between chemotherapy treatment and RS (p=not reported), implying that the HR for chemotherapy did not differ significantly between RS subgroups (within RS 11-25)
    - However, some subgroups (e.g. RS 21-25 for IDFS) did show a significant chemotherapy effect. It is difficult to be certain that there was no effect of chemotherapy in any RS group
    - Chemotherapy effect was not assessed outside RS 11-25 as patients were not randomised.
- Chemotherapy effect when varying other factors:
    - Significant interactions between chemotherapy treatment and **age** on IDFS and RFI but not on DRFI, with patients aged ≤50 years showing a significant effect of chemotherapy
    - No significant interactions between chemotherapy treatment and tumour size, grade, clinical risk, or menopausal status (but trend for greater effect in pre-menopausal patients)
    - Article notes that the greater effect of chemotherapy in younger / premenopausal patients may be partly due to anti-estrogenic effect of chemotherapy-induced menopause
- Chemotherapy effect when varying **RS** and **age/menopausal** status:
    - In women aged ≤50 years with higher RS, chemotherapy had a significant effect (RS 21-25 all outcomes and RS 16-20 some outcomes). The effect of chemotherapy varied significantly between combinations of age and RS group for IDFS (p=0.004) but not for DRFI or RFI. This suggests that for patients ≤50 years, there was some evidence for a difference in chemotherapy effect between RS subgroups, but this was based on exploratory analyses

- Similarly, for pre-menopausal patients, chemotherapy had a significant effect for RS 21-25 (on some outcomes) though this pattern was not consistent across outcomes
- In women older than 50 years, there was little effect of chemotherapy overall, but those with age 51-65 and RS 21-25 had HR for chemo of 1.38 (0.94 to 2.03) for IDFS

Implications

- Across all patients, there was no clinically relevant (or statistically significant) difference between endocrine and chemo-endocrine therapy in patients with RS 11-25. However, exploratory subgroup analyses suggest chemotherapy may have an effect in some subgroups, such as RS 21-25 and possibly RS 16-20, particularly in those aged ≤50 years. Some subgroups had upper CIs above the non-inferiority margin (though numbers were small)
- For patients with RS 11-15, there was no clear effect of chemotherapy in any subgroup shown
- In terms of prediction of differential chemotherapy benefit, there was no significant interaction between chemotherapy treatment and RS, implying that the HR for chemotherapy did not differ significantly between RS subgroups (within RS 11-25). However, subgroup analyses indicated significant effects in some higher RS groups
- 73-74% of randomised patients were clinically low-risk via modified AOL; it is likely a similar proportion may not be eligible for chemotherapy in UK. There was no chemotherapy effect in either low or high clinical risk (mAOL) subgroups, though these were not subgrouped by RS

Detailed summary

Population and treatment arms
The population and treatment arms in TAILORx are shown in Table 1.

Table 1: Population and treatment arms

Population	Arm	RS score	Treatment	N (ITT)	Low clinical risk (mAOL)
HR+ HER2- LN0 Met NCCN guidelines for recommendation or consideration of chemotherapy	A	0-10	ET	1619	78%
	B	11-25	ET	3399	74%
	C	11-25	ET+CT	3312	73%
	D	26+	ET+CT	1389	43%

Main results: No significant effect of chemotherapy overall, for patients with RS 11-25
When comparing patients with RS 11-25 randomised to chemo-endocrine therapy vs. endocrine therapy alone, there was no significant effect in terms of freedom from distant recurrence (DRFI), invasive disease-free survival (IDFS), freedom from distant or loco-regional recurrence (RFI) and overall survival (OS), at 5 years and at 9 years. Results at 9 years are shown in Table 2.

When including all four RS and treatment groups (randomised and non-randomised), there were significant differences in the rates of IDFS, recurrence, and death (P<0.001), driven largely by the higher likelihood of having an event in the cohort with RS 26+ (data not shown in this document).

Table 2: Main results

RS score	9yr (ITT)				9yr (as-treated)
	ET	ET+CT	Abs diff	HR (95% CI), p-value	HR (95% CI), p-value
Freedom from distant recurrence (DRFI) (%)					
0-10	96.8				
11-25	94.5	95.0	0.5%	1.10 (0.85 to 1.41), p=0.48	1.03 (0.80 to 1.33), p=0.81
26+		86.8			
Invasive disease-free survival (IDFS) (%)					
0-10	84.0				
11-25	83.3	84.3	1.0%	1.08 (0.94 to 1.24), p=0.26	1.14 (0.99 to 1.31), p=0.06
26+		75.7			
Freedom from distant or loco-regional recurrence (RFI) (%)					
0-10	95.0				
11-25	92.2	92.9	0.7%	1.11 (0.90 to 1.37), p=0.33	1.12 (0.91 to 1.38), p=0.28
26+		84.8			
Overall survival (OS) (%)					
0-10	93.7				
11-25	93.9	93.8	-0.1%	0.99 (0.79 to 1.22), p=0.89	0.97 (0.78 to 1.21), p=0.78
26+		89.3			

Prognostic ability of RS within the RS 11-25 subgroup
Distant recurrence was associated with RS as a continuous variable between RS 11 and 25 (no other information reported; not stated whether this is across patients receiving any treatment or separately for those receiving endocrine or chemo-endocrine therapy).

Differences in chemotherapy effect within subgroups

Differences in chemotherapy effect by RS subgroup: There was a significant effect of chemotherapy (vs. endocrine therapy alone) in the highest RS group for 9yr IDFS (RS 21-25; Figure 1). There was no significant effect of chemotherapy for DFRI or RFI in any RS subgroup. There may be a non-significant trend for greater chemotherapy effect with greater RS.

However, there were **no** significant interactions between **chemotherapy** treatment and **recurrence score** within the RS 11-25 range. (This was compared between RS 11 to 15 vs. 16 to 20 vs. 21 to 25 subgroups, and also between 11 to 17 vs. 18 to 25 subgroups.) In other words, the effect of chemotherapy (vs. endocrine therapy alone) did not differ significantly between RS subgroups.

These results are only for the RS 11-25 range. Chemotherapy effect could not be assessed for patients outside this range as they were not randomised.

Figure 1: Effect of chemotherapy by RS subgroup (within RS 11-25) (hazard ratios)

DRFI (9yr, ITT)
Figure removed for copyright but data can be found in Sparano et al. 2018.[1]

IDFS (9yr, ITT)
Figure removed for copyright but data can be found in Sparano et al. 2018.[1]

RFI (9yr, ITT)
Figure removed for copyright but data can be found in Sparano et al. 2018.[1]

Differences in chemotherapy effect by age: Chemotherapy (vs. endocrine therapy alone) showed a significant effect in patients aged ≤50 years for RFI and IDFS; this was borderline non-significant for DRFI (Figure 2). There **was** a significant interaction between **chemotherapy** treatment and **age** for IDFS (p=0.03) and RFI (p=0.02) but not DRFI (p=0.12).

Figure 2: Effect of chemotherapy by age (within RS 11-25) (hazard ratios)

DRFI (9yr, ITT)
Figure removed for copyright but data can be found in Sparano et al. 2018.[1]

IDFS (9yr, ITT)
Figure removed for copyright but data can be found in Sparano et al. 2018.[1]

RFI (9yr, ITT)
Figure removed for copyright but data can be found in Sparano et al. 2018.[1]

Differences in chemotherapy effect by other factors: There were no significant interactions between chemotherapy treatment and the following: tumour size, grade, clinical risk, menopausal status. However, pre-menopausal patients showed a trend towards greater chemotherapy effect than post-menopausal patients (data not shown in this document).

Differences in chemotherapy effect by age and RS subgroup: In women aged ≤50 years, chemotherapy had a significant effect for RS 21-25 for all three outcomes reported (Figure 3); a significant effect for RS 16-20 for IDFS and RFI (but not DRFI), but no significant effect for RS 11-15 on any outcome. Effects for overall survival were stated to be similar (no data reported).

In terms of statistical interactions, the effect of chemotherapy varied significantly between the nine combinations of age and RS group for IDFS (p=0.004) but not for DRFI or RFI (p=not reported). For the age ≤50 group alone, it is not reported whether there was a significant interaction between chemotherapy treatment and RS subgroup.

These results suggest that for the age ≤50 subgroup (which was the only age group showing a significant effect of chemotherapy), there was some evidence for a difference in chemotherapy effect between RS subgroups within the RS 11-25 range, but this was not conclusive. Chemotherapy effect could not be assessed for patients outside the 11-25 range as they were not randomised.

For the age older than 50 subgroup, the majority of patients showed little effect of chemotherapy; however those with age 51-65 and RS 21-25 had HR for chemo of 1.38 (0.94 to 2.03) for IDFS (not statistically significant, but point estimate HR above the non-inferiority margin).

Figure 3: Effect of chemotherapy by age and RS group (within RS 11-25) (hazard ratios)

DRFI (9yr, ITT)
Figure removed for copyright but data can be found in Sparano et al. 2018.[1]

IDFS (9yr, ITT)
Figure removed for copyright but data can be found in Sparano et al. 2018.[1]

RFI (9yr, ITT)
Figure removed for copyright but data can be found in Sparano et al. 2018.[1]

Differences in chemotherapy effect by menopausal status and RS subgroup:

Within pre-menopausal patients (who showed a trend towards greater chemotherapy effect than post-menopausal patients), chemotherapy effect was significant for RS 21-25 only (for DRFI and RFI); however for IDFS the chemotherapy effect was greater for RS 16-20 (data not shown within this document).

In terms of statistical interactions, effect of chemotherapy varied significantly over the six combinations of menopausal status and RS category for IDFS (p=0.02) but not for DRFI or RFI (p-not reported).

These results suggest that for the premenopausal subgroup, there was some evidence for a difference in chemotherapy effect between RS subgroups within the RS 11-25 range, but this was not conclusive.

[1] Sparano JA, Gray RJ, Makower DF, et al. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. N Engl J Med. 2018. doi: 10.1056/NEJMoa1804710.

Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer:

EAG Addendum following request from NICE after the 3rd committee meeting (13th June 2018)

Following the 3rd Committee Meeting on 13th June, NICE requested the following additional work from the EAG:

A. Repeating the sensitivity analysis which varied the relative risk of distant recurrence (0.76 in the base case) between 0.6 and 0.9 with the confidential access proposal test costs included (Oncotype DX, EndoPredict, Prosigna)
B. Summarising what data TAILORx does and doesn't provide in terms of the current EAG model
C. An exploratory analysis of Oncotype DX incorporating predictive benefit of chemotherapy using relative risks of recurrence informed by TAILORx
D. Summarise whether key studies in the DAR included or excluded patients with micrometastatic disease.

A: Sensitivity analyses around the relative risk of distant metastases for EPClin and Prosigna

(i) Additional analyses for EPClin – alternative sensitivity analyses around relative risk of distant metastases including access proposals

Table 1 presents the results of additional sensitivity analyses around the relative risk of distant metastases for chemotherapy versus no chemotherapy for EPClin (including the access proposals) versus usual practice.

Table 1: EPClin additional analyses – sensitivity analyses around relative risk of distant metastases, including access proposals

EPClin access proposal (central testing), test = [redacted]									
	LN0, NPI≤3.4			**LN0, NPI>3.4**			**LN+(1-3 nodes)**		
Relative risk	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**
0.76 (base case)	0.01	[redacted]	[redacted]	0.03	[redacted]	[redacted]	0.06	[redacted]	[redacted]
0.60	0.02	[redacted]	[redacted]	0.05	[redacted]	[redacted]	0.09	[redacted]	[redacted]
0.70	0.02	[redacted]	[redacted]	0.04	[redacted]	[redacted]	0.07	[redacted]	[redacted]
0.80	0.01	[redacted]	[redacted]	0.03	[redacted]	[redacted]	0.05	[redacted]	[redacted]
0.90	0.00	[redacted]	[redacted]	0.01	[redacted]	[redacted]	0.03	[redacted]	[redacted]
EPClin access proposal (local testing, 2 samples), test= [redacted] **, labour=** [redacted] **, total cost=** [redacted]									
	LN0, NPI≤3.4			**LN0, NPI>3.4**			**LN+(1-3 nodes)**		
Relative risk	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**
0.76 (base case)	0.01	[redacted]	[redacted]	0.03	[redacted]	[redacted]	0.06	[redacted]	[redacted]
0.60	0.02	[redacted]	[redacted]	0.05	[redacted]	[redacted]	0.09	[redacted]	[redacted]
0.70	0.02	[redacted]	[redacted]	0.04	[redacted]	[redacted]	0.07	[redacted]	[redacted]

0.80	0.01	[redacted]		0.03	[redacted]		0.05	[redacted]	
0.90	0.00	[redacted]		0.01	[redacted]		0.03	[redacted]	
EPClin access proposal (local testing, 6 samples), test= [redacted], labour= [redacted], total cost= [redacted]									
	LN0, NPI≤3.4			**LN0, NPI>3.4**			**LN+(1-3 nodes)**		
Relative risk	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**
0.76 (base case)	0.01	[redacted]		0.03	[redacted]		0.06	[redacted]	
0.60	0.02	[redacted]		0.05	[redacted]		0.09	[redacted]	
0.70	0.02	[redacted]		0.04	[redacted]		0.07	[redacted]	
0.80	0.01	[redacted]		0.03	[redacted]		0.05	[redacted]	
0.90	0.00	[redacted]		0.01	[redacted]		0.03	[redacted]	
EPClin access proposal (local testing, 12 samples), test= [redacted], labour= [redacted], total cost= [redacted]									
	LN0, NPI≤3.4			**LN0, NPI>3.4**			**LN+(1-3 nodes)**		
Relative risk	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**
0.76 (base case)	0.01	[redacted]		0.03	[redacted]		0.06	[redacted]	
0.60	0.02	[redacted]		0.05	[redacted]		0.09	[redacted]	
0.70	0.02	[redacted]		0.04	[redacted]		0.07	[redacted]	
0.80	0.01	[redacted]		0.03	[redacted]		0.05	[redacted]	
0.90	0.00	[redacted]		0.01	[redacted]		0.03	[redacted]	

As noted in the EAG addendum prepared after the second appraisal committee meeting, the use of alternative sources for post-test chemotherapy use for EPClin (Cusumano *et al* or Penault-Llorca *et al* rather than Bloomfield *et al*) results in lower ICERs for EPClin compared with the EAG base case. This analysis is reproduced in Table 2.

Table 2: Further analyses of EPClin access proposals using alternative chemotherapy use sources (assuming relative risk of distant metastases = 0.76)

Source of post-test chemotherapy use probabilities	**ICER (EPClin versus usual practice)**			
	Centralised testing (test cost= [redacted])	**Local testing (test cost= [redacted])**	**Local testing (test cost= [redacted])**	**Local testing (test cost= [redacted])**
LN0, NPI≤3.4				
Bloomfield *et al*	[redacted]	[redacted]	[redacted]	[redacted]
Penault-Llorca *et al*	[redacted]	[redacted]	[redacted]	[redacted]
Cusumano *et al*	[redacted]	[redacted]	[redacted]	[redacted]
LN0, NPI>3.4				
Bloomfield *et al*	[redacted]	[redacted]	[redacted]	[redacted]
Penault-Llorca *et al*	[redacted]	[redacted]	[redacted]	[redacted]
Cusumano *et al*	[redacted]	[redacted]	[redacted]	[redacted]
LN+ (1-4 nodes)				
Bloomfield *et al*	[redacted]	[redacted]	[redacted]	[redacted]
Penault-Llorca *et al*	[redacted]	[redacted]	[redacted]	[redacted]
Cusumano *et al*	[redacted]	[redacted]	[redacted]	[redacted]

(ii) Additional sensitivity analyses for Prosigna – relative risk of distant metastases including access proposals

Table 3 presents the results of sensitivity analyses around the relative risk of distant metastases for chemotherapy versus no chemotherapy for Prosigna (including the access proposals) versus usual practice. Scenario 1 relates to arrangements for labs that have a rental agreement with NanoString for the nCounter system. Scenario 2 is for labs that have existing instrumentation and do not need the rental part of the agreement.

Table 3: Prosigna additional analyses – sensitivity analyses around relative risk of distant metastases, including access proposals

Prosigna Access Scheme, Scenario 1 - test = [redacted], instrument rental fee= [redacted], labour=£240, total cost= [redacted]									
	LN0, NPI≤3.4			**LN0, NPI>3.4**			**LN+(1-3 nodes)**		
RR	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**
0.76 (base case)	0.02	[redacted]	[redacted]	0.07	[redacted]	[redacted]	0.07	[redacted]	[redacted]
0.60	0.03	[redacted]	[redacted]	0.11	[redacted]	[redacted]	0.12	[redacted]	[redacted]
0.70	0.03	[redacted]	[redacted]	0.08	[redacted]	[redacted]	0.09	[redacted]	[redacted]
0.80	0.02	[redacted]	[redacted]	0.06	[redacted]	[redacted]	0.06	[redacted]	[redacted]
0.90	0.01	[redacted]	[redacted]	0.03	[redacted]	[redacted]	0.02	[redacted]	[redacted]
Prosigna Access Scheme, Scenario 2 - test = [redacted], instrument rental fee= [redacted], labour=£240, total cost= [redacted]									
	LN0, NPI≤3.4			**LN0, NPI>3.4**			**LN+(1-3 nodes)**		
RR	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**	**Inc. QALYs**	**Inc. costs**	**ICER**
0.76 (base case)	0.02	[redacted]	[redacted]	0.07	[redacted]	[redacted]	0.07	[redacted]	[redacted]
0.60	0.03	[redacted]	[redacted]	0.11	[redacted]	[redacted]	0.12	[redacted]	[redacted]
0.70	0.03	[redacted]	[redacted]	0.08	[redacted]	[redacted]	0.09	[redacted]	[redacted]
0.80	0.02	[redacted]	[redacted]	0.06	[redacted]	[redacted]	0.06	[redacted]	[redacted]
0.90	0.01	[redacted]	[redacted]	0.03	[redacted]	[redacted]	0.02	[redacted]	[redacted]

(iii) **Sensitivity analyses for Oncotype DX - relative risk of distant metastases**

Analyses varying the relative risk of distant metastases for chemotherapy versus no chemotherapy for Oncotype DX (assuming no prediction of benefit from chemotherapy) were reported in Table 17 of the EAG's addendum dated 6th March 2018. ICERs ranged from £69,967 to Dominated.

B: Limitations of TAILORx in informing a health economic analysis of Oncotype DX in the LN0 population

Table 4 summarises the available evidence on Oncotype DX RS classification, distant metastases rates on endocrine therapy and relative risks of chemotherapy versus no chemotherapy provided from the TAILORx study.

Table 4: Relevant model inputs available from TAILORx

ODX RS score	% classification	DR rate on ET only	Chemotherapy HR
0-10	✓	✓	✗
11-15*	✓	✓ (only for age ≤50 years)	✓
16-20*	✓	✓ (only for age ≤50 years)	✓
20-25*	✓	✓ (only for age ≤50 years)	✓
26-30	✓	✗	✗
31+	✓	✗	✗

** DR rate for ET for all ages is only available for RS 11-25 as a whole*

The EAG notes that the use of TAILORx in informing a health economic analysis of Oncotype DX within a LN0 patient population who are eligible for chemotherapy according to NCCN guidelines is subject to the following limitations:

(1) TAILORx does not provide any information regarding distant metastases risk for patients receiving endocrine therapy only with an Oncotype DX risk score >25

(2) TAILORx does not provide hazard ratios for chemotherapy versus no chemotherapy for patients with an Oncotype DX risk score of <11 or >25

(3) Around 70% to 75% enrolled in the trial would likely be classified as clinically low-risk and would not be eligible for chemotherapy in the UK. The performance of the test in the population of patients who are eligible for chemotherapy in the UK may be different.

(4) Given the different RS cut-offs applied in TAILORx, this may change the way that clinicians interpret the Oncotype DX RS. Consequently, the NHS England Access Scheme dataset, which is used to inform pre- and post-test chemotherapy probabilities conditional on Oncotype DX risk score, is unlikely to represent clinical decision-making at the cut-offs of 11-25.

As a consequence of these limitations, it is unclear how the results of TAILORx could be used to directly inform a health economic analysis of Oncotype DX.

C: Additional sensitivity analyses for Oncotype DX – assuming zero benefit of chemotherapy for patients with Oncotype DX low-risk

Table 5 presents the results of two sets of analyses:

(1) The first set of analyses present the ICERs for Oncotype DX versus usual practice assuming a predictive benefit based on hazard ratios directly estimated from Paik *et al*, or indirectly estimated from naïve comparisons of B20 versus B14 and B20 versus TransATAC. These analyses were presented in an additional EAG addendum following the second appraisal committee meeting.

(2) The second set of analyses present the ICERs for the same scenario, with the inclusion of an additional assumption of zero chemotherapy benefit for patients in the Oncotype DX low RS category. The EAG notes that this analysis is based on the strong assumption that Oncotype DX not only identifies patients who will not relapse, but also identifies patients who will relapse but will not respond to chemotherapy.

Table 5: Oncotype DX additional analyses – excluding/including predictive benefit and excluding/including assumed hazard ratio for genomic low-risk of 1.0, LN0 NPI>3.4

Estimated HR for chemo vs. no chemo for distant recurrence based on direct/indirect comparisons				
Oncotype DX risk group	**Base case (no predictive effect)**	**B20 (Paik 2006)**	**B20 vs. B14 indirect comparison**	**B20 vs TransATAC indirect comparison**
Low	0.76	1.31	0.64	0.86
Intermediate	0.76	0.61	0.75	0.88
High	0.76	0.26	0.35	0.49
ICER	**Dominated**	**Dominating**	**£24,334**	**£8,150**
Estimated HR for chemo vs. no chemo for distant recurrence based on direct/indirect comparisons, including additional assumption of zero benefit for genomic low-risk patients based on TAILORx				
Oncotype DX risk group	**Base case (no predictive effect)**	**B20 (Paik 2006)**	**B20 vs. B14 indirect comparison**	**B20 vs TransATAC indirect comparison**
Low	0.76	1.00 (assumed from TAILORx)	1.00 (assumed from TAILORx)	1.00 (assumed from TAILORx)
Intermediate	0.76	0.61	0.75	0.88
High	0.76	0.26	0.35	0.49
ICER	**Dominated**	**£1,717**	**£2,425**	**£3,768**

As shown in the table, including an assumption of zero chemotherapy benefit for patients with low RS produces ICERs for Oncotype DX versus usual practice which are consistently less than £4,000 per QALY gained, irrespective of the source of the chemotherapy benefit parameters for the intermediate and high RS groups.

D: Inclusion of Micrometastases in key studies

(i) Oncotype DX

LN mixed: One study only reported data for a mixed population of LN0 and LN+ patients only. (South Florida study by Russell et al. 2016[1]) No information about micrometastases was provided.

LN0: Seven datasets reported data for LN0 patients.

- TransATAC ((data request)[2]Dowsett 2010[3])
- NSABP B-14 (Paik 2004;[4] Wolmark 2016[5])
- NSABP B-20 (Paik 2006[6])
- Sun Yat Sen China study (Gong 2016[7])
- Japanese study (Toi 2010[8])
- Beijing China study (Sun 2011[9])
- E2197 (ECOG tiral) Goldstein 2008 (5 year) ; [10]Sparano 2012[11] (10-year)

None reported whether micrometastases were included or not, apart from TransATAC where micrometastases were not assessed and were treated as LN0 (personal communication). It is unknown how many, if any, patients in the LN0 group had micrometastases in TransATAC. The expert member of the committee judged that the two NSABP B studies would have excluded micrometastases, but the EAG were not able to verify this from the published literature.

LN+: Six datasets reported data for LN+ patients.

- TansATAC ((data request)[2]Dowsett 2010[3 a])
- SWOGG-8814 (Albain 2010[12])
- NSABP B-28 (Wolmark 2016[5]; Mamounas 2012[13])
- PACS01 (Penault-Llorca 2014[14])
- Beijing China study (Sun 2011[9])
- E2197 (ECOG trial) Goldstein 2008 (5 year) ; [10]Sparano 2012[11] (10-year)

None reported whether micrometastases were included or not, apart from TransATAC where micrometastases were not assessed and were treated as LN0 (personal communication).

(ii) EndoPredict

Three reanalyses of RCTs:

- Two LN0 and LN+ (TransATAC[2 15] and ABCSG-6+8[16-18])
- One LN+ only (GEICAM 9906[19 20])

None mentioned micrometastases except TransATAC, where micrometastases were not assessed and were treated as LN0 (personal communication). It is unknown how many, if any, patients in the LN0 group had micrometastases in TransATAC.

(iii) Prosigna

Six reanalyses of RCTs:

- Four LN0 and LN+ (TransATAC,[2 21] ABCSG-8,[22 23] NCIC MA.12,[24] NCIC MA.21[25])
- Two LN+ only (GEICAM 9906,[19 20] CALGB 9741[26]

Two retrospective studies:

- Two LN0 and LN+ (DBCG[27-30] British Columbia[31 32])

None mentioned micrometastases except TransATAC, as above.

1. 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
2. Sestak I, Dowsett M, Cuzick J. NICE request - TransATAC data analysis (academic-in-confidence, data held on file), 2017.
3. Dowsett M, Cuzick J, Wale C, 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(11):1829-34.
4. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *New England Journal of Medicine* 2004;351(27):2817-26.
5. Wolmark N, Mamounas EP, Baehner FL, 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(20):2350-58. doi: https://dx.doi.org/10.1200/JCO.2015.62.6630
6. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *Journal of Clinical Oncology* 2006;24(23):3726-34.

7. 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
8. Toi M, Iwata H, Yamanaka T, 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(13):3112-18.
9. Sun B, Zhang F, Wu SK, et al. Gene expression profiling for breast cancer prognosis in Chinese populations. *Breast Journal* 2011;17(2):172-79. doi: https://dx.doi.org/10.1111/j.1524-4741.2010.01049.x
10. Goldstein LJ, Gray R, Badve 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(25):4063-71.
11. Sparano JA, O'Neill A, Gray RJ, 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(15 SUPPL. 1)
12. Albain K, Barlow W, Shak S, 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(1):55-65.
13. Mamounas EP, Tang G, Paik S, 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(27 SUPPL. 1)
14. Penault-Llorca FM, Filleron T, Asselain B, 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(15 SUPPL. 1)
15. Buus R, Sestak I, Kronenwett R, et al. Comparison of EndoPredict and EPclin with Oncotype DX recurrence score for prediction of risk of distant recurrence after endocrine therapy. *J Natl Cancer Inst* 2016;108(11) doi: https://dx.doi.org/10.1093/jnci/djw149
16. Dubsky P, Filipits M, Jakesz 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(3):640-47. doi: https://dx.doi.org/10.1093/annonc/mds334
17. Dubsky P, Brase JC, Jakesz 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(12):2959-64. doi: https://dx.doi.org/10.1038/bjc.2013.671
18. Myriad Genetics. EPclin in ABCSG-6+8: Subgroup analyses. 2015
19. Martin M, Brase JC, Ruiz A, 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(1):81-89. doi: https://dx.doi.org/10.1007/s10549-016-3725-z
20. Martin M, Brase JC, Calvo L, 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 Res* 2014;16(2):R38. doi: https://dx.doi.org/10.1186/bcr3642
21. Dowsett M, Sestak I, Lopez-Knowles E, 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(22):2783-90. doi: https://dx.doi.org/10.1200/JCO.2012.46.1558
22. Gnant M, Filipits M, Greil R, 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(2):339-45. doi: https://dx.doi.org/10.1093/annonc/mdt494

23. Filipits M, Nielsen TO, Rudas M, 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(5):1298-305. doi: https://dx.doi.org/10.1158/1078-0432.CCR-13-1845
24. Chia SK, Bramwell VH, Tu D, et al. A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. *Clinical Cancer Research* 2012;18(16):4465-72. doi: https://dx.doi.org/10.1158/1078-0432.CCR-12-0286
25. Liu S, Chapman JA, Burnell MJ, 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(2):439-48. doi: https://dx.doi.org/10.1007/s10549-014-3259-1
26. Liu MC, Pitcher B, Mardis E, et al. PAM50 gene signatures and breast cancer prognosis with adjuvant anthracycline- and taxane-based chemotherapy: correlative analysis of C9741 (Alliance). *Breast Cancer* 2016;2
27. Ejlertsen B, Jensen MB, Eriksen JO, 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(15 SUPPL. 1)
28. Laenkholm AV, Jensen MB, Eriksen JO, 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(15 SUPPL. 1):Abstract 546.
29. 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(Suppl):Abstract 544.
30. 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
31. Wallden B, Storhoff J, Nielsen T, et al. Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC Med Genomics* 2015;8:54. doi: https://dx.doi.org/10.1186/s12920-015-0129-6
32. Nielsen TO, Parker JS, Leung S, 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(21):5222-32.

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

Erratum to the EAG Diagnostic Assessment Report

Produced by: Sheffield University School of Health and Related Research Health Technology Assessment Group

Completed on 28th November 2017

This report was commissioned by the NIHR HTA Programme as project number 16/30/03

[redacted] and [redacted] data marked.

In response to the DAR consultation responses collated by NICE and sent to the EAG on 13th November 2017, the EAG provide the following erratum to the report. None of the amendments changed the overall conclusions of the report.

NB: this document of errata only contains changes that affected one page of the report. Where changes affected multiple pages, corrections were made in an addendum to the report also dated 28th November 2017.

Page 17: In response to Agendia comment #4, the EAG corrected the description of the MINDACT study results from:

"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."

To:

"The MINDACT randomised controlled trial (RCT) for MammaPrint reported that for patients who were high-mAOL, low-MammaPrint risk, chemotherapy gave a non-significant absolute benefit of 1.5% in 5 year DMFS (p=0.267). 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 finding was interpreted by the authors as implying that patients who were high-clinical but low-MammaPrint risk could potentially avoid chemotherapy."

Page 35: In response to Myriad Genetics comment #2, the EAG corrected the description of the EndoPredict Clinical (EPclin) score from:

"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."

To:

"The EP score is a number on a scale between 0 and 15. The EP score is the molecular score only and is not the final test result. 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. The EPclin score (cut-off 3.3) provides a single low/high risk cut-off; the threshold was set such that women with a low-risk result (EPclin <3.3.) have a lower than 10% risk of developing distant metastases over the next 10 years."

Page 36: In response to Agendia comment #10, the EAG corrected the description of Mammaprint as assessing the risk of distant metastatsis at 5 years to 5 and 10 years.

Page 62: In response to Agendia comment #26, the EAG corrected the description of the results from MINDACT to include the words "non-statistical" and "(p=0.267)", to read:

"For patients who were high-clinical, low-MammaPrint risk, 5-year DMFS was 95.9% with chemotherapy and 94.4% without chemotherapy, a non-significant absolute difference of 1.5% (p=0.267)."

Page 248: In response to Agendia comment #14, the EAG corrected the description of ABCSG6 and ABCSG8 as recruiting only LN0 patients to read:

"Data from other cohorts also have limitations: ABCSG6+857-59 only evaluated Prosigna for a proportion of patients (ABCSG-8);…"

Page 266: In response to Agendia comment #79, the EAG added to the description of the Oncotype DX analyses that the comparisons were between low/intermediate versus high risk groups:

"A further study180 reported increases in likelihood ratio χ^2 for Oncotype DX (low-/intermediate-risk group versus high-risk group) over MammaPrint and vice versa (see Table 84).

Page 355: Table 123 includes two rows which refer to clinical high-risk. The lower column should refer to clinical low-risk. This is a typographical error; the model calculations are not affected.

Page 370: Table 131 of the EAG report refers to the impact of AEs as a "disutility" – this should have stated that the parameter is applied in the model as a QALY loss (hence it reflects a full year impact, but is applied in the first cycle).

Page 409: Two changes were made on this page. The first was in response to Agendia's comment #103; DRFI was changed to DMFS, and the word "non-significant" was added to read:

"The MINDACT randomised controlled trial (RCT) for MammaPrint reported that for patients who were high-mAOL, low-MammaPrint risk, chemotherapy gave a non-significant absolute benefit of 1.5% in 5 year DMFS."

The second was in response to Myriad Genetics comment #1, where a typo was spotted relating to EndoPredict. The sentence was altered to read:

"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 Prosigna or IHC4)."

Page 434: in response to Agendia comment #123, reference 292 was corrected to read:

“van 't Veer LJ, Yau C, Yu NY, Benz CC, Nordenskjöld B, Fornander T, et al. Tamoxifen therapy benefit for patients with 70-gene signature high and low risk. Breast Cancer Research and Treatment 2017;166(2):593-601.”

The following pages are numbered in accordance with the version of the report sent by NICE for comments.

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 a non-significant absolute benefit of 1.5% in 5 year DMFS (p=0.267). 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 finding was interpreted by the authors as implying that patients who were high-clinical but low-MammaPrint risk could potentially avoid chemotherapy. 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.

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. The EP score is the molecular score only and is not the final test result. 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. The EPclin score (cut-off 3.3) provides a single low/high risk cut-off; the threshold was set such that women with a low-risk result (EPclin <3.3.) have a lower than 10% risk of developing distant metastases over 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 and 10 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

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, a non-significant absolute difference of 1.5% (p=0.267). 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

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 (**Error! Reference source not found.**), 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 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

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 (low-/intermediate- risk group versus high-risk group) over MammaPrint and vice versa (see **Error! Reference source not found.**). 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 (**Error! Reference source not found.**). 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 (**Error! Reference source not found.**), and MammaPrint was shown to have additional prognostic value in this subgroup of patients (adjusted for

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 **1**).

Table 1: 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 low-risk subgroup (n=3,324)	0.82	0.18

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 2 summarises the health utilities assumed in the EAG's base case analysis.

Table 2: 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	Once-only QALY loss applied on transition to distant recurrence state	-0.108	0.04 (assumed)	Campbell *et al*[263]
Chemotherapy AEs	Once-only QALY loss applied in first cycle	-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.

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 a non-significant absolute benefit of 1.5% in 5 year DMFS. 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 Prosigna or IHC4).

6.1.2 Cost-effectiveness – principal findings

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, Yau C, Yu NY, Benz CC, Nordenskjöld B, Fornander T, *et al*. Tamoxifen therapy benefit for patients with 70-gene signature high and low risk. *Breast Cancer Research and Treatment* 2017;166(2):593-601.

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

Erratum to the EAG Diagnostic Assessment Report

Produced by: Sheffield University School of Health and Related Research Health Technology Assessment Group

Completed on 28th November 2017

This report was commissioned by the NIHR HTA Programme as project number 16/30/03

[redacted] and [redacted] data marked.

In response to the DAR consultation responses collated by NICE and sent to the EAG on 13th November 2017, the EAG provide the following erratum to the report. None of the amendments changed the overall conclusions of the report.

NB: this document of errata only contains changes that affected one page of the report. Where changes affected multiple pages, corrections were made in an addendum to the report also dated 28th November 2017.

Page 17: In response to Agendia comment #4, the EAG corrected the description of the MINDACT study results from:

"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."

To:

"The MINDACT randomised controlled trial (RCT) for MammaPrint reported that for patients who were high-mAOL, low-MammaPrint risk, chemotherapy gave a non-significant absolute benefit of 1.5% in 5 year DMFS (p=0.267). 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 finding was interpreted by the authors as implying that patients who were high-clinical but low-MammaPrint risk could potentially avoid chemotherapy."

Page 35: In response to Myriad Genetics comment #2, the EAG corrected the description of the EndoPredict Clinical (EPclin) score from:

"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."

To:

"The EP score is a number on a scale between 0 and 15. The EP score is the molecular score only and is not the final test result. 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. The EPclin score (cut-off 3.3) provides a single low/high risk cut-off; the threshold was set such that women with a low-risk result (EPclin <3.3.) have a lower than 10% risk of developing distant metastases over the next 10 years."

Page 36: In response to Agendia comment #10, the EAG corrected the description of Mammaprint as assessing the risk of distant metastatsis at 5 years to 5 and 10 years.

Page 62: In response to Agendia comment #26, the EAG corrected the description of the results from MINDACT to include the words "non-statistical" and "(p=0.267)", to read:

"For patients who were high-clinical, low-MammaPrint risk, 5-year DMFS was 95.9% with chemotherapy and 94.4% without chemotherapy, a non-significant absolute difference of 1.5% (p=0.267)."

Page 248: In response to Agendia comment #14, the EAG corrected the description of ABCSG6 and ABCSG8 as recruiting only LN0 patients to read:

"Data from other cohorts also have limitations: ABCSG6+857-59 only evaluated Prosigna for a proportion of patients (ABCSG-8);…"

Page 266: In response to Agendia comment #79, the EAG added to the description of the Oncotype DX analyses that the comparisons were between low/intermediate versus high risk groups:

"A further study180 reported increases in likelihood ratio χ^2 for Oncotype DX (low-/intermediate-risk group versus high-risk group) over MammaPrint and vice versa (see Table 84).

Page 355: Table 123 includes two rows which refer to clinical high-risk. The lower column should refer to clinical low-risk. This is a typographical error; the model calculations are not affected.

Page 370: Table 131 of the EAG report refers to the impact of AEs as a "disutility" – this should have stated that the parameter is applied in the model as a QALY loss (hence it reflects a full year impact, but is applied in the first cycle).

Page 409: Two changes were made on this page. The first was in response to Agendia's comment #103; DRFI was changed to DMFS, and the word "non-significant" was added to read:

"The MINDACT randomised controlled trial (RCT) for MammaPrint reported that for patients who were high-mAOL, low-MammaPrint risk, chemotherapy gave a non-significant absolute benefit of 1.5% in 5 year DMFS."

The second was in response to Myriad Genetics comment #1, where a typo was spotted relating to EndoPredict. The sentence was altered to read:

"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 Prosigna or IHC4)."

Page 434: in response to Agendia comment #123, reference 292 was corrected to read:

“van 't Veer LJ, Yau C, Yu NY, Benz CC, Nordenskjöld B, Fornander T, et al. Tamoxifen therapy benefit for patients with 70-gene signature high and low risk. Breast Cancer Research and Treatment 2017;166(2):593-601.”

The following pages are numbered in accordance with the version of the report sent by NICE for comments.

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 a non-significant absolute benefit of 1.5% in 5 year DMFS (p=0.267). 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 finding was interpreted by the authors as implying that patients who were high-clinical but low-MammaPrint risk could potentially avoid chemotherapy. 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.

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. The EP score is the molecular score only and is not the final test result. 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. The EPclin score (cut-off 3.3) provides a single low/high risk cut-off; the threshold was set such that women with a low-risk result (EPclin <3.3.) have a lower than 10% risk of developing distant metastases over 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 and 10 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

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, a non-significant absolute difference of 1.5% (p=0.267). 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

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 (**Error! Reference source not found.**), 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 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, in LN0 patients in TransATAC,[43] EPClin categorised 73% as low-risk and these patients had a 10-year DRFI of 93.4%, whilst Prosigna categorised 54% as low-risk and these patients had a 10-year DRFI of 97%. This effect was more pronounced in LN+ patients in TransATAC, among whom Oncotype DX categorised 57% as low-risk and these patients had a 10-year DRFI of 80.6%, while Prosigna categorised 8% as low-risk and these patients had a 10-year DRFI of 100%. Another broad observation is that the tests generally perform differently in LN+ and LN0 patients. In TransATAC, both EPClin and IHC4+C tests reported lower HRs in the LN0 subgroup than in the LN+ subgroup at 10 years (EPClin LN0 HR 3.88 vs LN+ HR 6.58; IHC4+C LN0 6.06 vs LN+ 9.57), whilst Oncotype DX reported higher HRs in the LN0 subgroup than LN+ subgroup (Oncotype DX LN0 HR 5.83 vs LN+ HR 2.77). 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

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 (low-/intermediate- risk group versus high-risk group) over MammaPrint and vice versa (see **Error! Reference source not found.**). 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 (**Error! Reference source not found.**). 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 (**Error! Reference source not found.**), and MammaPrint was shown to have additional prognostic value in this subgroup of patients (adjusted for

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 **1**).

Table 1: 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 low-risk subgroup (n=3,324)	0.82	0.18

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 2 summarises the health utilities assumed in the EAG's base case analysis.

Table 2: 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	Once-only QALY loss applied on transition to distant recurrence state	-0.108	0.04 (assumed)	Campbell *et al*[263]
Chemotherapy AEs	Once-only QALY loss applied in first cycle	-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.

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 a non-significant absolute benefit of 1.5% in 5 year DMFS. 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 Prosigna or IHC4).

6.1.2 Cost-effectiveness – principal findings

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, Yau C, Yu NY, Benz CC, Nordenskjöld B, Fornander T, *et al*. Tamoxifen therapy benefit for patients with 70-gene signature high and low risk. *Breast Cancer Research and Treatment* 2017;166(2):593-601.

NICE National Institute for Health and Care Excellence

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Peony Breast Care Unit	1.	57	Risk classification	TAILORX may report as early as June 18. There is speculation that the delay in reporting is related to a relatively low event rate in patients with an RS below 25. This in turn suggests that Oncotype DX is picking out patients a good prognosis and that it is predicting limited chemotherapy benefit in patients with an RS below 25 (about 85% of patients) (as would be suggested by the B20 results). This much more closely corresponds to the clinical reality that only about 5% of patients benefit from chemotherapy in early node negative, ER positive, HER2 negative breast cancer. The reporting of TAILORX may well fundamentally change many of the metrics in this assessment.	TAILORx has currently not reported in full and has some limitations in relation to the decision problem: it uses different cut-off points for decision making than are currently recommended by the manufacturer (and form the focus of this assessment); patients recruited to the trial were indicated for adjuvant chemotherapy according to the National Comprehensive Cancer Network guidelines, which differ to UK guidelines. Until the results are reported, and the full implications of the trial in the context of the decision problem considered, it is not possible to take this into consideration.
Peony Breast Care Unit	2	74	Oncotype and RSPC prognosis	In large part the lack of correspondence between statistical methods of prognostic prediction and gene expression methods is related to the lack of reproducibility of standard histological parameters and confounding problems such as lack of controls, inconsistent fixation and no standardisation of methods for preparing and cutting sections. Any prognostic or predictive test which includes such data needs to contend with these significant confounders. This may explain the relative lack of benefit when adding back clinical data to Oncotype DX and makes any gene expression analysis technique which requires the inclusion of histopathologically derived data suspect. Please see our poster SABCS 2016	The study cited by the commentator describes a simulation where the values for grade were randomly varied, according to the variance observed in the WSG Plan-B study (between centrally vs. locally determined grade). Tumour size, ER and HER2 was also varied (but it is not clear if these were also based on variance seen in WSG Plan-B), and the correlation to 10 year predicted mortality calculated for the varied results. The EAG agree that theoretically variances in clinicopathological factor measurement can

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				P1-03-12, "A simulation study depicting the inconsistency of Adjuvant online compared to genomic testing when determining the benefit of chemotherapy." Thomas D et al	affect the usefulness of prediction tools that use them. However, it is unclear if WSG Plan-B variance is typical of UK labs and methods in current use. Comparisons between local and central labs are likely to be at the upper end of the spectrum of variance. Some insight into the scale of this problem for the tests that use some pathological data (EPClin, RSPC, Prosigna, IHC4) can be gained by comparing prognostic performance data obtained from studies at different sites. The impact of individual variables to a model depends on the weight given to them in the model, and the strength of the prognostic association of each factor. Weights are likely to be different for each test, meaning the data from Thomas et al 2017 is difficult to generalise. Variance in a single factor does not always lead to the same degree of variance in the prognostic performance.
Peony Breast Care Unit	3	104	Oncotype DX chemotherapy benefit	That Oncotype DX produces consistent predictions in the node negative, node positive, higher risk (Plan B and we await the RXsponder trial), and in the neo-adjuvant setting is strongly supportive of a prediction of chemotherapy benefit. This is supported by consistent results in daily clinical practice demonstrated by SEER and Clalit registry data all be it that there are limitations to all the methodologies.	It is not a given that chemotherapy benefit prediction in the neoadjuvant setting can be generalised to the adjuvant setting, as the profiles of tumours have been shown to change after neoadjuvant treatment. Neoadjuvant data was not within the scope of the assessment for this reason.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Peony Breast Care Unit	4	285	Oncotype DX decision impact	In the clinical situation of testing intermediate risk patients in NHS practice, I offer you the results of this local audit of 80 patients performed in our institution: 1. 29 decisions were unchanged (36.2%) 2. 51 decisions were changed (63.8%) 3. 6 changed to chemotherapy and hormone therapy 4. 45 changed to hormone therapy alone 5. Net chemotherapy saving in 39 of 80 patients (48.8%) **Final decision** **Initial decision** Hormone therapy only: 8 / 6 Hormone plus chemotherapy: 45 / 21	The review of decision impact studies included all UK and European studies which were published or available to the review team. The publication status of the data given here is unclear.
Peony Breast Care Unit	5	317	Oncotype DX cost effectiveness	In the node negative group, this analysis is likely to have to be revised once the TailorX trial reports. In the 1-3 node positive patients we have just initiated a decision impact trial of 260 patients which will report in about 18 months. This may inform the economic assessment in due course.	All available data from TAILORx was included in the assessment.
Myriad Genetics	1	18	Microarray studies	Contradictory statement: EndoPredict can discriminate between low and high risk versus 'no study'. There is one published microarray study by Bertucci et al. (2014) that showed prognostic power of EndoPredict [1].	Thank you for spotting this. The statement should read: "Microarray studies support conclusions from studies using the commercial versions of the assays in suggesting that Oncotype DX,

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
					MammaPrint and EndoPredict can discriminate between high- and low-risk patients regardless of LN status (there were no relevant microarray studies for Prosigna or IHC4)." Addenda provided to NICE
Myriad Genetics	2	35	3.3	Myriad Genetics considers the description of the EndoPredict test misleading with respect to the EPclin score description. Please consider rephrasing the below sentence, From: *'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.'* To: *From the EPclin score, the probability of metastasis formation within 10 years is estimated, assuming 5 years of hormonal treatment. The EPclin score (cut-off 3.3) provides a single low/high risk cut-off; the threshold was set such that women with a low-risk result (EPclin <3.3.) have a lower than 10% risk of developing distant metastases over the next 10 years.* In addition, within the description for EndoPredict, please consider adding the following description to add clarity for the reader:	The ERG is happy for this change to be made. Errata provided to NICE, so that this section reads (changes are underlined here): "The EP score is a number on a scale between 0 and 15. The EP score is the molecular score only and is not the final test result. 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. The EPclin score (cut-off 3.3) provides a single low/high risk cut-off; the threshold was set such that women with a low-risk result (EPclin <3.3.) have a lower than 10% risk of developing distant metastases over the next 10 years.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				• 'EndoPredict' is the name of the test (and is trademarked) • 'EP score' is the molecular score only. The EP score is a component of EPclin and it is not the final test result 1. 'EPclin score' and 'EPclin risk class' are the final test results	
Myriad Genetics	3	43	3.4.4	Analytic validity was not assessed via a full systematic literature review due to time constraints; however, the EAG mentions that a rapid review of IHC4 will follow as an addendum to this report. Myriad Genetics questions whether this process for evaluating analytical ability is appropriate. For example, how will the data identified in the rapid review for IHC4 be compared to other gene expression tests, if data for the other tests were not formally assessed? In addition, Myriad Genetics questions whether the omission of analytical validity from the assessment 'favours' tests that have not proven analytical ability thought CE marking or equivalent procedures?	All the tests except IHC4 have a CE mark (though for Oncotype DX the CE mark is for the collection kit as the test is performed centrally in the USA, at a lab with Clinical Laboratory Improvement Amendments certification), which is why we conducted a rapid review of analytical validity for IHC4.
Myriad Genetics	4	59	EndoPredict and EPclin	It is stated that there are no data on additional prognostic value of EPclin over clinicopathological variables, which is incorrect. Please see comment No. 9.	Although the GEICAM study (Martin 2014 Figure 2) reports that EPclin has a numerically higher c-index than clinical variables alone, no significance level is reported for this difference. Details of c-index for EPclin are reported in the main review section.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
					As noted in comment 9, p59 does report data on EPclin from TransATAC but this is redacted as AIC.
Myriad Genetics	5	65	LN+, variable ET/CT, column: "Additional value over CP factors or tests?"	For EndoPredict it has been shown in the GEICAM/9906 study by *c*-index analysis that EPclin has more prognostic power than clinical variables alone, and that the EndoPredict molecular score (EP score) provides prognostic information beyond the clinical parameters (Please see Figure 2 in Martin et al. 2014 [2])	Although the GEICAM study (Martin 2014 Figure 2) reports that EPclin has a numerically higher c-index than clinical variables alone, no significance level is reported for this difference. Details of c-index for EPclin are reported in the main review section.
Myriad Genetics	6	199	EndoPredict and EPclin 4.6.2	There is a discrepancy in how the clinical validation studies for EndoPredict are described. Clinical validation cohorts are available from four independent studies: • ABCSG6 [3] • ABCSG8 [3] • TransATAC [4] • GEICAM 9906 [2] Filipits et al. (2011) contains both the derivation cohort and the two validation trials ABCSG6 and ABCSG8, separately [3]. Therefore, this publication should be considered to evaluate the performance characteristics of EPclin in the two validation trials ABCSG6 and ABCSG8 (as two pooled cohorts of LN0 and LN+ patients). Subgroup analyses in clinical intermediate risk patients, with regard to different	The EAG report contains results from these four studies. However, data from ABCSG-6 and ABCSG-8 were available separately (Filipits 2011) or pooled (Dubsky 2013a, b and analyses provided to NICE by Myriad). Since the pooled analyses (Dubsky 2013a, b and Myriad analyses) contained the most data and reported on the subgroups most relevant to this assessment, these were included instead of the Filipits 2011 article, to avoid double-counting.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				tumour grading, nodal status, and early versus late metastasis, are published in two manuscripts by Dubsky et al. (2013) [5 6] and provided to NICE by Myriad Genetics [7].	
Myriad Genetics	7	199	4.6.2	Pooled validation analysis of N0 and N+ patients in the two studies ABCSG6 and ABCSG8 were published separately in Filipits et al. (2011) [3].	As comment 6: for ABCSG-6 and ABCSG-8 we used Dubsky 2013a,b and Myriad analyses rather than Filipits 2011. The sources we used reported separate data for N0 and N+ patients, in the population most relevant to this assessment (LN0-3).
Myriad Genetics	8	202	Additional prognostic value (First line)	It is stated that *c*-indices are only reported for Years 5–10. However, please note that *c*-indices for Years 0–10 are reported for ABCSG6 and ABCSG8 separately in the publication by Filipits et al. (2011) [3], and show that EndoPredict significantly adds information to a combination of clinical factors or Adjuvant!Online in both studies. EPclin showed the highest *c*-index in both studies (0.788 and 0.732 in Studies ABCSG6 and ABCSG8, respectively; Figure 4 in Filipits et al. 2011 [3]).	We agree that the increase in c-index for EP score over clinical factors or Adjuvant!Online applies in years 0-10 as well as years 5-10. This has been noted in the Addendum to NICE.
Myriad Genetics	9	202	Additional prognostic value (last line before next section)	It is stated that the additional prognostic value of EPclin over clinicopathological variables is not reported in published studies. However, please note that in the study by Buus et al. (2016) (TransATAC) it is shown that EPclin provides additional information beyond CTS in the mixed cohort and in the N0 and N+ subgroup (see likelihood ratio statistics in Table 1 of the publication by Buus et al. 2016) [4].	Buus et al is a TransATAC publication. Rather than report data from the original publication and the bespoke analysis using the ER+, HER2- subgroup we just extracted the data from the bespoke analysis, which is redacted. The data relating to EPClin is therefore redacted.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Myriad Genetics	10	203	Conclusions	It is stated that there are no data on additional prognostic information of EPclin. This is incorrect; see comment No. 9.	As for point 9, data on EPClin for TransATAC are reported here but redacted.
Myriad Genetics	11	204	Table 59	The two studies ABCSG6 and ABCSG8 are published separately (Filipits et al. 2011) as a mixed cohorts (N0/N+) with multivariate cox regression and *c*-index analysis for both studies [3]. It is requested that this publication be considered as part of the clinical validation evidence base for EndoPredict.	As noted above, data on these studies was taken from Dubsky 2013a, b and analyses provided to NICE by Myriad since these provided the most complete data in the population most relevant to this assessment (LN0-3).
Myriad Genetics	12	205	Table 60: line ABCSG 6&8	Both validation trials were performed blinded to outcome data. This is stated in the original publication by Filipits et al. (2011) describing the methods for validation (see methods section, paragraph 'validation and statistical analysis') [3].	Information noted. As this is a minor point, to avoid multiple further documents being created, we have not provided an erratum document for this.
Myriad Genetics	13	209	Table 63	Line ABCSG6 and ABCSG8: *c*-indices and multivariate analyses showing the added value of EndoPredict over clinicopathological parameters and Adjuvant!Online in Years 0–10 and the *c*-index of EPclin for Years 0–10 are published separately for both studies in Filipits et al. (2011) [3].	As point 8: We agree that the increase in c-index for EP score over clinical factors or Adjuvant!Online applies in years 0-10 as well as years 5-10. This has been noted in the Addendum to NICE.
Myriad Genetics	14	249	Discussion: studies assessing multiple tests	It is stated in the middle of the paragraph that ABCSG6 and ABCSG8 only recruited N0 patients. This statement is incorrect. These two trials recruited a mixed cohort of N0 and N+ patients.	Agreed; amended in an erratum to NICE

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Myriad Genetics	15	253	Table 77	Blinding for ABCSG6 and ABCSG8: see comment No. 12 above.	As point 12: Information noted. As this is a minor point, to avoid multiple further documents being created, we have not provided an erratum document for this.
Myriad Genetics	16	260	Table 81	ABCSG6 and ABCSG8: *c*-indices for Years 0–10 are published separately for ABCSG6 and ABCSG8 in Filipits et al. (2011) [3].	As point 8 and 13: We agree that the increase in c-index for EP score over clinical factors or Adjuvant!Online applies in years 0-10 as well as years 5-10. This has been noted in the Addendum to NICE.
Myriad Genetics	17	298	4.10	The DAR states that only six studies that reported outcomes relating to anxiety and HRQoL were identified. Myriad Genetics is aware of a UK trial that is complete, but not yet published, that included psychosocial outcomes (The Brighton Trial; ISRCTN69220108 [8]), which may provide further data for these outcomes of anxiety and HRQoL. Has the EAG sought a bespoke analysis and/or data to be shared under confidence from The Brighton Trial investigators to source UK data relevant to the decision problem for HRQoL, akin to the EAG's approach to contact the TransATAC trial team for data related to risk classification?	The data are included in the review of Anxiety and HRQoL, from the author Bloomfield et al 2017. See Table 97 of the EAG report
Myriad Genetics	18	303	Time to test result	There is a publication by Müller et al. (2013) showing the time to test result for EndoPredict [9]. In this study the median handling time was three working days [9].	Agreed; this is now provided in an Addendum to NICE.
Myriad Genetics	19	344	5.3	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)	Myriad's interpretation appears correct. NPI is only explicitly used in the EAG model to define risk subgroups in the node-negative population (NPI≤3.4 or NPI>3.4). The comparator is usual practice (which may include a mix of risk tools and/or other variables). We were unable to use

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				(3) node-positive (1–3 nodes). Thus, the analyses do not formally interrogate the use of other risk classification tools available in clinical practice as a means of identifying 'low risk' versus 'intermediate risk' patients, for example, NHS PREDICT. It is not clear from the DAR how the above subgrouping may impact on any possible future NICE decisions/recommendations with respect to the criteria/obligation of using a specific risk classification tool. As multiple risk classification tools are used in England, Myriad Genetics' interpretation of the analysis is to allow the NPI to be used as an example method for how 'low risk' versus 'intermediate risk' patients are identified, but decisions on the likely cost-effectiveness of interventions will not be wedded to using the NPI as the only risk classification tool, i.e. clinicians can use other risk classification tools for stratification into these two groups ('low risk' versus 'intermediate risk'). Myriad Genetics welcomes a response from the EAG/NICE to confirm whether this interpretation is correct? There is a concern that a NICE decision/recommendation that is reliant on the use of the NPI for risk classification only is inappropriate since clinical practice in England varies with respect to the type and number of risk classification tools used.	PREDICT or A!O to define risk subgroups due to data limitations.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Myriad Genetics	20	349	Key EAG model assumptions	Whilst CHF is also a potentially relevant long-term AE associated with chemotherapy, the EAG excluded it from the model due to a lack of evidence on the joint survival impact of CHF and metastatic breast cancer. Myriad Genetics questions the decision above, given that CHF is a potential driver of morbidity/mortality in those receiving chemotherapy, and hence is an important outcome to consider for the economic analysis. With respect to relevant literature on this topic, Myriad Genetics highlights the Mackay et al. (2010) study as an example of the published literature that reports on the frequency and associated mortality of CHF in breast cancer patients treated with chemotherapy [10]. The trial assessed standard adjuvant anthracycline chemotherapy with anthracycline–taxane combination chemotherapy in women with operable N+ breast cancer with a 10-year follow-up of the Phase III randomised BCIRG 001 trial. Results concerning CHF were: • Grade 3–4 heart failure occurred in 26 (3%) patients in the TAC group (docetaxel, doxorubicin, and cyclophosphamide) and 17 (2%) patients in the FAC group (fluorouracil, doxorubicin, and cyclophosphamide), and caused death in 2 patients in the TAC group and 4 patients in the FAC group.	The exclusion of CHF from the EAG model is documented within the EAG report (p410). *"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."* We note that the Mackay study relates to a trial of two different chemotherapy regimens and thus does not report the excess mortality associated with CHF in mBC. It is not obvious how these data could help with the inclusion of the impact of CHF on QALY losses.
Myriad Genetics	21	356	5.3, tapering of risk of recurre	The EAG notes that there is uncertainty with respect to the long-term risk of distant recurrence. The model base case assumes that the risk of distant metastases between 10–15 years is equal to half the risk during the preceding period (0–	We agree that there is uncertainty around this. This is discussed in the EAG report (p.362). Whilst there is some evidence which suggests that for some patients with particular disease

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
			nce over time	10 years) and beyond 15 years, the risk of distant recurrence is assumed to be zero. Myriad Genetics questions the appropriateness of this assumption in the context of other published estimates of long-term risk recurrence beyond 15 years. Whilst much of the published evidence shows that most recurrences occur in the first 5 years after treatment, risk of recurrence persists for ER+ disease beyond 15 years. For example, Colleoni et al. (2016) reported that among patients with ER+ disease, annualised hazards of recurrence remained elevated and fairly stable beyond 10 years, even for those with no axillary involvement (2.0%, 2.1%, and 1.1% for years 10–15, 15–20, and 20–25, respectively) and for those with 1–3 positive lymph nodes (3.0%, 3.5%, and 1.5%, respectively) [11]. Hence, the use of a 'zero' risk of recurrence beyond 15 years in the base case economic analysis may be an oversimplification, given the available published data on the long-term risk of recurrence.	subtypes, recurrence rates remain approximately constant between 5 and 20-years, 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. We undertook sensitivity analyses to test this assumption.
Myriad Genetics	22	360	Table 128 and associated text	There is a reliance on the recent Bloomfield et al. (2017) abstract reference to inform the probability of receiving chemotherapy conditional on test results for EndoPredict in the base case economic analysis [12]. Despite the study by Bloomfield et al. (2017) being UK-based, Myriad Genetics questions whether this is the most appropriate source for the base case, given that the nodal	We agree that there is substantial uncertainty around this set of parameters, particularly for the 2-level tests. We undertook a number of sensitivity analyses to consider other evidence not included in the base case (see EAG report, Table 148). We do not feel it is appropriate to take an average of the available studies and

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				status of the population is unclear in this study [12]. Myriad Genetics questions whether it would be more appropriate to take an average (for the probability of receiving chemotherapy conditional on tests results) from across all studies marked as relevant to EndoPredict by the EAG. These studies include the current base case source (Bloomfield et al. 2017 [12]) and sources probed in the sensitivity analysis (UKBCG survey, Penault-Llorca et al. 2016 [13] and Cusumano et al. 2014 [14]). Myriad Genetics suggests that this model parameter is investigated/probed further by the EAG, given that it is a key ICER driver.	therefore have not undertaken the further analysis suggested by the company.
Myriad Genetics	23	368	5.3.3	The price for EndoPredict is assumed to be £1,500 per assay, as the base case assumes all assays will be conducted at the centralised Myriad Genetics laboratory only. Myriad Genetics expects that, while the centralised test will indeed continue to be available to the NHS, the majority of NHS tests are likely to be run locally, to improve turnaround time and to streamline workflows. Accordingly, Myriad Genetics requests that NICE and the EAG consider both the central and local testing scenarios in its base case analyses. Regarding the local NHS testing scenario, the instrument placement model is on a reagent rental basis, with costs for the instrument fully absorbed into the test kit price. Test kits have already been supplied directly to the NHS on this basis, for small volumes of privately funded patients (40-50 per annum), with N+ disease only, at [redacted]. Going forward, Myriad Genetics is prepared to commit to a confidential discounted price across the NHS of [redacted]	Within the model, we assumed that testing would be centralised: • Local testing costs would vary by centre according to size and throughput • The clinicians and pathologist that we engaged with were happy with current centralised pathology and did not see there would be a benefit for local testing in terms of turnaround times. This new price was not available at the time the assessment was undertaken. This new price reduces the deterministic ICERs to: [redacted]

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				for such local testing, reflecting the much higher demand as the test becomes available nationally. Myriad Genetics acknowledges that, when compared with centralised testing, laboratories will incur additional modest direct costs associated with such local testing, but considers that the true costs will be less than the £240 per test figure used for the Prosigna (NanoString) economic analysis in the EAG report. Regarding centralised testing, the £1,500 per assay list price has indeed been paid by the low volume private sector. Going forward, Myriad Genetics is prepared to commit to a confidential discounted price across the NHS of [redacted] for centralised testing, on the robust internal assumption that Myriad Genetics will realise internal efficiencies associated with demand at national level. Abbreviations: ABCSG, Austrian Breast and Colorectal Cancer Study Group; AE, adverse event; BCIRG, Breast Cancer International Research Group; CE, European Conformity; CHF, congestive heart failure; CT, chemotherapy; CTS, clinical treatment score; DAR, diagnostics assessment report; EAG, external Assessment Group; EP, EndoPredict; ER, oestrogen receptor; ET, Endocrine therapy; HRQoL, health-related quality of life; ICER, incremental cost-effectiveness ratio; IHC4, immunohistochemistry 4; LN, lymph node; N+, node positive; N0, node negative; NHS, National Health Service; NICE, National Institute for Health and Care Excellence; NPI, Nottingham Prognostic Index; UK, United Kingdom; UKBCG, United Kingdom Breast Cancer Group.	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				References 1. Bertucci F, Finetti P, Viens P, et al. EndoPredict predicts for the response to neoadjuvant chemotherapy in ER-positive, HER2-negative breast cancer. *Cancer letters* 2014;355(1):70-75. 2. Martin M, Brase JC, Calvo L, 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(2) 3. Filipits M, Rudas M, Jakesz R, 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(18):6012-20. doi: https://dx.doi.org/10.1158/1078-0432.CCR-11-0926 4. Buus R, Sestak I, Kronenwett R, et al. Comparison of EndoPredict and EPclin with Oncotype DX recurrence score for prediction of risk of distant recurrence after endocrine therapy. *Journal of*	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				the National Cancer Institute 2016;108(11) doi: https://dx.doi.org/10.1093/jnci/djw149 5. Dubsky P, Brase JC, Jakesz R, et al. The EndoPredict score provides prognostic information on late distant metastases in ER+/HER2- breast cancer patients. *Br J Cancer* 2013;109(12):2959-64. doi: http://dx.doi.org/10.1038/bjc.2013.671 6. Dubsky P, Filipits M, Jakesz R, et al. EndoPredict improves the prognostic classification derived from common clinical guidelines in ER-positive, HER2-negative early breast cancer. *Ann Oncol* 2013;24(3):640-7. doi: http://dx.doi.org/10.1093/annonc/mds334 7. Markopoulos C, van de Velde C, Zarca D, et al. Clinical evidence supporting genomic tests in early breast cancer: Do all genomic tests provide the same information? *European Journal of Surgical Oncology* 2016;31:31. doi: https://dx.doi.org/10.1016/j.ejso.2016.08.012 8. ISRCTN registry. The use of a prognostic tool (EndoPredict®) to inform adjuvant chemotherapy decision in low to medium risk oestrogen receptor positive, Her-2 negative early breast cancer. Available at:	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				http://www.isrctn.com/ISRCTN69220108 [last accessed November 2017]. 9. Müller BM, Keil E, Lehmann A, et al. The EndoPredict Gene-Expression Assay in Clinical Practice - Performance and Impact on Clinical Decisions. *PLoS ONE* 2013;8(6):e68252. doi: http://dx.doi.org/10.1371/journal.pone.0068252 10. Mackey JR, Martin M, Pienkowski T, et al. Adjuvant docetaxel, doxorubicin, and cyclophosphamide in node-positive breast cancer: 10-year follow-up of the phase 3 randomised BCIRG 001 trial. *The Lancet Oncology* 2013;14(1):72-80. 11. Colleoni M, Sun Z, Price KN, et al. Annual hazard rates of recurrence for breast cancer during 24 years of follow-up: results from the international breast cancer study group trials I to V. *J Clin Oncol* 2016;34(9):927-35. 12. Bloomfield DJ, Arbon A, Cox J, et al. Patient/oncologist decisions about adjuvant chemotherapy in ER+ ve, HER2-ve early breast cancer following endopredict testing: American Society of Clinical Oncology, 2017.	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				13. A prospective multicenter non-randomized trial evaluating the effect of EndoPredict® (EPclin®) clinico-genomic test on treatment decision making among patients with intermediate clinical risk. San Antonio Breast Cancer Symposium; 2016; San Antonio, Texas, USA. 14. Cusumano P, Generali D, Ciruelos E, et al. European inter-institutional impact study of MammaPrint. *The Breast* 2014;23(4):423-28. 15. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *Journal of Clinical Oncology* 2006;24(23):3726-34.	
Nano String Technologies Inc.	1	n/a	5.3	A motivation in producing this updated diagnostic guidance is to address the emotional and psychological strain that breast cancer patients face when considering chemotherapy.[1] It does not then follow why the impact of gene expression profiling on this important patient health-related quality of life outcome is not included or explored in the cost-effectiveness analysis. Gene expression profiling tests have the potential to reduce anxiety and improve health-related quality of life.	It is unclear to the EAG why HRQoL (excluding impacts on clinical outcomes) should be any different for a patient using a genomic test compared with using A!O or NPI or PREDICT to predict recurrence risk. Contingent valuation is not part of NICE methods guide. NICE's decision-making approach, including details of relevant cost-

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Previous studies have attempted to measure how gene expression profiling is valued by patients. Marshall et al. show that, from a personal utility perspective, a gene expression profiling test result was the most important factor for determining chemotherapy treatment choice, over and above input from a clinical doctor, and that this value was greatest in the intermediate clinical risk group.[2] A definitive low or high-risk result may be preferred by patients over an intermediate classification. O'Neill et al. show that women are willing to pay, on average, $997 for genomic risk for recurrence testing.[3] To illustrate the impact of further information on this important aspect of health-related quality of life on the cost-effectiveness results, this value (approximately £760 at current exchange rates) can be converted onto the scale of health outcomes at a rate of £20,000 per QALY indicating a QALY benefit of 0.038 per test. Including this benefit to patients would drastically reduce the ICERs for Prosigna in all three subgroups, as shown in Table 1. The ICER for Prosigna would fall by approximately £10,000 in the 'LN0 NPI>3.4' and 'LN+' groups, while in the 'LN0 NP1≤3.4' group the ICER would fall by approximately £60,000. In fact, incorporating an additional QALY benefit at only 0.01 per test would reduce the ICERs by nearly £4,000 in the 'LN0 NPI>3.4' and 'LN+' subgroups, and by about £30,000 in the 'LN0 NPI≤3.4' group.	effectiveness thresholds for decision-making, are detailed in the NICE Diagnostic Programme Manual, pages 106-109.

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

Diagnostics Assessment Report (DAR) - Comments

<table>
<tr><th>Stakeholder</th><th>Comment no.</th><th>Page no.</th><th>Section no.</th><th>Comment</th><th>EAG response</th></tr>
<tr><td></td><td></td><td></td><td></td><td>
Table 1. Impact of including additional health benefit of reduced emotional and psychological strain - Prosigna versus current practice
<table>
<tr><th>Subgroup</th><th>Inc. QALYs*</th><th>Inc. costs</th><th>ICER (per QALY gained)</th></tr>
<tr><td colspan="4">Scenario 1. Additional QALY gain of 0.038 per test</td></tr>
<tr><td>LN0 NPI≤3.4</td><td>0.0587</td><td>£1,884</td><td>£32,070</td></tr>
<tr><td>LN0 NPI>3.4</td><td>0.1028</td><td>£1,686</td><td>£16,048</td></tr>
<tr><td>LN+ (1-3 nodes)</td><td>0.1054</td><td>£1,936</td><td>£18,362</td></tr>
<tr><td colspan="4">Scenario 2. Additional QALY gain of 0.01 per test</td></tr>
<tr><td>LN0 NPI≤3.4</td><td>0.0307</td><td>£1,884</td><td>£61,374</td></tr>
<tr><td>LN0 NPI>3.4</td><td>0.0747</td><td>£1,686</td><td>£22,570</td></tr>
<tr><td>LN+ (1-3 nodes)</td><td>0.0774</td><td>£1,936</td><td>£25,018</td></tr>
</table>
*Incremental QALY reported in Table 143, p384 EAG report estimated to precision required to match reported ICER using goal seek
References
</td><td></td></tr>
</table>

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				1. Gene expression profiling and expanded immunohistochemistry tests for guiding adjuvant chemotherapy decisions in early breast cancer management: MammaPrint, Oncotype DX, IHC4 and Mammostrat. NICE 2013 *2. 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:e010981. doi: 10.1136/bmjopen-2015-010981* *3. O'Neill SC, Brewer NT, Lillie SE, et al. Women's Interest in Gene Expression Analysis for Breast Cancer Recurrence Risk. Journal of Clinical Oncology 2007 25:29, 4628-4634*	
Nano String Technologies Inc.	2	33	3.3	Another factor that may be important to patients has been omitted, which is time spent waiting for a test result. In Section 3.3 of the EAG report information is provided on the anticipated time to test result for EndoPredict (2 days if processed in house and 4-5 days if sent to Munich), MammaPredict (10 days) and Oncotype Dx (7-10 days), IHC (average 1 week) but not for Prosigna. As the EAG assert in other sections of the document, Prosigna is anticipated to be processed in local NHS laboratories. This implies that the wait time for a Prosigna test result is likely to be shorter than for tests with central laboratory testing facilities, similar to EndoPredict. A shorter wait time may have benefits to patients in terms of reduced anxiety and improved health-related quality of life, and in reduced delay in accessing appropriate treatment.	Advice received by the EAG is that the turnaround time of the tests does not impact upon the time at which patients see their oncologist. In addition, for smaller centres, tests may need to be sent to larger centres due to low throughput, hence the turnaround times may not be any quicker for Prosigna.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				The cost-effectiveness model of the EAG report assumes that all tests are processed in central laboratories (p345 EAG report). However, the cost per test used for Prosigna is estimated based on local laboratory processing (Table 132, p168 EAG report). The EAG report should include comment on the expected time to test result for Prosigna and consider the benefits of its shorter time between test and treatment decision compared to other tests. Alternatively, the EAG should reduce the cost per Prosigna test to match the assumption of central processing. Processing of Prosigna test results at a central laboratory would involve higher throughput and less capital investment, which would reduce the cost per test and thereby reduce the ICER in all subgroups.	
Nano String Technologies Inc.	3	n/a	n/a	Another motivation for this guidance is to address the considerable variation in practice. The adoption of a gene expression profiling test could reduce current variation in approaches to risk profiling and variation in use of chemotherapy for women with similar characteristics. The adoption of a gene expression profiling test may even replace the use of the NPI, which has poorer prognostic value and is subject to reproducibility problems between clinicians due to the subjective assessment of tumour grade. The EAG does not comment on the potential value of this reduced practice variation, but one would anticipate that the Committee will take it into account in its deliberations.	A review of the analytical validity of NPI was not within the scope of the assessment. As such, the EAG cannot comment on the reproducibility of tumour grade. It should be noted that whilst use of a single gene test has the potential to reduce variability in treatment decisions, the same might (theoretically) be true were tumour grade performed centrally by fewer labs. In addition, Optima Prelim shows that different gene tests can return different risk categories for a given individual (but all returning roughly equally valid results at a population level), and

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
					as such use of different gene tests across England could (theoretically) also result in variation in treatment decisions.
Nano String Technologies Inc.	4	n/a	5.3	The value of the tests under review relates to the way in which they affect the decision to undergo chemotherapy compared to existing tools. The EAG cost-effectiveness analysis embodies a scenario in which the test is provided to all breast cancer patients within each subgroup, regardless of their preference. What is not accounted for is the ability for breast cancer patients to opt out of gene expression profiling due to an unwillingness or inability to undergo chemotherapy. DeFrank et al. show that genomic testing may be more common among patients who may benefit most from the information.[1] An alternative scenario in which use of the test was employed after discussion between patients and their clinicians could avoid the costs of testing for breast cancer patients who would not alter their choice regarding chemotherapy regardless of the test result. The current EAG assumption that all eligible patients will opt for or receive a gene profiling test may underestimate the value of the test, overestimate the corresponding ICERs compared to current practice, and overestimate the potential budget impact. **Reference** *1. DeFrank JT, Salz T, Reeder-Hayes K, Brewer NT. Who Gets Genomic Testing for Breast Cancer Recurrence Risk? Public Health Genomics 2013;16:215–222*	The population of the model reflects women who are able to receive chemotherapy who are also willing to undergo genomic testing. We agree that there would be no value in providing the test to people who are unwilling or unable to have chemotherapy – the value of the tests is in changing decision-making about the use of chemotherapy. For some of the studies relating to decision impact, it was not always clear whether the study populations reflected the model population (e.g. some women had already decided not to have chemotherapy). The limitations and uncertainties surrounding these studies are already discussed in the EAG report.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Nano String Technologies Inc.	5	362	5.3.3	The EAG cost-effectiveness model characterises the benefit of chemotherapy based on a relative risk reduction of 0.76 for anthracycline based chemotherapy regimens compared to no chemotherapy (Table 129, p362 EAG report). However, it is acknowledged that a proportion of the patients who undergo chemotherapy will be treated with taxanes, and the costs of taxane based regimens are included for 35% of treated patients (Table 133, p370 EAG report). The EBCTCG publication from which the relative risk is taken indicates that taxane based regimens are more efficacious compared to anthracycline based regimes, with a relative risk of 0.84 ($p<0.00001$) for distant recurrence in unconfounded trials at 8-year follow-up (Figure 1, left hand side).[1] In patients that receive taxanes, the relative risk of distant recurrence could be reduced to 0.84*0.76 = 0.64. A crude weighted average based on the proportions that receive taxanes in the EAG cost-effectiveness analysis would suggest that the relative risk of distant recurrence reduction would be closer to 0.72 rather than the 0.76 included in the base case. Potentially this relative risk could be reduced further in groups where there is a higher proportion of taxane use, for example in patients with nodal involvement. Applying a lower value for the relative risk of distant recurrence with chemotherapy is required to maintain consistency between	The EAG agrees that there is uncertainty surrounding the EBCTCG relative risk for adjuvant chemotherapy (derived from EBCTCG publication, extra web material, page 12, any anthracycline-based regimen versus no chemotherapy, distant recurrence). In reality, this relative risk could be affected by a number of factors including patient age, regimen used, lymph node involvement and a range of other potential treatment effect modifiers. It may also be time-varying. For simplicity, we selected the treatment effect estimate which appears to be the most relevant to the modelled population. The EAG has doubts about NanoString's alternative suggestion because the relative risk quoted relates to "any recurrence", rather than "distant recurrence" – the model uses the latter not the former. However, we note that the EAG report includes sensitivity analyses which explore the impact of alternative chemotherapy treatment effects.

<table>
<tr><th>Stakeholder</th><th>Comment no.</th><th>Page no.</th><th>Section no.</th><th>Comment</th><th>EAG response</th></tr>
<tr><td></td><td></td><td></td><td></td><td>the costs and effects in the EAG cost-effectiveness analysis. The EAG model is ACIC redacted, and so it is not possible to explore the impact of this on the ICERs directly. However, the deterministic sensitivity analyses include a scenario exploring a lower relative risk of 0.70, reproduced in Table 2 below, which reduced the ICERs for Prosigna in all three subgroups.

Table 2. Impact of lower relative risk of chemotherapy from EAG deterministic sensitivity analyses - Prosigna versus current practice
<table>
<tr><th>Subgroup</th><th>Inc. QALYs</th><th>Inc. costs</th><th>ICER (per QALY gained)</th></tr>
<tr><td colspan="4">Scenario 3. Chemotherapy RR=0.70</td></tr>
<tr><td>LN0 NPI≤3.4</td><td>0.03</td><td>£1,869</td><td>£71,107</td></tr>
<tr><td>LN0 NPI>3.4</td><td>0.08</td><td>£1,644</td><td>£19,926</td></tr>
<tr><td>LN+ (1-3 nodes)</td><td>0.09</td><td>£1,845</td><td>£21,508</td></tr>
</table>
Deterministic sensitivity analysis reported in Table 145, p386 EAG report

Reference
1. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Comparisons between different</td><td></td></tr>
</table>

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				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	
Nano String Technologies Inc.	6	358	5.3.3	In the EAG cost-effectiveness analysis the probability of chemotherapy post-test by category is applied as a simple proportion of all patients within category. In clinical practice, the probability of chemotherapy has been shown to increase with risk of recurrence within category.[1] This implies that the risk of recurrence among patients who receive chemotherapy within each category is greater than the simple average risk of recurrence across all patients in that category. Providing chemotherapy to patients with higher baseline risk will increase the absolute health benefit from chemotherapy. The implication of underestimating the health gains of those patients who receive chemotherapy is that the EAG model may have underestimated the health benefits of all gene expression profiling tests with continuous scores, and overestimated the corresponding ICERs compared to current practice. **Reference** 1. Enewold L, GeigerJo AM, Zujewski A, Harlan LC. Oncotype Dx assay and breast cancer in the United States: usage and concordance with chemotherapy. Breast Cancer Research and Treatment (2015) 151: 149	We agree that this is a simplification. With the exception of the NHS England Access Scheme Dataset, it was not possible to apply chemotherapy probabilities conditional on risk levels (without testing). In order to maintain consistency between the tests, this analysis was only undertaken as a sensitivity analysis for Oncotype DX (see EAG report, Table 139, "Baseline P(chemo) adjusted by Oncotype DX RS score"). We could not undertake the equivalent analysis for other tests due to a lack of evidence.
Nano String Technologies Inc.	7	360	5.3.3	The probability of chemotherapy pre- and post-test are key parameters in determining cost-effectiveness. The EAG identified a range of UK evidence on these proportions (Table 128, p360 EAG report). The deterministic sensitivity	The EAG notes that this may well be the case, but the NHS Access Scheme dataset was selected for use in the base case analysis as this is most likely to reflect how 3-level tests are

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				analyses results indicate that the NHS England Access Scheme dataset selected for the EAG base case is the most conservative of all the potential sources for these key parameters (Table 145, p386 EAG report), as QALY gains are increased for all deterministic analyses that use alternative sources. The EAG did not attempt to synthesise the multiple sources of evidence on these key parameters, but had they done so it would have improved the ICERs for Prosigna. The cost-effectiveness analysis employs a crude assumption that the post-test probability of chemotherapy is the same by category (low, intermediate or high risk) across all tests, i.e. that clinicians will interpret an 'intermediate risk' categorisation in the same way regardless of the test performed (p348 EAG report). The EAG cost-effectiveness analysis employs a post-test probability of chemotherapy that is determined by Oncotype Dx classification. In practice, clinicians and patients will work together in helping the patient come to a decision about whether or not to undergo chemotherapy in greater knowledge of the actual risk score and associated risk of recurrence. As the probability of chemotherapy has been shown to increase with the actual risk of recurrence within categories,[1] and as the risk of recurrence is not the same within category across all tests, this crude assumption in the EAG model misrepresents the post-test probability of chemotherapy for all tests apart from Oncotype Dx.	used in clinical practice in England in patients at clinical intermediate risk. There are a number of issues regarding the populations selected for inclusion in most of the other decision impact studies. Clinical advice received by the EAG was that it is reasonable to assume that clinicians would interpret the results of each of the 3-level tests in the same way. Further, it is unclear how the company propose that the alternative analysis should be implemented. We have not undertaken the sensitivity analysis proposed by the company.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				The EAG report and model are ACIC redacted, which makes it difficult to judge the implications of this assumption. Broadly it seems that the Prosigna test may have higher predicted disease-free survival within category compared to Oncotype Dx, particularly for the intermediate risk group. Patients with a lower risk of recurrence have less capacity to benefit from chemotherapy and may be more likely to forgo chemotherapy. The crude assumption that Oncotype Dx post-test probability of chemotherapy applies to all tests may have overestimated the use of chemotherapy in patients who receive a Prosigna test result. As the overestimation relates to patients who, on average, may have lower risk of recurrence and less capacity to benefit from chemotherapy, this may have biased downwards the expected health gains from Prosigna specifically. A sensitivity analysis should be provided that explores making the probability of post-test chemotherapy a function of risk score and associated risk of recurrence, in order to appropriately distinguish between the tests. The information required to conduct such a sensitivity analysis is available from the TransATAC study, and has for example been used by the EAG to estimate the risk of recurrence dependent on risk categorisation specific to each of the different three-level test types. This sensitivity analysis would help to address the potential biases in the estimated post-test probability of chemotherapy and the consequent health benefits from chemotherapy.	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Reference *1. Enewold L, GeigerJo AM, Zujewski A, Harlan LC. Oncotype Dx assay and breast cancer in the United States: usage and concordance with chemotherapy. Breast Cancer Research and Treatment (2015) 151: 149*	
Nano String Technologies Inc.	8	361	5.3.2	As noted by the Royal College of Physicians joint with National Cancer Research Institute, Royal College of Radiologists, Association of Cancer Physicians and Joint Collegiate Council for Oncology in response to previous guidance,[1] if one of these gene expression profiling tests has the ability to predict benefit from chemotherapy, this is likely to extend to all such tests. While currently no direct evidence exists as to the predictive benefit of Prosigna, one of the advantages of cost-effectiveness models is the ability to explore the sensitivity of the results to the potential that Prosigna does predict differential sensitivity to chemotherapy. The use of sensitivity analysis to explore this issue is especially important as the timescale required to generate direct evidence on the predictive benefit of Prosigna is very long, and denying access to patients in the meantime risks continued inappropriate targeting of chemotherapy and associated potential harms. The decision by the EAG to restrict exploration of predictive benefit to only one of the technologies (Oncotype Dx) is unbalanced. As the EAG model already applies a post-test probability of chemotherapy based on Oncotype Dx results to other tests, it would be simple to extend the predictive benefit	The cited suggestion that if one gene test has the ability to predict benefit from chemotherapy then this is likely to extend to all such tests is not supported by the available evidence; there are no data on the differential effect of chemotherapy in different risk groups for three of the tests (Prosigna, EndoPredict and IHC4), and the data available for two of the tests (MammaPrint, Oncotype DX) are subject to limitations and uncertainty. The EAG notes that the tests comprise different genes/markers to each other, and as such, the assertion that all the tests can predict chemotherapy benefit if one can does not necessarily follow. There are no empirical data to suggest Prosigna can predict benefit from chemotherapy, and without this, a sensitivity analysis would carry no importance.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				sensitivity analysis to all tests. It is anticipated that incorporating a predictive benefit for Prosigna would increase the estimated QALY gain markedly, which would significantly reduce the corresponding ICERs compared to current practice. **Reference** 2. https://www.nice.org.uk/guidance/dg10/documents/gene-expression-profiling-and-expanded-immunohistochemistry-tests-to-guide-the-use-of-adjuvant-chemotherapy-in-early-breast-cancer-management-mammaprint-oncotype-dx-ihc4-and-mammostrat-second-diagnos2	
Nano String Technologies Inc.	9	183	4.5.1	We respectfully request the inclusion of the following evidence sources supporting the analytical validity of Prosigna for use in a decentralized testing environment in the DAR: 1. *Nielsen T, Wallden B, Schaper C, Ferree S, Liu S, Gao D, et al. Analytical validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using Formalin-fixed paraffin-embedded breast tumor specimens. BMC Cancer. 2014 Mar 13; 14(1):177.* 2. *Martin M, Gonzalez-Rivera M, Morales S, Haba-Rodriguez J, Gonzalez-Cortijo L, Manso L et al.*	The EAG was not able to conduct a review of analytical validity for all the tests, due to time and resource constraints. Inclusion of these references without an independent, unbiased search strategy would introduce a high risk of bias and as such the EAG is not able to comply with Nano String's request.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Prospective study of the impact of the Prosigna assay on adjuvant clinical decision-making in unselected patients with estrogen receptor-positive, HER2-negative, node-negative early- stage breast cancer. Curr Med Res Opin. 2015 Jun; 31(6) 1129-37. 3. *Wuerstlein R, Sotlar K, Gluz O, Otremba B, von Schumann R, et al. The West German Study Group Breast Cancer Intrinsic Subtype study: a prospective multicentre decision impact study utilizing the Prosigna assay for adjuvant treatment decision-making in estrogen receptor-positive, HER2-negative early-stage breast cancer. Curr Med Res Opin. 2016 Jul; 32(7) 1217-24.* 4. *Hequet D, Callens C, Gentien D, Albaud B, Mouret-Reynier MA, et al. Prospective, multicentre French study evaluating the clinical impact of the Breast Cancer Intrinsic Subtype-Prosigna Test in the management of early-stage breast cancers. PLoS One. 2017 Oct 18; 12(10): e0185753. doi: 10.1371/journal.pone.0185753. eCollection 2017.* Nielsen et al. (#1 above) summarizes the analytical validation studies used to support the CE-IVD and FDA 510(k) clearance of the Prosigna test and instrument. Although the following three references (#2,3,4 above) report results from decision impact studies of Prosigna, each study contained an analytical reproducibility sub-study where each patient sample was tested a second time at second laboratory within	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				the country performing the study (Spain, Germany, and France). Each of these three 200 patient studies confirmed the analytical reproducibility performance characteristic as established in Nielsen et al (#1 above).	
Agendia N.V.	1a	**General**	Full Report	**We strongly disagree with the overall outcome for MammaPrint based on this assessment and the AEG model used.** We would like the authors to acknowledge throughout the report that MammaPrint is the only assay that has level 1A clinical evidence from prospective data on clinically high risk patients that addresses the question: which patients that are candidates for chemotherapy can safely forego chemotherapy.	The EAG agree that MammaPrint is the only one of the five tests to have reported evidence of a RCT (MINDACT) where patients were randomised to treatment guided by the test or by usual clinical practice, and patients with high-clinical but low-MammaPrint risk showed a non-significant effect of chemotherapy. This has been added as an addendum to NICE.
Agendia N.V.	1b	**General**	Full Report	All other tests have data on overall low risk patient groups but as NICE excludes patients with a NPI<3.4 for chemo and almost all mAOL High Risk have an NPI >3.4 MammaPrint is the only assay with prospective data in the intended use population.	We have added to the addendum to NICE that MINDACT randomised patients who were high-risk via either mAOL or MammaPrint. The statement is not entirely fair on the evidence base for other studies; the decision problem did not restrict to high risk patients only, though this was specified as a subgroup of interest.
Agendia N.V.	1c	**General**	Full Report	Although several clinical risk assessment tools are available and discussed in the report, the authors should recognise that there is only one such tool with level 1A evidence, which	mAOL has not been disregarded – rather, the EAG model for MammaPrint uses mAOL directly. The EAG considers that because A!OL

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				is Adjuvant Online. Although it is currently off line, the Cardoso et al., 2016 NEJM article provides a simple table format of mAOL (S13 appendix to the NEJM article) that allocates patients into clinically high or clinically low risk. This table is as easily accessible as for example NPI and could be used by anyone that needs to assess a patient's risk of recurrence. We argue that mAOL should thus not be disregarded due to the fact it is currently offline.	is currently offline, and other risk tools may be used to determine clinical risk, the economic analysis of MINDACT is limited in this respect.
Agendia N.V.	1d	**General**	Full Report	Another important issue we would like to point out is the fact that the authors make CT benefit obligatory for a positive assessment, which seems a deviation from the primary aim of this report. Patients gain from these tests without the proof of CT benefit, given that the outcome for patients with low risk results as sufficiently low to forego CT. As mentioned in the 'Aims and objectives of the assessment' section on P44 "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?" focus is on guiding adjuvant chemotherapy decision not on the predictive value of tumour profiling testing.	We disagree with this comment. The Assessment Report includes an assessment of prognostic performance and, where evidence allows (and claims are made), predictive benefit of chemotherapy. The model is informed by the systematic reviews.
Agendia N.V.	1e	**General**	Full Report	We also argue that the MINDACT data should be included in the prognostic evidence supporting MammaPrint. From the concordant groups (C-low/ G-low and C-high/ G-high) and in the discordant groups of the study (Clin low/ MP high or Clin high/MP low) where randomization took place, one arm of each discordant group was treated according to baseline	In the EAG report, we define Clinical Utility studies as those assessing effect of prospective use of tests on patient outcomes. Therefore MINDACT is included under Clinical Utility. We define Prognostic Performance studies as those assessing whether patients with high or low test

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				clinical parameters, thus not influenced by the use of MammaPrint and can therefore be seen as the control arm. The data from this RCT thus provides additional prognostic data that should not be ignored.	scores have different outcomes, where the test did not influence treatment. We take the point that it may be possible to generate prognostic performance data from MINDACT by comparing outcomes for low-MMP vs high-MMP patients using the concordant-risk groups plus the discordant-risk groups in which treatment was determined by mAOL rather than MammaPrint. However, we were not able to locate these data in the time available to respond to these comments. The EAG report notes (within Section 4.4 covering MINDACT) that, 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). This analysis does not omit the patients treated according to MammaPrint, but the adjustment for other factors may mitigate this. We have noted in the addendum to NICE that these data could potentially be considered prognostic data. This is consistent with the findings of other MammaPrint prognostic studies which showed that MammaPrint was

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
					statistically significantly prognostic in multivariable analyses.
Agendia N.V.	1f	**General**	Full Report	We also strongly disagree with the AEG model and the assumptions that were made in order to fit MammaPrint in this model. First of all, the AEG model and its variables are not taking the (long-term) side effects of chemotherapy appropriately into account. There is plenty of evidence available on the (long term) side effects of chemotherapy which the authors have ignored. We also question the values used regarding utilities in this model. The paper of Campbell et al. 2011 was used for the disutility for chemotherapy during 6 months in the model. In this paper itself however, they mention that the disutility for chemotherapy must be used for at least 1 year. We therefore recommend to use the chemotherapy decrement for at least 1 year instead of 6 months. This could heavily impact the incremental QALY in the final results. In fact, if we incorporate these utilities for the first year in our model, (for each risk group separately), the QALYs for the MammaPrint yield more compared to the mAOL and NPI in the clinical high risk group.	The EAG model assumes that the impact of chemotherapy is applied for 1 cycle. We note that Table 131 of the EAG report refers to this impact is as a "disutility" – this should have stated that the parameter is applied as a QALY loss (hence it reflects a full year impact, but is applied in the first cycle). This can be seen in the model formulae. We also note that the EAG sensitivity analyses include doubling this disutility. As shown in the results, this does have some impact for some of the MammaPrint analyses, but the ICERs remain high (>£70,000/QALY or MammaPrint remains dominated). The Agendia model uses the same AE disutility from Campbell et al. One difference between the EAG model and the Agendia model is that the Agendia model applies this decrement arbitrarily for 2 years. The additional evidence on long-term AEs is not referenced within the company's response, nor is it used in the Agendia model.
Agendia N.V.	1g	**General**	Full Report	We would like the authors to also recognize the limitations of the TransATAC study. Instead of making TransATAC the gold standard for the model, based on the level of existing	We have noted in the Addendum to NICE that MammaPrint is the only one of the five tests to have reported randomised controlled trial

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				evidence it should be the other way around: 'All tests except MammaPrint lacked level 1A evidence the information derived from a prospective randomized trial, hence an alternative source (TransATAC) was used.' The patient population included in the TransATAC study was limited to only postmenopausal and ER+ patients. This means that patients eligible for the trial had an indication for endocrine therapy and are therefore a group of patients with a lower risk in general. In other words, based on the TransATAC trial patients were of lower risk and can't be a representative starting point (bias). Those patients were no candidates for chemotherapy in the first place so are not suitable for an assessment to address the question whether patients can safely forego chemotherapy.	evidence (MINDACT) of treatment guided by the test versus usual practice, in patients who are high-risk via either mAOL or MammaPrint. We agree that the TransATAC trial selected patients who had not had chemotherapy, the majority of whom are likely to not have been indicated for chemotherapy. The decision to use TransATAC was taken pragmatically as four of the five tests had been conducted in the patient population, and we were able to split patients according to NPI and nodal status. Please see also the EAG response to Agendia comment 1b (comparison of MINDACT and TransATAC population level % tumour size, grade and LN status)
Agendia N.V.	1h	**General**	Full Report	We also want to point out that the input for the independent cost-effectiveness analysis is based on very different sources. For example, for current practice, the baseline probability of receiving adjuvant chemotherapy is based on the clinical judgement of one person, Professor Rob Stein. Furthermore, the probability of receiving chemotherapy conditional on results of the test is based on the UKBCG survey, which is based on the opinion of 11 breast cancer experts. The use of all these sources is not optimal, especially when there is raw survival data available at the EORTC which use would much better reflect current practice and outcome plus would be much less biased than	We agree that the sources for pre- and post-test chemotherapy probabilities were not perfect. We used the best evidence that we had access to, which in this case, was clinical opinion. This limitation is highlighted in the EAG report. It is unclear how the raw EORTC survival data can help inform chemotherapy probabilities in the model. In addition, we reiterate that the results of the Agendia model are difficult to interpret because the model assumes that all chemotherapy decisions are based on the test (i.e. all high risk receive chemo, no low risk

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				estimations based on personal expert opinions. Instead of the above mentioned expert opinions, we feel that raw survival data from MINDACT should be used. MammaPrint is cost-effective for the clinical high risk group as shown in our model provided that is based on the probabilities founded on the raw survival data from the MINDACT trial. Please find in addition to this document, confidential, the model with the proper probabilities and utilities derived from the raw data of the MINDACT. These numbers can be used in the EAG model. Besides, we also requested the raw survival data at the EORTC for the use of NICE, existing of the OS and DMFS survival rates for the concordant groups, clinical low-genomic high received chemotherapy, clinical low-genomic high not received chemotherapy, clinical high-genomic low received chemotherapy, and clinical high-genomic low not received chemotherapy. If this data will be used, we expect different outcomes than provided by the currently used AEG model. In particular, we expect the clinical high risk group to yield more (QA)LYs for the MammaPrint, as the quality adjusted survival for this groups turns out higher compared to the mAOL and the NPI in our own analyses.	receive chemo). This is not realistic and still appears to be the case in the company's new model submitted following consultation.
Agendia N.V.	1i	**General**	Full Report	If the conclusion of the report is that MammaPrint is not cost effective and not available for UK patients, mAOL high risk patients in the UK should all receive chemotherapy based on this assessment. The report argues that actual treatment in the UK is different but this is not evidence based. As none of the other assays that are evaluated have data in clinically high risk patients to show that it is safe to forego	This contrasts with the clinical opinion received by the EAG and the assumptions employed in the EAG model (see EAG report, page 364).

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				chemotherapy and that there is no significant or clinically meaningful benefit of chemotherapy, this would result in knowingly treating 46% of mAOL high risk patients with a very toxic therapy for which they do not receive any significant or clinically meaningful benefit.	
Agendia N.V.	1j	**General**	Full Report	With our model, we have shown that MammaPrint is cost effective in the UK setting, however the authors have not provided feedback on our model on why they conclude that the data is not correct. Moreover, the model and its calculations are black in the report so we cannot provide input and request visibility of this model so we can properly rebuttal the outcome. The MINDACT trial showed that MammaPrint in mAOL high risk patients can reduce chemotherapy with 46% without compromising outcome. Still the AEG model shows a higher QALY for patients without MammaPrint test than with a MammaPrint test. We strongly disagree with the outcome that the QALY without test is higher than the QALY with test in the mAOL high risk category. This would mean that withholding chemotherapy in 46% of mAOL high risk patients without compromising outcome has no health benefit.	The EAG notes that the new Agendia model makes the same fundamental mistake in the interpretation of Kaplan-Meier curves as the original submitted model. The results of the company's new analyses cannot be considered reliable. These issues are described on pages 326-331 of the EAG report. We do not have control over which information are redacted from the report. It appears that our critique and correction of the Agendia model has been redacted – this is because the information submitted to NICE by Agendia was provided in confidence.
Agendia N.V.	2	**17**	2.4.1	"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	The EAG were aware that B20 was the training set for Oncotype DX and state this in the report "Two analyses are presented, one of the tamoxifen monotherapy arm, which was also as a training set for Oncotype DX, and one of the

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

Diagnostics Assessment Report (DAR) - Comments

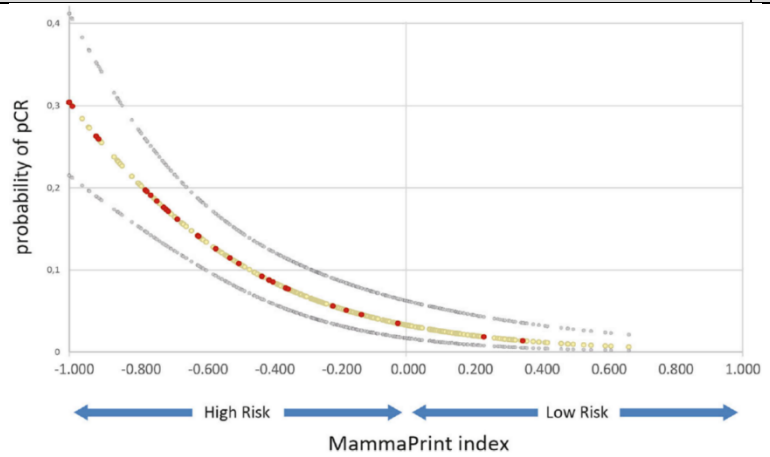
Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				interaction tests sometimes became non-significant when clinicopathological factors were adjusted for. Oncotype" **The B20 trial study should be excluded for predictive claim of Oncotype** **The data regarding the evidence of being predictive is flawed. The first publication of CT benefit for Oncotype selectively included patients from one arm of the study population used for the development of the Oncotype test, which leads to an inflated effect (Paik et al. 2006, ref 49 of DAR). The External Assessment group should omit the Paik et al 2006 study because it has severe flaws with major implications for the outcome of this study. It used 233 samples from the B20 tamoxifen treated arm for training, and used these same samples again in a comparative analysis for the chemotherapy prediction. The re-use of training samples is a methodological flaw and especially in this comparison where the re-use of this arm provides a selective advantage.** **<u>Use of B20 study as validation study: Paik 2004 NEJM; Paik 2006, Tang 2011</u>** **These studies use NSABP B20 study as validation but NSABP B-20 was used for training the Oncotype algorithm (See supplement to Paik 2004 NEJM "we weighted the NSABP B-20 study results most heavily in selecting the final gene list and algorithm").** **Therefore, this series cannot be used as validation or chemotherapy benefit studies as also explained by Ioannidis (2006, Nat Clin Pract Oncol): *"The greatest***	tamoxifen plus chemotherapy arm, which was not a training set for Oncotype DX. Patients were LN0." (Pg 66) Approximately 1/3 of the total number of patients in the trial was used as the training set. B20 was not the sole training data set for Oncotype DX, though it was more heavily weighted in the derivation of the algorithm. As such, the EAG decided to include the data, with the proviso that it was from the derivation set. However, the EAG agree that it would make sense to exclude the ET monotherapy arm from the prognostic dataset. This has been included in an addendum to the report. The EAG had included B-20 in the chemotherapy benefit analysis as it is the only data available in LN0 patients. The EAG agree that it was not clear in the report that this study carries a risk of bias, and that the analysis should be interpreted with caution. We have included this in the addendum. The use of the 50-point difference in the analysis of an interaction between RS and chemotherapy benefit in Albain et al.[1] does not indicate the clinical significance of the 18 -30 RS cut points. However, the study does

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				concern regarding Paik et al.'s study is that tamoxifen-treated patients from the NSABP B-20 study were used in the original development of the RS, and data from these patients were important in the selection of the 21-gene signature. RS is thus expected to (and does) differentiate the risk within the tamoxifen arm, since it has been trained purposely on these data. Conversely, RS does not appropriately differentiate." **The only true independent CT benefit study for Oncotype is Albain et al. 2010 (ref 68 of DAR) in LN+ patients, where the statistical significance is reported only for a 50 point increment of the Recurrence Score, which is not a clinically useful representation of the test. Clinically meaningful would be the hazard ratio between low and intermediate risk groups.**	conclude from the same analysis that there is little benefit from chemotherapy at RS<20 (though see other criticisms relating to lack of adjustment for other covariates in this analysis).
Agendia N.V.	3	17	2.4.1	"Evidence relating to the ability of MammaPrint to predict benefit from chemotherapy was extremely limited." **In the MINDACT design it was considered unethical to withhold chemotherapy in Clinically High Genomic High patients, the group where prospective predictiveness could have been established. MINDACT has shown that there is no significant benefit of chemotherapy in three of the 4 subgroups of the trial (if either Clinical is low or MammaPrint is low there is no significant benefit of chemotherapy). The authors could model the overall chemotherapy benefit from the EBCTCG overview to be**	We agree that MINDACT could not ethically randomise clinically-high MammaPrint-high risk patients to no chemotherapy. We also take the point that it is generally difficult to obtain data on high-risk patients randomised to chemotherapy or no chemotherapy, therefore it is difficult to assess chemotherapy benefit (we have noted this in the Addendum to NICE). However, this means that the remaining evidence was from pooled analyses of cohort studies. As stated in the ERG report, although

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				exclusively present in the Clinically High MammaPrint high risk group essentially enriching chemo benefit a 4 fold as only 25% of the ER+ patients fell into the Clinically High MammaPrint high category. **We also suggest the authors to regard the available neo-adjuvant data for the CT benefit of the MammaPrint test, given that patient samples from retrospective randomized studies for CT versus no CT are not available anymore which makes determining CT benefit from these type of study almost impossible. Therefore the assessment bodies should allow for reviewing alternative study set-ups for determining CT benefit.** **CT predictive evidence for MammaPrint in neoadjuvant setting:** ***Whitworth et al. Ann Surg Oncol 2016*** **Probability of pCR (ypT0/isN0) to NCT for the MammaPrint index (n = 405), and probability of pCR as a function of the MammaPrint index. The MammaPrint index is positively associated with the likelihood of pCR (p\0.001), suggesting that patients who are at the highest risk of recurrence are more likely to have chemotherapy benefit. pCR pathological complete response, NCT neoadjuvant chemotherapy.**	these showed that the effect of chemotherapy was significant in high-risk groups and not in low-risk groups, the interaction tests between risk groups and chemotherapy treatment were not significant, suggesting no statistically significant difference in effect of chemotherapy between risk groups. Unfortunately the assessment of neoadjuvant data was beyond the scope of this (already large) assessment. In addition, chemotherapy benefit prediction in the neoadjuvant setting may not be generalisable to the adjuvant setting, as the profiles of tumours have been shown to change after neoadjuvant treatment.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				 **Whitworth, P., Beitsch, P., Mislowsky, A. et al. Ann Surg Oncol (2016)**	
Agendia N.V.	4	17	2.4.1	'The MINDACT…. gave an absolute benefit of 1.5% in 5 year DRFI.' **Correct into The MINDACT trial did not show a significant benefit in 5 year DRFI. This is a crucial difference. The NEJM mentions the 1.5% difference but it is insignificant, meaning that one can not say there is a difference. Within the 95% confidence interval the opposite could be true.**	The EAG accepts this comment. An erratum has been provided to NICE to change the text as follows: "The MINDACT randomised controlled trial (RCT) for MammaPrint reported that for patients who were high-mAOL, low-MammaPrint risk, chemotherapy gave a non-significant absolute benefit of 1.5% in 5 year DMFS (p=0.267). This met the primary objective in that the lower

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				The non significant difference of 1.5% benefit was based on DMFS (instead of DRFI) and it should be mentioned that it is a non-significant difference. **In the development phase of MINDACT, a survey amongst women and their physicians was held to identify how much benefit chemotherapy must provide in order to be willing to undergo such therapy. At least 2% benefit turned out to be the minimal benefit needed to be worth the toxicity. MINDACT showed a non-significant difference of 1.5% between CT/No CT and is well below the at least 2% reduction in survival due to chemotherapy induced toxicities and below the 2% required benefit of CT as indicated by the survey. Additionally according to *Lippman et al Journal of the National Cancer Institute. 2001*, it was generally agreed by most physicians that an added absolute benefit of 3% survival is necessary to justify recommending chemotherapy.** **We believe that the sentence** 'This raises the possibility of avoiding chemotherapy in these patients.' **is an understatement and by far not covers the most important finding of the MINDACT study. MINDACT met its primary end point meaning that mAOL High MammaPrint Low risk can safely forego chemotherapy. The 5 year DMFI at 95% is so high that clinicians do not consider chemotherapy to be appropriate for this group. Furthermore there is no significant benefit of chemotherapy in this group and if the 1.5% difference**	bound of the 95% CI for 5-year DMFS in the no-chemotherapy group was at least 92%. This finding was interpreted by the authors as implying that patients who were high-clinical but low-MammaPrint risk could potentially avoid chemotherapy."

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				was significant, it is too low to justify chemotherapy based on the toxicity and side effects.	
Agendia N.V.	5	20	2.5	'..(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 …' **The fact that the current AEG model reflects outcome with previous model and that with Genomic Health's model does not proof this model to be correct or strong. Moreover, as Genomic Health is one of the comparators in this assessment it seems unfair to use a Genomic Health model to 'validate' results and therefore indicate the strengths of the EAG model.**	We do not claim that the EAG model is correct. To the contrary, we highlight nearly 2 pages of important limitations relating to the analysis (see EAG report, 409-410). What is relevant here is that when based on the same data, the original EAG model, the new EAG model and the new Genomic Health model all produce consistent conclusions.
Agendia N.V.	6	20	2.5	"..(iii) the analysis of MammaPrint is based on a different data source than the other four tests;…" **MammaPrint is the ONLY test with level 1A evidence for the group of patients (clinical high/ genomic low) where the question whether to give CT or not is most relevant. Authors should stress that the data available for MammaPrint is the highest possible level of evidence. (MammaPrint is the only test which has highest level of evidence based on a prospective RCT). The authors should also address the fact that the ATAC trial enrolled patients that were never candidates for chemotherapy so by far the ideal trial to identify patients for which it is safe to forego chemotherapy. The trial is also limited to post-menopausal woman and the validity of the data is only in post-menopausal woman.**	The EAG agree that MammaPrint is the only one of the five tests to have reported evidence of a RCT (MINDACT) where patients were randomised to treatment guided by the test or by usual clinical practice. Patients with high-clinical but low-MammaPrint risk showed a non-significant effect of chemotherapy. This has been added to an Addendum to NICE. With respect to the ATAC trial "Patients were ineligible if there was any clinical evidence of metastatic disease; if chemotherapy was started more than 8 weeks after surgery or completed more than 8 weeks before starting randomised

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
					treatment (neoadjuvant chemotherapy was not allowed) or, in patients not receiving chemotherapy, if primary surgery was completed more than 8 weeks before starting randomised treatment;" which implies that some patients had had adjuvant chemotherapy already. TransATAC then selected patients "who did not receive adjuvant chemotherapy, had the GHI-RS computed, and had adequate tissue for the four IHC measurements: ER, progesterone receptor (PgR), human epidermal growth factor receptor 2 (HER2), and Ki-67" As such, we agree that the TransATAC trial selected patients who had not had chemotherapy, the majority of whom are likely to not have been indicated for chemotherapy. The decision to use TransATAC was taken pragmatically as four of the five tests had been conducted in the patient population, and we were able to split patients according to NPI and nodal status. A quick comparison of the three factors that contribute to the NPI (Tumour size, tumour grade, number of lymph nodes) in MINDACT compared with TransATAC) shows that whilst

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
					TransATAC has smaller % with high risk characteristics, the difference is not massive, and some high risk patients who would have been indicated for chemotherapy in the UK appear to have been included (e.g. LN>3) For MINDACT versus TransATAC respectively: Tumour size <2: 71.6% vs 67% Tumour size 2 to 5: 27.2% vs 31% Tumour Grade 1 (well differentiated): 21.6% vs 27% Tumour Grade 2 (moderately differentiated): 49.1% vs 52% Tumour Grade 3 (poorly differentiated): 28.8% vs 16% LN0: 79% vs 71% LN+: 21% vs 29%
Agendia N.V.	7	21	2.7	"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." **The authors make CT benefit obligatory for a positive assessment, which seem unreasonable as these tests have all been developed to determine the risk of breast cancer recurrence not to determine benefit of CT. Most important is the evidence for a test to provide accurate risk classification and patients gain from these test**	The statement in the EAG report is fair. The company's statement is not accurate and their interpretation of the report is unreasonable. The EAG clinical review considers the evidence for both prognostic benefit and predictive benefit. The EAG model includes prognostic benefit in the base case, as well as a sensitivity analysis in which Oncotype DX is assumed to be predictive of chemotherapy benefit. We do not suggest at any point in the report that a tumour

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				without the proof of CT benefit, given that the outcome for patients with Low Risk results as sufficiently low to forego CT. **In addition, it is an improvement over current practice.** **As mentioned in the 'Aims and objectives of the assessment' section on P44 "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?" focus is on guiding adjuvant chemotherapy decision not on the predictive value of tumour profiling testing.** **When the MINDACT study is used and presented accurately MammaPrint does not necessarily need predictive data. The primary analysis of the MINDACT trial showed that withholding CT from Clinically-high risk/MammaPrint-low risk (C-high/MP-low) patients does not detrimentally impact outcome. This is a huge benefit to patients & impact on clinical practice as this is the case in 46% of the C-high/ MP-low patients.** **As discussed by Hudis in N Engl J Med 2016 "On the basis of the MINDACT study, clinicians may consider ordering the 70-gene signature for patients in line for chemotherapy who hope to forgo it on the basis of a possibly low genomic risk."**	profiling test can only be valuable if it is predictive of chemotherapy benefit.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Agendia N.V.	8	21	2.7	"There is limited evidence demonstrating long-term impacts resulting from the use of the five tumour profiling tests." **There is 10 year outcome data available for MammaPrint (Vliek et al ESMO 2017). Authors should acknowledge that most if not all chemotherapy benefit is in the first 5 years (EBCTCG overview) so for the chemotherapy decision 5 years follow up is sufficient.**	The EAG note the point that there are 10-year data (conference abstract only) from the prospective RASTER study of MammaPrint. This data do have some limitations; for example, some MammaPrint low-risk patients (15%) had chemotherapy, while some high-risk patients (9%) did not. We think that, as a general limitation, the point about limited long-term data on all tests still holds.
Agendia N.V.	9	36	3.3	"MammaPrint is a CE marked microarray test that is designed.." **The authors should mention the FDA clearances available for MammaPrint. MammaPrint has 6 FDA clearances.** **MammaPrint pre menopausal Fresh Frozen/2007 /K062694** **MammaPrint Ambient Temperature/2007/K70675** **Use of High Density Microarray Chip/2008/K08252** **MammaPrint in post menopausal women/2009/K81092** **MammaPrint in all Agendia controlled Laboratories/2011/K101454** **MammaPrint in Formalin Fixed Paraffin Embedded Tissue/2015/K141142**	It was not in the scope of the assessment to review or report this type of data for any of the tests.
Agendia N.V.	10	36	3.3	"…recurrence within 5 years and whether a woman would benefit from chemotherapy." **MammaPrint is designed to assess risk for patients at 10 years (see FDA clearance Code of Federal Regulations.**	See errata – page 36.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				2007. 21 CFR 866; Classification of Gene Expression Profiling Test System for Breast Cancer Prognosis. 72: 89, 26290-91). MammaPrint is designed to determine if a patient is at sufficiently low risk to forgo chemotherapy. **The safety and effectiveness of the MammaPrint test can be demonstrated by the fact that MammaPrint received 510(k) FDA clearance as well as CE marking. The first 510(k) IVDMIA clearance in 2007 by the Food and Drug Administration (FDA) in a De Novo Classification Process (Evaluation of Automatic Class III Designation).** **MammaPrint® FFPE received a Predicate Device 510(k) clearance in 2015 (U.S. Food and Drug Administration (FDA). MammaPrint 510(k) Substantial Equivalence Determination Decision Memorandum, May 20, 2015: https://www.accessdata.fda.gov/cdrh_docs/reviews/k141142.pdf). Agendia's FDA clearances for MammaPrint are publicly available at fda.gov (k062694, k070675, k080252, k081092, k101454, k141142). Link to website of Agendia licenses and accreditations:** http://www.agendia.com/our-science/accreditations-licenses/.	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Agendia N.V.	11	36	3.3	"In Europe, samples are sent for analysis at the Agendia laboratory in Amsterdam, the Netherlands." **Please provide Agendia's credentials similar to GH's (CAP, etc)** • **Food and drug administration (FDA, Link to website of Agendia licenses and accreditations:** http://www.agendia.com/our-science/accreditations-licenses/. • **ISO 13485:2003** • **21 CFR 820 – US FDA Quality System Regulation (QSR)** • **In vitro diagnostic medical devices 98/79/EC (for Agendia EU)** • **Clinical Laboratory Improvement (CLIA) since Sept. 2008** • **College of American Pathologists (CAP) since Dec. 2009** • **US State Requirements applicable for diagnostic testing laboratories (for Agendia Inc.)** • **21 CFR 803 – Medical Device Reporting**	Information noted. The report has not been amended.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				• **21 CFR 806 – Medical Devices – Reports of Corrections and Removals**	
Agendia N.V.	12	36	3.3	"Oncotype DX is designed to assess the risk of distant recurrence within 10 years and predict the likelihood of chemotherapy benefit." **Oncotype is not designed to predict CT benefit. This has not been proven significantly. So, the test is a prognostic, not a predictive test.** **As explained in an earlier comment no. 2 that the data regarding the evidence of being predictive is flawed. The first publication of CT benefit for Oncotype selectively included patients from one arm of the study population used for the development of the Oncotype test, which leads to an inflated effect (Paik et al. 2006, ref 49 of DAR). The only true independent CT benefit study for Oncotype is Albain et al. 2010 (ref 68 of DAR) in LN+ patients, where the statistical significance is reported only for a 50 point increment of the Recurrence Score, which is not a clinically useful representation of the test. The prospective TAILORX trial is designed to give chemotherapy to all patients with a recurrence score of 26 and higher. If this trial reports and the new high risk group is 26 and higher there will be no prospective or retrospective predictive (and retrospective prognostic) data for this group. The St Gallen guidelines already base their recommendation on a cut off of 25 in recurrence score.**	We acknowledge the point that Oncotype DX may not have been designed to predict chemotherapy benefit, but it does report data on chemotherapy benefit. The report has not been amended. Regarding the use of Paik et al 2006, please see response to question 2. Regarding the Albain et al. analysis, please see response to question 2 also. The use of the lower cut-off point to define high risk patients is not a matter for this assessment; TAILORx has not yet reported, and currently the cut-off points recommended by Genomic Health are RS 18 and 30.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Agendia N.V.	13	37	3.3	"The recurrence score may also predict the benefit of chemotherapy." **Again, this has not been proven, please remove from report.**	This is not a factual inaccuracy; we say "may also", and the evidence available does not show that there is no effect.
Agendia N.V.	14	42	3.4.3	"*the definition of this "intermediate group" is not clear-cut. Clinical advice suggests…. to be at intermediate-risk."* **The cut-off with an NPI of 3.4 is an essential (critical point) in the assessment. It seems unjustified to base these numbers only on clinical advice from a few physicians without real proven clinical evidence or utility. This whole section is now based on assumptions and probabilities of certain patients falling within a certain risk group. This seems scientifically unsound and not the right starting point for an assessment as such. The Authors should acknowledge that there is only one clinical risk assessment tool with level 1A evidence, which is Adjuvant Online. Although it is off line, the Cardoso et al NEJM provides a simple table that allocates patients into clinically high or clinically low risk and can be used.** **As AOL has the highest clinical evidence for prognosis it should be used as comparator for clinical chemotherapy decision.**	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 EBC) or based on the new tests and are unlikely to receive chemotherapy, therefore their management is unlikely to change. Few patients with ER- LN- HER2- EBC will have an NPI score above 5.4 and therefore those will an NPI above 3.4 can be considered as being at intermediate risk.
Agendia N.V.	15	47	4.1.2	"For IHC4, as there is no commercially available version of the test, any methodology was included." **This does not make any sense from an analytical and clinical validity point of view. Although authors seem to**	These issues are covered in the addendum providing a rapid review of analytical validity of IHC4.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				take consideration of this on page 48: "A rapid review of IHC4 will follow as an addendum to this report.." Still, it is worthwhile to stress the requirements that are expected from the other tests, so IHC4 would need to show analytical validity in the exact similar stringent ways that are required for the other tests. Or else exclude the non-centralized data from IHC4. It is generally accepted and should be mentioned in the report that the reproducibility of KI67 test is too low to be implemented clinically. ASCO guidelines for that reason state that KI67 staining should play no role in treatment decisions in breast cancer.	
Agendia N.V.	16	48	4.1.2	"Prognostic performance, …. Study designs include: " **The most important and robust study design is not included as study designs mentioned for prognostic performance, which is: a prospective randomized phase 3 study design. Data from prospective randomized phase 3 studies, when available should be taken into account in assessing the prognostic performance of a diagnostic test. MINDACT trial results should be included.**	As noted in response to comment 1e, we define Clinical Utility studies as those assessing effect of prospective use of tests on patient outcomes. Therefore MINDACT is included under Clinical Utility. However we take the point that it may be possible to generate prognostic performance data from MINDACT. This is noted in our response to 1e and in the Addendum to NICE.
Agendia N.V.	17	49	4.1.2	"Prediction of chemotherapy benefit, …. Study designs include: " **Authors should mention the difficulty that arises from the first study designs for chemotherapy benefit prediction. RCT where "some patients received CT"; please note that it should say patients were *randomized* to receive CT.**	The EAG accepts this point and has included it in an Addendum to NICE as follows: "The EAG agrees with Agendia (comment #17) that it is difficult to undertake further assessments of predictive ability for chemotherapy benefit, since there are few trials

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				This type of trial is very rare, and the few trials that are available have insufficient patient samples left. So, this type of study design is impossible to adhere to for these and future tests. **Authors should acknowledge this difficulty and are therefore strongly advised to also include other types of studies such as neo-adjuvant studies that are more and more being recognized as appropriate study design for determining benefit of treatment, especially in specifically stratified patient subgroups.** **We suggest the authors to regard the available neo-adjuvant data for the CT benefit of the MammaPrint test [Whitworth, Ann Surg Oncol 2014 and 2017; Baron, Ann Surg Oncol 2015; Beitsch, Ann Surg Oncol 2016 and 2017]. Therefore the assessment bodies should allow for reviewing alternative study set-ups for determining CT benefit.** ***Bhatt et. al N Engl J Med 2016*** **Randomized clinical trials serve as the standard for clinical research and have contributed immensely to advances in patient care. Nevertheless, several shortcomings of randomized clinical trials have been noted, including the need for a large sample size and long study duration, the lack of power to evaluate efficacy overall or in important subgroups, and cost.**	in which patients were randomised to chemotherapy versus no chemotherapy, and the few trials of this type that are available have insufficient tumour samples left on which to undertake tumour profiling tests." Unfortunately assessment of neoadjuvant studies was beyond the scope of this report since results for neoadjuvant treatment may not be generalizable to adjuvant treatment. Please see Peony Breast Care Unit comment #3

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Hudis in N Engl J Med 2016 **further supported this notion that this type of trial and level of evidence is very rare and challenging. He made comments specific to MINDACT explaining that "the stated difference does not precisely exclude a benefit that clinicians and patients might find meaningful. An adequately powered randomization or a higher threshold for 5-year metastasis-free survival might have provided a more convincing result but would have raised other major challenges for the investigators."**	
Agendia N.V.	18	52	4.1.5	"for studies assessing…or published. **As this is a model specified for Prediction we question the relevance for determining the prognostic quality (risk of bias). As mentioned in the comment 2 and 7 concerning the importance of predictiveness of the test.**	Items from the PROBAST tool were selected based on their relevance to studies assessing prognostic or predictive benefit. The review team were not aware of any better quality assessment tool to use. No studies were excluded based on the quality assessment. Personal communication with the authors of PROBAST confirmed at the outset that the tool could assess risk of bias in studies assessing prognostic or predictive performance.
Agendia N.V.	19	57	4.2	"For MammaPrint, there were no LN+ endocrine monotherapy studies, but in studies with variable endocrine and chemotherapy use, 59-62% were high-risk (2 studies60, 61); similar to LN0." **For MammaPrint, MINDACT data is available; from Rutgers et al ESMO 2013, it can be inferred that % LR is 65% for LN+ and 64% for LN0. So indeed similar.**	Please see response to comment 1e and Addendum to NICE.

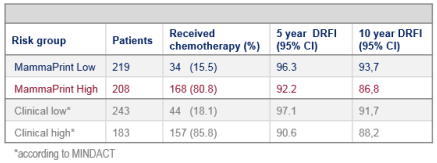
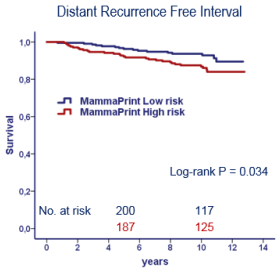
Tumour profiling tests to guide adjuvant chemotherapy decisions in people with breast cancer (update of DG10)

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				However, different from numbers mentioned above. Authors should include MINDACT data on 1405 LN+ patients here.	
Agendia N.V.	20	58	4.2	"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)." **The EORTC considered the MINDACT trial mature at 5 years for the chemotherapy decision. (Bogaerts et al. Nature Clinical Practice Oncology 2005). 6693 MINDACT patients' outcome should be included in this section.**	Please see response to comment 1e and Addendum to NICE.
Agendia N.V.	21	60	4.2	**LN0:** **5288 MINDACT LN0 patients should be included. MINDACT constitutes the highest level of evidence available for any prognostic gene assay for breast cancer.** **Also:** "There were no studies of MammaPrint in this population." **This is contradicting the just mentioned study. Please add poster presented at ESMO 2017 on 10-year FU from the RASTER data on LN0.**	As noted above, in the EAG report, studies assessing prospective use of the test on patient outcomes were defined as Clinical Utility studies rather than Prognostic Performance studies. Therefore, MINDACT and RASTER are summarised under Clinical Utility. Please see response to comment 1e and Addendum to NICE regarding MINDACT and prognostic data. For the Yao study, not all patients received endocrine monotherapy (even in the subgroup with no chemotherapy); this study is included, but not in the summary statement on studies of LN0 patients receiving endocrine monotherapy.

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

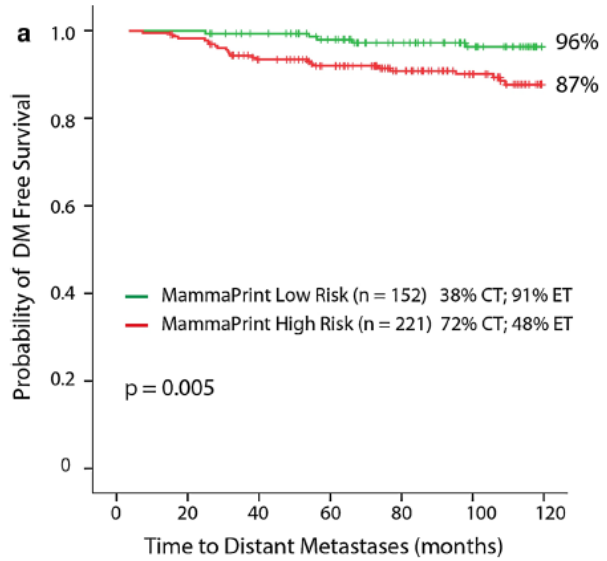
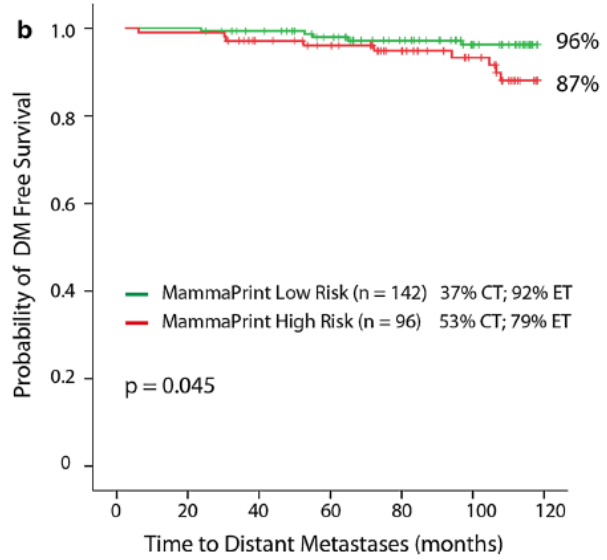
Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				10 year follow-up of the RASTER study (2004-2006) • Prospective evaluation of the MammaPrint® in community based hospitals. • 427 breast cancer patients of 60 years or younger with cT1-3N0M0. • Decision on adjuvant systemic treatment was based on: ✓ Dutch guideline (CBO 2004) ✓ Preference of patient and physician ✓ High or Low Genomic risk of distant recurrence (MammaPrint)  *according to MINDACT  **Also, Yao et al, BCRT 2015 can be used here. See figure below and also: MammaPrint low-risk patients (n = 93) who did not receive adjuvant CT had a 10-year DMFS of 98 % (95 %CI 94.0–100) and MammaPrint high-risk patients (n = 60) who did not receive adjuvant CT had a 10-year DMFS of 85 % (95 %CI 74.8–95.6), data not shown.**	

Risk group	Patients	Received chemotherapy (%)	5 year DRFI (95% CI)	10 year DRFI (95% CI)
MammaPrint Low	219	34 (15.5)	96.3	93,7
MammaPrint High	208	168 (80.8)	92.2	86,8
Clinical low*	243	44 (18.1)	97.1	91,7
Clinical high*	183	157 (85.8)	90.6	88,2

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

Diagnostics Assessment Report (DAR) - Comments

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				LN+: **MINDACT data should be included. MINDACT constitutes the highest level of evidence available for any prognostic gene assay for breast cancer.** **Please include:** **M*ook et al, (2009)* The 70-gene prognosis-signature predicts disease outcome in breast cancer patients with 1–3 positive lymph nodes in an independent validation study. BrCResTr;116(2):295-302** **Mook et al. (2010) Metastatic potential of T1 breast cancer can be predicted by the 70-gene MammaPrint signature. Ann Surg Onc; 17(5):1406-13**	
Agendia N.V.	22	60	4.2	"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." **Authors forget to mention the flaws in the predictive data: The first publication of CT benefit for Oncotype selectively included patients from one arm of the study population used for the development of the Oncotype test, which leads to an inflated effect. The only true independent CT benefit study for Oncotype is ALbain et al in LN+ patients, where the statistical significance is reported only for a 50 point increment of the Recurrence Score, which is not a clinically useful representation of the test.**	Please see response to comment 2 from Agendia.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Agendia N.V.	23	61	4.2	"The evidence for the ability of MammaPrint to predict chemotherapy benefit is therefore extremely limited;.." **Again, authors are urged to consider the plethora of neo-adjuvant data for predicting CT benefit by MammaPrint. See comment no. 17**	Unfortunately, assessment of neoadjuvant studies was beyond the scope of this report, since results for neoadjuvant treatment may not be generalisable to adjuvant treatment.
Agendia N.V.	24	61-62	Paragraphs 'Oncotype' and 'MammaPrint'	**In the conclusion of ODx it is specifically stated that** "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." **However in this section and in MammaPrint conclusion, the authors never use this direct language to propose the alternate: "With the highest level of evidence, it is possible to conclude that patient outcomes would be affected by use of the MammaPrint test in a clinical setting." Authors should use the same kind of language for all tests, please re-phrase this paragraph of MammaPrint in equal manner.**	This logic does not necessarily follow. While it is true to say that without the highest level of evidence it is not possible to be certain about something, it doesn't follow that the presence of some evidence means we are totally certain that a result is true. We feel that we have fairly and comprehensively summarised the results of the MINDACT and RASTER studies.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Agendia N.V.	25	62	2.4	"MammaPrint: Two studies reported evidence relating to the clinical utility of MammaPrint. MINDACT is an RCT of MammaPrint versus clinical practice." **To assign the MINDACT study 'only' to be of use in the Clinical Utility assessment is downgrading the MINDACT study. It indeed assesses MammaPrint utility, but also provides the highest level of evidence for prognosis. "However, the comparator was mAOL, and it is unclear whether the same would be true for other clinical risk scores."** **In the study by Viale et al, BCRT 2017, results clearly show that the MINDACT results also apply when using more contemporary comparators.**	We chose to define studies assessing prospective use of tests as Clinical Utility studies and report them in a separate section so they would not be lost among the many retrospective studies. We do not feel that this is downgrading the evidence. We feel that we have given MINDACT sufficient prominence in the EAG report. We did not identify the study by Viale 2017 as it was published in September 2017.
Agendia N.V.	26	62	4.2	"For patients…an absolute difference of 1.5%." **Authors should mention that this 1.5% absolute difference was <u>not significant</u> and therefore patients could safely forego chemotherapy.** **In the development phase of MINDACT, a survey amongst women and their physicians was held to identify how much benefit chemotherapy must provide in order to be willing to undergo such therapy. At least 2% benefit turned out to be the minimal benefit needed to be worth the toxicity. MINDACT showed a non-significant difference of 1.5% between CT/No CT and is well below the at least 2% reduction in survival due to chemotherapy induced toxicities and below the 2% required benefit of CT as indicated by the survey.**	The EAG accepts the comment about noting the non-significant difference. An erratum has been provided to NICE to change the text as follows: "For patients who were high-clinical, low-MammaPrint risk, 5-year DMFS was 95.9% with chemotherapy and 94.4% without chemotherapy, a non-significant absolute difference of 1.5% (p=0.267)."

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Additionally according to *Lippman et al, Journal of the National Cancer Institute. 2001*, it was generally agreed by most physicians that an added absolute benefit of 3% survival is necessary to justify recommending chemotherapy. **We believe that the sentence** 'This raises the possibility of avoiding chemotherapy in these patients.' **is an understatement and is by far not covering the most important finding of the MINDACT study. MINDACT met its primary end point meaning that mAOL High MammaPrint Low risk can safely forego chemotherapy. The 5 year DMFI at 95% is so high that clinicians do not consider chemotherapy to be appropriate for this group. Furthermore there is no significant benefit of chemotherapy in this group and if the 1.5% difference was significant, it is too low to justify chemotherapy based on the toxicity and side effects.**	
Agendia N.V.	27	63	2.4	**Comment on whole paragraph "Concordance":** Shouldn't concordance be defined by the degree of a test compared to current clinical risk assessment to assign the same patients to the same risk groups?	Concordance as we have used it in the EAG report has been defined here. Concordance can refer to the agreement between any two risk-assigning scores.
Agendia N.V.	28	63	2.4	**Quality of life data was assessed in the first 800 patients enrolled in MINDACT is published (Retel et al. BMC Cancer 2013).**	The EAG cannot find the data referred to in the BMC publication, which states "Of 566 patients we invited to participate, 347 returned completed questionnaires…" These data are reported in Table 97 of the EAG report.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Agendia N.V.	29	64	Table 7	**Add MINDACT study, Mook et al., Yao et al. and RASTER 10-years data.**	Tables 7 and 8 provide an overview of comparable data and are restricted to DRFS/DRFI outcomes. Complete data are presented in the main report. As noted earlier, MINDACT and RASTER are covered in the Clinical Utility sections.
Agendia N.V.	30	71-72	Oncotype	**Indeed, a statistical significant HR is reported for several studies for Oncotype for a "50-point difference in RS". Authors should address the clinical significance for this. A 50-point increment for RS is not of any clinical significance, since it does not reflect the risk groups for Oncotype, it generously overrides it. Also, a significant HR for the low risk group compared to the high risk does not suffice to report the clinical prognostic validity of the Oncotype test.** **The only meaningful HR would be a statistical significant difference between low risk and intermediate risk, which is often not significant, or not reported.** **Please see the numerous reporting of the above:** **"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 **********.[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 ******* ********* ********** **********[35, 43] while in a third study**	Whilst HRs for 50-point differences have been reported in many of the studies, all studies reporting unadjusted analyses also reported p values for risk rates between categories as well, usually for high versus low patients. Of those reporting adjusted analyses (additional prognostic value), several did rely on a 50-point difference analysis, or did not report whether it was for a 50-point difference. The EAG agree this does not provide any information about which cut point to use, but disagree that the studies are irrelevant – the statistical significance of a 50-point differences, implies a statistically significant change for a 1 point differences, and therefore implies that there would be a statistically significant difference between risk groups, but does not indicate which cut points are optimal, or how clinically meaningful the difference would be.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				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 **"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]"** **"One study [68] in LN+ patients reported a statistically significant 10-year HR for a 50-point difference in RS.."** **"..whilst Albain et al. 201068 (LN+) reported an HR for 10-year OS for a 50-point difference in RS of 4.42.."** **"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…"** **"…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.."** **"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]"** **"The NSABP B-14 analysis45 of LN0 patients treated with endocrine monotherapy showed that Oncotype DX was**	The EAG do not agree that the only meaningful comparison is between low versus intermediate patients. The most meaningful comparison depends entirely upon what is done with intermediate patients in clinical practice, and is more likely to be a comparison of low/intermediate versus high, or low versus high/intermediate. The EAG state that "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", which alludes to the fact that it is currently unclear how intermediate patients will be handled, and therefore what constitutes the correct comparison is unknown. As such, the EAG present the available data for deliberation by the Committee.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				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).”	
Agendia N.V.	31	74 and Table 9 P81	Oncotype	“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).” **The use of NSABP B-20 for validation should not be accepted as valid material.** **This is because the tamoxifen treatment samples in that analysis were also used to select the 21 genes and to develop the Recurrence Score used in the Oncotype DX assay. Therefore, the observed difference in outcomes between the chemotherapy and tamoxifen arms in the B-20 analysis might be exaggerated by training bias.** **And any use of these samples should be avoided or disregarded.**	The EAG agree with this point, and have made a correction in the Addendum.
Agendia N.V.	32	82	Table 10	**Applicability of ‘Test as per decision problem?’ should be NO for the following studies:** - **Albain et al 2010 – only significant per 50-point RS** - **Paik et al 2006 – Tam treated samples used for development of ODx** - **Tang – see comment no 31**	We defined the scoring of this item as relating to the performance of the test, not the performance of the analysis. Therefore, these changes have not been made.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Agendia N.V.	33	93	Table 17	**For all studies reporting "RS 50 point difference" – the clinical insignificance of this outcome should be addressed.**	Please see response to Agendia comment #30
Agendia N.V.	34	95	Table 18	**For all studies reporting "RS 50 point difference" – the clinical insignificance of this outcome should be addressed.**	Please see response to Agendia comment #30
Agendia N.V.	35	97	Table 19	**NSABP B20 data should be dismissed as per reason mentioned above, comment no 31 (Tam arm used for ODx development)**	Please see response to Agendia comment #2
Agendia N.V.	36	98	Oncotype NSABP B20	"It should be noted that some of the patients of the B-20 trial were used to derive the Oncotype DX score.[49]" **This sentence does not cover the essential data flaw for this study.** **Not "some" patients were used, but specifically the patients from the tamoxifen treatment samples in that analysis were also used to select the 21 genes and to develop the Recurrence Score used in the Oncotype DX assay. In the study these patients are then compared to the other patients.** **Therefore, the observed difference in outcomes between the chemotherapy and tamoxifen arms in the B-20 analysis is very likely exaggerated by training bias.**	Please see response to Agendia comment #2

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Agendia N.V.	37	100	Albain et al	"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" **Please note that this study only reports HR for a RS 50 point increment, which is not a clinically meaningful result**	Please see response to Agendia comment #30
Agendia N.V.	38	102	Tang et al	"Whilst the results from Tang et al. 2011a suggest that Oncotype DX is better at identifying individuals who would benefit from chemotherapy than AOL,.." **Tang et al includes the NSABP-B20 study, and cannot be used to determine CT benefit as per the above mentioned flaw. See comment no. 31**	Please see response to Agendia comment #2
Agendia N.V.	39	104	CT benefit	**The issues with the CT data should be mentioned (HR only for 50 point increment, and studies using NSABP B20 arm flawed.** **"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."**	Please see response to Agendia comment #30 and #2

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Agendia N.V.	40	106	Conclusion	**The conclusion section concerning Oncotype and CT benefit should be adjusted to reflect the remarks above. See comment no. 2 and 31**	Please see response to Agendia comment #30, #31 and #2
Agendia N.V.	41	108	Table 21	**Applicability of 'Test as per decision problem?' should be NO for the following studies:** - **Albain et al 2010 – only significant per 50-point RS** - **Paik et al 2006 – Tam treated samples used for development of ODx**	Please see response to Agendia comment #32
Agendia N.V.	42	109	Table 22	**Paik et al 2006 should not be included for reporting of CT benefit, given the fact that one arm of this study were the training samples for the ODx test.** **Given this notion, the ALbain et al. 2010 dataset is better reflective of CT benefit prediction of the OT test, which is non-significant.**	Please see response to Agendia comment #2
Agendia N.V.	43	114	Study design..	"… (TAILORx),106 randomises patients to treatment guided by the test or treatment according to usual practice." **Authors should make readers aware that TAILORx does not truly assess clinical utility; it only does so for Oncotype Intermediate patients.** **A study for 'true' measurement of clinical utility would compare with current clinical practice. Which is difficult for any such test, given that clinical practice changes over time.**	We believe our description is close enough to not require amending, especially as there are no results yet for the randomised cohort of TAILORx.
Agendia N.V.	44	116	Outcome	"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."	We defined Clinical Utility studies as any studies assessing prospective use of the tests and the effect on patient outcomes. Therefore these studies meet this definition. They are

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				So on the one hand authors conclude that no studies exist for ODx to determine clinical utility, on the other hand, these non-qualifying studies are being used to "reveal something". **This is a very dual message, and begs for the non-qualifying datasets of the other tests to also be included in the NICE assessment.**	limited by being single-armed in nature; this limitation is clearly stated. The same definitions and inclusion criteria were used for studies of all five tests.
Agendia N.V.	45	119	Race	"and showed generally similar rates across race categories,.." **Doesn't it worry authors that there were no differences in risk categories across race categories, whereas breast cancer recurrence rates are known to be different across race categories?**	The data included are those reported in the study. There were limited available data by race.
Agendia N.V.	46	121	Table 24	**Even though mentioned earlier in the text correctly, that the TAILORx results as reported by Sparano et al cannot be used as clinical utility data since the reporting of the low risk patients is merely an observational study, it is surprising to see the TAILORx Sparano study in this table.** **Similarly true for the Plan B study.**	As noted above, we defined Clinical Utility studies as any studies assessing prospective use of the tests and the effect on patient outcomes. Therefore these studies meet this definition. They are limited by being single-armed in nature; this limitation is clearly stated.
Agendia N.V.	47	125	WSG PlanB	**HR reported for percentiles of the RS: this is not clinically meaningful.** **Authors should make readers aware of this.**	The inclusion criteria for the review did not restrict by cut-off used. The percentile data is therefore still eligible for inclusion. See response to Agendia question 30.
Agendia N.V.	48	128	4.4.1	"The initial validation cohort in the same article (n=19)114 incorrectly identified 2/19 patients (whether these were recurrences or nonrecurrences was not reported)" **They were non-recurrences. Based on Figure 2c in Van't Veer 2002, the test incorrectly identified 2 patients who did not recur as poor prognosis.**	Information noted. The report has not been amended since this relates to a very small subset of data.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Agendia N.V.	49	128	4.4.1	"A multivariable logistic regression analysis that included "classical prognostic factors" (variables not reported)…" **Multivariable model is the combination of microarray and clinical parameters. The complete list of clinical parameters (ER, PR status, tumor grade and size, age, angioinvasion) included in supplementary section of Van't Veer 2002. Outcome of disease was the dependent variable.**	Information noted. The report has not been amended since this relates to a very small subset of data.
Agendia N.V.	50	128	4.4.1	Last sentence……"*it was unclear whether patients included were from the derivation cohort or validation cohort*……." **The supplementary section (Van't Veer 2002) has full detail on how the odds ratios were calculated. The cross-validation process is based on the derivation cohort (n=78).**	Information noted. The report has not been amended since this relates to a very small subset of data.
Agendia N.V.	51	129	Overlap	"Therefore, it should be noted that there is some overlap between patient cohorts within the references included here." **This is indeed true, however we note a very different language use and underscoring for this test in comparison to the ODx test, where the NSABP-B20 & B14 studies are also analysed in multiple studies, and no clear attention is drawn to this fact. Also, the very concerning use of the NSABP-B20 tam only treated samples in the predictiveness study for ODx is not clearly being underscored. But for the MP test this**	There was no intention to use different language for the different tests. As in any review, we felt it important to note that there may be some overlap between studies reported in this section. For Oncotype DX, the B-20 and B-14 studies were reused to derive and validate RSPC, which is treated within this assessment as a different test. Therefore, no double counting of

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				is done immediately in the first paragraph of the prognostic performance. **Also in the original Van de Vijver paper the use of some of the same samples has been separately analysed to check whether this would interfere with the validation.** ***Please see below text how this was done (taken directly from the NEJM 2002 vd Vijver paper): We wished to investigate the prognostic value of the gene-expression profile in a consecutive series of patients with breast cancer. We included 61 of the 78 patients with lymph-node–negative disease who were involved in the previous study that determined the 70-gene prognosis profile (vh veer 2002).*** ***Leaving them out would have resulted in selection bias, since the previous study included a disproportionately large number of patients in whom distant metastases developed within five years. We included these 61 patients in the study, but we used the "leave-one-out" cross-validated classification established in our previous study to predict the outcomes among these patients. In this approach, the classification of the left-out sample was based*** ***on its correlation with the mean levels of expression of the remaining samples from the patients with a good-prognosis signature, with the sample in question excluded from the gene-selection process.*** ***This approach minimizes to some extent the possibility of overestimating the value of the prognosis profile while***	patients occurred and no correction is required. The use of B20 has been addressed above; see response to Agendia comment #2.

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				it keeps the consecutive series complete. We also provide validation results taking only the new samples into account. … we first calculated the estimated odds ratio for the development of metastases within five years for the patients with lymph-node–negative disease in the present series (thus excluding the 61 patients who were also part of the previous study9) (Table 2). This analysis included only patients in whom distant metastases developed within five years and patients who remained disease-free for at least five years. The odds ratio for the development of distant metastases within five years in this group was similar to the ratio in our previous study (15.3 and 15, respectively) (Table 2). The prognosis signature was also highly predictive of the risk of distant metastases among the subgroup of patients with lymph-node– positive disease and among the subgroup of all new patients (Table 2)	
Agendia N.V.	52	129	Prognostic	"Prognostic data on MammaPrint mainly consists of retrospective analyses.." **MammaPrint is the only test for which level 1A prognostic data is available.** **This effort and the results should be mentioned here!** **The test is being poorly represented by only mentioning the retrospective consecutive patient series at this point.** **Also, the STO data series (vh Veer 2017 and Esserman 2017) is a level 1B validation series, and should be mentioned here.**	As explained above, MINDACT is extensively covered in the Clinical Utility sections. Please see response to 1e and Addendum to NICE regarding MINDACT and prognostic ability. The STO study is clearly described in this section.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				The test is being poorly represented by only mentioning the retrospective consecutive patient series at this point. The reader has to wait till the next page where it says "in addition, there is one retrospective analysis of an RCT." It is unclear why prognostic performance of MammaPrint is based on studies excluding the most informative ones, namely the prospective randomized trial MINDACT. Data from this trial gives the most valuable prognostic information, providing level 1A clinical evidence and should be clearly stated in this section of the assessment as well. **The difference in reporting between the tests is huge.**	
Agendia N.V.	53	130	4.4.1	**Reference 119 needs to be updated. Esserman et al is published in JAMA 2017**	The references identified by our systematic search were included in the review, unless the EAG became aware of a more recent publication. We had already updated this study with the recently published Van't Veer 2017 study, identified by Agendia after the deadline for submission of evidence. The EAG were not aware of Esserman et al. 2017 and it was therefore not included.
Agendia N.V.	54	131	4.4.2	**Criticism Van de Vijver paper that used a small proportion of patients derived from the derivation set. It should be mentioned here that to avoid bias, a correction for this was performed.** **The small proportion (n=61) were included to avoid selection bias, since the previous study included a disproportionately large number of patients in whom**	Information noted. This has been included in an Addendum to NICE.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				distant metastases developed within five years. **The correction in analysis was made using the "leave-one-out" cross-validated classification to predict the outcomes among these patients. This approach minimizes to some extent the possibility of overestimating the value of the prognosis profile while it keeps the consecutive series complete. The study also provides validation results taking only the new samples into account. See also comment no. 51**	
Agendia N.V.	55	131	4.4.2	"*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*". **ref 114 (Van de Vijver 2002): distant metastases as a first event to be a treatment failure; death from causes other than breast cancer was censored.** **ref 53 (van't Veer 2017): analysis on breast cancer-specific survival and DMFS, however information on metastasis is less complete as compared to information on death.** **ref 63 (Bueno-de-Mesquita 2009): only included distant metastasis as first failure. They refer to it as distant**	Information noted. The report has not been amended since it was too complex to do so at this stage. RASTER is included in the Clinical Utility sections.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				metastasis free percentage. So it excludes all death? **ref 64 (Buyse 2006): time to distant metastases, excluding all other events.** **ref 126 (Knauer 2010): time from surgery to any distant metastasis (DMFS). They also measured BCSS.** **ref 128 (Beumer 2016): measured both DMFI defined as the time to distant recurrence and DMFS defined as the time to distant metastasis or death by any cause** **ref 63 and 64 specifically mention that they designed the endpoint similar to the original study, which is time to distant metastasis (censoring death from other causes). But it is unclear whether that includes death from breast cancer**	
Agendia N.V.	56	132	4.4.2	**Esserman and vh veer 2017 is missing in the evidence of long term follow-up.** **Esserman 2017: In a secondary analysis of the STO-3 RCT of tamoxifen treatment compared with no systemic therapy in node-negative post-menopausal women, MP scoring identified 'ultralow' risk patients with exceptional long-term survival rates. Tamoxifen-treated ultralow risk patients had 100% BCSS at 15 years and 97% BCSS at 20 years. Untreated ultralow risk patients had 97% BCSS at 10 years and 94% at 20 years.** **Van't Veer 2017:** **In the STO-3 RCT, in which post-menopausal node-negative patients were randomized to tamoxifen or no systemic treatment, patients were retrospectively assessed by MP risk classification. Tamoxifen-treated**	Data from the STO-3 RCT (van't Veer 2017) is included in this section (p133 in circulated PDF). Data from Mook et al. and Yao et al. are also included (text and tables). RASTER is covered in detail in the Clinical Utility section.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				MP low- and high-risk patients had 20-year BCSS of 90% or 83%, whereas untreated patients had 20-year BCSS of 80% and 65%, respectively. **In addition please add Mook et al., Yao et al. and RASTER 10-years data.**	
Agendia N.V.	57	136	4.4.2	**Authors of the report state that prognostic value is mainly based on nine small retrospective analyses, that these were a mixed population, they consisted of pooled cohorts occasionally; it is for obvious reasons that the MINDACT study will outperform these 9 studies in terms of prognostic value and was therefore also performed. MINDACT should be taken into account into this analysis and the prognostic performance of MP needs to be evaluated higher by the authors.** **The MINDACT trial provides level 1A outcome for contemporary patients. The study inclusion represents the higher compliance to screening and early detection as well as third generation chemotherapy for high risk patients. As such MammaPrint provides the highest level evidence for the prognosis of early stage breast cancer for both patients receiving endocrine therapy alone or endocrine plus chemotherapy. For example, MammaPrint Low Risk patients ER+ LN- HER2- have a 96.7% DMFS at 5 year without chemotherapy (figure S4 appendix Cardoso et al NEJM 2016)**	As explained above, MINDACT is extensively covered in the Clinical Utility sections. Please see response to 1e and Addendum to NICE regarding MINDACT and prognostic ability.
Agendia N.V.	58	136	4.4.2	*…"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*	We understand that the percentage categorised as low/high risk depends on the population studied. These populations are described in

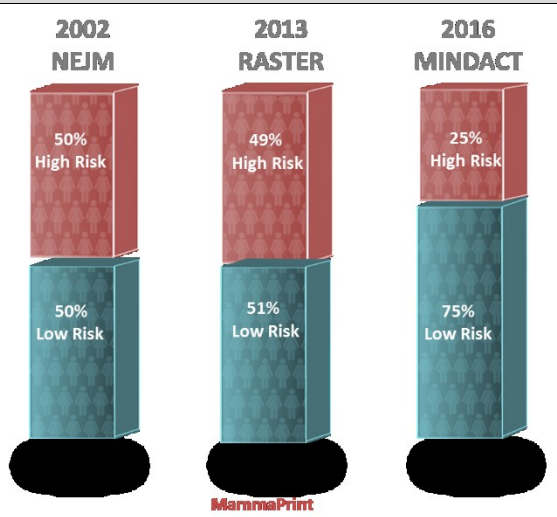
Tumour profiling tests to guide adjuvant chemotherapy decisions in people with breast cancer (update of DG10)

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				analyses of LN+ patients,60, 61 percentage categorised as low-risk was 38% and 41%, while percentage high-risk was 59% and 62%" **….This statement if presented this way can be interpreted as if MP is not stable within its ability to stratify patients into low or high risk of recurrence. This risk stratification, however, very much depends on the investigated population. It can be assumed that in a clinically low risk population the chance of identifying more MP low risk patients is of course higher than when looking at clinically high risk populations. It can be seen from previous publications however, that when assessing similar populations the percentage of low and high risk identified patients stays very constant between Buyse et al 2006, Bueno de Mesquita 2007, Cardosos 2016. Van't Veer 2017 shows a slightly higher % of low risk because this was the STO 3 trial which was designed for low risk breast cancer.**	more detail in the main results section; the page alluded to here is a Discussion summary. In the Overview of Main Results across all tests (Section 4.2) we summarise this risk categorisation by LN status and separately for studies where patients received endocrine monotherapy, to allow a more consistent comparison across tests.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				 **As you can see from the figure, over time, MammaPrint has been able to identify a large number of patients to be at Low Risk of recurrence. In fact, even when clinically high risk patients are included, MammaPrint can safely identify those that are at low risk of recurrence and have no clinically meaningful or significant benefit of chemotherapy. Unlike with the ODx test where the number of low risk patients has decreased over time, with the prospective MINDACT study, Mammaprint has shown to in fact safely increase this group of low risk patients with the highest level of clinical evidence.**	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Agendia N.V.	59	137		…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)""….. **As most other tests have been validated on clinically low risk populations, indeed this could be a reason for this finding as MammaPrint validations have been performed including high risk populations as well. In fact it is the only test that has provided prospective evidence that it can identify clinically high risk patients that can safely forego chemotherapy without compromising outcome. When comparing outcome in low risk patients however, the table below shows that DRFI rates are similarly low for MP vs ODx patients but that MP is able to identify many more low risk patients.**	Information noted; no response required.
Agendia N.V.	60	140	Table 27	**Esserman 2016, could now be replaced with Esserman publication in JAMA Oncology 2017**	This study was not identified by our systematic searches or submitted by the company in time for inclusion in the report. However, data for this

	5-year DMFI	% of patients
MammaPrint Low-Clin Low	98%	58.3%
Oncotype < 11	99%	15.9%

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
					study (STO-3) is included from van't Veer 2017 (full paper) and conference abstracts by Esserman 2016 and Lindstrom 2015.
Agendia N.V.	61	159	4.4.3	Discussion: "…it was unclear whether the interaction test was unadjusted or adjusted, and if so for which factors." **Based on the description in Statistical analysis section in Knauer et al paper, the interaction test was adjusted for clinico-pathological variables.** **"Co-variates used in adjusted models included age at diagnosis, tumor size, number of positive lymph nodes, histological grade, ER and PR status, hormonal therapy, and CT. Relative differences between treatment effects by 70- gene risk groups were assessed by adding an interaction term to the model"**	Two reviewers and a statistician examined this article carefully and were unable to determine whether the interaction test was conducted within the adjusted or unadjusted analysis. As noted, the methods section in the article mentions the interaction term next to the information on adjustment. However, in the results section of the article, the information on the interaction test follows the paragraph on unadjusted analyses. Therefore, this information was noted as unclear in the EAG report.
Agendia N.V.	62	160	4.4.3	**We disagree with the final conclusion regarding CT predictiveness. Please see also comment no. 2,3 and 17**	Please see our response to comment #3.
Agendia N.V.	63	165	4.4.4	"…*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*."…. **It should be noted here that the sample size was modified from 6000 to 6600 to compensate for these**	Information noted. The EAG report uses the ITT results (where available) and has not been amended.

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

Diagnostics Assessment Report (DAR) - Comments

<table>
<tr><th>Stakeholder</th><th>Comment no.</th><th>Page no.</th><th>Section no.</th><th>Comment</th><th>EAG response</th></tr>
<tr><td></td><td></td><td></td><td></td><td>changes and that no patient that required chemotherapy was left untreated.

Moreover, a sensitivity analysis was performed to test whether leaving out the patients enrolled during the so called “G-shift period” (due to change in RNA extraction solution) of the MINDACT trial would have any impact on outcome for the total population. As you can see from the table below derived from Cardoso et al., 2016, the PP, PPS and ITT analyses come to essentially the same conclusions, for which reason confidence in the study results is excellent. Consistency between the results of primary analysis and the results of sensitivity analysis strengthens the conclusions or credibility of the findings.
<table>
<tr><th>High C-risk, Low G-risk</th><th colspan="2">PP-Per Protocol (tabel 2)</th><th colspan="2">PPS-Sensitivity (tabel S5)</th><th colspan="2">ITT – IntentionToTreat (tabel S14)</th></tr>
<tr><td>DMFS</td><td>0.65 (0.38 – 1.10)</td><td>0.11</td><td>0.60 (0.34 – 1.06)</td><td>0.080</td><td>0.78 (0.50 – 1.21)</td><td>0.267</td></tr>
<tr><td>DFS</td><td>0.64 (0.43 – 0.95)</td><td>0.03</td><td>0.57 (0.37 – 0.87)</td><td>0.009</td><td>0.71 (0.50 – 1.01)</td><td>0.055</td></tr>
<tr><td>OS</td><td>0.63 (0.29 – 1.37)</td><td>0.25</td><td>0.54 (0.23 – 1.26)</td><td>0.154</td><td>0.69 (0.35 – 1.35)</td><td>0.278</td></tr>
</table></td><td></td></tr>
<tr><td>Agendia N.V.</td><td>64</td><td>166</td><td>4.4.4</td><td>“Low clinical, high MammaPrint Group: Given that low clinical risk patients…result in patients receiving chemotherapy by not gaining any benefit.”
I think it should be noted somewhere that this study was not powered to assess chemotherapy benefit in this group of patients. The clinical-high, MP low-risk group was the smallest group in the study, and would have required about a 1000 additional patients in order to</td><td>We think the EAG report essentially agrees with Agendia’s comment here (see last sentence of comment): that low-clinical risk patients are unlikely to benefit from MammaPrint.

Even if there was under-powering in the low-clinical high-MammaPrint group, the small effect size (0.8% difference in 5-year DMFS between</td></tr>
</table>

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				properly assess chemotherapy benefit. This is exemplified in the supplementary Table S14 of the Cardoso et al., NEJM 2016 paper, in which outcome according to discordant risk group and treatment strategy is provided for the ITT population. In this table, DMFS is compared in each group between patients that received CT according to either genomic or clinical risk. In the high clinical risk/low genomic risk group, there is a 1.5% absolute difference in DMFS between those that received CT and those that did not, which is not statistically significant (p=0.267). In the clinical low/high genomic risk group, there is a 0.8% difference in DMFS between those that received CT and those that did not, which is also not significant (p=0.657). The p-values suggest the size of the group is not sufficient to accurately assess benefit of chemotherapy. This is also stated by the authors in the Discussion section on page 167: "the primary aim was to determine whether patients who were high-clinical but low-MammaPrint risk could avoid chemotherapy." The converse is also true: the study was not designed to evaluate the benefit of chemotherapy in the other discordant group. This also indicates that clinically low patients where there is no doubt about treatment genomic tests have no added benefit and it is mainly the clinically high risk patients that can benefit.	CT and no CT) suggests that the result for this patient group is non-clinically significant as well as non-statistically significant.
Agendia N.V.	65	167	4.4.4	"This analysis also assumes that in the MammaPrint strategy, all patients would be treated according to MammaPrint,	We believe this statement still holds (please see response to comment #64).

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				whereas the results above indicate this may not be justified for low-clinical, high-MammaPrint patients." **Same argument as above; the MINDACT was not powered to assess CT benefit in the C-low/ G-High group. See comment no. 64**	
Agendia N.V.	66	168	4.4.4	Conclusions: "This could be interpreted as showing that MammaPrint may not be useful in the group as it would increase chemotherapy rates without improving outcomes." **In addition to the study not being designed to answer this question (see comment no. 64), genomic testing would generally not be ordered on clinically low-risk patients. This is in line with the exclusion of patients with a NPI <3.4 as candidate for a genomic assay as described in the report. It is also in line with the recently updated ASCO guidelines, where MammaPrint is recommended only for clinically high risk patients and the only genomic tests that is in fact recommended for LN+ patients.**	Again we think the EAG report is in agreement with Agendia here.
Agendia N.V.	67	175	RASTER	**"**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). 5-year overall survival was not statistically significantly different between MammaPrint groups (p=0.35, HR NR; Table 49) **In addition to providing p value, it need to state that DRFI was indeed statistically significant. The wording in this paragraph underestimate the power of Mammaprint. It's highlighting the non-significance in overall survival, but not mentioning that the DRFI, which is a more accurate**	We feel that citing the p-value of p=0.03 clearly shows that this is statistically significant. To avoid multiple errata, we have not amended the report here.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				endpoint assessing distant relapses and breast cancer specific deaths only, was statistically significant.	
Agendia N.V.	68	175-176	***Results for clinical risk tools:***	**This section is focused on comparing NPI and PREDICT to AOL. The paper they are referencing (Drukker et al, 2014) also used St. Gallen and Dutch national guidelines (2004 and 2012) prediction algorithm. Both of those showed no statistical significance in 5 year DRFI between low and high risk group, similar to AOL. Why are St. Gallen and Dutch guidelines (which is what the actual treatment decision is based on) excluded from the report?**	The St Gallen and Dutch guidelines were excluded because they are not of relevance to the decision problem as they are not used in the UK. We state in our Methods (section 4.1): "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."

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Agendia N.V.	69	175		**"**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)." **As stated in Drukker et al, 2014: Among the low risk systemically untreated patients, no significant difference was seen for most clinical risk algorithms (p = 0.29 for AOL, p = 0.66 for NPI, p = 0.37 for St. Gallen, p = 0.65 for the 2004, and p = 0.14 for the 2012 Dutch national guidelines) between patients with a concordant low risk assessment and patients with a 70-gene signature low risk result but a high risk assessment by one or more of the clinical indexes. Please add the p-values.**	These p-values were not considered relevant here since we were reporting on the difference between MammaPrint low-risk and high-risk patients (within patients who were clinically high-risk), not the difference between low/high risk on a clinical tool within MammaPrint risk group (as quoted in the comment). The report has not been amended.
Agendia N.V.	70	176		"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 categorize fewer patients as high-risk." **From table 2 in Drukker et al, 2014, we can calculate the number of NPI or PREDICT high risk patients who received no chemotherapy and those that were MammaPrint low:** • **NPI: 28 high risk NPI without CT, 68% Mammaprint low** • **PREDICT: 43 high risk without CT, 67% Mammaprint low**	We do not think we have outcomes for these subgroups of patients. The report has not been amended.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				• **St. Gallen: 155 high risk without CT, 80% Mammaprint low** • **Dutch 2004: 27 high risk without CT, 78% Mammaprint low** • **Dutch 2012: 119 high risk without CT, 82% Mammaprint low**	
Agendia N.V.	71	177		**"**MammaPrint provided additional prognostic information over AOL and NPI, but not over PREDICT plus." **Mammaprint also provided additional prognostic information over St. Gallen and Dutch guidelines.**	Please see response to comment 68 – these comparators were not in scope when other UK-relevant comparators were reported in the same study.
Agendia N.V.	72	177	4.4.4	"Estimates of prognostic performance between risk groups are likely to be affected by the different rates of chemotherapy..." **Although the rates of chemotherapy are different in the different risk groups (81% in the HR group and 15% in the LR group), it could be noted here that the expectation would be that differences in DRFI between the two groups would only become larger if equivalent numbers of patients in each group were treated with chemotherapy. One would presume that DRFI would decline in HR patients with a lower rate of chemotherapy treatment. Thus, although treatment rates likely affected DRFI, prognostic performance of MP would not likely be diminished if chemotherapy treatment rates were equivalent in the low and high risk groups.**	We note the point that higher chemotherapy in the high-risk group would be likely to reduce recurrences in this group and therefore underestimate prognostic performance. However we feel that this statement still holds for any study in which chemotherapy use was influenced by the test; therefore the report has not been amended.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Agendia N.V.	73	200		**"**Both analyses of LN0 patients (TransATAC and ABCSG-6+8) showed that EPClin was statistically significantly prognostic for 10-year DRFS/DRFI." **It might be important to note that these studies did not have the same endpoint. TransATAC and GEICAM 9906 primary endpoints were DRFS, while ABCGS 6+8 primary end point was DRFI, the latter excluding all non-breast cancer related deaths, while all deaths are included in DRFS!**	Due to the inconsistent reporting of outcome definitions across the evidence base it was not always possible to make a distinction between DRFS and DRFI. Therefore, these outcomes have been grouped for consistency.
Agendia N.V.	74	199, 201		"All three data sets included LN+ patients, all of whom had 1-3 positive nodes (LN1-3)" **"**analyses of LN+ patients showed that EPClin was statistically significantly prognostic for 10-year DMFS/DRFS/DRFI." **Note that only one-third of patients in TransATAC and ABCSG 6+8 had involved lymph nodes and of those in ABCGS only 5% with > 3 metastatic nodes**	The LN0 and LN+ data are presented separately in the EAG report. For ABCSG-6+8, subgroup analyses were provided by the company for patients with LN1-3.
Agendia N.V.	75	201, 202		**"**In LN+ patients in GEICAM 9906, adding EndoPredict to a combination of clinico-pathological 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). In ABCSG-6+8 (two-thirds LN0), the C-index was only reported for years 5-10 (no data for years 0-5)"	The fact that GEICAM patients were chemotherapy-treated has been noted in our text and tables. Data for years 0-10 incorporates the benefits from chemotherapy in years 0-5, as long as the benefit is not lost during years 5-10.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				It should be noted that in GEICAM 9906 trial we are looking at chemotherapy-treated patients, while in ABCSG trials (as well as TransATAC) we are analyzing endocrine-treated only patients. Additionally, in GEICAM 9906, only 10 year DMFS is reported. However, based on Oxford Meta-analysis of chemotherapy benefit in early stage breast cancer, the benefit of chemo is limited to the first 5 years. The first 5 years is where we see the greatest difference.	
Agendia N.V.	76	243		"As the TransATAC analysis is key to this assessment… are lacking." **We would like the authors to recognize the limitations of the TransATAC study and should NOT be the key in this assessment.** **The ATAC trial was designed for post-menopausal patients that were not considered candidates for chemotherapy. This is a sub optimal data set to address which candidates for chemotherapy can safely forego chemotherapy as addressed in MINDACT.**	The TransATAC analysis has been useful in this assessment because a) it compares four of the five tests and b) data was available to allow subgrouping by NPI status. See also previous responses to similar comments. We agree that the MINDACT study is also useful and have covered this extensively in the EAG report.
Agendia N.V.	77	247	4.8.1	"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." **Although it is true that both tests are significant between high- and low-risk groups, it should be noted that, for Oncotype DX, there is not a significant difference in DFS between low and intermediate (p=0.76) risk groups or between intermediate and high (p=0.072) risk groups.**	The point is noted but no amendment to the report is required as an error has not been made.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				This is an important point because a large portion of tumor samples tested by Oncotype DX return intermediate scores.	
Agendia N.V.	78	249	4.8.1	"Another broad observation is that the tests generally perform differently in LN+ and LN0 patients." **It could be noted here that MammaPrint is the only test currently recommended by the ASCO guidelines to inform treatment decisions in LN+ patients (Krop et al. *JCO* 2017)**	The point is noted but no amendment to the report is required as an error has not been made. The statement has not been verified by the EAG.
Agendia N.V.	79	266	4.8.2	"A further study…added a little more." **It should be noted that in study summarized in this section, Oncotype DX low and intermediate risk groups were combined for the analyses. This is an important note, because as mentioned by others referenced in this section (Ahn et al. 2014), "the selection of chemotherapy for patients with intermediate RSs remains controversial".**	The EAG agree that the comparison is low/intermediate versus high. An erratum has been made to reflect this. However, as it is unclear how intermediate risk patients will be handled in a clinical setting, it is not possible to make any comment on whether the analysis is appropriate or not.
Agendia N.V.	80	268	4.8.2	"One study showed that MammaPrint could further categorize Oncotype DX intermediate-risk patients…however…it is not possible to conclude that MammaPrint outperforms Oncotype DX" **Although it may not be possible to make this conclusion based on this study; however, it may be worth noting that MammaPrint provides additional prognostic value, especially in patients with intermediate Oncotype DX recurrence scores, for whom treatment recommendations require additional information. Also, the Prospective Study of MammaPrint in Breast Cancer Patients with an Intermediate Recurrence Score**	The EAG do not feel an omission has been made and no change to the report has been made. Tsai 2017 is a decision impact study conducted in the USA and as such did not meet the inclusion criteria for the review, as chemotherapy is prescribed more frequently in the USA compared to the UK. As such, its results have low relevance to the decision problem. It is not possible to conclude whether either test is resulting in over- or under-treatment on the basis of this study design as no long term outcomes have been reported.

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				(PROMIS) trial, recently published in JAMA Oncology (Tsai 2017) showed the impact of performing a MammaPrint test on 840 women who had early-stage breast cancer and an ODx intermediate recurrence score 18-30. Each woman had her sample re-tested with MammaPrint and treatment recommendations were recorded before and after receipt of the MammaPrint results. **45% of intermediate risk patients had a Low Risk result with MammaPrint and 55% had a High Risk result. MammaPrint Low and High Risk results were found at every score across the entire intermediate results range (RS 18 to 30) with 50% of MammaPrint High Risk results found between a RS of 18 and 25. This highlights the lack of correlation between the two tests. 29% of patients (108) had chemotherapy removed from their treatment after receiving a MammaPrint Low Risk result. More importantly, for patients classified as MammaPrint High Risk, 37% of patients (171) were recommended to receive chemotherapy, potentially preventing under-treatment. This suggests that the results of the 21-gene assay have the potential to cause over- and under-treatment of patients whose risk-of-recurrence prognosis is unclear. Physicians changed their treatment decisions in alignment with the MammaPrint treatment guidance by recommending chemotherapy in 88% of High Risk patients and recommending no chemotherapy in 91% of Low Risk patients.**	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Long-term outcomes were not measured in this study, however MammaPrint is currently the only assay that has published prospective randomized clinical utility evidence supporting the lack of significant chemotherapy benefit in genomically Low Risk patients.	
Agendia N.V.	81	286	4.9	"MammaPrint. There were no UK studies…insufficient data to assess results by LN status." **Please add Tsai et al., JAMA Oncology 2017 with results from the PROMIS study stating change in treatment recommendation based on MammaPrint.**	The Tsai study is a US decision impact study. Only UK and European decision impact studies were included, due to time constraints and the very different levels of chemotherapy use in the US.
Agendia N.V.	82	291	Table 91	**Please add MINDACT to this table.**	Table 91 includes decision impact studies only.
Agendia N.V.	83	296	Table 96	**Concerning Wuerstlein 2016:** ' Post-test Recomm (unclear) '; **This study demonstrates that the use of the gene expression profiles MammaPrint and BluePrint has a strong impact on therapy decisions as shown by the physicians' change between pre- and post-test treatment recommendations and their increased confidence in their therapy advice.** • **MammaPrint and the corresponding molecular subtype BluePrint strongly impacted clinical therapy decisions (28.4% switch) in early breast cancer patients with up to 3 involved lymph nodes.**	The quoted text mentions both recommendations and decisions which is why this was noted as unclear in the table.

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

Diagnostics Assessment Report (DAR) - Comments

<table>
<tr><th>Stakeholder</th><th>Comment no.</th><th>Page no.</th><th>Section no.</th><th>Comment</th><th>EAG response</th></tr>
<tr><td></td><td></td><td></td><td></td><td>
Table 2: Switch in CT decision based on MammaPrint
<table>
<tr><th colspan="4">MammaPrint Low Risk</th></tr>
<tr><th rowspan="2">Pre test recommendation</th><th colspan="3">Post test recommendation</th></tr>
<tr><th>CT</th><th>no CT</th><th>Total</th></tr>
<tr><td>CT</td><td>24 (30.0%)</td><td>56 (70.0%)</td><td>80</td></tr>
<tr><td>no CT</td><td>3 (1.5%)</td><td>191 (98.5%)</td><td>194</td></tr>
<tr><td></td><td></td><td></td><td>274 (63.7%)</td></tr>
</table>
<table>
<tr><th colspan="4">MammaPrint High Risk</th></tr>
<tr><th rowspan="2">Pre test recommendation</th><th colspan="3">Post test recommendation</th></tr>
<tr><th>CT</th><th>no CT</th><th>Total</th></tr>
<tr><td>CT</td><td>83 (98.8%)</td><td>1 (1.2%)</td><td>84</td></tr>
<tr><td>no CT</td><td>62 (86.1%)</td><td>10 (13.9%)</td><td>72</td></tr>
<tr><td></td><td></td><td></td><td>156 (36.3%)</td></tr>
</table>
Poster SABCS 2016
</td><td></td></tr>
<tr><td>Agendia N.V.</td><td>84</td><td>307</td><td></td><td>“ Most of the models that evaluated Oncotype DX against current practice assumed that the test was associated with a predictive benefit of chemotherapy.”
We argue that Oncotype has sufficient evidence to prove predictive value and that this can be used as an</td><td>The wording of the company’s comment is unclear – we assume that they mean “dispute the argument” rather than “argue.” Notwithstanding the lack of clarity in this meaning of the comment, the assumptions</td></tr>
</table>

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				assumption in the cost-effectiveness analysis. Please discard the studies including these studies with predictive assumptions in their analysis as evidence in this report. See further comment no. 2	made within other health economic models about predictive benefit are not relevant in determining whether this aspect of the value of a test is true. The base case assumptions employed in the EAG model regarding predictive benefit have been reached on consideration of the findings of the clinical review, not what others have assumed in economic models.
Agendia N.V.	85	317-324	5.2.1	**All text concerning the Agendia cost effectiveness report – MammaPrint versus current practice and elaboration on this is a black box. As it is made invisible we are not given the opportunity to respond to this part. Therefore, we request openness for the section and additional time to be able to properly respond to this section.**	This issue should be taken up with NICE.
Agendia N.V.	86	326		**Regarding the utilities, we think that the disutility for chemotherapy used in the analysis has large impact on the outcomes. The paper of Campbell, 2011 was used for the disutility for chemotherapy during 6 months in the model. In the paper itself, they mention that the disutility for chemotherapy must be used for at least 1 year [Campbell].Campbell: *"Analyses of HRQoL data collected during ABC, NEAT, and TACT (with the regression analysis described above again used to predict EQ-5D scores in the NEAT trial) suggested that the negative impact of chemotherapy on underlying HRQoL persisted for at least a year following completion of treatment."***	This same source is used in the Agendia model. As noted in response to earlier comments, this is not applied as a disutility – it is a QALY loss (relating to a year's decrement). We acknowledge that this is not clear in the EAG report, but is clear from scrutiny of the model. We have included a sensitivity analysis in which a larger disutility is applied. This does not change the economic conclusions of the analysis. The EAG is unclear how the company has produced EQ-5D estimates from MINDACT as this instrument does not appear to have been

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				We recommend to use the chemotherapy decrement for at least 1 year instead of 6 months. This could change the incremental QALY in the final results. **Furthermore, also in the earlier send model, you can find confidential data regarding utilities measured by means of the EQ-5D during the first 800 patients of the MINDACT trial. See for detailed methods in the published paper where we report on the QoL measurements [Retel, BMC Cancer, 2013]. We will publish this data on a poster for the San Antonio Breast Cancer Conference, December 8th 2017. If we incorporate these utilities for the first year, for each risk group separately, the QALYs for the MammaPrint yield more compared to the mAOL and NPI in the clinical high risk group. Please find in addition to document, confidential, the model with the proper probabilities and utilities derived from the raw data of the MINDACT. These numbers can be used in the EAG model and when raw data on patient level is required we have to await the approval of the EORTC (request is submitted).**	included in the trial. The EQ-5D is also not mentioned in the Retel 2013 paper mentioned in the company's comment. No details are provided in the Agendia cost-effectiveness paper provided by the company. The EAG also notes that the utility for distant metastases employed in the original Agendia model and the new Agendia model is based on Ward et al, not MINDACT.
Agendia N.V.	87	335		"The EAG considers it unlikely that patients would suffer the adverse effects of adjuvant chemotherapy years after they have completed their treatment." **We think authors underestimate the long-term effects of chemotherapy treatment.** **Please find here the reference for longterm toxic effects:**	The EAG model considers long-term AEs (AML) separate to short-term AEs. As noted in the EAG report, whilst we recognise that CHF is also a potentially relevant long-term AE associated with chemotherapy, this was excluded from the model due to a lack of

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				H. A. Azim Jr, E. de Azambuja, M. Colozza, J. Bines & M. J. Piccart. Long-term toxic effects of adjuvant chemotherapy in breast cancer. Annals of Oncology 2011, 22: 1939–1947	evidence on the joint survival impact of CHF and metastatic breast cancer.
Agendia N.V.	88	344	population	" The modelled population for these four tests reflects that of the TransATAC study, .." **We would like the authors to recognize the limitations of the TransATAC study.** **The patient population included in the TransATAC study were limited to only postmenopausal patients with hormone receptor–positive primary breast cancer from the tamoxifen- or anastrozole-alone arms. This means that patients eligible for the trial had an indication for endocrine therapy and are therefore a group of patients with a lower risk in general. In other words, based on the TransATAC trial patients were of lower risk and can't be a representative starting point (bias). Those patients were no candidates for chemotherapy in the first place so are not suitable for an assessment to address the question whether patients can safely forego chemotherapy.**	Please see previous responses.
Agendia N.V.	89	345		"MammaPrint was not included in the TransATAC study, hence an alternative source was required." **Based on the level of existing evidence the information of this sentence should be the other way around, like: 'All tests except MammaPrint lacked level 1A evidence the information derived from a prospective randomized**	The report has not been amended. The point regarding level 1A evidence has been added to our addendum.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				trial, hence an alternative source (TransATAC) was used.'	
Agendia N.V.	90	345	Comparator	"AOL is currently being updated and has been temporarily disabled." **Please re-phrase this sentence to: 'Although AOL is currently being updated and has been temporarily disabled the decision tree is publicly available and allows for risk classification of clinical low and clinical high risk patients and can be found in the supplementary information (Table S13) of the MINDACT trial [Cardoso et al., NEJM 2016].' Moreover, this table version of mAOL is as easily accessible or used as NPI e.g. and the offline status of AOL is thus not a reason not to use this tool in current clinical practice or be included in this assessment.**	This sentence is accurate. Adjuvant online is currently offline because it is being updated with new risk information, meaning the previous version is not the best available tool. The developers of Adjuvant! Online currently (21st November 2017) direct users to PREDICT until Adjuvant becomes available (https://www.adjuvantonline.com/). The report has not been amended.
Agendia N.V.	91	345	Comparator	"Owing to the use of a different evidence source for MammaPrint134 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." Please see comment no. 88 and 89	The sentence in the EAG report is accurate. The point the company are trying to make is unclear. We therefore cannot provide a response.
Agendia N.V.	92	349	5.3.2	" 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. "	In the absence of data with which to estimate the impact of CHF on mortality and HRQoL, it is difficult to see how the company would like this to be implemented in our model. We have

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Moreover, congestive heart failure is mostly included in breast cancer models, the lack of data should not be the reason to not include this important input parameter [Joensuu et al. 2006].	highlighted this as a limitation of the EAG model in the report.
Agendia N.V.	93	349-350, 358	Table 121	As AOL has the highest clinical evidence for prognosis due to the MINDACT study, it is clear this is the best available comparator for chemotherapy decision making and should be used as such. **The input for the independent cost-effectiveness analysis is based on very different sources. For example, for current practice, the baseline probability of receiving adjuvant chemotherapy is based on the clinical judgement of one person, Professor Rob Stein. Furthermore, the probability of receiving chemotherapy conditional on results of the test is based on the UKBCG survey, which is based on the expert opinion of 11 breast cancer experts, where the questions were asked concerning ER+/Her2- breast cancer (ATAC trial, selection of post-menopausal women). Although nice to have, the input does by no means reflect the clinical evidence level of prognostic tools such as mAOL or PREDICT and therefore also shouldn't be weighted in a similar way. Moreover, compared to the MINDACT population, which includes also ER- and Her2+ patients, the ATAC population is a more favorable group in terms of outcome. The use of all these sources is not optimal, in the light that there is raw survival data available at the**	The model is based on the best available evidence. With respect to the mAOL baseline chemotherapy probabilities, we used clinical judgement as this was the best source available to us. We are unclear how using survival data can inform baseline chemotherapy use parameters.

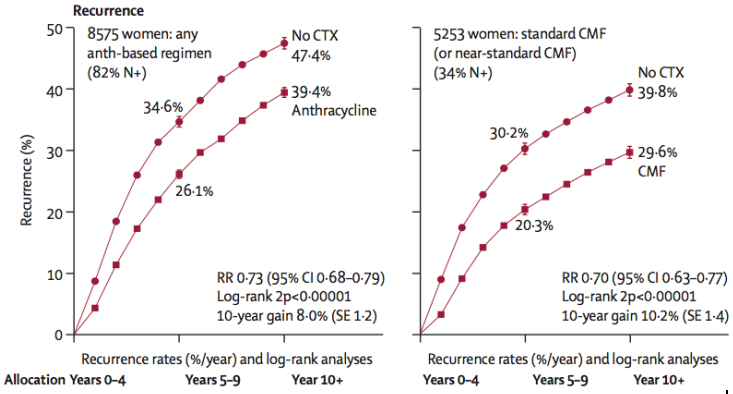
Tumour profiling tests to guide adjuvant chemotherapy decisions in people with breast cancer (update of DG10)

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				EORTC which use would much better reflect the real use and outcome plus would be much less biased than estimations based on personal expert opinions. We do think that the above sources are not the correct reflection of the case of the MammaPrint. Instead we feel that by using the raw survival data we have shown that the MammaPrint is cost-effective for the clinical high risk group as shown in the model provided previously. **With the submission, we sent a model including the probabilities based on the raw survival data from the MINDACT trial. Besides, we also requested the raw survival data at the EORTC for the use of NICE, existing of the OS and DMFS survival rates for the concordant groups, clinical low-genomic high received chemotherapy, clinical low-genomic high not received chemotherapy, clinical high-genomic low received chemotherapy, and clinical high-genomic low not received chemotherapy. If this data will be used, we expect different outcomes than provided by the currently used AEG model. In particular, we expect the clinical high risk group to yield more (QA)Lys for the MammaPrint, as the quality adjusted survival for this groups turns out higher compared to the mAOL and the NPI in our analyses. Please find in addition to document, confidential, the model with the proper probabilities and utilities derived from the raw data of the MINDACT. These numbers can be used in the EAG model and when**	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				raw data on patient level is required we have to await the approval of the EORTC (request is submitted).	
Agendia N.V.	94	354		**Extrapolation from the event rate from 0-5 years to 5-10 years is not done properly. The way it was done you assume that the risk of having an event is identical between 0-5 and 5-10 years which is an incorrect assumption. As can also be seen from the figure below as presented in the Oxford meta-analysis of chemotherapy benefit in early stage breast cancer [Oxford analysis 2012 Early Breast Cancer Trialists' Collaborative Group (EBCTCG): Lancet. 2012 Feb 4; 379(9814): 432–444], the event rate is highest within the first 5 years and slows down after that.**  	We agree that the exponential approach used to model distant recurrence reflects a simplifying assumption which was necessary given the data available. Given the data available on recurrence rates, we consider this to be reasonable (for example, we did not have Kaplan-Meier curves to 5 years for every concordant and discordant risk group in MINDACT). We note that the EBCTCG data presented in the company's comment relate to any recurrence, rather than distant recurrence. In addition, we note that whilst the EAG's extrapolation is imperfect, the Agendia model did not include any extrapolation – this makes it very difficult to interpret the results of the Agendia model as it excludes all long-term costs and health impacts.

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

Diagnostics Assessment Report (DAR) - Comments

<table>
<tr><th>Stakeholder</th><th>Comment no.</th><th>Page no.</th><th>Section no.</th><th>Comment</th><th>EAG response</th></tr>
<tr><td></td><td></td><td></td><td></td><td>This conclusion can also be drawn from our own 10y FU date from the RASTER data in the table below, where the event rate is lower from year 5-10. This means that 1% per year as assumed currently in the assessment is not representative.

<table>
<tr><th>Risk group</th><th>Patients</th><th>Received chemotherapy (%)</th><th>5 year DRFI (95% CI)</th><th>10 year DRFI (95% CI)</th></tr>
<tr><td>MammaPrint Low</td><td>219</td><td>34 (15.5)</td><td>96.3</td><td>93,7</td></tr>
<tr><td>MammaPrint High</td><td>208</td><td>168 (80.8)</td><td>92.2</td><td>86,8</td></tr>
</table>

<u>Difference in yearly event rate between 0-5 years and 5-10 years:</u>
MammaPrint low patients: event rate drops from 0.74 to 0.52
MammaPrint high risk: event rate drops from 1.56 to 1.08

This means that extrapolation of event rates between 0-5 years to 5-10 years based on the first 5 years is arbitrary. Authors should refine and improve this based on the RASTER data as presented in the table.</td><td></td></tr>
<tr><td>Agendia N.V.</td><td>95</td><td>362</td><td>Table 129</td><td>Please note our earlier remark in comment no. 2 concerning the flaws of the Paik et al study. Please note that it is not the use of some samples, but that the B20 study uses the exact same arm of patients used for training of the Oncotype profile. The use of this one arm that provides a selective advantage in the chemotherapy</td><td>This is not a reasonable request. We have presented a range of alternative analyses such that the Appraisal Committee can select which, if any, scenarios they consider to be most reliable, given the available evidence.</td></tr>
</table>

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				benefit study which could potentially explain the study outcome. **Please remove this second column from table 129 and exclude this predictive part of Oncotype from the cost-effectiveness analysis.**	
Agendia N.V.	96	366	Table 130	**In the earlier sent model provided by Agendia, you can find confidential data regarding utilities measured by means of the EQ-5D during the first 800 patients of the MINDACT trial. See for detailed methods in the published paper where we report on the QoL measurements [Retel, BMC Cancer, 2012]. Please add this data to table 130.**	The Retel paper provided does not mention the EQ-5D. It is therefore unclear how Agendia have produced health utilities for use in the model based on this source. The value or reliability of using the company's utility estimates is unclear. The EAG also note that the distant metastases utility value used in the original and new Agendia models is from the Lidgren *et al* study. Our review of utility evidence required studies to report utilities for both relapse-free and distant metastases states – a study which provides only relapse-free utilities would not be included.
Agendia N.V.	97	367	Section ' Resource use and costs'	**Finally, societal costs were not included in the analysis. This will also have large impact on the outcomes. It is well known that patients undergoing chemotherapy treatment have productivity loss and can be on sick leave for days and more occasionally months during and after their treatment. This has a huge impact on societal costs. For example, in the Netherlands, this is estimated for around 7,800 euros**	Societal costs do not form part of the NICE Reference Case. The EAG notes that the Agendia model does not included these costs either.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				per patient, using the Friction cost method. [Hanly et al. Value in Health 2012, Mewes et al. BMC Cancer2015]	
Agendia N.V.	98	370	Table 133	**Regarding the costs, we think some values are missing in the analysis. Currently, an increased use of third regimen chemotherapy is being observed, for example the use of paclitaxel. In the Netherlands, Paclitaxel is used in around 25% of early breast cancer cases. In the current model, only 10% is used. We argue that it would enhance the accuracy of costs by taking into account the use of more advanced regimens in this analysis that more likely reflect current practice use of such regimen. By doing so, it is likely going to lead to change the total –and incremental- costs substantially.** **Also the use of GCSF can be different in many hospitals; at least a sensitivity analysis should be performed on the variation [Retel, JCO 2015].**	Chemotherapy costs were based on a recently published analysis (Hall et al), with assumptions regarding the proportionate usage of alternative regimens. We have already provided sensitivity analyses which include higher and lower chemotherapy costs.
Agendia N.V.	99	392	Table 149	**How are the QALYs between MammaPrint and no test established? We find it highly surprising to see that in mAOL high-risk patients (in which use of MammaPrint causes a 46% reduction of CT based on MINDACT) use of no test has a higher QALY. For No test, was the same group of patients used as for mAOL high-risk and was NPI>3.4 used as a cut-off for CT use? This is unclear to us. The result in table 149 is also exactly the other way**	We have fully explained our methods in the EAG report. We note a number of important criticisms with the Agendia model which lead to problems in the interpretation of the results presented: (1) Errors in the interpretation of the Kaplan-Meier curves which mean that the model does not predict its own data

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				around from what we found in our submitted cost eff analysis report. **We challenge the outcome of the model and conclude that the model that was independently developed based on the MINDACT data by Retel more accurately represents the cost effectiveness of MammaPrint in the UK setting.**	(2) Use of a short time horizon without consideration of long-term health impacts and costs (3) Questionable assumption that risk exclusively determines whether patients receive adjuvant chemotherapy
Agendia N.V.	100	392	Table 149	**We assume that the statistically non-significant 1,5% difference of CT benefit in this mAOL high-risk group used in this cost effectiveness analysis. This survival difference should not be used in the model. There is no significant difference in distant metastasis free survival between chemotherapy and no chemotherapy in mAOL High/MammaPrint Low risk group.**	The model methods are clearly explained in the EAG report. There is a difference between the curves. The model captures this. The absence of a significant difference does not mean that such differences should not be included in a model.
Agendia N.V.	101	403		**….** "In addition, the follow-up period for this study was limited to a duration of 5-years"…. **We argue that 5 years follow up of the MINDACT trial is a limitation of the study. 5-years follow was a predefined endpoint of the trial as no additional benefit from chemotherapy is expected beyond this point as is also described elsewhere [OXFORD overview EBCTCG 2012]. Oxford overview; the entire body of peer-reviewed randomized trials in Adjuvant Therapy for Breast Cancer has been periodically reviewed by the Early Breast Cancer Trialists Collaborative Group (EBCTCG) in the so-called "Oxford Overview" and has established the standard of care for early breast cancer. It has been**	The wording of the EAG report is accurate. The trial follow-up was limited to 5-years.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				established as fact by the published data from the Oxford Overview conducted by the EBCTCG that the benefit of adjuvant therapy in early breast cancer varies by the type of intervention and the time period of risk. Specifically, the benefit of chemotherapy, in both ER+ and ER- breast cancer, is limited to reducing recurrences within the first 5 years, with no later effect. This has been documented in both the 2005 and 2012 overview summaries, which reviewed trials involving over 30,000 patients, with 15-20 years of follow-up, treated with chemotherapy regimens ranging from CMF to anthracycline and taxane containing regimens (*EBCTCG (2005), EBCTCG (2012)*). The specific observations relevant to the benefit of chemotherapy are stated here (Ref 1, p 1699): **"Among younger women the main divergence in recurrence [between chemotherapy and no chemotherapy] takes place just during the first 5 years, when the absolute recurrence rate is high and the recurrence rate ratio is most favorable. This produces an absolute difference of 12% (37% vs 25%) in the 5-year recurrence probability, and "*this absolute difference of about 12% then persists after year 5 …Among older women, the main divergence in recurrence takes place just within the first 2 years of starting chemotherapy…".*** **It has therefore been established by extensive data that the benefit of adjuvant chemotherapy for reducing breast cancer recurrence is seen only within the first 5 years, with no additional increase in benefit**	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				observed beyond 5 years. The MINDACT study identified, in a prospective, randomized trial of nearly 7000 women, a cohort of women with a Low Genomic Risk in the MammaPrint Assay, who show no evidence of benefit from chemotherapy within the first 5 years. The data from the Oxford Overview confirm that no further benefit from chemotherapy will be observed beyond 5 years, for both women under 50 and those from 50 to 69 years, and therefore, no more than 5 years of follow-up is needed to establish the clinical utility of the MammaPrint assay for identifying this cohort. It is also recognized, however, that late recurrences after 5 years continue to occur in ER+ breast cancer, but it is only endocrine therapy, not chemotherapy, which affects the incidence of late recurrences, from years 5 to 10, and 10 to 15. Therefore, in the case of the MammaPrint assay, the principle area of clinical utility is to determine the potential benefit of chemotherapy, a benefit which, if present, will only be observed in the first 5 years. **Moreover, both also other highly respected organizations such as the ASCO (3) and AJCC consider 5 years as a mature end point for DMFS outcome in early stage breast cancer in relation to the decision to recommend or withhold chemotherapy.** *References: EBCTCG (2005) Lancet 2005; 365: 1687–1717; EBCTCG (2012) Lancet 2012; 379: 432–44 (3)* Krop I et al.: American Society of Clinical Oncology Clinical Practice Guideline Focused Update. *JCO.* 2017	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Agendia N.V.	102	406		"Evidence relating …between risk groups." **First we would like to argue that the authors make CT benefit obligatory for a positive assessment, which seems a deviation from the primary aim of this assessment. Patients gain from these test without the proof of CT benefit, given that the outcome for patients with Low Risk results as sufficiently low to forego CT. In addition, it is an improvement over current practice. As mentioned in the 'Aims and objectives of the assessment' section on P44 "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?" focus is on guiding adjuvant chemotherapy decision not on the predictive value of tumour profiling testing.** **Secondly, authors should mention the difficulty that arises from the first study designs for chemotherapy benefit prediction (as stated in comment no. 3 of this report). RCT where "some patients received CT"; please note that it should say patients were *randomized* to receive CT.** **This type of trial is very rare, and the few trials that are available have insufficient patient samples left. So, this type of study design is impossible to adhere to for these and future tests.** **Authors should acknowledge this difficulty and are therefore strongly advised to also include other type of studies such as neo-adjuvant studies that are more and**	The company's statement is not accurate and their interpretation of the report is unreasonable. The EAG clinical review considers the evidence for both prognostic benefit and predictive benefit. The EAG model includes prognostic benefit in the base case, as well as a sensitivity analysis in which Oncotype DX is assumed to be predictive of chemotherapy benefit. We do not suggest at any point in the report that a tumour profiling test can only be valuable if it is predictive of chemotherapy benefit.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				more being recognized as appropriate study design for determining benefit of treatment, especially in specifically stratified patient subgroups. **We suggest the authors to regard the available neo-adjuvant data for the CT benefit of the MammaPrint test [Whitworth, Ann Surg Oncol 2014 and 2017; Baron, Ann Surg Oncol 2015; Beitsch, Ann Surg Oncol 2016 and 2017]. Therefore the assessment bodies should allow for reviewing alternative study set-ups for determining CT benefit.**	
Agendia N.V.	103	407	6.1.1	"the MINDACT…. 5 year DRFI." **The difference of 1.5% benefit was based on DMFS (instead of DRFI) and it should be mentioned that it is a non-significant difference. The absolute non-significant difference in Clin High/ Genomic low patients in terms of DRFI (DMFI) was 1.3%. see earlier comment no. 4.**	This has been noted in an erratum to NICE.
Agendia N.V.	104	407	6.1.1	Paragraph "The MINDACT study…. alter treatment decisions." **The only study proving clinical utility for a genomic test is the MINDACT study where MammaPrint is being used. As Clinical Utility is dominating prospective studies MammaPrint should be valued as higher compared to the other tests. Authors should highlight that such a level of evidence for patients that are candidate for chemotherapy is only available for MammaPrint. This level of evidence is available for the ODx test only for clinically low risk patients that would not be candidates for chemotherapy anyhow as most if not all Oncotype**	We have noted in the addendum to NICE that MammaPrint is the only one of the five tests to have reported randomised controlled trial evidence (MINDACT) of treatment guided by the test versus usual practice, in patients who are high-risk via either mAOL or MammaPrint.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				patients with a score <11 have a NPI < 3.4 (100% LN-, >92% grade 1 and 2, >93% <3cm.	
Agendia N.V.	105	408-409	'MammaPrint'	**The difference in effects ((QA)LYs) are so small (which is in fact the concept of the MammaPrint), that each parameter on the effect side will have major impact. We consider the cost effectiveness study based on MINDACT and submitted to NICE to be more accurate. We argue that the authors underestimate the short and long term side effects and cost of chemotherapy. We strongly disagree with the outcome that the QALY without test is higher than the QALY with test in the mAOL high risk category. This would mean that withholding chemotherapy in 46% of mAOL high risk patients without compromising outcome has no health benefit.**	Please refer to the EAG's major concerns with the Agendia model, as noted above.
Agendia N.V.	106	409	6.2.1	"the evidence base was large, ….considered to be a high quality source of data." **Highest level of evidence of clinical utility is not being mentioned. Authors should acknowledge that the MINDACT study is the only study generating evidence of clinical utility for MammaPrint in this assessment. TailorX has only been able to present outcome of one of its trial arms (Low risk) thus far.**	Please see response to comment # 104.
Agendia N.V.	107	409	6.2.1	"There were some key gaps in the literature for IHC4+C and RSPC. …similar across centres." **Again this paragraph is confirming the extreme limitations of the clinical evidence of IHC4 or IHC4+**	The EAG do not feel any adjustment to the report is required as the limitations of the IHC4 evidence base are clearly reported.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				indicating that this tool is far from having the same level of evidence compared to the other tests. Authors should realize that for IHC4 or IHC4+ with such a limited scientific body of evidence and not being commercially available this tool first needs further research and commercialization to be a reasonable party (intervention) in this assessment. Currently it has no concrete added value for patients and until proven differently, might do more harm than good **It seems unfair and unjustified to work with different levels of requirements within this assessment just because the required information is not available.**	
Agendia N.V.	108	409	6.2.1	"There were relatively limited data relating to the ability of Oncotype…. Adjustment for all relevant variables." **We agree with the fact that there is relatively limited data related to predict benefit from chemotherapy derived from prospective trials on adjuvant treatment, seen the aim of this assessment; "*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?"* the focus is on guiding adjuvant chemotherapy decision not on the predictive value of tumour profiling testing.** **The authors make CT benefit obligatory for a positive assessment, which is not reasonable. Patients gain from these test without the proof of CT benefit, given that the outcome for patients with Low Risk results as sufficiently low to forego CT.**	The EAG disagree that the Assessment report only focusses on predictive benefit of chemotherapy. The report clearly includes a wealth of evidence relating to the prognostic performance of the five tests. No change to the report has been made.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Please see earlier comment no. 2,3 and 7	
Agendia N.V.	109	410	6.2.1	"Data relating to the …test in real clinical practice." **We argue that there is limited evidence of prognostic value for MammaPrint. We wonder why the authors did not use the 'control groups' of the MINDACT study as prognostic evidence. This would mean a reanalysis of an RCT MINDACT where, in case of the 'control groups' (MINDACT; C-low/ G-low, C-low/ G-high following clinical risk assessment, + C-high/G-low following clinical risk assessment and C-high/ G-high), the test did not influence the treatment but test results are available.**	We have added an addendum relating to the potential use of MINDACT in prognostic assessment.
Agendia N.V.	110	410	6.2.1	"Retrospective observational studies….compared to the evidence base for most other tests." **Again we argue that the RCT MINDACT study should be used to define the prognostic value of MammPrint. Although not used as a validation study (as validation was already in place), it still shows the prognostic value of the test and should therefore not be excluded from the evidence provided by MammaPrint. Also we think that it is unjustified that only MammaPrint is highlighted in this paragraph as most studies on MammaPrint were in fact not observational and other tests do not even have prospective data from an RCT to present.**	As noted above, we have added an addendum relating to the potential use of MINDACT in prognostic assessment.
Agendia N.V.	111	410		"These problems were particularly relevant to the MammaPrint evidence base, where most studies were observational in nature rather than reanalyses of RCTs."	The paragraph in question relates specifically to prognostic performance data, and as such the statement is correct.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				This statement is not properly reflecting MammaPrint evidence base. Since MammaPrint is the only test for which a prospective randomized Phase 3 study (MINDACT) is available. **The authors explain that their informed decision was based on "evidence" in these broad categories: Development (validation pubs), Prognostic performance (survival pubs), Chemotherapy benefit (re-analysed RCT pubs), Decision impact (decisions only with no LTFup pubs--which have additional value in this category with Tsai et al.,JAMA Oncology 2017, also previously presented as poster publications and therefore already in the public domain before the submission deadline of this report) and Clinical utility (the authors explain this ideally should be randomized and prospective)--in which case, evidence in this section would be exclusive to us.** **Our Point; MammaPrint is the only test with valid evidence that satisfies criteria for our data to be placed in the category of Clinical Utility but there is very limited appreciation of this fact in this report.**	However, the EAG agree that it fails to highlight the higher level of evidence provided by MammaPrint in MINDACT. We have noted in the addendum to NICE that MammaPrint is the only one of the five tests to have reported randomised controlled trial evidence (MINDACT) of treatment guided by the test versus usual practice, in patients who are high-risk via either mAOL or MammaPrint.
Agendia N.V.	112	411	6.2.2	"(iii) the model structure…and the Genomic Health model, and…" **As Genomic Health is one of the comparators in this assessment it seems unfair to use a Genomic Health model to 'validate' results and therefore indicate the strengths of the EAG model. Moreover, elsewhere in the**	This criticism does not make sense. The EAG identified errors in the Genomic Health model. It is the EAG-corrected version of the Genomic Health model which is compared against the previous and current EAG models. This is clear from the text.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				same report the GH model is being criticized, which seems odd if it has been used as a validation for the AEG model.	As noted in response to comment 5, we do not claim that the EAG model is "correct". Rather, we highlight nearly 2 pages of important limitations (see EAG report, 409-410) that should be considered when interpreting the EAG model results. What is relevant here is that when based on the same data, the original EAG model, the new EAG model and the new (EAG-corrected) Genomic Health model all produce consistent conclusions.
Agendia N.V.	113	411	6.2.2	"However, ….(ii)…influence economic conclusions drawn from the analysis." **See earlier comment no. 108. If limited data on predictive data is strongly influencing economic conclusions from the analysis we are interested how authors deal with this issue in relation to the other tests and why this is not mentioned. We wonder why no comments were made concerning the other tests of this assessment and their 'predictive value'.**	The other tests did not have any evidence relating to predictive benefit. Predictive benefit was considered for Oncotype DX within the sensitivity analyses only, due to uncertainty in the clinical evidence.
Agendia N.V.	114	411	6.2.2	"However, …(iii)…the analysis of MammaPrint is based on a different data source that the other four tests; " **Whilst other tests were all conducted within the same population (transATAC), MammaPrint was the only test of which results came from extrapolating the numbers of another study (MINDACT). Which gives these differences: other population, 5 yrs FU vs 10 yrs FU, and other Clinical Risk parameters. MammaPrint results are**	See previous responses. This is not discussed as a limitation of MINDACT – it is a limitation of our overall economic analyses. The report has not been amended.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				all based upon assumptions, and are difficult to compare to the rest of the test results. **Important to realize; in basis most essential part of the total assessment is to acknowledge the study with the highest level of clinical evidence instead of the inclusion of all tests in one study to compare the interventions.** **Because the added value of a risk classifier test lies within the group of patients in which CT recommendation is not clear based on clinico-pathological factors alone authors should acknowledge the level 1A evidence of MammaPrint by the MINDACT study is more important than the fact that a direct comparison of tests is possible.** **If different data sources for the different tests are considered to be a limitation in the model this should be stated as a 'neutral' limitation and not that this is a specific limitation due to MammaPrint. We request the authors to re-phrase this sentence.**	
Agendia N.V.	115	411	6.3	"The evidence … from clinically high-risk patients. " **Authors should acknowledge the available clinical utility data from the MINDACT study.**	The sentence in question is an overarching statement about the evidence base as a whole, not about each test individually. The EAG do not agree that this statement should be added and no amendment has been made to the report.
Agendia N.V.	116	412	6.5	'IHC4 is not currently commercially available….within the NHS." **This implicates that IHC4 is not mature enough and far behind compared to other tests. This may indicate that**	No change to the report required.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				this test is not yet a serious intervention in this assessment at this moment.	
Agendia N.V.	117	412	6.6	"there is uncertainty regarding…..for all relevant clinic-pathological factors." **We argue that MammaPrint doesn't have sufficient data to prove clinical utility as we have acquired highest level of evidence with the results of the MINDACT study. In this study MammaPrint has proven to be of additional value on top of clinico-pathological factors.**	This point in the report relates to prediction of chemotherapy benefit, which is not addressed in MINDACT. No change to the report has been made.
Agendia N.V.	118	412	6.6	"There is limited evidence …. Would be valuable." **See earlier comment no. 56 not all long term evidence on MammaPrint is taken in to account.**	The sentence states "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". The EAG believe this statement is valid for all tests, as MammaPrint only compares to mAOL, which is not used in clinical practice currently, and currently only reports 5 year outcomes. No change to the report has been made.
Agendia N.V.	119	413	6.6	"there is uncertainty…. Of these tests." **For the cost eff analysis of MammaPrint we disagree with several assumptions and disagree with the incorporated influence of predictive value of the test. (aanvulling Christa nodig)**	Predictive benefit has been included in sensitivity analyses for Oncotype DX.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Agendia N.V.	120	405-413	6	**Besides the suggested research priorities, no final (clear) conclusions are given concerning the main aim of this assessment. Described 'Aims and objectives of the assessment' section on P44 "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?"**	Due to the large and heterogenous evidence base, the EAG have purposefully not made any direct conclusions, as all conclusions will require assumptions to be made about the generalisability of the evidence base.
Agendia N.V.	121	421	References	Ref 114 **Esserman 2016, could now be replaced with Esserman et al. publication in JAMA Oncology 2017**	This evidence was not submitted to the EAG in time to be included in the report.
Agendia N.V.	122	428	References	**Ref 220 Kuijer et al.2016, could now be replaced with Kuijer publication in JCO 2017**	This evidence was not submitted to the EAG in time to be included in the report.
Agendia N.V.	123	432	References	**Ref 292. Incorrect order of authors. First author is Cardoso F.** Azim Jr H.A., de Azambuja E., Colozza M., Bines J. & Piccart M. J.. Long-term toxic effects of adjuvant chemotherapy in breast cancer. Annals of Oncology 2011, 22: 1939–1947 Baron P. , Beitsch P., Boselli D., et al. Impact of Tumor Size on Probability of Pathologic Complete Response After Neoadjuvant Chemotherapy. Ann Surg Oncol. 2015	We agree. This reference should have been: van't Veer, L.J., Yau, C., Nancy, Y.Y., Benz, C.C., Nordenskjöld, B., Fornander, T., Stål, O., Esserman, L.J. and Lindström, L.S. Tamoxifen therapy benefit for patients with 70-gene signature high and low risk. Breast Cancer Research and Treatment 2017. This has been included in the errata.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				May;23(5):1522-9.Bhatt D, Mehta C. Adaptive design for clinical trials. N engl J Med. 2016;375(1):65-74. Beitsch P, Whitworth P, Baron P, et al. Genomic Impact of Neoadjuvant Therapy on Breast Cancer: Incomplete Response is Associated with Altered Diagnostic Gene Signatures. Ann Surg Oncol. 2016 Oct;23(10):3317-23. Beitsch P. et al. Pertuzumab/Trastuzumab/Ct Versus Trastuzumab/Ct Therapy for HER2+ Breast Cancer: Results From the Prospective Neoadjuvant Breast Registry Symphony Trial (NBRST). Ann Surg Oncol 24 (9), 2539-2546. 2017 Apr 26. Bogaerts J., Fatima Cardoso, Marc Buyse, Sofia Braga, Sherene Loi, Jillian A Harrison, Jacques Bines, Stella Mook, Nuria Decker, Peter Ravdin, Patrick Therasse, Emiel Rutgers, Laura J van 't Veer and Martine Piccart on behalf of the TRANSBIG consortiumGene signature evaluation as a prognostic tool: challenges in the design of the MINDACT trial. Nature Clinical Practice Oncology 2005 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.	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Cardoso F, van 't Veer L, Bogaerts J, et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer - supplement. *N Engl J Med*. 2016;375(8):717-729. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Comparisons between different polychemotherapy regimens for early breast cancer: Meta-analyses of long-term outcome among 100 000 women in 123 randomised trials. *Lancet*. 2012;379:432-444. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). EBCTCG (2005) Lancet 2005; 365: 1687–1717 Esserman L, Yau C, Thompson C, et al. Use of Molecular Tools to Identify Patients With Indolent Breast Cancers With Ultralow Risk Over 2 Decades. *JAMA Oncol*. June 2017. Hanly P. BA, MA, PhD, Aileen Timmons BSc, MSc, PhD, Paul M. Walsh MSc, PhD, Linda Sharp BSc, MSc, PhD. Breast and Prostate Cancer Productivity Costs: A Comparison of the Human Capital Approach and the Friction Cost Approach Value in Health, Volume 15, Issue 3, May 2012, Pages 429-436.	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Hudis C, Dickler M. Increasing Precision in Adjuvant Therapy for Breast Cancer. *N Engl J Med*. 2016;375(8):790-791. Joensuu, H., Kellokumpu-Lehtinen, P.L., Bono, P. et al, Adjuvant docetaxel or vinorelbine with or without Trastuzumab for breast cancer. N Engl J Med. 2006;354:809–820. Kuijer A, Straver M, den Dekker B, et al. Impact of 70-Gene Signature Use on Adjuvant Chemotherapy Decisions in Patients With Estrogen Receptor-Positive Early Breast Cancer : Results of a Prospective Cohort Study. *J Clin Oncol*. 2017;35. Knauer M, Mook S, Rutgers E, et al. The predictive value of the 70-gene signature for adjuvant chemotherapy in early breast cancer. *Breast Cancer Res Treat*. 2010;120(3):655-661. Krop I., Ismaila N., Andre F., Bast R.C, Barlow W. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy forWomenWith Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline Focused Update. JCO 2017 Lippman M, Hayes D. Adjuvant therapy for all patients with breast cancer? *J Natl Cancer Inst.* 2001;93(2):80-82. Mewes Janne C., Lotte M. G. Steuten, Iris F. Groeneveld, Angela G. E. M. de Boer, Monique H. W.	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Frings-Dresen, Maarten J. IJzerman and Wim H. van Harten. Return-to-work intervention for cancer survivors: budget impact and allocation of costs and returns in the Netherlands and six major EU-countries BMC Cancer2015, 15:899. Mook S, Schmidt M, Viale G, et al. The 70-gene prognosis-signature predicts disease outcome in breast cancer patients with 1-3 positive lymph nodes in an independent validation study. *Breast Cancer Res Treat*. 2009;116(2):295-302. Mook S, Knauer M, Bueno-de-Mesquita J, et al. Metastatic potential of T1 breast cancer can be predicted by the 70-gene MammaPrint signature. *Ann Surg Oncol*. 2010;17:1406-1413. http://www.agendia.com/our-science/accreditations-licenses/ Retèl VP, Linn SC, van Harten WH. Molecular profiling rather likely to be cost-effective. JCO 2015 May 10;33(14):1626-7. Retèl V, Groothuis-Oudshoorn C, Aaronson N et al. Association between genomic recurrence risk and well-being among breast cancer patients. *BMC Cancer*. 2013;13(1):295. Rutgers E, Piccart-Gebhart M, Bogaerts J, et al. Baseline results of the EORTC 10041/MINDACT TRIAL (Microarray In Node 0-3 positive Disease may Avoid	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				ChemoTherapy) on behalf of the MINDACT TRANSBIG Studygroup. *ESMO Annu Meet*. 2013:Abstract. Tsai M., Lo S., Audeh W, Qamar R. Association of 70-Gene Signature Assay Findings With Physicians' Treatment Guidance for Patients With Early Breast Cancer Classified as Intermediate Risk by the 21-Gene Assay JAMA Oncol. doi:10.1001/jamaoncol.2017.3470 van 't Veer L, Dai H, van de Vijver M, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415:530-536. van de Vijver M, He Y, van 't Veer L, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002;347(25):1999-2009. van 't Veer L, Yau C, Yu N, et al. Tamoxifen therapy benefit for patients with 70-gene signature high and low risk. *Breast Cancer Res Treat*. 2017:1-9. Viale G, de Snoo F, Slaets L, et al. Immunohistochemical versus molecular (BluePrint and MammaPrint) subtyping of breast carcinoma. Outcome results from the EORTC 10041/BIG 3-04 MINDACT trial. *Breast Cancer Res Treat*. 2017. Vliek S., Retel V., Drukker C., Bueno-De-Mesquita J.M., Rutgers E., van Tinteren H.,. van de Vijver M.J, Wesseling J., van Harten W., Linn S.C. The 70-gene signature in node positive breast cancer: 10-year follow-up of the observational RASTER study. ESMO 2017	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Congress http://oncologypro.esmo.org/Meeting-Resources/ESMO-2017-Congress/The-70-gene-signature-in-node-positive-breast-cancer-10-year-follow-up-of-the-observational-RASTER-study Yao K, Goldschmidt R, Turk M, et al. Molecular subtyping improves diagnostic stratification of patients with primary breast cancer into prognostically defined risk groups. *Breast Cancer Res Treat*. 2015;154(1):81-88. Whitworth P, Stork-Sloots L, de Snoo FA, et al. Chemosensitivity predicted by BluePrint 80-gene Functional Subtype and MammaPrint in the prospective Neo-adjuvant Breast Registry Symphony Trial (NBRST). Ann Surg Oncol. 2014 Oct;21(10):3261-7. Whitworth P, Beitsch P, Mislowsky A, et al. Chemosensitivity and Endocrine Sensitivity in Clinical Luminal Breast Cancer Patients in the Prospective Neoadjuvant Breast Registry Symphony Trial (NBRST) Predicted by Molecular Subtyping. *Ann Surg Oncol*. 2017;24(3):669-675.	
NHS Professional	1	44	3.5	"*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?"* There are two main ways in which tumour profiling testing could potentially represent clinically effective and cost	No response required.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				effective use of NHS resources. These would apply in certain patient groups. In the first a risk score is provided from genomic profiling which guides a clinical decision whether or not to use adjuvant chemotherapy when standard clinical parameters are uncertain. The main group in which this is most relevant in NHS practice is the group of patients detailed in the recommendation of the DG10 i.e. LN negative patients with intermediate risk scores from clinical parameters. This is a group of patients in the UK which more often than not avoids adjuvant chemotherapy. In this group, a low genomic risk score would indicate no chemotherapy which would contribute to clinical and cost-effectiveness. A high score in this group would lead to adjuvant chemotherapy, which it is assumed would improve outcomes in the individual and stratified patient group and therefore contribute significantly to clinical effectiveness. For those genomic profiles which have intermediate scores, then the individual decision would come back to the clinician and patient to make together. In the second, genomic profiling provides a risk score in a situation where adjuvant chemotherapy is generally given in the UK. The main group in which this is most relevant is the ER +ve HER2-ve LN + group. In this group a low risk genomic profiling score could result in patients avoiding chemotherapy and this would contribute to both clinical and cost-effectiveness.	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				This question is being asked prospectively in the NIHR HTA funded OPTIMA study. Patients with ER+ve, HER2-ve, early breast cancer who are LN positive (1-9 nodes), or LN negative with a tumour size >/= 30mm are included. Patients are randomised to chemotherapy (standard treatment) or test-directed therapy (high risk – chemotherapy followed by hormonal treatment: low risk – no chemotherapy, immediate hormone treatment) and the trial has a non-inferiority endpoint.	
NHS Professional	2.	16	2.4	*"In LN+ patients, three tests (Prosigna/ROR-PT, EPClin [EndoPredict Clinical] and IHC4+C [IHC4 + clinical score]) categorised far more lymph node positive (LN+) than lymph node negative (LN0) patients as high-risk among studies of endocrine monotherapy."* Presumably the influence of the addition of the clinical score to the genomic score for these 3 tests, drives the finding of higher scores in LN positive patients. *"However, Oncotype DX categorised more patients as low-risk in LN+ than other tests (57% in Oncotype DX versus 4% to **% 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)."* This would support the inclusion of clinical parameters into the genomic risk scores which is now what all these multi-parameter tests have done. In Oncotype Dx without the	The EAG are not able to speculate on the likely cause of the categorisation of more LN+ patients as high-risk by these tests, though it is an interesting suggestion, and an interesting research question posed by the commentator.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				clinical parameters, the test appears to produce more similar frequencies of risk categorisation independent of LN status than the other tests which have incorporated them. From a research point of view there is an interesting question here – Does the tumour which demonstrates LN metastases, but is similar in other clinical / pathological parameters to tumours which do not have LN metastases, have a different genomic profile which drives this LN spread ?	
NHS Professional	3	18	2.4	The data with regards to Oncotype Dx in comparison with other genomic scores also flags up comparisons between the different genomic tests performed in the same group of patients. Concordance between tests was not systematically examined (p.18) but was reported for OPTIMA prelim. These demonstrate that comparing Oncotype DX, MammaPrint, Prosigna and IHC4 although there are similar numbers of patients assigned to each risk category, the test results for an individual patient can vary significantly between the 4 tests.	No response required
NHS Professional	4	17	2.4	Prognostic performance for all the tests was good, although not fully validated in RSPC and IHC4+C.	No response required
NHS Professional	5	17	2.4	Prediction of chemotherapy effectiveness The genomic profiling tests have been developed mainly to provide additional prognostic information to provide risk predictions. Clinical decision making tools in terms of the need or not for adjuvant chemotherapy are based on both risk and trial evidence of benefit of adjuvant chemotherapy treatment.	No response required, though the EAG do not fully support the statement “Oncotype DX is the only test which has demonstrated a definite positive prediction for the effect of adjuvant chemotherapy in the high risk group”, as described in our report and the addendum generated in response to these comments.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				OncotypeDx is the only test which has demonstrated a definite positive prediction for the effect of adjuvant chemotherapy in the high risk group. Interestingly when the clinical parameters are added to form the Oncotype Dx RSPC, the prediction of benefit from adjuvant chemotherapy is lost. My comment here would be that this is a finding which I find difficult to explain? Prognostic and predictive tests. Historically, for the ER positive population decision-making to add chemotherapy has been based more on risk prediction. Unlike for adjuvant hormonal therapy (where ER is the target), there are no defined 'targets' for standard chemotherapy treatment except high proliferation rates because of the mechanism of action of chemotherapy.	
NHS Professional	6	17	2.4	MINDACT Study This study set out to look at the 70-gene assay signature (Mammaprint) and looked at genomic risk and clinical risk groups. Where there was concordance between clinical and genomic parameters, the groups either received or did not chemotherapy. Where there was lack of concordance then adjuvant chemotherapy was randomised. There were significant numbers of patients who did not receive the allocated treatment for two main reasons. First there were alterations in the genomic risk score after the original allocation to high or	No response required. The EAG agree with the commentator. It should be noted that Agendia's comment #64 relates to the 0.8% benefit in the low clinical/high genomic group, and our response to their comment is relevant here.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				low risk, and after the randomised allocation had been made. Secondly, a number of patients who were randomised to chemotherapy did not receive it, and who were randomised to no chemotherapy did receive it. The most interesting group and that focussed on in the NEJM publication, was the high clinical risk / low risk genomic score population. A total of 1550 patients were in this group randomised between chemotherapy and no chemotherapy. At 5 years the rate of survival without distant metastasis in this group was 94.7% (95% confidence interval, 92.5 to 96.2) among those not receiving chemotherapy. The absolute difference in this survival rate between these patients and those who received chemotherapy was 1.5%, with the rate being lower without chemotherapy. This result was within the non-inferiority boundaries set for the trial, and the trial concluded in the high clinical risk group who would all have received chemotherapy, 46% of patients would have avoided chemotherapy by having a low genomic risk score. The non-inferiority was 1.5% reduction in 5 year distant disease-free survival which was judged acceptable in clinical terms to avoid the short and long term toxicity and costs of chemotherapy. This result would be in the second category of result noted in comment 1, above. Adding adjuvant chemotherapy for the low clinical risk / high genomic risk group, only resulted in 0.8% benefit in 5 year	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				distant disease-free survival. Which would suggest that for the Mammaprint test there is no advantage in terms of clinical or cost-effectiveness in testing low clinical risk patients.	
NHS Professional	7	18	2.4	Decision Impact The decision impact has a wider range in Europe than in the UK. This may have something to do with the fact that the bench mark rates of adjuvant chemotherapy are lower in the UK than in Europe of the US. Therefore in the UK, a low risk genomic test, would be more likely to confirm a pre-test decision not to use chemotherapy.	No response required
NHS Professional	8	21	2.7	Will the data from NHS England Access Scheme Dataset for Oncotype Dx accessed following the DG10 guidance be available to add to the evidence?	This information was provided as commercial in confidence. The EAG has no control over the release of this dataset.
Genomic Health	1	102, 105, 115, 116, 117, 118, 120, 361	4.3.3, 4.3.4	Executive Summary: The Assessment of Chemotherapy Benefit • The Oncotype DX breast Recurrence Score ® assay is the only assay to demonstrate a statistical interaction for chemotherapy benefit in two independent well-designed prospective/retrospective clinical trials (NSABP B20 and SWOG-8814). o Both studies reported that there was little or no chemotherapy benefit derived for patients with a Recurrence Score result <18. o Both studies reported that there was a statistically significant benefit of the addition	The EAG base their description of the evidence to support Oncotype DX's ability to predict benefit from chemotherapy as being weak on the basis p-values around 0.05 are associated with a high false positive rate and interaction tests often had *p*>0.05 when clinicopathological variables were adjusted for, which suggests the observed differences between risk groups could be due to confounding by clinicopathological variables, rather than due to a real effect. In addition, the cohort used to test for chemotherapy benefit in LN0 patients (NSABP B20) was the derivation cohort for Oncotype DX, and is therefore at high risk of bias.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				of chemotherapy for patients with a Recurrence Score result ≥ 31. • With such proof of chemotherapy prediction, it has since been considered unethical to design and conduct studies which would randomise patients with low Recurrence Score results to chemotherapy, and conversely randomise patients with a high Recurrence Score to no chemotherapy. Therefore, all subsequent prospective trials were/are designed based on the proof that the Recurrence Score® identifies both patients who may safely be spared chemotherapy, and those who will benefit from chemotherapy. • There are misunderstandings regarding the strength of the evidence around the prediction within the Diagnostics Assessment Report. Genomic Health have addressed all misunderstandings individually in detail within the appendix of this document. • Genomic Health can demonstrate that the evidence for Oncotype DX ®; o is of high quality, o is generated from well-designed studies which minimise bias, o complies with the EGAP and Simon et al. biomarker development assessment criteria.	The EAG included all data found through our rigorous systematic review and believe we have included all relevant data. Whether the assessment should be postponed is not a matter for the EAG to consider and this comment should be brought to NICE directly. TAILORx uses different cut-off points than are currently recommended (11 and 25, versus 18 and 31), and it is unclear whether this trial will provide evidence relating to the prediction of chemotherapy benefit. It also recruited only LN0 patients who met the NCCN guidelines for chemotherapy, and tested them all with Oncotype DX. It is unclear to the EAG whether NCCN guidelines result in a group of patients who would be indicated for chemotherapy in the UK, according to usual clinical practice. As such, the generalisability of findings from TAILORx may be limited.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				• It seems that the EAG have not considered the evidence in its entirety, and as a result have failed to draw a conclusion on the evidence as a whole. • The prospective evidence from over 60,000 patients from randomised controlled trials and real-world evidence, consistently show that the Recurrence Score® can reliably identify patients who do, and do not benefit from chemotherapy. This indisputable Oncotype DX® evidence in its entirety is unsurpassable by any other assay's evidence under review. Recommendation Summary • The EAG should consider the evidence related to the Oncotype DX Breast Recurrence Score assay in its entirety. This evidence as a whole is indisputable, and demonstrates that the Recurrence Score can predict benefit from chemotherapy, and should be modelled using the standard 18 and 31 Recurrence Score cut points as predictive in the base case cost-effective analysis. • Genomic Health strongly recommends that NICE take the decision to postpone the assessment until the upcoming TAILORx RCT trial results can be incorporated. This prospective randomized clinical trial will provide additional precision on the effect of chemotherapy, if any, for patients with Recurrence Score results between 11 and 25 in almost 7000	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				patients randomized to treatment. [redacted] Detailed Response in appendix: The Assessment of Chemotherapy Benefit	
Genomic Health	2			The Assessment of Chemotherapy Benefit ***EAG comment -"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, (DAR ref 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."*** The Paik et al publication reported that the Oncotype DX Breast Recurrence Score® assay was shown to be predictive of chemotherapy benefit and reported a significant assay	"*The authors at NSABP continue to stand by their conclusion*." – The EAG agree that in Paik 2006 that in the unadjusted analyses, Oncotype DX was predictive of chemotherapy benefit. However, the EAG also note that interaction tests presented that adjusted for individual clinicopathological factors were not always statistically significant, and no analysis was presented that adjusted for all clinicopathological variables at the same time, or for all randomisation stratification factors. This leads to the possibility that the observed difference between risk of recurrence in the Oncotype DX RS groups is confounded by differential distributions of clinicopathological variables. The EAG note that Genomic Health do not offer an alternative interpretation of this evidence, but rather do not mention it in these comments at all.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				result by treatment interaction. The authors at NSABP continue to stand by their conclusion. Evidence from the NSABP B20 study suggests that the effect of chemotherapy was clinically equivalent to the effect of no chemotherapy for Recurrence Score results up to approximately 18. The authors reported that patients with low Recurrence Score results (RS < 18) experienced minimal if any benefit from the addition of chemotherapy (relative risk, 1.31; 95% CI, 0.46 to 3.78). The mean absolute decrease in distant recurrence rate at 10 years is -1.1% (SE, 2.2%). These results are in stark contrast to those for patients with a high RS result (RS ≥ 31) who experienced a clear and large benefit, i.e., a reduction in risk, with the addition of chemotherapy (relative risk, 0.26; 95% CI, 0.13 to 0.53), with a mean absolute decrease in the rate of distant recurrence at 10 years of 27.6% (SE, 8.0%). Based on these findings, in the United Kingdom, United States, and in other countries, the cut-off of 18 is routinely used in clinical practice to recommend treatment with hormonal therapy alone and the cut-off of 31 is routinely used to recommend treatment with chemo-hormonal therapy.	"*Evidence from the NSABP B20 study suggests that the effect of chemotherapy was clinically equivalent to the effect of no chemotherapy for Recurrence Score results up to approximately 18.* " – Whilst the categorical analysis does show no benefit from chemotherapy in the low risk group, Paik et al. 2006 also a perform an analysis of RS as a continuous variable, and state "a clear cut-off point for RS, below which there is no demonstrable benefit from chemotherapy, cannot be accurately defined."

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Genomic Health	3			***EAG comment- "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; (DAR ref 68) analyses in NSABP B-20 appeared to only include each covariate separately), (DAR ref 49, 50) or whether all stratification factors used in randomising patients to treatment were included as well."*** The ability of the Recurrence Score result to predict chemotherapy benefit was formally tested, as per the guidance in Ballman et al 2015, "A biomarker is predictive if the treatment effect (experimental compared with control) is different for biomarker-positive patients compared with biomarker-negative patients. As will be described shortly, there must be at least two comparison groups available (eg, two different treatment arms in a randomized trial) to make this determination." Further in the Statistical Considerations section, it states, "To determine whether a biomarker is potentially predictive or prognostic, a formal test for an	The EAG agree that interaction tests were performed, but note the limitations of these analysis as described in response to Genomic Health's comment #2. The issue is not whether the individual covariates interact with treatment but that they should be included in the Cox regression irrespective of whether they are balanced across treatments and whether their main effect is statistically significant. Non-linear models such as Cox regression must include all relevant covariates (regardless of statistical significance) and it is not unusual for apparent interactions on the log-hazard ratio scale to be explained by the inclusion of such covariates. The quotes from Paik et al. 2006: "… "the anticipated benefit of adding chemotherapy to hormonal therapy may not exceed the risks" for many women with low RS result. Alternatively, "the anticipated benefit of adding chemotherapy appears to be very favourable when compared with the risks" for patients with high RS result." This statement is equally valid whether chemotherapy benefit is predicted or not, as patients in low risk groups are expected to gain

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				interaction between the biomarker and treatment group needs to be performed." Indeed, these conditions were met in both the B20 and SWOG-8814 data analyses (interaction p-values = 0. 038 and 0.029, respectively). As reported in the NSABP B20 publication, there was no evidence that age, tumour size, or tumour grade predicted chemotherapy benefit. In fact, the interaction of chemotherapy treatment with a comprehensive set of clinicopathological variables, including age, tumour size, tumour grade, quantitative ER, quantitative PR, and individual gene expression variables were first evaluated in a series of separate models for B20. This covers all stratification variables in B20 except for type of surgery, which is known not to impact outcomes. Although hazard ratios for the interactions were in the expected directions, there was no significant interaction between clinical variables and treatment with chemotherapy, nor was there expected to be as the inability of these characteristics to predict chemotherapy benefit is consistent with the conclusion of the vast literature of breast cancer studies, including the authoritative Oxford overview [Early Breast Cancer Trialists' Collaborative Group (EBCTCG). *Lancet*. 2005]. Also of note, the vast literature does not support the reviewer's speculation that the level of ER by IHC would contribute to the prediction of chemotherapy benefit. Thus, the Recurrence Score was clearly shown in NSABP B20 to predict chemotherapy benefit. A similar approach was taken for the analyses of SWOG-8814 and showed consistent results.	less absolute benefit in survival compared to high risk groups.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				The discussion section of the Paik paper underscores the practice-changing clinical implications of these results: “the anticipated benefit of adding chemotherapy to hormonal therapy may not exceed the risks” for many women with low RS result. Alternatively, “the anticipated benefit of adding chemotherapy appears to be very favourable when compared with the risks” for patients with high RS result. The analysis presented in this manuscript subsequently changed clinical practice, led to the design of the TAILORx trial (which considered it unethical to randomize patients with RS<11 or RS>25) and, across the ensuing supportive evidence, patients with low RS results are spared chemotherapy and experience very low rates of distant recurrence. This is further supported by long-term outcomes evidence now from over 50,000 patients (2) (3). Finally, the prospective randomized clinical trial TAILORx will provide additional precision on the effect of chemotherapy, if any, for patients with Recurrence Score results between 11 and 25 in almost 7,000 patients randomized to treatment. [redacted]	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Genomic Health	4			***EAG comment- "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."*** The NSABP in the published manuscript reported analysis as pre-specified in the protocol both for the continuous Recurrence Score and for pre-specified RS risk group precisely to be able to present as much information as possible to physicians and to patients. In fact, the main analysis of treatment interaction used the Recurrence as a continuous variable. The NSABP has confirmed (data on file, NSABP and Genomic Health) that there are no significant imbalances in clinical pathologic variables included in Paik et al by Recurrence Score risk group for patients with and without chemotherapy treatment across the set of patients analysed in NSABP B20.	We accept that the primary analysis reported by Paik et al (2006) analysed continuous recurrence score. However, all stratification factors used in the randomisation of patients to treatments should be included in the Cox regression irrespective of whether they are balanced across treatments and whether their main effect is statistically significant. We accept that the categorisation of recurrence score into risk groups was pre-specified prior to protocol finalisation, which made it possible to present Kaplan-Meier plots. However, we assert that these should only be used for descriptive purposes and that the unadjusted results for any stratification variables or other important covariates have a different interpretation to those from a full Cox regression model. This is in accordance with Harrell et al. 2017 Thus, the main inferences should be based on the results from a Cox regression with recurrence score as a continuous variable and

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				All statistical models underwent thorough diagnostic testing by independent statisticians at the NSABP. The categorization of the Recurrence Score result into risk groups was pre-specified prior to protocol finalization as it made possible the presentation of Kaplan-Meier (KM) plots that are so useful to clinical interpretation. Finally, none of the clinicopathologic characteristics were statistically significant predictors of chemotherapy benefit.	with the inclusion of all stratification and other important covariates.
Genomic Health	5			***EAG comment- "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)."*** The speculation by the reviewers that attrition of samples caused bias is not supported by the NSABP analysis indicating, as noted above, that there are no significant imbalances in clinical pathologic variables included in Paik et al by Recurrence Score risk group for patients with and without chemotherapy treatment across the set of patients analysed in NSABP B20. [redacted]	**NSABP B20 Attrition of samples:** The NSABP B20 analysis of baseline characteristics showed a statistically significant difference between analysed and non-analysed patients from the trial in tumour grade (significant at p=0.03). Even were no statistically significant differences apparent, this would not ensure that there were no differences in unknown or unreported confounders, so the risk of bias from attrition is a valid concern. **NSABP B20 HER2- patients:** Genomic Health have not presented an analysis which tests for an interaction between RS and chemotherapy benefit in the HER2- population; a simple presentation of KM curves does not show

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				[redacted] There is no evidence to support these biases. The fact that the independent SWOG 8818 study also clearly showed that chemotherapy benefit was mostly observed with a high Recurrence Score result is evidence that no known or unknown biases impact the interpretation of the results. Distribution of patient age, tumour size, tumour grade, and hormone status in the 651 patients assessable for the reanalysis of B20 resembled those in all 2,299 clinically eligible NSABP B20 patients [DAR ref 49]. Missing data was kept to a minimum due to the strict monitoring of the source trial data. For SWOG-8814, not only was there no bias in tumour sample availability, but the subset of patients available for analysis was also overwhelmingly representative of the parent trial. This was true for age, ethnic origin, progesterone-receptor status, and duration of follow-up. Patients in the subset used for reanalysis did have a slightly lower number of positive nodes and a smaller tumour size. This would serve to make the cohort slightly more homogenous. This further supports the strength of the Recurrence Score test because, despite the clinicopathologic homogeneity, the Recurrence Score result was still able to categorize patients into risk groups and predict chemotherapy benefit.	whether confounders between groups may account for the differences in risk rates in RS risk groups. **SWOG-8814 Attrition of samples:** Genomic Health appear to provide contradictory information when stating that there was no bias in tumour sample availability, but noting a lower number of patients with positive nodes and smaller tumours. The EAGs assessment of risk of bias therefore appears to be supported. The loss of smaller samples means generalisability to this patient group may be limited. However, the EAG did not exclude any studies on the basis of risk of bias, and make these points for the sake of transparency for the committee. *Homogeneity of evidence base:* The EAG are not able to verify Genomic Health's assertion that the Oncotype DX evidence base is clinically more heterogeneous than the other tests due to time constraints. Homogeneity would only be desirable if this homogeneity is also representative of the population of interest.

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				We concur with the authors of the diagnostic assessment report that a homogeneous patient sample is ideal when evaluating the performance of a genomic assay; it ensures that the distinction in risk groups is identifiable by the assay itself instead of other underlying clinicopathologic characteristics. The research performed on the Oncotype DX Breast Recurrence Score® has used a more homogenous population than all alternative assays. This should be noted when evaluating the other assays, as the variation in risk groups may be due to known prognostic clinicopathologic characteristics rather than the genomic component itself.	
Genomic Health	6			***EAG comment- "From the three observational cohort studies,(DAR refs 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). DAR refs 73, 74)"*** The reviewers correctly cite that the large population-based SEER study found that patients with high Recurrence Score results who did not receive chemotherapy experienced worse outcomes than those who did receive chemotherapy, and that there was a positive interaction. However, the patients who received chemotherapy in clinical practice had, as would	The EAG agree that in these observational studies the likely direction of effect of patients treated with chemotherapy having worse clinicopathological features would be to reduce the apparent difference between chemotherapy and no chemotherapy. Therefore, the risk of bias is high in not detecting a difference where there is one, so the assessment of risk of bias is correct, and should be read in conjunction with the mixed findings. However, the lack of interaction tests in two out of three studies is still problematic within this data set.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				be expected, worse clinical pathologic features. Thus, it was especially notable that the patients who were treated with chemotherapy with worse prognostic features had more favourable outcomes. These results are consistent with the observation in both NSABP B20 and SWOG 8814 that high Recurrence Score patients have greater chemotherapy benefit, that the Recurrence Score is predictive. The SEER analysis did not dismiss or omit any information that was available in the SEER dataset. Information of estrogen receptor status (positive/negative) is obtained from the sites. There is no information in the SEER dataset on IHC intensity or percent of positive cells. The evidence from the observational studies is overwhelmingly consistent and a benefit to the collective body of clinical evidence because the studies represent real world clinical practice. Test results are intended to guide chemotherapy recommendation. Although other factors can influence chemotherapy recommendation, results from the observational studies show that chemotherapy is used sparingly in the low RS result group and increases with higher RS result. Furthermore, sensitivity analyses incorporating propensity score adjustments have been employed in select analyses to account for the lack of randomization.	Whether the SEER analysis omitted ER and PR status due to lack of data does not change the fact that these variables were not included in the interaction model. A question the committee might wish to explore, however, is whether these variables would be available as a quantified values in practice in England, for potential inclusion in a clinical model. The EAG were not able to ascertain this within the timescales available. The evidence from the observational studies is not overwhelmingly consistent. The MD Anderson study (N=1424) showed no statistically significant difference between chemotherapy and no chemotherapy groups in any risk category both before and after adjustment for clinicopathological variables. The Clalit Health study (N<2700) only reported data in the intermediate group, [redacted], and statistical significance was not reported in LN1micro to LN3 patients. The SEER registry (n=40,134) reported a statistically significant interaction test, but no HRs between chemotherapy and no chemotherapy groups.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Genomic Health	7			***EAG comment- "In practice, it is unlikely that chemotherapy decisions would be made on Oncotype DX scores independent of clinicopathological variables."*** Clinicopathologic variables are important prognostic risk factors and should be used to define an intermediate prognostic risk group of patients for whom adjuvant chemotherapy treatment decisions have greater uncertainty (as per the existing NICE Diagnostic Guidance 10). Oncotype DX testing of this patient sub-group adds additional information, based on the underlying biology of a tumour, regarding the likelihood that a tumour will be responsive to chemotherapy. It has been shown through robust real-world evidence that chemotherapy is used sparingly in patients who have a low RS result. It increases accordingly by RS result risk classification.	The EAG agree that the likely best use of Oncotype DX would be to identify a group of clinically intermediate patients for testing, but note that Genomic Health's statement is not referenced, or supported by empirical data. The extent to which Oncotype DX can add additional prognostic value in a clinically intermediate risk group is included in the economic model of this assessment. The extent to which it can predict chemotherapy benefit in such a group is unknown.
Genomic Health	8			***EAG comment- "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. (DAR refs 49, 50, 68) Interestingly, Tang et al. 2011a50 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,***	We do not accept that we are over-interpreting the p-value for the interaction effect. It is often the case, particularly in non-linear models that apparent interactions between treatment and some covariate can be explained by the inclusion of important omitted covariates. Nevertheless, we do not assert that RSPC is not predictive of chemotherapy benefit on the log-hazard ratio

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				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.(DAR ref 50)" The hazard ratio and confidence intervals for RS and RSPC as a predictive factor for distant recurrence are overlapping. There is no evidence of any "loss" of predictive value of RS by the addition of the clinicopathological variables alongside RS. This supports the conclusion that the RS is predictive and is not confounded by the clinic-pathologic variables. As is often the case, the reviewers are overinterpreting small differences in p-values.	scale, only that the inclusion of clinicopathological variables in the RSPC algorithm may explain the observed interaction between treatment and RS score; there may be additional omitted important covariates and we would like more investigation of these. $P = 0.10$ is weak evidence of an interaction and is associated with a high probability of a false-positive effect, in which case it would be wrong to over-interpret the result as evidence of an interaction.
Genomic Health	9			***EAG comments- "Only one study, the Trial Assigning Individualized Options for Treatment (TAILORx), randomizes patients to treatment guided by the test or treatment according to usual practice… 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) trial (n = 3198) is also a prospective RCT, but does not aim to assess the clinical utility of Oncotype DX, as it randomizes patients with RS>=12 to two different sorts***	The EAG are not contending that Oncotype DX does not have prognostic power, but we do think the evidence on chemotherapy benefit is less robust (see main report and response to previous comments). The EAG note the cut-off quoted by Genomic Health as having been sufficiently evidenced is RS<11 and RS<12, whereas currently Genomic Health recommend use of RS<18 as the cut off

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				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." It should be noted that with TAILORx and RS result <11, the accompanying Editorial by Dr. Cliff Hudis concluded there is no chance for any benefit of chemotherapy for patients with RS result <11 [Hudis. *NEJM*. 2015]. The question has been definitively asked and answered with prospective outcomes in contemporary patients. The TAILORx and WSG Plan B trials were carefully designed by independent cancer consortiums to evaluate unanswered questions. Their study designs further underscore that the question of sparing chemotherapy in the low RS result group and adding chemotherapy in the high RS result group have already been sufficiently answered by the research preceding the TAILORx and WSG Plan B trials. Furthermore, the results from the low RS result group (RS < 11) from TAILORx and the recently reported 5-year outcomes from WSG Plan B (RS ≤ 11) [Nitz. *Breast Cancer Res Treat.* 2017] are consistent with real-world clinical evidence studies and add to the ever-growing body of evidence for the responsible use of the RS test to guide treatment recommendations. While the portions of the TAILORx and WSG Plan B trials that focused on low RS result groups were not randomized, they were carefully controlled, protocol-driven, and had stringent data monitoring	for low risk. This disparity is not explained by the company. The EAG note that the opinion of other researchers, whilst of interest, is not empirical evidence. We have consulted the empirical evidence, rather than the interpretation of that by others. Even well conducted observational studies have inherent limitations; the description of these studies as effectively prospective observational cohort studies is accurate and made to indicate that the data is not comparative. The EAG do not agree that there is consistency of results over 60,000 patients for both prognostic and predictive abilities of Oncotype DX, as the data relating to chemotherapy benefit was not consistent, and the definition of low risk patients in the evidence base is not consistent.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				and completeness standards, more so than the typical observational or registry study. The consistency of results across differing study designs involving over 60,000 patients lends additional credibility to the prognostic and predictive abilities of the Oncotype DX Breast Recurrence Score® test (3) (4) (5) (6) (7).	
Genomic Health	10			***Assessment with the Cochrane risk of bias tool for RCTs*** Here again, it is important to understand that the independent SWOG-8814 study found the same results as the B20 study. Thus, more than one study supports the biological conclusion that low Recurrence Score disease does not benefit from the addition of chemotherapy and high Recurrence Score disease does benefit from the addition of chemotherapy. The Cochrane risk of bias tool, while pure in its aim, is "frequently implemented in non-standard ways" and, in over 85% of included RCTs, at least one risk of bias domain was judged as "unclear" [Jørgensen. *Systematic Reviews.* 2016]. We propose that an alternative approach is to assess the consistency of the randomized clinical trials and collective body of evidence supporting the Oncotype DX Breast Recurrence Score® test. It performs consistently as both prognostic and predictive of chemotherapy benefit across multiple study designs in over 60,000 patients. This is a remarkable volume of evidence and far exceeds that of the	As noted by the EAG in our addenda (in response to comments on our report), approximately 1/3 of patients from the NSABP B20 study were used as the majority of the derivation set for Oncotype DX. As such, results from B20 (LN0 patients) should be considered to be at high risk of bias. As such, SWOG-8814 (which is in LN+ patients) is the only independent (of the derivation set) trial data available which analyses the ability of Oncotype DX to predict chemotherapy benefit. The Cochrane RoB tool was not used to assess SWOG-8814 or NSABP B20 with respect to their chemotherapy benefit analyses. The PROBAST tool was used instead, as per the systematic review protocol, and in accordance with the developers of the tools intended uses (personal communication Dr Robert Wolff).

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				other genomic assays under evaluation, as well as the current practice comparator.	
Genomic Health	11			***EAG comment- "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."*** We have spoken to leading breast cancer clinical researchers around the world. Although we and they all agree it is theoretically attractive to randomize patients to test versus no test, none of the experts thought that a new prospective trial going forward could be enrolled with that design. Given the body of evidence generated for patients who underwent Recurrence Score testing, there are also ethical concerns with performing prospective, randomized trials comparing assay-directed treatment with usual practice. In addition, this would be prohibitively resource intensive due to	The EAG agree that such a trial would be very difficult to conduct, and state in the report that "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." In our description of Clinical Utility (pg 55). However it is still valid to point out that the highest possible level of evidence has not been reached, even where this seems infeasible, to highlight that there are uncertainties in the evidence base. The EAG have presented the evidence base as transparently as possible for the committee to draw its own interpretations. The EAG agree that the clinical utility of using Oncotype DX recurrence score can be modelled, but this will always necessitate assumptions, and always generate uncertainties where the available evidence has limitations (as is the case for this DAR). The economic modelling section of our report provides our

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				both the expected size and duration of studies. A large body of decision impact evidence for Recurrence Score testing demonstrates the significant change in treatment decisions (in both directions) following Recurrence Score testing, versus usual practice. Based on this, combined with the extensive validation and outcomes data for the Recurrence Score assay, the clinical utility (improvement in patient outcomes) is clear and can be modelled without a problem for health economic analyses. Because of the above challenges with prospective tested vs. untested studies, the concept of prospective-retrospective studies, like that performed for NSABP B14, Kaiser, NSABP B20, and SWOG-8814 have gained increased credibility in the clinical research and regulatory fields (Simon et al. *J Natl Cancer Inst.* 2009, U.S. Food and Drug Administration. (August 18, 2016). Principles for Codevelopment of an In Vitro Companion Diagnostic Device with a Therapeutic Product [Webinar]. Retrieved from https://www.fda.gov/downloads/Training/CDRHLearn/UCM517159.pdf). These studies both allow practice changing technology to move into the clinic more rapidly leveraging robust and well-defined trials. In fact, an EGFG inhibitor drug was approved by the EMEA based on evaluation of the tumour biomarker KRAS in "prospective retrospective" studies. Furthermore, as identified by Friedlin et al JNCI 2010, the statistical properties of the biomarker-strategy design are	independent work to assess the cost-effectiveness of the use of Oncotype DX in clinical practice, using data that are of most relevance to the decision problem in England. As far as the EAG are aware, NSABP B20 and SWOG-8814 did not stratify treatment groups by the biomarker value (RS), as the assay was not available when these trials were conducted (1988 to 1993 and 1989 to 1995 respectively). As such, how the quote from Fiedlin et al supports Genomic Health's approach is unclear to the EAG.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				problematic. The authors conclude that *"An adequately sized design that randomly assigns patients to treatment A or treatment B stratified by the biomarker value will provide rigorous evidence for determining the best treatment in the biomarker-positive and biomarker-negative subgroups"*.	
Genomic Health	12			***EAG comment- "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."*** Standard methodology allows for the calculation of 5 year estimates and confidence intervals even when the median follow-up is shorter than 5 years.	Our consultation with clinicians on the clinical relevance of survival data suggested that studies with less than 5 years median follow-up would be immature for survival outcomes. WSG Plan B had less than 5 years median follow-up.
Genomic Health	13			***EAG comment- "Clalit… but again, surprisingly, DRFI was lower in the RS<11 analyses than the RS<18 (Table 26)."*** The final analyses have since been published [Clalit et al. *npj Breast Cancer.* 2017] and show that the confidence intervals are overlapping. The span of the confidence interval for the RS < 18 cohort is narrower because the group is much larger and therefore the estimate is more precise.	This data was not available in time for the EAG to include it in the review.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Genomic Health	14			*EAG comment- "Clinical utility Oncotype summary, 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."*** In the published NICE Diagnostics Guidance 10, NICE already acknowledged that patient outcomes would be affected by use of the Oncotype DX Breast Recurrence Score in a clinical setting. Indeed, it would seem only logical that this assumption was a prerequisite for NICE to make a positive recommendation for use of the test in NHS clinical practice.	The EAG stand by their statement that we cannot conclude whether patient outcomes would be affected, or even in which direction they would be affected. This conclusion is based on the clinical evidence, not on the economic modelling. Any conclusion would be speculative, and dependent on assumptions which the EAG are not in a position to make. This is especially true given the lack of clarity over how intermediate patients would be treated in clinical practice; whether the test would be used in isolation of clinicopathological factors, and how clinicopathological factors would be used in clinical practice; the weak evidence relating to chemotherapy benefit, which is based on only one independent (from the derivation set) set of trial data, in LN+ patients; and the uncertainty around the magnitude of the change in chemotherapy treatment decisions, and any associated impact this may have on recurrence and survival. The EAG are not obliged to agree with other commentators on the evidence base. The reliance on the results of NSABP B20 to design subsequent trials are of concern given the

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Recommendation 1.1: *Oncotype DX is recommended as an option for guiding adjuvant chemotherapy decisions for people with oestrogen receptor positive (ER+), lymph node negative (LN−) and human epidermal growth factor receptor 2 negative (HER2−) early breast cancer if:* • *the person is assessed as being at intermediate risk and* • ***information on the biological features of the cancer provided by Oncotype DX is likely to help in predicting the course of the disease*** *and would therefore help when making the decision about prescribing chemotherapy* We would not agree that it is not possible to draw conclusions as to whether patient outcomes would be affected by use of the test in a clinical setting. Evidence from B20 and SWOG-8814 studies clearly shows prediction of chemotherapy benefit, as well as risk reclassification vs. 'current practice.' The independent editorial in 2006 from Dr. Sandra Swain in JCO 2006 that accompanied the Paik paper supports the conclusions of the study and makes the additional point that the TAILORx trial was designed based on the ability of the 21-gene assay to predict chemotherapy benefit. It was considered unethical to randomize patients in TAILORx with a Recurrence Score of 31 or greater to chemotherapy or not.	overlap with the derivation set for Oncotype DX meaning results are at high risk of bias.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				All patients with high Recurrence Score according to protocol were treated with chemotherapy because the benefit of chemotherapy was clear. All patients with low Recurrence Score according to protocol were treated with hormone therapy alone. The design was directly based on the Paik et NSABP B-20 results. At this point, given the existing body of evidence and research, it would not be ethical to randomize both high and low Recurrence Score result patients to hormonal therapy alone vs. chemo-hormonal combination therapy.	
Genomic Health	15	52 ,54, 69,99,	4.1.4, 4.2 ,4.3.2, 4.3.3,	Executive Summary: The Assessment of Clinical Evidence Quality • We question the decision to use PROBAST as the evidence criteria to assess genomic classifiers' prognostic and predictive capabilities and believe this decision unacceptable as these criteria are unpublished, not peer reviewed, and deviate from broadly accepted published peer reviewed criteria. • The criteria selection process is not transparent and indeed biases the outcome of the assessment. • Subsequent conclusions drawn about the level of evidence supporting the Oncotype DX® assay are not valid as a result of this selection of evidence criteria.	The EAG state in the study protocol "The PROBAST tool has been developed specifically for use in systematic reviews of prediction models by the Cochrane Prognosis Methods Group. Whilst this tool is not yet validated or published, it has been designed using robust methods including 42 topic experts and a Delphi process,14 and is freely available from the lead author (Dr Robert Wolff)." As such, the EAG were confident that the tool represented an excellent, up to date option for assessing risk of bias. We are not obliged to agree with the conclusions drawn by other commentators or

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				• Published consensus criteria show that the Oncotype DX® assay is the only genomic assay to satisfy clinical validity and clinical utility with robust data from validation studies, prospective randomised controlled studies, and real-world evidence. Recommendation Summary • Genomic Health suggest that the EAG conduct a full unbiased literature review on the criteria for assessing genomic classifiers and then use the most widely accepted criteria from peer reviewed published articles to conduct the current assessment this should include EGAP 2015 and Simon 2009. Detailed Response in appendix: The Assessment of Clinical Evidence Quality	published works. That is the nature of an independent assessment. However, the EAG feel that the conclusions drawn by EGAPP on Oncotype are broadly in concordance with their own. From the abstract of their most recent report:[2] "The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group found insufficient evidence to recommend for or against the use of Oncotype DX testing to guide chemotherapy treatment decisions in women with hormone receptor-positive, lymph node-negative, or lymph node-positive early breast cancer who are receiving endocrine therapy. This recommendation statement updates a 2009 EGAPP statement on the use of gene expression profiling tests in breast cancer. Evidence of clinical validity for Oncotype DX was confirmed as adequate. With regard to clinical utility, although there was evidence from prospective retrospective studies that the Oncotype DX test predicts benefit from chemotherapy, and there was adequate evidence that the use of Oncotype DX gene expression profiling in clinical practice changes treatment decisions regarding chemotherapy, no direct evidence was found that the use of Oncotype DX testing leads to improved clinical outcomes."[2]

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
					A full literature review of risk of bias assessment tools for prognostic and prediction studies was not within the scope of this work.
Genomic Health				Appendix: The Assessment of Clinical Evidence Quality The EAG should attempt to formalise a standard approach for the assessment of genomic classifiers by conducting and documenting a more extensive review of publications. For transparency, it is important to document the reason for selecting one set of criteria over another. The current search has excluded key references that propose roadmaps for developing and validating therapeutically relevant genomic classifiers which have been accepted and validated by peer review and are cited in the majority of subsequent review articles (8) (9). The EGAP guidance have been cited by the EAG, but updates to this review have not been included which should be a priority in the evolving genomic landscape (10). The EAG need to detail why this important study has been excluded.	A full literature review of risk of bias assessment tools for prognostic and prediction studies was not within the scope of this work.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Genomic Health				On balance, the proposed unpublished POBAST criteria deviate from the broadly accepted criteria on a few crucial points. a) The inclusion of all available sample blocks from all eligible patients from trials prospective retrospective randomised controlled trials (RCTs) is not a necessary requirement if a threshold is set or the vast majority of samples available are used (8) (9) (11). b) It is also possible to overcome block availability issues by the randomisation of specimens to select a sample of specimens for study that mirror the known important prognostic and predictive factors of the population as a whole (11) c) The limitations of the number of covariates in multivariable analyses are not exceeded, thereby helping to ensure that the results generated are stable and reliable. d) The findings from the retrospective/prospective studies are confirmed by prospective trials (9).	a) The EAG maintain that it is not wrong to point out the limitations of the evidence base with regard to patient spectrum and loss of samples. Whilst this may be an intractable problem, it should be made clear that it is a limitation of the evidence base. b) The EAG are not aware that this was conducted in any of the studies, so do not think this item is relevant to this review c) This item is included in PROBAST, but was not included in the short list of questions selected by the EAG from the tool. This was because we are not concerned with the development of the assays, but in their performance as they are currently marketed or published (in the case of IHC4) d) The relevance of this item is unclear: There is only one prospective trial in the evidence base that has reported all results, and this is for MammaPrint. TAILORx has only reported for the low risk group, and the limitations of this study in that it does not have a comparator arm remain.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Genomic Health				It has been proposed by the EAG that the highest level of evidence can only be generated by a RCT of test-directed therapy versus control in which standard prognostic factors are used to inform treatment decisions. However, results can be particularly confounded and diluted in cases where the standard of care is variable among physicians, making it difficult to detect a difference of test directed treatment versus standard of care (9). The TAILORx trial has been designed to overcome this limitation as all patients are tested but treatment is assigned based on the Recurrence Score® (RS) result (12). At the time of design of the TAILORx trial in the year 2005/6, the panel of highly respected collaborative research groups considered it unethical to use chemotherapy in the group of patients identified as low risk by RS, and conversely withhold chemotherapy in patients with high RS. To further minimise risk of under or over treatment the Recurrence Score cut points for the TAILORx trial arms were based on the Paik et NSABP B-20 results (12) (4). It is for this reason that the EAG statement about TAILORx and Plan B low-Recurrence Score arms are deemed as observational and of low value in terms of assessing clinical utility is not correct. Both trials are well designed prospective randomised controlled trials that have demonstrated that patients with a low RS across multiple clinicopathological risk groups can be safely spared chemotherapy and have excellent 5 year DRFI (4) (5). Both TAILORx and Plan B provide level 1A evidence that Oncotype DX is validated for	The EAG have acknowledged the difficulties with conducting RCTs in this topic: "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." (Page 58 of EAG report) The EAG have stated objectively the available evidence and maintain that it is important to highlight the limitations of the evidence base, as decisions are to be made on the basis of it. "These data support the findings from NSABP 20 and SWOG 8814 validation studies showing that there is a statistical interaction between the RS and the benefit of chemotherapy (13) (14)." The EAG do not agree that these data support or refute the claim of chemotherapy benefit, as even if Oncotype DX only has prognostic value,

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				providing clinical benefit because it enabled the identification of patients whose prognosis was so good with tamoxifen monotherapy that they could be spared the toxicity (potential mortality and secondary malignancies), inconvenience, and expense of chemotherapy (9). These data support the findings from NSABP 20 and SWOG 8814 validation studies showing that there is a statistical interaction between the RS and the benefit of chemotherapy (13) (14). The updated EGAP recommendation statement suggests that potential limitations from the validation studies 'maybe ameliorated' with the availability of more studies 'evaluating health-outcome benefits beyond risk reclassification, such as toxicity of treatment and survival outcomes following testing and differential treatment' (10). Since publication of these recommendations the evidence base for Oncotype DX is further strengthened by long term health outcome benefits from over 50,000 patients who have broadly been treated in line with the Recurrence Score result. In these studies patients had good outcomes based on Recurrence Score-directed adjuvant treatment, despite clear discordance between clinicopathological criteria and Recurrence Score group classification. These studies confirm that in real world clinical practice the current way in which clinicians use Oncotype DX retains good five year BCSM or DRFI rates from test directed treatment decisions (15) (3) (6) (7).	it would be expected to identify a group at low risk of recurrence. Also, both NSABP B20 and SWOG-8814 used the cut-off for RS18 for low risk, not RS11. The EAG have included the study of 50,000 patients in the report and can be found in the section on clinical utility of Oncotype DX (page 116 to 129 of the EAG report). This study remains an observational study, and as such has limitations, as described by the EAG. In agreement with the conclusions drawn by EGAPP, the EAG conclude "it is not possible to conclude whether patient outcomes would be affected by use of the test in a clinical setting." Whilst EGAPP did not include the study of 50,000 patients, they did note that the results from the TAILORx low risk group did not change their conclusions.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				In summary, the Oncotype DX Breast Recurrence Score ® assay is the **only** assay in the current EAG report that has robust evidence of clinical validity and clinical utility. The weight of evidence for Oncotype DX is consistent from all validation studies, prospective randomised controlled studies, and real-world evidence. As a result, the Oncotype DX assay is the **only** assay with sufficient evidence for safe and effective use within the NHS.	
Genomic Health	16	342 -	5.3, 5.4	Executive Summary: Clinical Relevance of the Cost-effectiveness Analysis • Genomic Health would argue that the clinical relevance of the EAG's cost-effectiveness analysis is very limited and that key input assumptions used in the analysis are fundamentally flawed. This is evident in the results and conclusions the analysis yields: o The EAG's analysis supports indiscriminate use of chemotherapy for all patients vs. current practice or use of genomic tests such as Oncotype DX, as tested by Genomic Health by modelling this theoretical scenario, using the EAG's input assumptions.	The EAG note that both the EAG model and the Genomic Health model are based on Ward et al. When the unequivocal errors in the Genomic Health model were corrected by the EAG, and the same assumptions were used (regarding predictive benefit and AEs), the two models produce consistent results. The base case analysis assumes no predictive benefit of chemotherapy; in the absence of robust evidence to support this assumption, this was considered to be an appropriate approach. An analysis assuming a predictive benefit was presented in the sensitivity analyses. The EAG notes the following text from the DG 10 guidance: *"The Committee accepted an analysis*

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				○ This alone shows that the EAG's assumptions and analysis are fundamentally unsound, as it is, in large part, contradictory to the very purpose of gene expression profiling tests in early stage, ER+, HER2- breast cancer; to avoid over-treatment. • Indeed, the analysis is biased in favour of certain tests which classify a larger proportion of patients as high-risk and are assumed to lead to high chemotherapy rates. ○ The Oncotype DX Breast Recurrence Score® test, which has an important net chemotherapy-sparing impact, is penalized in the current analysis, despite having by far the greatest body of supporting clinical validation and utility evidence. ○ The EPClin test, for which the published algorithm places the greatest weighting on lymph node status, is 'rewarded' in the current analysis for classifying the majority of LN+ patients as high-risk, leading to high chemotherapy rates. Based on the approach used in the analysis, the EAG's conclusion regarding the cost-effectiveness of EPClin in the LN+ patient population is fundamentally unsound. ○ It is altogether unclear how the more favourable cost-effectiveness results were arrived at for the	*performed by the External Assessment Group, which showed that the ICER for Oncotype DX (compared with current practice) in this group of patients was £22,600 per QALY gained, assuming prognostic benefits of the test but no predictive effect. The Committee also noted the ICER could be significantly lower if Oncotype DX was shown to predict the benefit of chemotherapy by robust evidence from future research."* Since DG10, no additional robust evidence are available to change this viewpoint. The EAG model does not penalise any patient group: it is the prognostic value in addition to the benefit (and use) of chemotherapy which drives the cost-effectiveness conclusions. A test which identifies a true high-risk (or true low-risk) will fare better than a test which classifies fewer true high-risk and low-risk patients. Excluding the assumption regarding the predictive benefit of chemotherapy, the EAG model operates in the same way as the Genomic Health model. The model is based on the prognostic data available. Regarding the criticism of transparency, we note that data have been redacted (including

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Prosigna test for both LN0 NPI>3.4 and LN+ populations in comparison to other tests. There is a lack of transparency in the assessment. [redacted] Recommendation Summary • Based on the above comments, Genomic Health believes the current cost-effectiveness analysis to be ill-founded and unreliable and would make the following recommendations: o We respectfully recommend that the EAG's assessment should be substantially reworked using clinically relevant assumptions which are supported by published evidence. o Genomic Health strongly encourages NICE to take the decision to postpone the assessment until the upcoming TAILORx trial results can be incorporated to inform the important assumption regarding the differential relative risk reduction of adjuvant chemotherapy [redacted]	the NHS England Access Dataset) – inevitably this will reduce transparency. TAILORx is not yet available. Whilst this study may provide useful information, the cut-offs are different and will not be comparable to the current evidence included in the EAG report or the Genomic Health submission. The model can only be based on the evidence currently available. If the EAG model is invalidated by TAILORx, the Genomic Health model will be as well. The RR for chemotherapy benefit used in the EAG model is very similar to that assumed in the previous EAG model. We are unclear whether the additional analyses presented by the company take into account the major errors identified in the original Genomic Health model. We also note the fact that giving chemotherapy to everyone is a cost-effective option compared with giving chemotherapy to a subset of patients does not invalidate the model as it is plausible that chemotherapy is cost-effective. In addition, the EAG replicated the company's new analysis including the assumption of predictive effect for everyone who receives chemotherapy using the corrected

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				[redacted] Detailed Response in appendix: Clinical Relevance of the Cost-effectiveness Analysis Appendix: Clinical Relevance of the Cost-effectiveness Analysis The purpose of genomic assays is to identify the relatively small subset of ER+, HER2- patients who will benefit from the addition of chemotherapy to their endocrine therapy. The assumption of the current EAG model of a large uniform reduction in chemotherapy benefit runs counter to this purpose by favouring assays that identify a large proportion of patients as high risk. This assumption poses both clinical and mathematical challenges that call into question the validity of a EAGs analysis. The use of a high (24%) uniform reduction in risk of distant recurrence for chemotherapy is a major source of bias in the approach by the EAG, as tests that identify greater numbers of patients at high risk and lead to high chemotherapy rates will inevitably lead to the greatest QALY improvements and will very likely be found to be more cost-effective. This approach does not reflect the clinical reality and in large part runs counter to the purpose of gene expression profiling tests, as defined in NICE's Final Scope for the assessment, treatment guidelines and the tests themselves.	Genomic Health model – this analysis also suggests that chemotherapy remains cost-effective. The EAG model is necessarily complex as it had to include 5 tests and a large number of datasets and scenarios. We have double-programmed the model to ensure its correct implementation – the double-programmed model produced the same results as the full EAG model. We also had an external peer reviewer who managed to scrutinise the model without any communication with the EAG.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				This assumption is also directly contradictory to the published evidence which shows that a relatively small proportion of early stage ER+, HER2- patients derive a benefit from adjuvant chemotherapy treatment i.e. patients with tumours which are sensitive to chemotherapy. So, far from supporting improved care for breast cancer patients, by avoiding the already widely acknowledged over-treatment based on current practice (traditional clinicopathological criteria alone) and allocating chemotherapy only to the minority of patients likely to benefit, the EAG's analysis in fact promotes further increases in over-treatment of breast cancer patients with chemotherapy. Indeed, based on the EAG's analysis, there is a strong argument for indiscriminate use of chemotherapy for all node-negative NPI >3.4 patients. To test this, we used as much of the EAG input data as possible in the Oncotype DX Cost-Effectiveness Model to evaluate the cost-effectiveness of all patients receiving chemotherapy (as the EAG acknowledged that the models produce similar outcomes when the same input data are used). When a uniform 24% distant recurrence risk reduction in the Oncotype DX Cost-Effectiveness Model was applied, treating	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				all patients with chemotherapy was found to be cost-effective versus both Oncotype DX testing and standard care.  Specifically, the following key values were used: • Clinical	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				○ Distant recurrence: The 10-year risk of distant recurrence was based on Dowsett *et al.* 2010 ○ AML: The annual probability of AML was set to 0.025% • Chemotherapy ○ Chemotherapy allocation: In the *All chemo* arm, all NHS England patients in the low-risk (48% of all patients), intermediate-risk (39%) and high-risk (14%) groups were assumed to receive chemotherapy. Consequently, no decision cost of assigning a patient to chemotherapy was used ○ Chemotherapy benefit: In all treatment arms, a 24% reduction in distant recurrence risk was associated with chemotherapy (EBCTCG, 2012) • Costs ○ Chemotherapy: GBP 3,145.19 ○ Distant recurrence: GBP 4,540.65 ○ Local recurrence: GBP 13,911.92 ○ AML: GBP 10,400.34 ○ End-of-life care cost: GBP 0	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				○ Onco*type* DX test cost: [redacted] • Utility ○ Chemotherapy decrement: −0.038 (applied during the first cycle only) ○ End-of-life utility: not included in the EAG model, set to recurrence-free utility of 0.824 so end-of-life /hospice care is not associated with a utility decrement (results changed only marginally if a utility of 0 was used, with treating all patients with chemotherapy continuing to be cost-effectiveness Onco*type* DX and standard care) This finding is an astonishing implication of the modelling choice by the EAG and contradicts the published literature on physiology of breast cancer, and distant recurrence, and on patients' and physicians' treatment preferences and decisions, as well as widespread concerns about overtreatment with chemotherapy (Cardoso et al. 2016; Curigliano et al. 2017; Fey et al. 2014; Marshall et al. 2016; Paik et al. 2006; Sparano et al. 2015; Tao et al. 2015). Results were obtained from setting Oncotype DX Cost-Effectiveness Model parameters as close as possible to those of the EAG model. (Of note, the layout and interface of the EAG model are very convoluted and violate basic user interface design rules, for example with regard to interface	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				simplicity, spatial relations of elements, hierarchical structuring of information and user [HHS, 2017]), making it very difficult for stakeholders to evaluate the model structure and assumptions). This finding alone, demonstrates that the approach used by the EAG for the cost-effectiveness analysis is highly unsound and is not in support of improved care for breast cancer patients or indeed efficient use of NHS resources.  The EAG states that *"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"*. Based on current practice (clinicopathological criteria alone), there is known to be substantial over-treatment with adjuvant chemotherapy. The described reduction in the use of adjuvant chemotherapy following Oncotype DX testing	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				represents patients avoiding unnecessary chemotherapy, and the associated side-effects and the waste of healthcare resources. Based on the EAG's current approach to modelling cost-effectiveness, this very meaningful benefit to patients from Oncotype DX testing is not only under-valued, but entirely penalized.	
Genomic Health	17	348, 350, 361	5.3, 5.3.3	Executive Summary: Choice of Chemotherapy Benefit Assumption for the Cost-effectiveness Analysis • Genomic Health disagrees with the assumption used by the EAG of a uniform 10-year risk reduction due to chemotherapy of 24% to all risk groups. o Not only is this assumption unsupported by published evidence, it leads to a large bias in the analysis. • The EAG indicate in the DAR that this assumption is based on a meta-analysis (EBCTCG, 2012), however, the authors of the meta-analysis discuss limitations that make this study a poor choice on which to base this fundamental assumption: o Few of the patients in the trials had the types of tumours that are within scope for this assessment i.e. early stage screen-detected breast cancers, a relatively high proportion of	We agree that there is uncertainty around the treatment effect for relevant chemotherapy options in terms of distant recurrence. We selected what we believe to be the most appropriate estimate from the updated 2012 EBCTCG meta-analysis. This estimate is very similar to that used in the DG10 model. The company suggests a number of criticisms with the EBCTCG study but does not suggest an unbiased alternative. We have tested alternative values in the sensitivity analyses (RR=0.70 and RR=0.80). These did not change the economic conclusions for Oncotype DX. We note that the thrust of the company's criticism is really about predictive benefit of chemotherapy. We have already presented an analysis which considers this, and have discussed the importance of this assumption

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				which have low disease burden, low proliferative index, and hence a high probability of being endocrine-responsive luminal-A tumours. Therefore, the meta-analyses were not directly informative about the effects of chemotherapy in the relevant patient population for this assessment. ○ The trials included in the meta-analysis do not reflect current clinical practice for several reasons, not least because in half of the studies, no endocrine therapy was given. • Furthermore, the RR assumption used by the EAG is not in line with the meta-analysis results (in fact, may be considered its very opposite): ○ The authors of the meta-analysis clearly state that the benefit of chemotherapy could not be assessed by risk group but that the benefit is almost certainly different between high- and low-risk groups. ○ The authors of the meta-analysis commented that [quote] "…in low-risk ER-positive disease treated with effective endocrine therapy any further risk reduction from adding chemotherapy cannot, in absolute terms, be large, and patients not helped by chemotherapy are harmed by its toxicity".	with respect to the economic conclusions, as highlighted in previous responses.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				• The uniform application of such a high reduction in distant recurrence risk is difficult to understand in light of the published evidence (some of which is included in the EAG model) that suggests that the benefit of chemotherapy differs by risk of distant recurrence (Paik et al. 2006; Stemmer et al. 2017). • Genomic Health believe that the EAG's assumption of universal chemotherapy benefit in the economic model is invalid and ethically questionable. Recommendation Summary • We respectfully recommend that the EAG's assessment should be substantially reworked using clinically relevant assumptions which are supported by published evidence. • Considering the weight of the published evidence, Genomic Health recommend that the EAG should assume a differential relative risk across the Recurrence Score® risk groups, in the base case cost-effectiveness analysis. • Genomic Health strongly encourages NICE to take the decision to postpone the assessment until the upcoming TAILORx study results can be incorporated to inform the important assumption regarding the differential relative risk reduction of adjuvant chemotherapy.	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Detailed Response in appendix: Choice of Chemotherapy Benefit Assumption for the Cost-effectiveness Analysis, and Universal Chemotherapy Benefit in all Patients Appendix: Universal Chemotherapy Benefit in all Patients The EAG have assumed that there is a universal benefit from chemotherapy for all patients which is referenced in the EBCTCG meta-analyses (DAR ref 262 & 274). However, the meta-analysis authors discuss limitations that make this study unsuitable for inclusion in the model (DAR ref 247). a) Trials comparing anthracycline-based (or standard or near-standard CMF) regimens with no chemotherapy do not fully reflect current clinical practice in the UK. b) The median start day of trials was 1986 (interquartile range: 1980−90) and therefore does not reflect a contemporary cohort. c) In half of the studies, no endocrine therapy was given which does not represent current clinical practice, and is not the population of the current decision problem. d) Supportive care during treatment was considered “suboptimal” by EBCTCG authors e) Dosage was likely to be limited due to concerns about toxicity (and, at the time when trials were begun, about chemotherapy in general)	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				f) There were no data on modern markers of tumour biology (neither quantitative immunohistochemical markers, nor multigene assays) and how they may predict prognosis or benefit from treatment in ER positive tumours across risks groups in the trials (DAR ref 247). g) The authors of the meta-analysis specifically discuss that they were unable to assess chemotherapy benefit by risk (DAR ref 247, p. 443). From this meta-analysis, no statement about chemotherapy benefits on distant recurrence for any subgroup should be derived as this issue was not investigated. There is now a significant amount of data that show it is no longer appropriate to extrapolate chemotherapy benefit to all patients when Oncotype DX is used to stratify patients by risk of distant recurrence. Published trial evidence includes early results from the TAILORx trial which show that patients with a Recurrence Score result <11 have favourable 5-year outcomes (4). Two prospective studies have been published that show across different risk groups (as determined by clinical covariates age, tumour grade and size, and node status {LN0-LN3}) the Recurrence Score is able to identify patients who can safely be spared chemotherapy (4) (5). For	

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				example, the Recurrence Score is able to identify patients with high grade and node positive tumours who can safely be spared chemotherapy and have excellent outcomes with endocrine therapy alone (5). These data from well-designed prospective trials (providing level 1a evidence) support the findings from the NSABP -B20 and SWOG 8814 studies that chemotherapy benefit is not universal to all patients (13) (14) (9). It is therefore no longer appropriate to assume that chemotherapy benefit is universal in all early ER positive HER2 negative patients. Additionally, data from real world evidence of the Recurrence Score used to determine treatment in over 50,000 patients clearly shows that patients with low and intermediate scores have little if any benefit from chemotherapy (15) (3) (6) (7). Taken together, these existing data sets provide strong support that the benefit of chemotherapy is not universal across Recurrence Score risk groups. [redacted]	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				[redacted] Numerous studies using the Recurrence Score to stratify patients to neoadjuvant chemotherapy have enabled testing of the chemotherapy response as predicted by the Recurrence Score in tumours pre-surgery. All of these studies have shown that tumours with a high Recurrence Score result respond to chemotherapy enabling breast conserving surgery. Whereas tumours with low Recurrence Scores do not respond to neoadjuvant chemotherapy but respond well to endocrine therapy (17) (18) (19). These neoadjuvant studies using the Recurrence Score clearly	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				show that tumour response to chemotherapy can be predicted by the Recurrence Score. The global clinical consensus is that early invasive ER positive, HER2 negative breast cancers have a very good prognosis, and treatment needs to be deescalated to avoid overtreatment (20) (21) (22) (23). The aim of all the risk tools, algorithms, and genomic assays is to enable the identification of patients who would experience all of the harms of chemotherapy without any of the benefit. Even the ATAC study cited by the EAG shows that there are patients who have very good outcomes with endocrine therapy alone (24). The weight of evidence from Oncotype DX and Mamaprint clearly show that not all patients experience the same benefit from chemotherapy (13) (14) (2) (5) (4) (6) (7) (25). It is therefore both invalid and unethical for the EAG to assume universal chemotherapy benefit for all patients in the economic model. Appendix: Choice of Chemotherapy Benefit Assumption for the Cost-effectiveness Analysis The EAG used a relative 10-year risk of distant recurrence of 0.76 for chemotherapy versus no chemotherapy, based on the estimate derived for any anthracycline-based regimen versus no chemotherapy obtained in a meta-analysis (EBCTCG, 2012) [of note, in the EAG report, the	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				corresponding page in the web appendix of the EBCTCG 2012 study is referenced as page 12 while, in the model, it is referenced as page 13. In fact, it is page 11 by pagination and page 12 by page order in the web appendix.] The meta-analysis was conducted to identify efficacy differences between different polychemotherapy regimens for breast cancer and used individual patient data from various trials to investigate mortality and recurrence outcomes. While the study was well designed and performed, its authors discuss limitations that make this study a less than ideal choice for inclusion in the model. • Trials comparing anthracycline-based (or standard or near-standard CMF) regimens with no chemotherapy do not fully reflect current clinical practice in the UK (EBCTCG, 2012, p. 438): • The median start date of trials was 1986 (interquartile range: 1980−90) • In half of the studies, no endocrine therapy was given • Supportive care during treatment was considered "suboptimal" by EBCTCG authors	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				• Dosage was likely to be limited due to concerns about toxicity (and, at the time when trials were begun, about chemotherapy in general) • While few differences of chemotherapy effect were observed across age, nodal or endocrine status and tumor differentiation or diameter, it should be acknowledged that the effect of chemotherapy on distant recurrence was not investigated by these subgroups. From this meta-analysis, no statement about chemotherapy benefits on distant recurrence for any subgroup should be derived as this issue was not investigated. Note that this argument does not imply that chemotherapy has no distant recurrence benefit or that this benefit is the same/different in different subgroups – rather, no conclusions can be drawn from this meta-analysis. • Relating to the previous point, the authors of the meta-analysis specifically discuss that they were unable to assess chemotherapy benefit by risk (EBCTCG, 2012, p. 443): "Relatively few patients in these trials (and even fewer of those with recurrence) had small, well differentiated tumours. By contrast, widespread mammographic screening finds many breast cancers with low disease burden, low proliferative index, and hence a high probability of being	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				endocrine-responsive luminal-A tumours. The present meta-analyses were not directly informative about the effects of chemotherapy on such low-risk tumours, but in low-risk ER-positive disease treated with effective endocrine therapy any further risk reduction from adding chemotherapy cannot, in absolute terms, be large, and patients not helped by chemotherapy are harmed by its toxicity. This includes not only acute toxicity and leukaemogenicity but also any persistent neurotoxicity and anthracycline cardiotoxicity. Longer follow-up of the trials will help to assess the eventual risks and benefits more reliably." (EBCTCG, 2012) Given that the authors clearly state that the benefit of chemotherapy could not be assessed by risk group but that the benefit is almost certainly different between high- and low-risk groups, applying a uniform risk reduction of 24% to all risk groups is not in line with the meta-analysis results (in fact, may be considered its very opposite). The uniform application of such a high reduction in distant recurrence risk is difficult to understand in light of the published evidence (some of which is included in the EAG model) that suggests that the benefit of chemotherapy differs by risk of distant recurrence (Paik et al. 2006; Stemmer et al. 2017). The TAILORx trial, for example demonstrated a low risk of distant recurrence in patients classified at low risk by Oncotype DX (RS<11), with 99.3% free of distant recurrence	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				at 5 years. In a recent analysis of prospectively registered, real-world data, 5-year Kaplan-Meier estimates for patients with RS<11 and 11≤RS≤25 indicated distant recurrence risks of 1.0% and 1.3%, respectively (Stemmer et al. 2017). These findings are confirmed by classifications from different gene expression tests, including, for example, MINDACT, which was also used in the EAG model (Cardoso et al. 2016). Of note, the EBCTCG authors specifically suggested that large-scale trials such as MINDACT and TAILORx would be necessary and able to evaluate differences in risk reductions by subgroup (EBCTCG, 2012, p. 443): "Certain trials have suggested that in ER-positive disease the levels of expression of various genes (including those related to proliferation) might correlate not only with prognosis but also with chemosensitivity, so they might help to predict benefit, or identify some higher-risk patients who would gain little from chemotherapy. We could not test such hypotheses. Three new trials (MINDACT, TAILORx, RxPONDER) have included more than patients with ER-positive disease and measurements of gene expression profile who have been randomly allocated chemoendocrine therapy versus the same endocrine therapy alone. Their combined results will be able to assess reliably the prognostic relevance of such measurements (and of other measurements, including quantitative immunohistochemistry) and will help to assess	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				any differences in chemotherapy RRs between subgroups." (EBCTCG, 2012)	
Genomic Health	18	350, 358	5.3.3	Executive Summary: Chemotherapy Allocation Assumptions for the Cost-effectiveness Analysis • As acknowledged by the EAG in the current DAR, a robust assessment of decision-impact evidence for the technologies under evaluation has not been conducted to inform chemotherapy allocation assumptions. As a principal driver of the results of the cost-effectiveness analysis, this is highly surprising. • The EAG applied the, seemingly arbitrary, assumption that chemotherapy allocation by test risk group is the same for all 3-level and 2-level tests respectively. It was assumed that the evidence for Oncotype DX from the NHS England Patient Access Scheme can be applied directly to the Prosigna and IHC4 tests. Considering the very different impact of these 3-level tests on treatment decisions and chemotherapy use, as demonstrated by published decision-impact evidence, and that the OPTIMA Prelim study which showed that the tests are not interchangeable, this assumption is fundamentally unsound and non-transparent.	Given time constraints, the EAG could only undertake an assessment of decision impact studies undertaken in the UK and elsewhere in Europe (see EAG report, Section 4.9, pages 284-297). We are unclear what the company's proposed additional assessment should involve. With respect to the model, we could only use those studies which gave proportions of patients receiving chemotherapy in each risk group – this limits the number of available studies which could be considered for inclusion in the model. We undertook sensitivity analyses to explore the impact of using other studies not included in the base case analysis. Clinical expert opinion suggested that each 3-level test would be interpreted in the same way.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				• Whilst UK decision impact study evidence for Oncotype DX shows a significant net chemotherapy-sparing impact from testing (by avoiding unnecessary chemotherapy in the majority of patients unlikely to benefit, whilst identifying the smaller sub-group of patients who are most likely to benefit), the other tests have the opposite impact. • Furthermore, to directly apply the Oncotype DX decision-impact data to other 3-level tests, it must be assumed that clinicians' treatment decisions are equally in line with test risk classification by each of the tests. However, in the largest (European) decision-impact study for Prosigna, for over 1/3 of patients with a Prosigna high-risk score who had a change in treatment recommendation, the change was from CHT to HT. This is not in line with the published decision-impact evidence for Oncotype DX and may indicate a low level of clinician confidence in the actionability of a Prosigna high-risk score. • Prosigna and MammaPrint have no evidence from studies of UK patients and as previously highlighted by NICE and the EAG, treatment practices vary between geographies and so applying evidence from studies in other countries to the use of these tests in the UK would be speculation only. • The EAG concluded that "There was insufficient [decision-impact] data to assess results by LN status	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				[for Endopredict]". Given that the EPClin test incorporates, and places the greatest weighting on, lymph node status, it is even more important that UK decision impact evidence be available by lymph node status to inform the analysis. Recommendation Summary • We recommend that the EAG conduct a full assessment of the published UK decision-impact evidence for the technologies under evaluation and that this be used to inform the cost-effectiveness analysis. Detailed Response in appendix: Chemotherapy Allocation Assumptions for the Cost-effectiveness Analysis Appendix: Chemotherapy Allocation Assumptions for the Cost-effectiveness Analysis A robust assessment of decision impact evidence for all tests under evaluation was not conducted by the EAG. This creates an assumption set that all assays will have the same effect on physician and patient decision making. Published data on both distribution of assay results and decision impact data do not support such an assumption. As a principal driver of the outcome of the cost-effectiveness analysis, this surprising departure from following a robust and rigorous approach, means there is a considerable risk of bias to the	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				analysis and a large degree of doubt as to the credibility of the results and conclusions. The EAG states that "Studies assessing decision impact, analytic validity and HRQoL/anxiety were not quality-assessed due to time constraints". This is concerning as, as acknowledged in the DAR, these considerations have a substantial impact on the conclusions of the assessment. It would seem that the EAG / NICE have prioritised expediency of the assessment over quality. The reasons for this are unclear to Genomic Health, but considering some of the surprising and unsupported conclusions made in the DAR, it is very questionable whether the shortcuts have allowed for a high-quality assessment. Evidence from studies of UK patients of the impact of each test on treatment decision-making is not used in the cost-effectiveness analysis, but rather the impact of all 3-level tests on chemotherapy allocation, is taken from evidence of the Oncotype DX test (data from the NHSE Patient Access Scheme for Oncotype DX). The impact of all 2-level tests on chemotherapy allocation is taken from the Bloomfield et al 2017 study for EndoPredict. The assumption used in the analysis, therefore, is that chemotherapy allocation by test risk group is the same	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				following all 3 and 2-level tests respectively. We would argue that this is not supported by published evidence. Considering the vastly different impact of the three tests on chemotherapy use, as shown in decision-impact studies, it is very surprising that the EAG chose to apply the same assumption to each of the 3-level tests regarding their impact on chemotherapy use by test risk group. Whilst the UK decision impact study evidence for Oncotype DX shows a significant net chemotherapy-sparing impact from testing (by identifying the smaller sub-group of patients who are most likely to derive benefit from chemotherapy, whilst avoiding unnecessary chemotherapy in the larger sub-group not likely to benefit), the other tests have the opposite impact. The UK decision-impact study for IHC4 showed an increase in chemotherapy use of 11% (see below re: incorrect figures in the DAR). There is no UK decision impact evidence for Prosigna. The EU studies show between a 2% reduction and a 9% increase in chemotherapy use. Error Regarding the Decision-impact Study for IHC4: There is an error in Table 94 of the DAR, relating to the UK decision-impact study for IHC4. It is stated that a pre-test chemotherapy recommendation occurred in 45 (36%) patients but this was in fact the cases recorded as 'Discuss chemotherapy'. 'Recommend chemotherapy' was the option	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				recorded for 29 (23%). The decrease of 2% in chemotherapy use is therefore incorrect. There was an overall increase of 11% (23% to 34%). "Discuss chemotherapy" is not the same as a "recommendation" for chemotherapy. If it were, respondents would logically have selected "recommend chemotherapy", given that this was an option https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4522631/ Test Risk Group vs. Clinician Decision Discordance for the Prosigna Test: Furthermore, in the largest (European) decision-impact study for Prosigna, the authors highlighted that "Among the 33 patients with Prosigna high risk, 14 (42.4%) had a change in the treatment recommendation: 9 (64.3%) of them from HT to CHT. Five patients in this high-risk group received only HT after Prosigna". This means for over 1/3 of patients with a Prosigna high-risk score who had a change in treatment recommendation, the change was from CHT to HT. This may indicate a low level of clinician confidence in the actionability of a Prosigna high-risk score.	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Lack of UK Decision-impact Evidence for the Prosigna and Mammaprint Tests: Prosigna and Mammaprint have no evidence from studies of UK patients and as previously highlighted by NICE and the EAG, treatment practices may vary between geographies and so applying evidence from studies in other countries to the use of these tests in the UK would be speculation only. Distribution of Test Risk Groups for Prosigna: The distribution of test risk groups is noticeably different between the two largest (European) decision-impact studies for Prosigna; 51%, 33%, 17% vs. 43%, 35%, 22% low, intermediate and high-risk for Martin et al 2015 and Wuerstlein et al 2016 respectively. This variation in test risk group distribution is also observed across clinical validation studies (ref). The distribution of test risk groups has a large influence on the results of the cost-effectiveness analysis, so the observed variation across studies raises doubts about the reliability of these evidence sources or perhaps the test itself. Decision-impact by Nodal Status for EPClin: The EAG conclude that “There was insufficient [decision-impact] data to assess results by LN status [for Endopredict]”. This fact appears does not seem to be mentioned in the cost-effectiveness analysis conclusions.	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Given that the EPClin test incorporates, and places the greatest weight on, lymph node status in it's published algorithm* for the test score, it is even more important that UK decision impact evidence be available by lymph node status to inform the analysis. * (EPClin = 0.35x Tumour size + 0.65x Node status + 0.28 x EP score) Sestak et al 2016 showed that a large majority (~80%) of lymph node positive patients are found to be high risk by EPClin. It is likely that the baseline treatment would be chemotherapy for many of these patients, which raises significant questions about the utility of EPClin testing in lymph node positive patients. The EAG conclude in the DAR that "Within the LN+ (1-3 nodes) subgroup, the ICER for EPClin versus current practice is expected to be £21,458 per QALY gained". Given the current analysis is biased towards tests which allocate more chemotherapy and the EPClin score is driven most by lymph node status and so the majority of these patients are classified as high risk and assumed to receive chemotherapy, we would argue that it is not surprising that the EAG reached this conclusion regarding the cost-effectiveness of EPClin.	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Genomic Health	19	16,17,20,53,56,57,66,67,70,72,78,242,342-344,349,351	2.3.2, 2.4.1, 2.5,4.2, 4.3.2, 4.8.1,5.3	Executive Summary: Use of a Bespoke TransATAC Dataset for the Cost-effectiveness Analysis • TransATAC is not an appropriate study as the sole foundation of the model, as it is an outdated study that does not represent contemporary treatment or all patients included in the decision problem. • The bespoke TransATAC analysis was not made available for review by stakeholders which prevented a full review of the validity of the approach taken. This again highlights a lack of transparency of this assessment and the repeated deviation from NICE's own policies on transparency of the review process. • The TransATAC bespoke analysis does not meet the PROBAST criteria or the most accepted criteria for assessment of genomic classifiers, as the study is unlikely to include sufficient tumour samples to mirror the distribution of clinical covariates in the parent ATAC trial. The resulting selection bias invalidates the conclusions drawn. • Using TransATAC biases the outcome of the economic model as mRNA extraction for all assays was carried out by Genomic Health in its central laboratory. The mRNA extraction process is an important first step which is at risk of large inter-laboratory variability. Therefore, the findings are not generalisable to the commercial assays assessed	This is an unusual criticism. The Genomic Health model also used the TransATAC data to characterise recurrence risk. We have noted limitations of this trial in the report; however, this was a large UK trial providing a direct comparison between four of the five in-scope tests. We have undertaken some sensitivity analyses using other data sources where applicable. The TransATAC data were held as academic-in-confidence. We do not own the data and have no control over its release.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				and bias the outcome in favour of Prosigna and Endopredict. • The comparison of pure genomic signatures with composite genomic signatures containing clinical covariates in a selected population using NPI (tumour size, grade and nodal status) leads to over fitting of the data (double counting of the significance of the clinical covariates) which biases the outcome to favour the hybrid signatures. Recommendation Summary • The TransATAC data should not be the only data source used in the model. It is important to include more recent data that reflects current treatment and assay performance to increase the credibility of the model. For the Oncotype DX assay the more recent data include TAILORx, Plan B, SEER and Clalit (see also above recommendation for assuming a differential relative risk across the Recurrence Score® risk groups, in the base case cost-effectiveness analysis).	

				Detailed Response in appendix: Use of a Bespoke TransATAC Dataset for the Cost-effectiveness Analysis. Appendix: Use of a Bespoke TransATAC Dataset for the Cost-effectiveness Analysis The mRNA material for all samples used in the TransATAC analysis was extracted by the central lab at Genomic Health in Redwood City (26). The crucial first step of mRNA extraction affects the results of all the multiparameter tests in this study. Therefore, the extrapolation and generalization of the performance of the MammaPrint, Prosigna, EndoPredict, and EPClin tests to their current day commercially available test is invalid. The commercial tests' mRNA is extracted in diverse central and/or local laboratories using different reagents batches with varying quality assurance measures. The original ATAC trial was closed to recruitment in 2006 (27). Over this 11-year period treatment has evolved considerably including advancement in diagnostic techniques, patient classification, surgical techniques, drug therapy use, drug therapy selection, and duration of therapy (28) (20) (21) (23) (29). It is therefore very unlikely for the original cohort of patients in the study to be representative of the outcomes of patients treated in 2017. It is therefore vital that any retrospective data form old randomised trials are complemented by the inclusion of data form more recent trials. For Oncotype this would include TAILORx in which the mean patient follow-up is 8.5 years. TAILORx represents one	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				of the largest clinical studies investigating treatment of early ER positive HER2 negative breast cancer, when it reports it will change the treatment of this population (12) (4). Therefore, any guidelines produced that exclude these pivotal data will be invalidated by the publication of TAILORx. The bespoke retrospective analysis of a subset of tissue blocks from the TransATAC subset of the original ATAC trial cohort fundamentally undermines the validity of conclusions drawn from this analysis; a) The original ATAC analysis was not designed or powered for subset analysis of this kind, which will likely result in small patient numbers in each subset (27). b) It is very unlikely that current sample size of this bespoke analysis after the NPI criteria application contains a sufficient number of the original ATAC trial tumour blocks to ensure the same distribution of clinical covariates as the parent trial. This analysis is not valid and does not meet the criteria of EGAP, Simon, or PROBAST cited by the EAG (DAR ref 28) (10) (9). c) As the sample size has decreased substantially a balanced distribution of patient and tumour characteristics are unlikely between subgroups analysed which will bias outcomes.	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				d) It is likely that there is a selection bias towards large tumours as tumour specimens have been depleted by multiple analysis. e) Restricting the population to UK only samples could cause a selection bias which in turn could lead to a treatment bias and subsequent outcome bias. f) A selection bias is caused by only including patient samples that had the necessary information to perform a retrospective NPI calculation. g) It is not clear why the aromatase inhibitor population have not been included. h) Premenopausal patients are not included in the TransATAC data, these patients experience life changing side effects from chemotherapy treatment. The providers of the bespoke analysis have not been blinded to the outcomes of all prior analysis of the TransATAC data which they have been implicit in producing, leading to potential for bias in the outcome of the bespoke analysis as well as the overall DAR. It would appear that the TransATAC bespoke analysis has been used as a convenience sample instead of exploring the utility of more current and robust clinical trial and registry data. This could have been addressed through sensitivity analysis, at minimum. No sensitivity analyses have been performed to evaluate assay specific risk distribution and	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				patient outcomes from the rich body of evidence for Oncotype DX (4) (6) (7) (5) (3). In addition, the appropriateness of comparing the MINDACT and TransATAC analyses has not been addressed, nor have sensitivity analyses been performed on the use of MINDACT endpoints instead of TransATAC in the model. It is also crucial for the EAG to consider more recent data sets such as Clalit, TAILORx and Plan B, as these study populations are more contemporary and represent current clinical practice (6) (7) (4) (5). The TransATAC study used to estimate test risk classification and distant recurrence probabilities was the derivation study for IHC4/ IHC4+C (30). Therefore, there is potential for the overestimation of prognostic performance for these tests. The Recurrence Score result® reflects pure tumour biology and provides independent information from clinicopathologic characteristics (26), grade and patient age (14) (4) (26). Therefore, the comparison between a pure genomic marker such as the Recurrence Score result and composite markers such as EPClin, Prosigna and IHC4 +C is skewed by the composite markers inclusion of clinicopathologic characteristics, including nodal status, tumour size, and/or grade in results (31). TransATAC only included postmenopausal patients so is not generalisable to the premenopausal population, in whom	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				chemotherapy causes life changing morbidity (early menopause) (30). Clearly the need to identify patients who can be spared chemotherapy is a high priority in this setting. The current analysis discriminates against this group. Use of the likelihood ratio chi-square (LR χ^2) statistic to claim that one assay is more prognostic than another is not clinically meaningful in this setting. The LR χ^2 analysis comparing prognostication of the EPClin, IHC4/IHC4+C, Prosigna and the Recurrence Score ® is invalid because it does not solely assess the biological component of each test, nor does it illustrate whether a test result is predictive of treatment benefit. It simply provides support that all models of recurrence are all a reasonable fit for the data, and are all statistically significantly prognostic (32).	
Genomic Health	20	363 - 372	5.3.3	Executive Summary: Unit Cost and Utility Assumptions for the Cost-effectiveness Analysis • Costs and utilities used in the current EAG model differ markedly from those used in the previous EAG assessment (Ward et al. 2013). The differences predominantly have a negative impact on the results of the analysis for Oncotype DX. They are also largely unexplained, leading to a distinct lack of transparency regarding the assessment.	The utility estimates are very similar to the original EAG analysis as they are based on the same source. We did not include the end of life decrement, but otherwise they are identical. The costs have been updated. The costs for chemotherapy and associated short-term toxicities have been updated using more robust published evidence which is more reflective of chemotherapy regimens used in the UK.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				• Detailed comments are included in the appendix, with key highlights provided below, as follows: o As the OPTIMA prelim data spreadsheet was not shared by EAG, the calculation of resource use and costs cannot be evaluated. A comparison of chemotherapy costs used by EAG vs. costs reported by Stein et al. (2016, Table 15) in the OPTIMA prelim trial, showed discrepancies. Overall, the changes made by EAG substantially reduce the cost associated with chemotherapy but changes compared with neither the previous NICE assessment nor the OPTIMA prelim trial were explained or justified. o Significant changes to chemotherapy regimen use assumptions vs. Ward et al. (2013, p. 106) were unexplained. o Costs of cancer recurrence and long-term toxicities, supposedly sourced from a "bespoke costing study of NHS patients" (Walkington et al. 2012), a conference abstract, supposedly used to source the, cannot be traced to the Walkington et al. 2012 (or Hall et al. 2017) study. o If the correct inflation factor is used, the inflated cost of local recurrence is GBP 15,459.01, approximately GBP 1,500 more than that used in the EAG assessment.	The OPTIMA prelim spreadsheet is not ours to share. The company can contact the study authors for access. All of our assumptions are clearly reported in the EAG report. We have not used the Walkington study directly and are unclear about the basis of this comment. We used the same source for costing distant recurrence as the previous EAG model (Thomas et al, 2009). We agree that the price year should have been 2004 for local recurrence and therefore this cost may be slightly underestimated. Please note that we did undertake sensitivity analyses around doubled & halved recurrence costs – using a higher local recurrence cost as suggested has a negligible impact on the ICERs for Oncotype DX. We used Wolff et al to estimate the probability of AML based on clinical advice. We excluded costs of death as these apply to all patients. The company is mistaken in how the HRQoL impact of chemotherapy is applied. This is applied as a QALY loss (applied in the first

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				○ If the correct index is used, the inflated cost of distant recurrence is GBP 4,514, i.e. lower than reported by EAG. ○ No explanation was offered for the lower 6-month probability of acute myeloid leukemia (AML) assumed in the current EAG assessment vs. Ward et al. 2013. ○ No explanation was offered to explain why neither the disutility nor the cost associated with end-of-life care where used in the current EAG assessment. ○ The paper by Campbell et al. 2011 reported an annual disutility for chemotherapy; it may be appropriate to apply this beyond the 6-month cycle length of the first model cycle (as per the current EAG assessment). Recommendation Summary • Genomic Health requests that the necessary corrections are made and that justification is given for all differences in cost and utility assumptions used in the current vs. previous EAG analyses. Detailed Response in appendix: Unit Cost and Utility Assumptions for the Cost-effectiveness Analysis	cycle); it is not a utility decrement. This reporting error is discussed in previous responses. The same assumption was applied in the previous EAG model.

				Appendix: Unit Cost and Utility Assumptions for the Cost-effectiveness Analysis Tracing costs For the current assessment, EAG used data based on the OPTIMA prelim trial, specifically from a cost publication based on this trial which is accepted for publication in Value and Health (Hall et al. 2017; Stein et al. 2016). In the Hall et al. 2017 study itself, no detailed cost breakdown was provided but the OPTIMA prelim study was referenced instead: "Chemotherapy procurement, delivery, and toxicity costs were taken from the British National Formulary, the NHS Commercial Medicines Unit, and the NHS Reference Costs. The proportions, case mix, and test selection of patients treated with anthracycline plus taxane, anthracycline alone, or taxane alone were modelled directly from the OPTIMA prelim data. Costs of cancer recurrence and long-term toxicities were taken from a bespoke costing study of NHS patients and the published literature" (Hall et al. 2017, p. 3) Of note, the "bespoke costing study of NHS patients" (Walkington et al. 2012), a conference abstract, does not include any costs so cancer recurrence and long-term toxicity cost were either taken from data not published in this abstract (Peter Hall was an author on both the abstract and the Hall et al. 2017 study) or from the published literature.	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				This, however, cannot be traced to the Walkington et al. 2012 or Hall et al. 2017 study. It seems more likely costing in the current EAG assessment was based on 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), the British National formulary (BNF) and NHS Reference Costs 2013/14. Unit costs associated with the management of chemotherapy-related Grade 3/4 toxicity were based on NHS Reference Costs 2013/14. 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." (EAG, 2017, p. 368) Chemotherapy regimens Chemotherapy regimen use was said to be derived from the OPTIMA prelim data and was estimated at 25% for FEC100-T (3+3 cycles), 20% for TC (4 cycles), 45% for FEC75 (6 cycles) and 10% for FEC100-Pw (3+3 cycles) (EAG, 2017, Table 133). Compared with the previous NICE assessment, this is a drastic reduction in patients treated with FEC75, which was the only treatment considered by Ward et al.	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				(2013, p. 106) as it was the most commonly used chemotherapy regimen in ER-positive, node-negative, HER2-negative patients. No explanation is offered for this change in the type of chemotherapy regimen used. Costs of chemotherapy As the OPTIMA prelim data spreadsheet was not shared by EAG, the calculation of resource use and costs cannot be evaluated. However, costs used by EAG can be compared with costs reported by Stein et al. (2016, Table 15) in the OPTIMA prelim trial. A comparison of chemotherapy costs showed discrepancies between chemotherapy costs reported by Stein et al. 2016 and the EAG 2017. While the cost of FEC100-PW (3+3 cycles) was almost the same (the small difference of GBP 0.02 is not explained but probably has little influence), the cost of FEC75 (6 cycles) was higher in the EAG 2017 assessment, by approximately GBP 7. However, substantial cost decreases were observed for FEC100-T (3+3 cycles) and TC (4 cycles) in the EAG assessment versus the Stein et al. 2016 study (Table 2). Of note, cost decreases for FEC100-T and TC far outweigh the cost increase for FEC75. Interestingly, the reason for cost increases is neither explained nor consistent as, for FEC100-T, much lower costs of supportive medication were	

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				assumed by EAG while the discrepancy for TC is due to differences in drug costs. Overall, the changes made by EAG substantially reduce the cost associated with chemotherapy but changes compared with neither the previous NICE assessment nor the OPTIMA prelim trial were explained or justified. Costs of local recurrence In the current EAG report, as in the previous NICE assessment, costs of local recurrence were sourced from Karnon et al. 2007. In this study, costs of local recurrence were reported at GBP 11,701. The year of costs is not stated explicitly but most likely to be 2004 (as 2003 values in the analysis were inflated but 2004 values were not). The previous NICE assessment inflated the cost to 2010 values using PSSRU inflation indices (Curtis et al. 2010), yielding GBP 14,132, while the current EAG assessment inflated costs, from 2006/07 to 2015/16, yielding GBP 13,912 (HCHS index). If the correct inflation factor of 1.32 (index for 2015/16=297, index for 2003/04=224.8) is used, the inflated cost of local recurrence is GBP 15,459.01, approximately GBP 1,500 more than that used in the EAG assessment. Cost of distant recurrence	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Costs of distant recurrence were sourced from Thomas et al. 2009 as in the previous NICE assessment. In the previous NICE assessment, inflated 2010 costs were calculated at GBP 4,082 (Ward et al. 2013). Of note, these costs appear to have been incorrectly inflated in the current EAG assessment as they were inflated from 2008/09 costs, not 2010 costs, to GBP 4,541. If the correct index is used (2009/10 to 2015/16), costs are GBP 4,514, i.e. lower than reported by EAG. Probability of acute myeloid leukemia (AML): In the previous NICE assessment, the 6-monthly probability of AML was 0.046%, based on Praga et al. 2005. In the current EAG assessment, the 6-month probability was 0.02456%, based on a study by Wolff et al. 2014 in stage I−III breast cancer patients. No explanation was offered for this change versus Ward et al. 2013 Disutility and cost of end-of-life care: In the previous NICE assessment, end-of-life care was associated with a disutility and additional cost, based on a study by Campbell et al. 2011 (from which the chemotherapy decrement of −0.038 was derived). Neither the disutility nor the cost associated with end-of-life care where used in the current EAG assessment. No explanation was offered for this change versus Ward et al. 2013	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				Implementation of chemotherapy disutilities in the GHI ODX model: The EAG criticized the GHI ODX model for applying a utility decrement associated with adverse events of adjuvant chemotherapy in every model cycle over the remainder of patients' lifetimes (p. 327). The EAG assessment is correct as the chemotherapy disutility was applied in all model cycles. Applying the disutility only to the first model cycle (i.e. the first 6 months), made little difference to outcomes in the modelling analysis. Of note, the paper by Campbell et al. 2011 reported an annual disutility associated with chemotherapy. It may be appropriate to apply this disutility beyond the 6-month cycle length of the first model cycle. The quality of 'annotation' of the cost-effectiveness analysis section is poor. Critical elements are omitted that prevent quality scientific review and discourse. This group of evaluators seemed to have deviated from BMPs (Best Modelling Practices), which is concerning.	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Genomic Health	21	344	5.3	Executive Summary: Additional Cost-effectiveness Analyses NPI Sub-group Analysis: • By conducting the cost-effectiveness assessment for the broad patient group, LN0 NPI≤3.4, it is likely that a sub-group of patients, for which gene expression profiling testing could be of considerable value, is entirely missed. • At the upper range of NPI scores in this broad patient group, there are likely to be patients for whom there is treatment uncertainty. For example, LN0 patients with 1-2cm grade 2 tumours and other inconclusive clinicopathological factors, corresponding to NPI 3.2 to 3.4. [redacted] Analysis of Added Value of Genomic Signature Components of Tests Incorporating Clinicopathological Variables: • The EPClin and Prosigna tests incorporate prognostic information from clinicopathological variables which are already routinely available to inform risk stratification and treatment selection. It is important that the cost-effectiveness of the genomic	We could only undertake analyses based on the data we had access to. Regarding how the tests would be used, we based this on how the manufacturers intend these to be used.

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				signature components of such tests be isolated in the cost-effectiveness analysis to determine their added-value. Recommendation Summary • Genomic Health would recommend that a cost-effectiveness analysis be conducted for the NPI 3.2 to 3.4 patient sub-population. • Genomic Health would recommend the cost-effectiveness analysis be conducted to identify the additional clinical and economic value added specifically by the genomic signature components of each test under evaluation. Detailed Response in appendix: Additional Cost-effectiveness Analyses Appendix: Additional Cost-effectiveness Analyses The cost-effectiveness analysis was conducted for three broad patient groups; LN0 NPI≤3.4, LN0 NPI>3.4 and LN+. Whilst Genomic Health considers the latter two sub-groups to be reasonable as they reflect patients for whom adjuvant chemotherapy treatment decisions can have considerable uncertainty, we would suggest that modelling LN0 NPI≤3.4 patients as a broad group is less ideal. This group contains patients at the lower range of NPI scores with a very low risk of recurrence, for whom gene expression profiling tests are unlikely to change treatment decisions. However, at the	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				upper range of NPI scores in this group, there may be patients for whom there is treatment uncertainty. For example, LN0 patients with 1-2cm grade 2 tumours and with other inconclusive clinicopathological factors, corresponding to NPI 3.2 to 3.4. [redacted]	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				[redacted] It is important that a group of patients is not missed in this assessment by using, what now can be considered somewhat outdated cut-points based on NPI. The clinicopathological criteria which form the basis of the NPI tool, when used alone without gene expression profiling tests, have been shown to have more limited prognostic value. Significant caution is therefore needed when using tools like NPI to stratify the analysis sub-groups. In clinical practice, such tools may be the most practical way to stratify patients for gene expression profile testing, but the analysis conducted by NICE to inform such clinical eligibility criteria, should not be limited by the outdated tools themselves. Genomic Health would therefore recommend that a sub-analysis be conducted for the NPI 3.2 to 3.4 patient population, as we feel, based on the available evidence, that it is likely that there is value to be gained from gene expression profiling testing in this patient group.	

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Genomic Health	22a	52,53	4.1.4	Executive Summary - Analytic Validity • It is vital that analytic validity is not assumed for any assay as this could have severe consequences for patients whose treatment is planned based on a test that has not been fully validated analytically or otherwise. • The OPTIMA prelim study showed discordance between assay risk stratification of the same patients, it is therefore dangerous to assume analytic validity. • The clinical validity data for each assay is of varying quality and breadth, it is therefore vital to ensure analytic validity is strong. Recommendation Summary • Genomic Health strongly suggest that the EAG should assess analytic validity of all assays under study to ensure they are fit for purpose and can be used safely in the NHS. Detailed Response in appendix: Analytic Validity	All the tests except IHC4 have a CE mark (though for Oncotype DX the CE mark is for the collection kit as the test is performed centrally in the USA, at a lab with Clinical Laboratory Improvement Amendments certification), which is why we conducted a rapid review of analytical validity for IHC4. The difference between risk stratification by the different tests is not necessarily due to assay analytical validity, as each test includes different genes and different algorithms.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
Genomic Health	22b			Appendix: Analytic Validity All assays need to undergo the same assessments to ensure they are fit for purpose and safe and effective to be used in routine clinical practice. It is vital that analytic validity is not assumed for any assay as this could have severe consequences for patients whose treatment is planned based on a test that has not been fully validated analytically or otherwise.	See response to Genomic Health comment 22a
Genomic Health	22c			The importance of conducting full analytic validation can be illustrated using the IHC based assay. For example, there is known methodological variability inherent in IHC techniques; a) Pre-analytic factors such as, time to fixation, different fixatives, and duration of fixation, as well as, analytic factors such as the clone of Ki-67 (two different clones of Ki-67 were used in the IHC4 publication), methods of Ki-67 assessment (two different methods of assessment were used in the IHC4 publication: manual morphometry and digital pathology with image analysis) and Ki-67 cut-points all influence results. b) It is also well documented that the concordance/test results in different laboratories is far from optimal	See response to Genomic Health comment 22a. Many of these issues are addressed in the EAG's addendum to the report on IHC4 analytical validity.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				even for standard analysis such as ER, PR (33) (34) (35) (36). c) It has been suggested that the antibody used for Ki67 analysis in IHC4 (SP6) was not standard in routine practice at the time of study by the International Ki67 in Breast Cancer Working Group (37). In contrast RT-PCR techniques enables the use of normalization to compensate for pre-analytical sources of variability in a way that is not possible with IHC. Automation processes can also minimize operator dependence.	
Genomic Health	22d			Tumour sample preparation is a vital first step influencing test results which needs to be assessed. In the case of Oncotype DX ®, samples with biopsy cavities and extra-tumoural material are manually micro-dissected by board-certified surgical pathologists in order to avoid contamination with non-tumour tissue.	The EAG are unable to comment on the accuracy of this statement.
Genomic Health	22e			The EAG have assumed that all tests are performed in a central lab, whereas in reality some are carried out in local labs. It is therefore vital that all test undergo full analytic validity assessment comparing central lab and local lab	An assessment of analytic validity other than for IHC4 was beyond the scope of this report.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				results where appropriate, as this could lead to variability in results.	
Genomic Health	22f			It seems highly unusual to include a test in which the derivation dataset for the tests provides the foundation of the EAG assessment and model without undertaking a full analytic validity assessment of said test. This approach is at risk of biasing the outcome in favour of the IHC4/IHC4+C assay over all other tests.	The fact that the derivation cohort for IHC4 is used does not appear to have any logical bearing on whether a full review of analytical validity should have been conducted. The EAG had conducted a rapid review based on unbiased, focused searches relating to IHC4 analytical validity as an addendum to the original report.
Genomic Health	22g			The OPTIMA prelim study compared the risk stratification of the same patient samples using four of the tests under review. It is a concern that there was considerable discordance between assays which could have huge implication on subsequent treatment. (EAG DAR ref 75) It is therefore a risk to patients to assume analytic validity for all assays.	This appears to be repetition. Please see response to 22a
Genomic Health	22h			The data currently included on the prognostic performance of IHC4/IHC4+C, Prosigna, Endopredict and Mamaprint are of poor quality, making it difficult to reliable ascertain if these assays are prognostic in the target population. For example, the Prosigna assay was found not to be prognostic in one	The general point made by Genomic Health about the evidence base for the other tests is unfair given the limitations of the evidence base for Oncotype DX (see Table 10 in the EAG report), and how this compares to the quality

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				validation study and not predictive in another, which is a concern if the analytic validity is also unknown (DAR EAG ref 141,140) . It is therefore essential that full analysis of analytic validity is undertaken citing peer reviewed publications assessing if these tests are fit for clinical practice.	assessment for the other tests, which is roughly equivalent. The example given relating to Prosigna is correct for prognosis, though the study in question was at high risk of confounding due to the administration of chemotherapy to all patients, which alter estimates of prognostic performance. It also included 28% of patients with >3 positive lymph nodes, and may therefore have low generalisability. The data relating to prediction of chemotherapy benefit from Liu 2015[3] was not for Prosigna, but for PAM50 and was therefore not included in this review. Prosigna currently do not claim to be able to predict chemotherapy benefit.
Genomic Health	23	66, 67 ,68,71, 72	4.3.2	Executive Summary Inclusion of Studies of a 21 Gene Assay from China • The EAG have extensively cited two studies from China in the clinical validity analysis of Oncotype DX Breast Recurrence Score®. These analyses did not use the Oncotype DX® Recurrence Score assay which is performed by Genomic Health, Inc but instead reference an assays and companies that are not related or affiliated in any way with Genomic Health, Inc or the Recurrence Score® assay (DAR refs 85 and 89).	This is not an oversight. The EAG clearly state that the assays used in these studies are not performed by Genomic Health on page 72 “The three exceptions were the two studies from China where the test was not performed by Genomic Health,[4 5] 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.” Both studies use the RS algorithm.

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

Diagnostics Assessment Report (DAR) - Comments

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				• Therefore, the inclusion of these data by EAG is not valid and any assumptions or conclusions drawn are of no relevance to the performance of the Oncotype DX® assay. • This oversight undermines the confidence in the accuracy of the report and the relevance of the publications cited and the conclusions drawn. Recommendation Summary • Genomic Health suggest that the EAG remove all data cited from DAR refs 85 and 89, and delete all conclusions and assumptions from the report and model. The EAG should rewrite the section of the report that was heavily informed by these two references and resubmit the report for review by all stakeholders. Detailed Response in appendix: Inclusion of Studies of a 21 Gene Assay from China Appendix: Inclusion of Studies of a 21 Gene Assay from China The first analysis by Gong C et al. 2016 uses a constructed gene signature that by coincidence has 21 genes and is referenced to a company called Surexam® which is registered in Guangzhou, China. (DAR ref 85) The second study by Sun B et al. 2011 is another constructed gene	If we were to exclude the studies from China, we would also have had to exclude Paik et al. 2004 on the same grounds that the test was not the same as Oncotype DX as it is currently offered. As we also included studies of Prosigna both as the commercial assay and the ROR-PT research algorithm, and of IHC4 regardless of who conducted the tests, these Chinese 21-gene assays were included for consistency across reviews. Also, it was thought they may be able to provide some insight into how the test might operate in patients with a different ethnicity, though the EAG are reluctant to extrapolate in this way as clinical practice may also affect results in different countries. In fact, the EAG were careful to exclude one of the Chinese studies (Sun et al. 2011)[5] from the summary of results as it was a clear outlier (see footnotes to Table 7 of the EAG report). The other study[4] appeared consistent with other data. As such, the inclusion of these studies has not impacted on the results or conclusions drawn about Oncotype DX. **Gong et al.[4]** state: “among all the HR-positive, lymph node-negative breast cancer patients enrolled in this study, 21-gene assay

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				signature which contains 21 genes by coincidence (DAR ref 89). Neither of these assays are related to the official Oncotype DX ® product in any way, and therefore will not represent the performance of the Oncotype DX assay. The Oncotype DX assay is only performed at the Genomic Health Laboratory in Redwood City California. The Genomic Health laboratory has strict SOPs in place and quality assurance process that ensure the consistent performance of assay. There are numerous unpublished steps that constitute Genomics Health intellectual property on assay processing that cannot be replicated by anyone else. Therefore, the findings of the two studies cited by the EAG are irrelevant and undermine the confidence in the overall report.	(Surexam®, Guangzhou, China) (Paik et al., 2004; Zhang et al., 2015) was performed in 153 cases to generate a 21-gene Recurrence Score (21-gene RS) in the paraffin-embedded tumor tissue samples and the results were compared with those generated with our 10 BCSC-associated miRNA classifier". As Paik 2004 (the derivation study for Oncotype DX) and a study Zhang 2015[6] (a study of RS in Chinese patients) were referenced, the EAG are confident that the assay performed used the Oncotype DX algorithm. **Sun et al. 2011** state: "Currently, RS assay (oncotype DX) is commercially available with Genomic Health (Redwood City, CA) (8). There have a few different characteristic of breast cancer in Chinese compared with other populations, such as more premenopausal, less HR positive patients. The training set of RS assay was from patients in NSABP B-14 and B-20 trials (9,10) which did not include Chinese patients. And the assay cannot be performed in China and it is very expensive, we therefore have sought to develop a low-cost method through some adjustments of experiment processes to assess the predictive value of RS in Chinese patients. In this study, we used QRT-PCR to quantify the expression of 21-gene

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
					and calculate RS in formalin fixed, paraffin-embedded (FFPE) specimens obtained from women with HR positive, LN negative or positive breast cancer conducted at the Department of Breast Cancer at Affiliated Hospital of Academy of Military Medical Science." As RS is defined in this excerpt as referring to Oncotype DX, the EAG are confident that this study used the RS algorithm. The EAG therefore remain confident that the inclusion of these studies was in accordance with the inclusion criteria applied across all tests, and maintain that their inclusion did not impact on the conclusions drawn about Oncotype DX in any detrimental way, as the outlier (Sun et al.)[5] was excluded as an outlier. These studies were not included in the EAG model.
Genomic Health	24	18,262	2.4.1, 4.8.2	Tissue microarray studies • Genomic Health requests that the EAG explains why a tissue microarray (TMA) review, that was not	The EAG included the review of Microarray studies after clarification from NICE. The EAG are clear in our write-up that these studies are of limited value: "These studies differ from studies that used the commercially

Stakeholder	Comment no.	Page no.	Section no.	Comment	EAG response
				included at the initial scoping meeting, has been included in the report. • The tissue microarray analysis section is irrelevant to the analysis of the Oncotype DX Recurrence Score commercial assay. The TMA assay does not follow pre-analytic sample preparation steps fundamental to clinical sample analysis. These steps require analysis of whole sections of patient tumor tissue in order to mitigate the effects of tumor heterogeneity. They do not provide any evidence for the analytic or clinical validity of the assays under review. • The Oncotype DX Recurrence Score® assay follows numerous pre-analytic and quality assurance SOPS that are proprietary and provide enriched tumor tissue devoid of tissue contaminants in order to generate an accurate Recurrence Score® result. Tissue microarrays are insufficient to provide these accurate results (1). Recommendation Summary • The EAG should remove the entire tissue microarray review from the report, along with any reference to it, and conclusion drawn from it, as the whole section is irrelevant.	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." (section 4.8.2, page 265) Results from this section of the report are not drawn upon in our conclusions to any great extent, and are only interpreted with reference to the studies using the commercial tests. Our only statement in the conclusions of the report (page 412) is "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)." This statement does not seem contentious! For these reasons, the EAG make no amendment to the report.

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

Diagnostics Assessment Report (DAR) - Comments

1. Albain K, Barlow W, Shak S, 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(1):55-65.
2. Evaluation of Genomic Applications in P, Prevention Working G. Recommendations from the EGAPP Working Group: does the use of Oncotype DX tumor gene expression profiling to guide treatment decisions improve outcomes in patients with breast cancer? *Genetics in Medicine* 2016;18(8):770-79. doi: https://dx.doi.org/10.1038/gim.2015.173
3. Liu S, Chapman JA, Burnell MJ, 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(2):439-48. doi: https://dx.doi.org/10.1007/s10549-014-3259-1
4. 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
5. Sun B, Zhang F, Wu SK, et al. Gene expression profiling for breast cancer prognosis in Chinese populations. *Breast Journal* 2011;17(2):172-79. doi: https://dx.doi.org/10.1111/j.1524-4741.2010.01049.x
6. 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

Genomic Health

Genomic Health UK, Ltd.
[redacted]
Office: [redacted]
Fax: [redacted]
www.genomichealth.co.uk

12 November 2017

Via e-mail only to: donna.barnes@nice.org.uk, diagnostics@nice.org.uk

Donna Barnes
Project Manager – Diagnostics Assessment Programme
National Institute for Health and Care Excellence
Level 1A, City Tower
Piccadilly Plaza
Manchester M1 4BT

Re: Tumour Profiling Tests to Guide Adjuvant Chemotherapy Decisions in People with Breast Cancer (Update of DG 10) – Comments on the Diagnostic Assessment Report

Dear Donna,

Please find attached Genomic Health's comments on the Diagnostic Assessment Report (DAR).

We submit these comments to avoid missing the opportunity entirely to provide input into the assessment. However, for the reasons stated in Genomic Health's letters dated 3 November 2017, we have major concerns that the assessment procedure is fundamentally flawed and has indeed resulted in a DAR that is unbalanced and, for the reasons we explain in the comments, fails properly to take account of the purpose of tumour profiling tests.

To summarize our concerns:

- Authors of and contributors to the DAR have major conflicts of interest. One of the reference authors of the DAR was previously removed from the NICE Specialist Committee due to an acknowledged conflict of interest.
- The current assessment appears to be biased in favour of certain technologies under assessment.
- Principal data inputs for the cost-effectiveness model, identified in the Appendix to this letter, have not been disclosed to stakeholders, despite repeated requests, making it impossible for an effective and transparent consultation to take place.
- Several crucial assumptions used in the cost-effectiveness analysis are not supported by published evidence and have very limited clinical relevance.
- The time allowed for stakeholders to comment on the DAR was wholly insufficient given the volume and complexity of the report, rendering the consultation ineffective.

Based on the above considerations, we again respectfully request that NICE withdraw the current DAR altogether and commission a new assessment from independent and unbiased experts, which properly addresses the concerns identified above, in keeping with its obligations. To the extent that you are unwilling to agree to this request, we fully reserve our right to take such further action as may be appropriate.

Genomic Health UK, Ltd.
[redacted]
Office: [redacted]
Fax:
www.genomichealth.co.uk

We look forward to your prompt reply regarding this recommendation.

Yours truly

[redacted]

APPENDIX - INPUTS NECESSARY TO UNDERSTAND COST-EFFECTIVENESS MODEL
(LIST FROM TABLE 121 OF THE DAR)

- Risk classification probabilities for Oncotype DX, EPClin, Prosigna, IHC4+C
  - SOURCE: 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 (Oncotype DX, EPClin, Prosigna, IHC4+C)
  - SOURCE: TransATAC bespoke data request.43 Analysed by subgroup (LN0 NPI≤3.4, LN0 NPI>3.4 and LN+ [1-3 nodes])
- Baseline probability of receiving adjuvant chemotherapy (current practice)
  - SOURCE: LN0 NPI≤3.4 subgroup NCRAS bespoke data request
  - SOURCE: LN0 NPI>3.4 subgroup: NHS England Access Scheme dataset
  - SOURCE: LN+ (1-3 nodes) subgroup: NCRAS bespoke data request
- Probability of receiving chemotherapy conditional on results of test (3-level tests – Oncotype DX, IHC4+C and Prosigna)
  - SOURCE: LN0 NPI>3.4 subgroup: NHS England Access Scheme dataset

Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer Addendum: EAG responses to key themes within the Comments on the Diagnostics Consultation Document

As part of the Diagnostic Assessment Programme topic "Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer", following the 1st Diagnostics Appraisal Committee meeting on 30 November 2017, NICE produced a Diagnostics Consultation Document (DCD, dated 10 January 2018).[1] Commentators provided comments on the DCD, and the EAG has responded to these comments in a separate document. This addendum provides responses to key themes within the comments document.

1. Use of TransATAC data in the economic model

1.1. Rationale for using TransATAC data in the EAG health economic model

All studies reporting prognostic ability or prediction of chemotherapy benefit and meeting the inclusion criteria were included in the clinical review. The rationale for using the TransATAC data in the EAG model was that it could be restricted to the population in the NICE scope (ER+ HER2- 0-3 positive nodes) and it was possible to split the node-negative patients into clinically low-risk and clinically intermediate-risk (according to NPI score above or below 3.4).

1.2. The TransATAC analysis is unreported and has not been subjected to scientific peer review

Several analyses of TransATAC focussing on different tumour profiling tests have been published in peer-reviewed journals. On behalf of the EAG, the TransATAC authors produced a bespoke analysis[2] which covered four of the five tests included in the DAR (Oncotype DX, EndoPredict, Prosigna and IHC4+C) and which was restricted to the relevant population as above.

Subsequent to the publication of the EAG report, the TransATAC authors have published a pre-planned analysis of these data in a peer-reviewed journal (Sestak *et al.,* 2018[3]).

Table **1** presents some key data from the bespoke analysis for the EAG[2] alongside the data from Sestak *et al*., 2018.[3] Whilst there are some small differences, these data are largely consistent. It is not possible to use the newly-published data[3] in our model since LN0 patients are not stratified into clinically low-risk and clinically intermediate-risk, and hazard ratios (HRs) are reported for a 1 standard deviation (1SD) change rather than between risk groups.

Table 1: A comparison of key analyses reported in the data request analysis[2] and in Sestak 2018[3]

Test	LN0 HR (95% CI) for 1SD 10 year		LN1-3 HR (95% CI) for 1SD 10 year		ΔLR-χ2 to CTS 10 year		
	Data request[2]	**Sestak 2018[3]**	**Data request[2]**	**Sestak 2018[3]**	**Data request[2] LN0**	**Data request LN1-3**	**Sestak 2018[3] LN0-3**
Oncotype DX	1.67 (1.39-2.01)	1.69 (1.40-2.03)	1.42 (1.05-1.91)	1.39 (1.05-1.85)	22.78 *p*<0.0001	4.75 *p*=0.023	15.2
IHC4+C	2.56 (1.98-3.33)	NR	1.83 (1.31-2.56)	NR	48.55 *p*<0.0001	12.60 *p*<0.001	NR
IHC4	NR	1.95 (1.55-2.45)	NR	1.33 (0.99-1.78)	NR	NR	20.1
Prosigna	2.58 (1.97-3.38)	2.56 (1.96-3.35)	1.59 (1.16-2.17)	1.58 (1.16-2.15)	50.77 *p*<0.0001	8.51 *p*=0.004	26.3
EPClin	2.34 (1.82-3.02)	2.14 (1.71-2.68)	1.84 (1.34-2.53)	1.69 (1.29-2.22)	40.60 *p*<0.0001	12.91 *p*<0.001	24.4

1.3 Patient numbers per subgroup are small

The number of patients per subgroup were: at least 410 for LN0 NPI<3.4 (more for some tests), at least 253 for LN0 NPI>3.4, and at least 192 for LN1-3. The EAG do not consider the subgroups to be unreliably small.

1.4. Overlapping confidence intervals for recurrence rates between risk groups and between tests

The EAG agrees that there is some overlap between confidence intervals. However, this does not prevent the data from being useable. The point estimates for recurrence per test risk group (for LN0 and LN+ patients) are consistent with estimates from other studies (see point 2 of this addendum, distant recurrence rates by risk classification). The EAG's probabilistic sensitivity analysis fully characterises the uncertainty surrounding these estimates.

1.5. Bias in the patient spectrum due to exclusion of small tumours with insufficient tissue

The EAG report noted this limitation. This is a limitation of most analyses using stored tumour samples and is not limited to TransATAC. A comparison of some basic population-level statistics between the MINDACT trial and the TransATAC data population was provided for the previous round of comments on the DAR, and no major differences were observed.

1.6. TransATAC includes postmenopausal women who were not suitable for chemotherapy

TransATAC selected patients who had not received chemotherapy in order to assess prognostic ability of tumour profiling tests, which required calculation of distant recurrence rates in the absence of

chemotherapy. The EAG report noted this limitation. Many other prognostic studies included in the systematic review also included patients receiving no chemotherapy, to allow a consistent assessment of prognostic ability. TransATAC does appear to include some patients who would be currently indicated for chemotherapy in the UK (e.g. LN>3).

2. Distant recurrence rates by risk classification

2.4. Consistency of Oncotype 10-yr outcomes across re-analyses of RCTs included in the review

Table 2 shows distant recurrence-free rates at 10 years across re-analyses of RCTs with endocrine monotherapy. Distant recurrence-free rates at 10 years in LN0 Oncotype DX low-risk patients (not subgrouped by clinical risk) are consistent across TransATAC publications (94.9% in the bespoke analysis;[2] 94.1% in the Sestak 2016 SABCS presentation;[4] 96% in Dowsett et al. 2010,[5] the latter being measured at 9 years rather than 10 years). These rates are also consistent with those from other studies: B14[6] (93.2%) and B20[7] (96.8%), for patients in the no-chemotherapy arms. Outcomes for other risk groups were also consistent across studies (Table 2).

Table 2: 10-year distant recurrence for Oncotype DX (RCT re-analyses; endocrine monotherapy)

Nodal status	**Oncotype DX risk group**	**Percent of patients distant recurrence-free at 10 years (95% CI)**				
		TransATAC data request[2]	**TransATAC (Sestak 2016 SABCS[4])**	**TransATAC (Dowsett 2010;[5] 9yr recurrence)**	**B14 (Paik 2004,[6] Tang 2011a[8])**	**B20 (Paik 2006[7])**
LN0	ODX low	94.9	94.1	96 (93 to 97)	93.2 (90.4, 96.0)	96.8 (93.7, 99.9)
LN0	ODX int	87.7	83.3	88 (82 to 92)	85.7 (79.7, 91.7)	90.9 (82.5, 99.4)
LN0	ODX high	77.2	72.8	75 (66 to 83)	69.5 (62.6, 76.4)	60.5 (46.2, 74.8)
		LN1-3 only	**Incl LN4+**	**Incl LN4+**		
LN+	ODX low	81.8 (72.7-88.0)	73.8	83 (76 to 88)		
LN+	ODX int	75.4 (63.0-84.2)	65.3	72 (61 to 80)		
LN+	ODX high	68.6 (44.7-83.9)	51.2	51 (36 to 65)		

Data from Table 12 in EAG report. No additional RCTs of endocrine monotherapy reported distant recurrence in LN+ patients.

2.5. Consistency of outcomes across studies: Oncotype low-risk patients subgrouped by clinical risk

There are several comments referring to the 10-year distant recurrence rate of 15% in the LN0 Oncotype DX low-risk group in the TransATAC analysis (i.e. 85.4% distant recurrence-free). It is vital to point out that this does not represent the Oncotype DX low-risk group as a whole (see response 2.1 and Table 2 for the whole Oncotype DX low risk group). Instead, it represents the LN0 NPI>3.4 subgroup (i.e. LN0 and clinically intermediate-risk).

Table 3 shows distant recurrence-free rates at 10 years for LN0 patients, subgrouped by clinical risk. For TransATAC, these were subgrouped according to NPI score (which includes nodal status, tumour grade and tumour size). For the Oncotype DX low-risk, clinically intermediate subgroup (NPI>3.4), the distant recurrence-free rate at 10 years was 85.4%. We could not identify any other studies subgrouping by NPI score. However, the B14 analysis subgrouped by various other measures of clinical risk: tumour size, grade and Adjuvant! Online (AOL).[6, 8] B14 results appeared consistent with TransATAC, with similar 10-year distant recurrence-free rates for Oncotype DX low-risk, clinically intermediate-risk patients (tumour >4cm, 87%; grade poor-differentiated, 86%; AOL intermediate-risk, 86.6%, AOL high-risk, 95.0%). Outcomes for other Oncotype DX risk groups sub-grouped by clinical status were also consistent across studies (Table 3).

Table 3: 10-year distant recurrence for Oncotype DX by clinical risk group (RCT re-analyses)

Oncotype DX risk group	**Clinical risk**	**TransATAC data request[2] LN0**		**B14 (Paik 2004,[6] Tang 2011a[8]) LN0**	
		Definition of clinical risk	**% DRF at 10yr (95% CI)**	**Definition of clinical risk**	**% DRF at 10yr**
ODX low	Clinical low	NPI≤3.4	98.3 (96.3-99.2)	Tumour <1cm	100
				Grade well-diff	96
				AOL low-risk	94.4
	Clinical intermediate	NPI>3.4	85.4 (77.6-90.7)	Tumour >4cm	87
				Grade poor-diff	86
				AOL int-risk	86.6
				AOL high-risk	95.0
ODX int	Clinical low	NPI≤3.4	93.1 (86.7-96.5)	Tumour <1cm	87
				Grade well-diff	91
				AOL low-risk	90.0
	Clinical intermediate	NPI>3.4	79.8 (69.4-86.9)	Tumour >4cm	88
				Grade poor-diff	76
				AOL int-risk	86.1
				AOL high-risk	76.6
ODX high	Clinical low	NPI≤3.4	83.8 (57.7-94.5)	Tumour <1cm	83
				Grade well-diff	69
				AOL low-risk	81.8
	Clinical intermediate	NPI>3.4	74.9 (59.8-85.1)	Tumour >4cm	47
				Grade poor-diff	60
				AOL int-risk	56.8
				AOL high-risk	68.5

TransATAC data from Table 124 in EAG report. B14 data by size/grade estimated from graphs in Paik 2004.[6] DRF, distant recurrence-free

2.6. Consistency of Oncotype 5yr outcomes between TransATAC and observational studies

There were several comments suggesting that the TransATAC recurrence rates used in the EAG model were less favourable than the recurrence rates from observational studies of Oncotype DX. Table 4 shows outcomes at 5 years for TransATAC and for observational studies of Oncotype DX (no 5-year data were available for other reanalyses of RCTs). Outcomes at 5 years were similar between

TransATAC and observational studies of Oncotype DX. It should be noted that some patients in the observational studies received chemotherapy; this may have improved observed outcomes.

The differences between the TransATAC recurrence rates used in the EAG model and the recurrence rates reported in observational studies appear to be due to: (a) the model data being stratified by clinical risk (those with NPI >3.4 had less favourable outcomes), and (b) the observational data being reported at a 5-year rather than 10-year follow-up.

Table 4: 5-year outcomes for Oncotype DX (RCTs and observational studies; some chemotherapy use)

Oncotype DX risk group			**LN0-mic**						**LN0-3, clin high risk**
	TransATAC data request[2] (LN0) N=829		**CT use in obs. studies**	TAILORx (Sparano 2015[9]) N=1626	MD Anderson (Le Du 2015[10]) N=1030	Clalit (Stemmer 2016[11]) N=1594	Memorial Sloan Kettering (Wen 2017[12]) N=1406	SEER (Petkov 2016,[13] Roberts 2016[14]) N=38,568	WSG PlanB (Nitz 2017 [15-17]) N=2646
	CT use	**DRFI 5yr**		**DRFS 5yr**	**DRFS 5yr**	**DRFI 5yr**	**DRFI 5yr**	**BCSS 5yr**	**IDFS 5yr**
ODX very low (<11/12)	None		0%	99.3 (98.7, 99.6)			99.9%	99.6 (99.4, 99.8)	94.2 (91.2, 97.3)
ODX low (RS<18)	None	99.1	1-12%	-	95.9 (93.0, 97.6)	99.5 (98.4, 99.8)	99.6%	99.6 (99.4, 99.7)	
ODX int (RS 18-30)	None	94.0	26-43%		-	98.8 (97.2, 99.4)		98.6 (98.3, 98.9)	94.3 (92.8, 95.8) **(RS 12-25)**
ODX high (RS >30)	None	88.9	89-90%		76.4 (59.2, 87.1)	93.1 (87.1, 96.3)		95.6 (94.4, 96.6)	84.2 (80.6, 87.8) **(RS ≥25)**

Data from Table 26 in EAG report. CT, chemotherapy; DRFS, distant recurrence-free survival; DRFI, distant recurrence-free interval; IDFS, invasive disease-free survival; BCSS, breast cancer-specific survival

3. Ability of Oncotype DX to predict differential relative benefit from adjuvant chemotherapy

3.1. Clarification on the difference between absolute and relative benefit

A key issue for clinical and cost-effectiveness of tumour profiling tests is whether the **relative** benefit from chemotherapy differs between test risk groups. It is important to note that this relates to relative rather than absolute benefit. We concluded in our EAG report that all the tests have additional prognostic ability over clinicopathological factors, at least in LN0 patients, i.e. that recurrence rates are higher in higher-risk groups. This means that the **absolute** benefit of chemotherapy is also higher in higher-risk groups. However, this does not necessarily mean that the relative benefit differs between groups.

As an example, if distant recurrence rates in the test high-risk group were 30% without chemotherapy and 20% with chemotherapy, the absolute benefit of chemotherapy would be 10%. Likewise, if distant recurrence rates in the test low-risk group were 3% without chemotherapy and 2% with chemotherapy, the absolute benefit of chemotherapy would be 1% (i.e. much smaller). However, the relative benefit would be the same in both groups (relative risk of 0.67, i.e. chemotherapy reduces recurrence by one-third).

3.2. Summary of data on the ability of Oncotype DX to predict benefit from chemotherapy

Data on ability of Oncotype DX to predict differential relative chemotherapy benefit is summarised in this section. Limitations of the chemotherapy benefit studies are summarised in Section 3.3. The EAG's overall view on chemotherapy benefit data is provided in Section 3.4.

Data on the ability of Oncotype DX to predict chemotherapy benefit comes mainly from two re-analyses of RCTs: one in LN0 patients (NSABP-B20; Paik 2006,[7] Tang 2011a[8]) and one in LN+ (SWOG-8814, Albain 2010[7, 8, 18]). In both, patients were randomised to endocrine monotherapy or endocrine plus chemotherapy. Summary results are provided in Table 5.

Relative and absolute benefit per risk group (adjusted and unadjusted): Both studies showed that unadjusted HRs for the effect of chemotherapy vs. no chemotherapy on survival and recurrence outcomes were most favourable in the higher-risk groups. HRs were generally statistically significant in high-risk groups but not in low- or intermediate-risk (). In the B20[7, 8] study (LN0), unadjusted HRs for 10-year distant recurrence-free interval (DRFI) in the low, intermediate and high-risk groups were 1.31, 0.61 and 0.26. HRs restricted to HER2- patients (adjusted and unadjusted) showed the same pattern (Table 5; not reported in journal article - provided via personal communication with Dr Tang via NICE). However, it is interesting to note that absolute differences (for chemotherapy vs. no chemotherapy) were very small in the low and intermediate-risk groups (1.1% and 1.8%, both favouring no chemotherapy), though greater in the high-risk group (27.6% favouring chemotherapy).

In SWOG-8814 (LN+),[18] DRFI was not reported. HRs for 10-year disease-free survival (DFS) for low, intermediate and high-risk groups, adjusted for number of positive nodes, were 1.02, 0.72 and 0.59.

Interaction tests (adjusted and unadjusted): Interaction tests indicate whether the difference in chemotherapy effect for a change in RS score is statistically significant. In B20 (LN0), the unadjusted interaction test for 10-year DRFI (for continuous RS score by chemotherapy) was reported as p=0.031[8] or p=0.038,[7] indicating a statistically significant difference in chemotherapy benefit as RS changes (Table 5). Interaction tests adjusted for clinicopathological factors were borderline significant for the full cohort (p=0.035, p=0.039 and p=0.068; difference due to method of assessing grade), while for the HER2- subgroup they were statistically significant (p=0.007, p=0.018 and p=0.022). The EAG report stated that it was unclear whether all factors were adjusted for simultaneously in B20; however, personal communication with the biostatistician (via NICE) confirms that this was the case.

In SWOG-8814 (LN+), the interaction test for 10-year DFS (for continuous RS score by chemotherapy; adjusted for number of nodes) was p=0.053 for all years and p=0.029 for years 0-5. Interaction tests adjusted individually for each of age, ethnicity, tumour size, grade, PR, P53 and HER2 were also statistically significant (p=not reported). Initially, the EAG interpreted this as a model including all clinicopathological variables; however, clarification from the authors in a personal communication to the EAG stated that each variable was included in a separate model. However, an interaction test adjusted for Allred-scored ER status was not significant (p=0.15). No interaction test was available that included all clinicopathological variables together.

Observational studies: Three observational studies had some data on chemotherapy benefit: two studies in patients with LN0 disease (MD Anderson[10, 19] and SEER[14, 20]) and one study in patients with LN+ disease (Clalit Health[21, 22]). Evidence was mixed and at high risk from confounding, since receipt of chemotherapy was influenced by Oncotype DX score, and patients receiving chemotherapy were likely to be at higher risk. Only one study (SEER) reported an interaction test; this was statistically significant (p=0.03), but only adjusted for grade, tumour size, age and race (omitting ER and PR).[13, 14] The other two studies only reported HRs for chemotherapy versus no chemotherapy in intermediate (MD Anderson and Clalit Health)[10, 11, 19, 21, 22] and high-risk patients (MD Anderson),[10, 19] and these were statistically non-significant, even after adjustment for confounders in one study.[10, 19]

Table 5: Prediction of chemotherapy benefit by Oncotype DX – Reanalyses of RCT data

Study	Outcome	% recurrence-free; absolute benefit			Hazard ratio for CT vs no CT (95% CI)			Interaction tests	Adjusted interaction tests
		Low	Intermediate	High	Low	Intermediate	High		
NSABP-B20 **LN0** ER+ N=651 Paik 2006[7] Tang 2011a[8] Personal comm.	**DRFI 10yr Unadjusted** **HER2- Unadjusted** **HER2- Adjusted[a]**	CT: 95.6% No CT: 96.8% Abs diff -1.1%	CT: 89.1% No CT: 90.9% Abs diff -1.8%	CT: 88.1% No CT: 60.5% Abs diff 27.6%	1.31 (0.46, 3.78), *p*=0.61 1.21 (0.41, 3.55), *p*=0.73 1.18 (0.40, 3.53), *p*=0.76[a]	0.61 (0.24, 1.59), *p*=0.39 0.78 (0.29, 2.11), *p*=0.62 0.67 (0.24, 1.87), *p*=0.44[a]	**0.26 (0.13, 0.53), p<0.001** **0.21 (0.08, 0.53), p<0.001** **0.20 (0.07, 0.52), p=0.001[a]**	Interaction (continuous RS) ***p*=0.031** or ***p*=0.038** (Tang 2011a[8] and Paik 2006[7])	Interaction[a] (continuous RS) adjusted for age, tumour size, grade, ER and PR: - All pts: ***p*=0.035, 0.039**, 0.068[b] - HER2-: ***p*=0.007, 0.018, 0.022** [b]
	DFS 10yr				0.91 (0.57, 1.45)	0.79 (0.43, 1.47)	**0.41 (0.23, 0.71)**	p=0.082	
	OS 10yr				1.37 (0.63, 3.01)	0.94 (0.4, 2.25)	**0.31 (0.16, 0.60)**	**p=0.011**	
SWOG-8814 **LN+** HR+ HER2+/- N=367 Albain 2010[18]	**DFS 10yr**	CT: 64% No CT: 60% Abs diff 4%		CT: 55% No CT: 43% Abs diff 12%	1·02 (0·54, 1·93); *p*=0·97 [c]	0·72 (0·39, 1·31); *p*=0·48[c]	0·59 (0·35, 1·01); ***p*=0·033**[c]		- Interaction (continuous RS) adjusted for positive nodes: All years: *p*=0.053 [c] 0-5 years: ***p*=0.029** [c] 5-10 years: *p*=0.58 [c] - Interaction (continuous RS) adjusted for each of age, ethnicity, size, grade, PR, P53, HER2: significant (p=NR). - Interaction adjusted for Allred-scored ER: p=0·15
	BCSS 10yr			CT: 73% No CT: 54% Abs diff 19%	*p*=0.56	*p*=0.89	***p*=0.033** [c]		
	OS 10yr			CT: 68% No CT: 51% Abs diff 17%	1·18 (0·55, 2·54, p=0·68)[c] p=0.63 log-rank	0·84 (0·40, 1·78, p=0·65) [c] p=0.85 log-rank	0·56 (0·31, 1·02, *p*=0·057) [c] ***p*=0.027** log-rank		Interaction (continuous RS) [c] All yrs: ***p*=0.026** 0-5 yrs: ***p*=0.016** 5-10 yrs: *p*=0.87

Data from Table 22 in EAG report. [a]Adjusted for age, tumour size, grade, ER and PR.[b]p-values correspond to analyses using different assessments of tumour grade. [C]Adjusted for number of positive nodes (1 to 3 vs. 4 or more)

3.3. Key limitations of studies assessing chemotherapy benefit

a) *Lack of data on chemotherapy benefit for the clinically intermediate-risk group:* NICE currently recommends Oncotype DX only for patients who are clinically intermediate-risk, for whom the chemotherapy decision is uncertain. This is a key subgroup for the economic modelling (defined as NPI>3.4). There are no data on the chemotherapy effect in patients who are Oncotype DX low-risk but clinically intermediate-risk. It is plausible that even if there is no chemotherapy benefit for clinically-low Oncotype DX-low patients, there could be benefit for clinically-intermediate (NPI>3.4) Oncotype DX-low patients.

b) *Statistical significance of interaction tests:* Most unadjusted interaction tests were statistically significant (Table 5). In terms of adjusted interaction tests, these were significant or borderline significant in B20 (LN0); and more clearly significant for the new HER2- subgroup (personal communication via NICE). One of the key concerns in the EAG report was that it was unclear whether all factors were adjusted for simultaneously in B20; however, personal communication with the biostatistician confirms that this was the case. This, along with the new HER2- subgroup analysis, provides stronger evidence for an interaction than presented in the EAG report.

However, in SWOG-8814 (LN+), it is now apparent after clarification from the lead biostatistician that interaction tests were adjusted for each clinicopathological factor individually (not all together, as initially thought by the EAG). All were individually significant except for the interaction test adjusted for Allred-scored ER status (p=0.15). As such, it remains unclear whether the interaction test would remain significant after adjustment for all relevant clinicopathological variables.

This also raises an interesting point as to whether results should be adjusted for ER status. On the one hand, test results should be adjusted to account for the effect of clinicopathological factors for which data are available in routine practice. On the other hand, it is not clear to what extent quantitative ER results are routinely available in UK practice, or their level of analytic validity; the SWOG-8814 author noted in his personal communication that performance of the Allred score is subject to some variability between pathologists. The author further stated that *"It is certainly possible that by including other measures of HER2, ER degree, Ki-67, grade, nodal size etc that one could make the interaction nonsignificant. However … you do get the benefit of most of those in a single well controlled measure (RS) rather than relying on separate assays for each with high known variability."* In other words, the benefit of Oncotype DX could be more accurate prognosis, rather than the prediction of chemotherapy benefit.

c) *Possible overestimation of chemotherapy benefit due to B20 being derivation study:* Patients from the no-chemotherapy arm of B20 were used to derive the Oncotype DX score. Therefore, Oncotype DX

may be overfitted in this study arm (i.e. recurrence rates may be artificially low in Oncotype low-risk patients and artificially high in Oncotype DX high-risk patients). This could lead to an overestimate of chemotherapy benefit since the chemotherapy arm was not used in derivation, therefore recurrence rates in this arm may show less separation between the low and high risk groups.

B14 (Paik 2004)[6] is a validation study of Oncotype DX (tamoxifen only; no chemotherapy arm). Comment 162 notes that the prognostic effect of Oncotype DX in the no-chemotherapy arm of B20 is greater than that in B14. As shown in Table 6, in the absence of chemotherapy, there is greater separation in B20 than B14; in other words, low-risk patients have a better 10-year recurrence-free rate in B20 (96.8%) than B14 (93.2%), while high-risk patients have a worse recurrence-free rate in B20 (60.5%) than B14 (69.5%).

In terms of prediction of chemotherapy benefit, B20 has a worse recurrence-free rate in the chemotherapy arm in low-risk patients (95.6% with chemotherapy vs. 96.8% without). This is counter-intuitive, and gives a corresponding HR greater than 1 (HR=1.31). However, comparing the chemotherapy arm of B20 (95.6% recurrence-free) with the no-chemotherapy arm of B14 (93.2% recurrence-free) indicates a small benefit in low-risk patients, though this breaks randomisation and may be affected by population differences between trials.

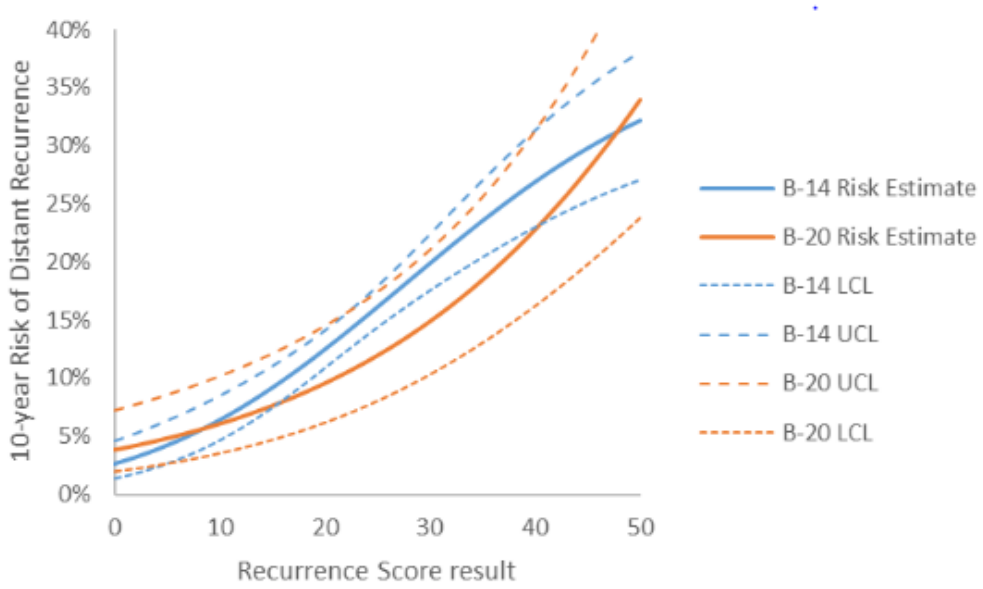
Additional data (personal communication with Dr Tang) compares the recurrence rates for a range of Oncotype DX scores in B14 and B20 (Figure 1). This analysis (which uses continuous Oncotype DX scores) is interpreted by Dr Tang as suggesting that the range of distant recurrence risk estimates, and slopes, are very similar between B20 and B14. However, the EAG still note that recurrence rates per risk group do appear to show greater separation in B20 than B14 (Table 6).

Table 6: Comparison of Oncotype prognostic ability in B14 and B20

Oncotype risk group	NSABP-B14 (Paik 2004)[6]		NSABP-B20 (Paik 2006[7])			
	Tamoxifen		Tamoxifen		Tamoxifen + chemotherapy	
	% patients per risk group (n)	% recurrence-free 10yr	% patients per risk group (n)	% recurrence-free 10yr	% patients per risk group (n)	% recurrence-free 10yr
Low	51% (388)	93.2%	60% (135)	96.8%	51% (218)	95.6%
Intermediate	22% (149)	85.7%	20% (45)	90.9%	21% (89)	89.1%
High	27% (181)	69.5%	21% (47)	60.5%	28% (117)	88.1%

Data from Table 12 in EAG report (also comment 161a in Comments on Diagnostics Consultation Document)

Figure 1: 10yr risk of distant recurrence in tamoxifen-alone groups: B20 and B14 (personal communication with Dr Tang, B20 study)



d) *Clinical relevance of chemotherapy benefit is unclear for the Oncotype DX intermediate-risk group:* Hazard ratios for chemotherapy benefit are available for this group, but it is unclear how they should be interpreted in clinical practice, i.e., would patients be treated, not treated, or would other clinicopathological variables be taken into consideration when making a decision?

e) *The number of events per subgroup is relatively low*, particularly for the B20 study (Table 7). Confidence intervals for the hazard ratios in low-risk and intermediate-risk groups are very wide in both B20 and SWOG-8814 (Table 5).

Table 7: Event rates for B14, B20 and SWOG-8814

Oncotype risk group	**Treatment**	**N events / N patients**		
		B14 (Paik 2004)[6] LN0	**B20 (Paik 2006)[7] LN0**	**SWOG-8814 (Albain 2010),[18] LN+**
Low	Chemo	-	10 / 218	26 / 91
Low	No chemo	28 / 338	5 / 135	15 / 55
Intermediate	Chemo	-	9 / 89	20 / 57
Intermediate	No chemo	25 / 149	7 / 45	22 / 46
High	Chemo	-	13 / 117	28 / 71
High	No chemo	56 / 181	18 / 47	26 / 47

3.4. EAG summary of evidence and limitations for prediction of chemotherapy benefit by Oncotype

Both B20 (LN0) and SWOG-8814 (LN+) showed that hazard ratios for chemotherapy vs. no chemotherapy were most favourable in the higher-risk groups, and were generally statistically significant in high-risk groups but not in low- or intermediate-risk groups. Unadjusted interaction tests

were statistically significant. Adjusted interaction tests were borderline significant in B20 (significant in HER2- patients), while in SWOG-8814 they were significant when adjusted for some clinicopathological variables individually, but not when adjusting for ER determined by Allred status.

Considering the limitations discussed above, the EAG considers that there remains uncertainty surrounding whether Oncotype DX is associated with a predictive benefit of chemotherapy (i.e. a difference in relative effect by genomic risk group), and if so, that there is uncertainty in the likely magnitude of this predictive effect within the clinical subgroups considered in this appraisal.

3.5. Observational studies showing low recurrence rates in test low-risk groups: to what extent does this bypass the issue of whether tests are predictive for chemotherapy benefit?

Some comments have noted the low recurrence rates within Oncotype low-risk groups in large observational studies. These are summarised in Table 4. LN0 patients with RS<18 have been reported as having a 5-year DRFS of 95.9%[10] and a 5-year DRFI of 99.5-99.6%.[10-14] For LN0-mic patients with RS<11/12, reported rates of 5-year DRFS, DRFI and BCSS range from 99.3-99.9%.[9, 12-14] The fact that TAILORx has not yet reported final results also indicates that recurrence rates are likely to be low.

Some commentators question whether these low recurrence rates in low-risk patients bypass the issue of whether tests are predictive for chemotherapy benefit. This is an important consideration. However, the EAG consider the following points to be important here:

a) The low-risk RS cut-off is currently 18 rather than 11 or 12, according to the NICE scope, the manufacturers, UK clinical practice, and NHS England Access Scheme data. Despite this, data using the RS<11/12 cut-point were included in the EAG clinical review for completeness.

b) NICE currently recommends Oncotype DX only for patients who are clinically intermediate-risk, for whom the chemotherapy decision is uncertain. This clinically-intermediate subgroup is a key subgroup for the economic modelling (defined as NPI>3.4). Conversely, the observational studies (as well as the reanalyses of RCTs) include a range of clinically low- and intermediate-risk patients. Patients who are RS low-risk but clinically intermediate-risk have a higher recurrence rate than the wider RS low-risk group, as shown in both TransATAC and B14 (see Table 3). The observational evidence may include patients who would not require an Oncotype DX test in UK clinical practice due to their low clinical risk, and may mask a subgroup of clinically-intermediate risk patients with higher recurrence rates.

c) The issue of predictive performance remains important for the modelling, because whether to accept the very different **relative** chemotherapy benefits between high-risk and low-risk patients (e.g. from the B20 study, with its limitations as discussed above) has a large impact on cost-effectiveness.

4. Risk of recurrence after 5 years

As noted in the EAG report, the assumptions employed in the model regarding the long-term risk of distant recurrence and the impact of chemotherapy are based on the earlier model reported by Ward *et al*[23] used in NICE DG10.[24] These assumptions are also applied in the Genomic Health model. As noted in the EAG's response to consultation on the assessment report, 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, 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. We undertook sensitivity analyses in which the risk tapering assumption is removed (see EAG report, Tables 139, 142, 145, 148 and 151); these sensitivity analyses indicate that removing the assumption of capped recurrence risk does not significantly impact upon the conclusions drawn from the analysis.

5. Adverse effects of chemotherapy

5.1 Additional EAG sensitivity analysis - Inclusion of additional adverse events

In response to the DCD, several commentators have criticised the EAG model for excluding long-term adverse events (AEs) associated with chemotherapy, for example, chronic heart failure (CHF), permanent alopecia and peripheral neuropathy. As noted in the original EAG report, CHF was excluded from the EAG model due to a lack of evidence on the joint survival impact of CHF and metastatic breast cancer.

Within this addendum, the EAG has undertaken exploratory analyses to assess the potential impact of including these potential late effects of chemotherapy on the cost-effectiveness of the tumour profiling tests.

Estimated lifetime QALY losses and costs associated with CHF were obtained from a re-analysis of the model previously developed as part of the OPTIMA-Prelim study (Hall *et al*[25]); this was one of a minority of studies identified within the EAG's review which included this late effect of chemotherapy. The lifetime impact of CHF was estimated using the Hall *et al* model by comparing two scenarios: (i) all patients receive adjuvant chemotherapy (including excess CHF risk), and; (ii) the excess CHF risk is set equal to zero (although background levels of CHF are still included).

In addition, the EAG has included additional disutilities associated with permanent alopecia and peripheral neuropathy, based on studies identified within a systematic review of studies reporting utility values associated with AEs of chemotherapy (Shabaruddin *et al*[26]). Of the range of potentially relevant disutilities reported in the review, studies were considered potentially relevant for inclusion in the exploratory analysis if they: (a) included a counterfactual state for comparison (i.e. the same state

without the AE), and (b) if the valuations were elicited from the general public (rather than from patients experiencing the AE or from health care practitioners acting as proxy for patients). The selected disutility for alopecia was based on a general population time trade-off (TTO) study of lung cancer states reported by Nafees *et al.*[27] The disutility for peripheral neuropathy was based on a general population TTO study of colorectal cancer states reported by Shiroiwa *et al.*[28]

These additional HRQoL and cost impacts were included in the EAG's model, based on the assumptions set out in Table 8. The results of the analysis are shown in Table 9.

Table 8: Additional assumptions included in EAG's sensitivity analysis

Adverse event	Incidence	Health loss	Cost
Acute myeloid leukaemia (AML)	0.49% at 10-years (Wolff *et al*[29])	Health state utility = 0.26	Lifetime cost £10,400
CHF	Based on excess CHF risk relative to that of the general population	Net lifetime QALY loss -0.0385 QALYs (Hall *et al*[25])	Net lifetime cost -£2 (Hall *et al*[25])
Alopecia	15% of all patients receiving chemotherapy (commentator opinion)	Disutility = -0.04495 (Nafees *et al*[27])	Cost not included in analysis
Peripheral neuropathy	12% of all patients receiving chemotherapy (commentator opinion)	Disutility = -0.02 (Shiroiwa *et al*[28])	Cost not included in analysis

Table 9: Central estimates of cost-effectiveness

Test	Scenario	NPI≤3.4	NPI>3.4	LN+ (1-3 nodes)
Oncotype DX	EAG base case	£120,144	Dominated	Dominated
	Additional AEs included	£121,270	£548,524	Dominated
IHC4+C	EAG base case	£2,752	Dominating	Dominating
	Additional AEs included	£1,735	Dominating	Dominating
Prosigna	EAG base case	£89,693	£25,857	£28,666
	Additional AEs included	£88,114	£25,277	£31,807
EPClin	EAG base case	£141,848	£46,482	£21,489
	Additional AEs included	£350,042	£46,310	£19,911
Test	**Scenario**	**MINDACT ITT**	**MINDACT high-risk**	**MINDACT low-risk**
MammaPrint	EAG base case	£134,059	Dominated	£399,182
	Additional AEs included	£59,193	Dominated	£848,869

As shown in Table 9, the economic conclusions drawn from the analyses are largely unchanged by the inclusion of these additional AEs, although the inclusion of alternative disutilities may lead to different results. The EAG has a number of concerns regarding the reliability of this additional exploratory analysis:

- The QALY losses and costs associated with CHF have been derived from a separate model (Hall *et al*[25]).
- The baseline health state utilities for the relapse-free and post-relapse states included in the EAG model (taken from Lidgren *et al*[30]) may already include a proportion of patients who are experiencing AEs at the time of HRQoL assessment.
- The Lidgren *et al* study[30] and the AE utility studies identified from the Shabaruddin *et al* review[26] relate to different hypothetical populations; the selected utility estimates for peripheral neuropathy and alopecia do not relate to breast cancer states.
- The available AE utility studies[26] typically use stated preference elicitation techniques (standard gamble or time trade-off), hence both the measurement and valuation of AEs within these studies are from individuals who do not have breast cancer and who have not experienced the AE under consideration. This is not ideal.
- As they are based on comparisons of hypothetical health state scenarios, it is unlikely that the disutilities from the AE utility studies include the possibility of amelioration or resolution of the AE under consideration. It is also unclear how to quantify the distribution of severity of the AEs resulting from chemotherapy within the analysis.

5.2 QALY shortfall analysis

In light of the uncertainties associated with the analysis presented in Section 5.1, the EAG undertook a further analysis which presents the QALY shortfall associated with each test achieving an ICER of £20,000 and £30,000 per QALY gained, based on the deterministic version of the EAG model (see Table 10, Table 11, Table 12, Table 13, Table 14 and Table 15). Other things being equal, this additional analysis may further inform the Appraisal Committee's deliberations around whether other factors which cannot be reliably quantified might have a sufficient impact on the ICERs of the tumour profiling tests to change the interpretation of the model results.

Within each analysis, the QALY shortfall represents the additional number of incremental QALYs that would need to be accrued, given the currently quantified estimates of the incremental QALYs gained for the test and its incremental cost, in order for each test to achieve an ICER at a particular threshold (λ=£20,000 per QALY gained or λ=£30,000 per QALY gained). In health economic terms, this QALY shortfall is equivalent to net clinical benefit. The Committee may find it useful to consider whether the expected magnitude of the health losses avoided by reducing chemotherapy use via tumour profiling tests which are not captured in the EAG model is likely to be equal to or greater than this estimated QALY shortfall. It should be noted that this analysis is predicated on the commentators' assumption that the adverse effects of chemotherapy have been underestimated in the EAG's model. However, the EAG model suggests that with the exception of IHC4+C, all tests increase chemotherapy use at least in

some subgroups (see EAG report, Appendix 7); where this is the case, changing the balance of the net health gains and losses of chemotherapy will produce less favourable ICERs for the tumour profiling tests. It should also be noted that any potential underestimation of QALY losses only apply to those patients who would have received chemotherapy and who would have experienced associated late effects who now do not receive chemotherapy due to the tumour profiling test result and thus avoid these late effects.

The QALY shortfall analysis operates as follows. As shown in Table 10, within the LN0 NPI>3.4 group, Oncotype DX (assuming prognostic benefit only) is estimated to lead to -0.02 QALYs and additional costs of £869 compared with no testing, hence it is expected to be dominated by no testing. In this subgroup, Oncotype DX would need to make up a further 0.06 QALYs in order to achieve an ICER of £20,000 per QALY gained given its incremental cost (£869 / [0.06+-0.02] = £20,000). Within this subgroup, the EAG model suggests that the probability of receiving chemotherapy is reduced by 16% due to the use of Oncotype DX. Assuming that 25% of these patients experience late effects of chemotherapy which are not accounted for within the EAG model, this means that 4% (0.16 x 0.25) of those forgoing chemotherapy will avoid late effects. Given the overall QALY shortfall of 0.06 QALYs and the probability of avoiding late effects of 0.04, this means that each patient who would have experienced a late effect of chemotherapy would have had to have lost 1.49 QALYs (0.06/0.04) due to that AE in order for Oncotype DX to be cost-effective at a threshold of £20,000 per QALY gained.

The results for this analysis are summarised below.

Oncotype DX (prognostic benefit assumed) – refer to Table 10

LN0, NPI≤3.4 – Analysis not relevant as more patients receive chemotherapy in the test group.

LN0, NPI>3.4 – Each patient who avoids chemotherapy and avoids experiencing a late AE not quantified in the EAG model would have to save 1.49 QALYs due to the unquantified AE in order for Oncotype DX to have an ICER of £20,000 per QALY gained. Assuming a threshold of £30,000 per QALY gained, the equivalent value is 1.12 QALYs per patient.

LN+ (1-3 nodes) – Each patient who avoids chemotherapy and avoids experiencing a late AE not quantified in the EAG model would have to save 1.44 QALYs due to the unquantified AE in order for Oncotype DX to have an ICER of £20,000 per QALY gained. Assuming a threshold of £30,000 per QALY gained, the equivalent value is 1.29 QALYs per patient.

Oncotype DX (predictive benefit assumed) – refer to Table 11

LN0, NPI≤3.4 – Analysis not relevant as more patients receive chemotherapy in the test group.

LN0, NPI>3.4 – Analysis not relevant as test dominates.

LN+ (1-3 nodes) – Analysis not relevant as test dominates.

IHC4+C – refer to Table 12

LN0, NPI≤3.4 – Analysis not relevant as ICER already below £20,000 per QALY gained.

LN0, NPI>3.4 – Analysis not relevant as test dominates.

LN+ (1-3 nodes) – Analysis not relevant as test dominates.

Prosigna – refer to Table 13

LN0, NPI≤3.4 – Analysis not relevant test increases chemotherapy use.

LN0, NPI>3.4 – Analysis not relevant test increases chemotherapy use.

LN+ (1-3 nodes) – Analysis not relevant test increases chemotherapy use.

EPClin – refer to Table 14

LN0, NPI≤3.4 – Analysis not relevant test increases chemotherapy use.

LN0, NPI>3.4 – Analysis not relevant test increases chemotherapy use.

LN+ (1-3 nodes) – Analysis not relevant at threshold of £30,000 per QALY gained as ICER is below this. Each patient who avoids chemotherapy and avoids experiencing a late AE not quantified in the EAG model would have to save 0.69 due to the unquantified AE in order for EPClin to have an ICER of £20,000 per QALY gained.

MammaPrint – refer to Table 15

MINDACT ITT - Each patient who avoids chemotherapy and avoids experiencing a late AE not quantified in the EAG model would have to save 2.03 QALYs due to the unquantified AE in order for MammaPrint DX to have an ICER of £20,000 per QALY gained. Assuming a threshold of £30,000 per QALY gained, the equivalent value is 1.23 QALYs per patient.

MINDACT high-risk - Each patient who avoids chemotherapy and avoids experiencing a late AE not quantified in the EAG model would have to save 1.39 QALYs due to the unquantified AE in order for MammaPrint to have an ICER of £20,000 per QALY gained. Assuming a threshold of £30,000 per QALY gained, the equivalent value is 1.11 QALYs per patient.

MINDACT low-risk - Analysis not relevant test increases chemotherapy use.

Table 10: QALY shortfall analysis - Oncotype DX (prognostic benefit only)

Oncotype DX (prognostic)	**LN0, NPI<3.4**	**LN0, NPI>3.4**	**LN+ (1-3 nodes)**
Inc. QALYs	0.01	-0.02	-0.07
Inc. costs	£1,317	£869	£647
ICER	£120,144	Dominated	Dominated
QALY shortfall to achieve ICER=£20,000/QALY gained	**0.05**	**0.06**	**0.10**
QALY shortfall to achieve ICER=£30,000/QALY gained	**0.03**	**0.04**	**0.09**
Proportion patients avoiding chemo due to testing	0.00	0.16	0.29
Proportion patients unaccounted AEs (assumption based on consultation responses)	0.25	0.25	0.25
Proportion patients tested avoiding chemo with unaccounted AEs	n/a - more get chemo in test group	0.04	0.07
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£20,000/QALY	**n/a - more get chemo in test group**	**1.49**	**1.44**
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£30,000/QALY	**n/a - more get chemo in test group**	**1.12**	**1.29**

Table 11: QALY shortfall analysis - Oncotype DX (predictive benefit)

Oncotype DX (predictive)	**LN0, NPI<3.4**	**LN0, NPI>3.4**	**LN+ (1-3 nodes)**
Inc. QALYs	0.04	0.27	0.09
Inc. costs	£1,211	-£364	-£68
ICER	£34,245	Dominating	Dominating
QALY shortfall to achieve ICER=£20,000/QALY gained	**0.03**	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**
QALY shortfall to achieve ICER=£30,000/QALY gained	**0.01**	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**
Proportion patients avoiding chemo due to testing	0.00	0.16	0.29
Proportion patients unaccounted AEs (assumption based on consultation responses)	0.25	0.25	0.25
Proportion patients tested avoiding chemo with unaccounted AEs	n/a - more get chemo in test group	0.04	0.07
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£20,000/QALY	**n/a - more get chemo in test group**	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£30,000/QALY	**n/a - more get chemo in test group**	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**

Table 12: QALY shortfall analysis - IHC4+C

IHC4+C	**LN0, NPI<3.4**	**LN0, NPI>3.4**	**LN+ (1-3 nodes)**
Inc. QALYs	0.01	0.01	0.05
Inc. costs	£22	-£89	-£269
ICER	£2,752	Dominating	Dominating
QALY shortfall to achieve ICER=£20,000/QALY gained	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**
QALY shortfall to achieve ICER=£30,000/QALY gained	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**
Proportion patients avoiding chemo due to testing	0.04	0.08	0.07
Proportion patients unaccounted AEs (assumption based on consultation responses)	0.25	0.25	0.25
Proportion patients tested avoiding chemo with unaccounted AEs	0.01	0.02	0.02
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£20,000/QALY	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£30,000/QALY	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**

Table 13: QALY shortfall analysis - Prosigna

Prosigna	**LN0, NPI<3.4**	**LN0, NPI>3.4**	**LN+ (1-3 nodes)**
Inc. QALYs	0.02	0.07	0.07
Inc. costs	£1,891	£1,713	£1,967
ICER	£89,693	£25,857	£28,666
QALY shortfall to achieve ICER=£20,000/QALY gained	**0.07**	**0.02**	**0.03**
QALY shortfall to achieve ICER=£30,000/QALY gained	**0.04**	**n/a - ICER already below threshold**	**n/a - ICER already below threshold**
Proportion patients avoiding chemo due to testing	0.00	-0.01	-0.08
Proportion patients unaccounted AEs (assumption based on consultation responses)	0.25	0.25	0.25
Proportion patients tested avoiding chemo with unaccounted AEs	n/a - more get chemo in test group	n/a - more get chemo in test group	n/a - more get chemo in test group
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£20,000/QALY	**n/a - more get chemo in test group**	**n/a - more get chemo in test group**	**n/a - more get chemo in test group**
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£30,000/QALY	**n/a - more get chemo in test group**	**n/a - more get chemo in test group**	**n/a - more get chemo in test group**

Table 14: QALY shortfall analysis - EPClin

EPClin	**LN0, NPI<3.4**	**LN0, NPI>3.4**	**LN+ (1-3 nodes)**
Inc. QALYs	0.01	0.03	0.06
Inc. costs	£1,686	£1,401	£1,185
ICER	£141,848	£46,482	£21,489
QALY shortfall to achieve ICER=£20,000/QALY gained	**0.07**	**0.04**	**0.00**
QALY shortfall to achieve ICER=£30,000/QALY gained	**0.04**	**0.02**	**n/a - ICER already below threshold**
Proportion patients avoiding chemo due to testing	-0.07	-0.01	0.02
Proportion patients unaccounted AEs (assumption based on consultation responses)	0.25	0.25	0.25
Proportion patients tested avoiding chemo with unaccounted AEs	n/a - more get chemo in test group	n/a - more get chemo in test group	0.01
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£20,000/QALY	**n/a - more get chemo in test group**	**n/a - more get chemo in test group**	**0.69**
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£30,000/QALY	**n/a - more get chemo in test group**	**n/a - more get chemo in test group**	**n/a - ICER already below threshold**

Table 15: QALY shortfall analysis - MammaPrint

MammaPrint	**MINDACT ITT**	**MINDACT high-risk**	**MINDACT low-risk**
Inc. QALYs	0.01	-0.04	0.01
Inc. costs	£1,757	£1,380	£2,415
ICER	£134,059	Dominated	£399,182
QALY shortfall to achieve ICER=£20,000/QALY gained	**0.07**	**0.11**	**0.11**
QALY shortfall to achieve ICER=£30,000/QALY gained	**0.05**	**0.09**	**0.07**
Proportion patients avoiding chemo due to testing	0.15	0.33	-0.03
Proportion patients unaccounted AEs (assumption based on consultation responses)	0.25	0.25	0.25
Proportion patients tested avoiding chemo with unaccounted AEs	0.04	0.08	n/a - more get chemo in test group
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£20,000/QALY	**2.03**	**1.39**	**n/a - more get chemo in test group**
QALY loss for patients avoiding chemo with unaccounted AEs required to achieve shortfall at λ=£30,000/QALY	**1.23**	**1.11**	**n/a - more get chemo in test group**

6. Probability of having chemotherapy

Several commentators have suggested other potentially relevant decision impact studies could or should have been included in the EAG report. However, the studies suggested are either already included in the EAG report, or were excluded from the report with justification. The only exception to this is a study reported by Rodriguez *et al*; this study was not identified by the EAG searches, however, the results appear to be consistent with other Prosigna decision impact studies already included in the EAG review.

7. EAG systematic review and meta-analysis

All major comments relating to this theme are discussed in the EAG's table of responses.

8. EAG economic model

8.1. Re-analysis of MammaPrint by Agendia within the EAG model

Agendia have undertaken a re-analysis of the cost-effectiveness of MammaPrint using the EAG model *"with corrected usage of available MammaPrint data in those instances where we [Agendia] strongly disagree with the chosen inputs in the current model."* With respect to this analysis, the company claims that on the basis of altered model inputs, the ICER for MammaPrint is now less than £30,000 per QALY gained. However, the EAG notes that within the company's re-analysis, chemotherapy is assumed to be associated with no additional benefit in terms of DRFS for any patient population (including those with clinical-high MammaPrint-high risk). If this was the case, genomic testing would have no value as clinicians would never give chemotherapy to any patient. The EAG considers Agendia's re-analysis of the EAG model to be inappropriate and believes that the results are not meaningful.

8.2. Additional EAG sensitivity analysis - Cost-effectiveness of adjuvant chemotherapy by subgroup

During the consultation on the EAG report and the DCD, it has been suggested that the EAG model is predisposed to find giving chemotherapy to all patients a clinically effective and cost-effective use of resources. This interpretation of the model is inaccurate. In the interests of clarity, Table 16 presents the results of an analysis comparing 100% chemotherapy versus 0% chemotherapy using the EAG model. As shown in the table, the strategy involving the indiscriminate use of chemotherapy is dominated by the no chemotherapy option for patients with NPI≤3.4 (i.e. chemotherapy generates fewer QALYs at a greater cost). Chemotherapy appears to have a favourable clinical and cost-effectiveness profile within the LN0, NPI>3.4 and LN+ subgroups.

Table 16: Cost-effectiveness of chemotherapy versus no chemotherapy

Subgroup	Option	QALYs	Costs	Inc. QALYs	Inc. costs	ICER
LN0, NPI≤3.4	100% chemotherapy	13.83	£7,454	-0.04	£3,670	**Dominated**
	No chemotherapy	13.87	£3,784	-	-	-
LN0, NPI>3.4	100% chemotherapy	12.85	£11,700	0.27	£2,316	**£8,449**
	No chemotherapy	12.58	£9,384	-	-	-
LN+	100% chemotherapy	12.63	£12,668	0.35	£2,011	**£5,787**
	No chemotherapy	12.28	£10,658	-	-	-

8.3. Additional EAG sensitivity analysis - Alternative estimates of chemotherapy benefit within clinical risk subgroups

Several commentators have raised issues regarding the estimated relative risk of distant recurrence associated with chemotherapy. The original EAG report acknowledged that there is uncertainty around this estimate and notes that the estimated relative risk of 0.76 was calculated using the most relevant data reported within the EBCTCG 2011 meta-analysis paper[31] (data specifically relating to distant recurrence). The EAG notes that it is possible that the relative benefit of chemotherapy could be different between clinical risk groups, although the EBCTCG meta-analysis does not provide sufficient information to determine the relative risk of distant recurrence within each of the three model subgroups (LN-, NPI≤3.4; LN- NPI>3.4, and LN+[1-3 nodes]). Tables 139, 142, 145, 148 and 151 of the EAG report presented sensitivity analyses using values of 0.70 and 0.80 to explore the impact of this uncertainty on the cost-effectiveness of the tests; these limits are similar to reported rate ratios for any recurrence (including local and regional) for ER+ patients with N0/N- and N1-3 within the EBCTCG meta-analysis paper.

Within this addendum, the EAG has expanded this existing sensitivity analysis to reflect a broader range of relative risk estimates. As shown in Table 17, the economic conclusions drawn from the model for Oncotype DX, IHC4+C and MammaPrint are unaffected by these alternative values. Conversely, within the scenarios in which chemotherapy is assumed to be less favourable, the ICERs for Prosigna and EPClin are markedly less favourable in the LN0 NPI>3.4 and LN+ subgroups.

Table 17: Additional EAG sensitivity analysis - Alternative estimates of chemotherapy benefit within subgroups

Test	Scenario	ICER (per QALY gained)		
		LN0 NPI≤3.4	**LN0 NPI>3.4**	**LN+**
Oncotype	**Chemotherapy RR = 0.76 (EAG base case)**	**£120,144**	**Dominated**	**Dominated**
	Chemotherapy RR = 0.6	£69,967	Dominated	Dominated
	Chemotherapy RR = 0.7	£94,920	Dominated	Dominated
	Chemotherapy RR = 0.8	£145,102	Dominated	Dominated
	Chemotherapy RR = 0.9	£297,925	£201,602	Dominated
IHC4+C	**Chemotherapy RR = 0.76 (EAG base case)**	**£2,752**	**Dominating**	**Dominating**
	Chemotherapy RR = 0.6	£1,326	Dominating	Dominating
	Chemotherapy RR = 0.7	£2,138	Dominating	Dominating
	Chemotherapy RR = 0.8	£3,223	Dominating	Dominating
	Chemotherapy RR = 0.9	£4,745	Dominating	Dominating
Prosigna	**Chemotherapy RR = 0.76 (EAG base case)**	**£89,693**	**£25,857**	**£28,666**
	Chemotherapy RR = 0.6	£52,504	£13,975	£14,678
	Chemotherapy RR = 0.7	£71,107	£19,926	£21,508
	Chemotherapy RR = 0.8	£107,875	£31,645	£36,018
	Chemotherapy RR = 0.9	£214,907	£65,467	£87,917
EPClin	**Chemotherapy RR = 0.76 (EAG base case)**	**£141,848**	**£46,482**	**£21,489**
	Chemotherapy RR = 0.6	£65,750	£26,202	£11,702
	Chemotherapy RR = 0.7	£99,445	£36,317	£16,663
	Chemotherapy RR = 0.8	£195,508	£56,485	£26,089
	Chemotherapy RR = 0.9	£2,680,967	£116,586	£50,984
MammaPrint	**Scenario**	**MINDACT ITT**	**mAOL High risk**	**mAOL Low risk**
	Chemotherapy RR = 0.76 (EAG base case)	**£134,059**	**Dominated**	**£399,182**
	Chemotherapy RR = 0.6	£176,352	Dominated	£113,124
	Chemotherapy RR = 0.7	£148,424	Dominated	£161,338
	Chemotherapy RR = 0.8	£127,971	Dominated	£276,670
	Chemotherapy RR = 0.9	£112,346	£216,964	£920,361

RR – relative risk

9. Company economic models - new model submitted by Agendia

In response to the diagnostic consultation document, Agendia submitted a revised version of their model based on the MINDACT trial. The EAG has scrutinised this new analysis. The EAG notes that the model trace shows that the proportion of patients remaining alive and recurrence-free increases over time, whilst the proportion of the modelled cohort who are dead is allowed to decrease over time (see Figure 2 and Figure 3); this is clearly incorrect and as such the model lacks any face validity. In addition, whilst the company states that extrapolation has now been included in the model in order to account for longer-term costs and health impacts (assuming a constant event rate), the model trace indicates that no additional events occur between years 7 and 10. This also indicates major programming errors. On the basis of these errors, the EAG does not consider the company's new analyses to be reliable.

Figure 2: Probability of being alive, genomic test group, new Agendia model

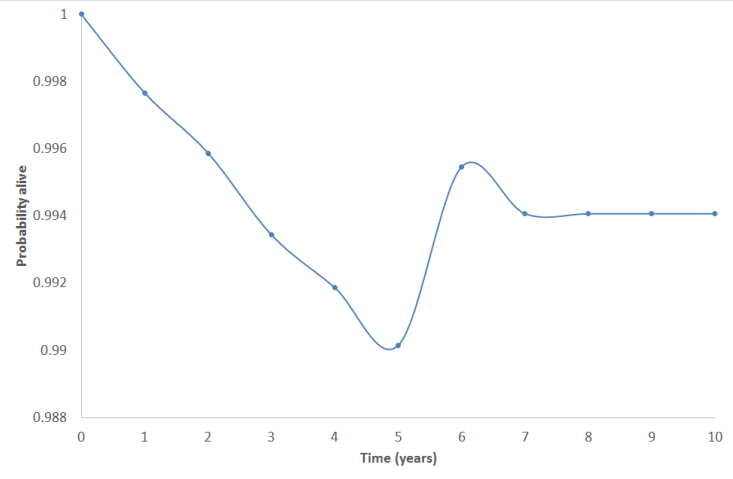
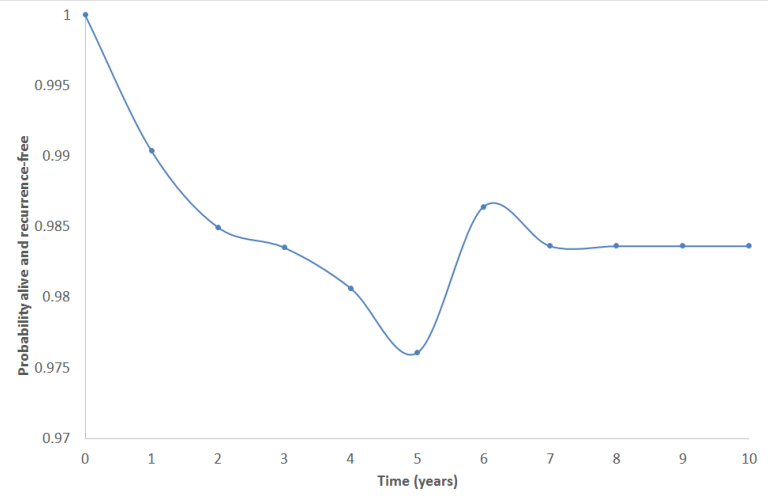


Figure 3: Probability of being alive and recurrence-free, genomic test group, new Agendia model



10. New commercial access schemes

Analyses based on company access proposals are included in a confidential addendum.

REFERENCES

1. National Institute for Health and Care Excellence. Diagnostics Consultation Document: Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer.; 2018.
2. Sestak I, Dowsett M, Cuzick J. NICE request - TransATAC data analysis. In; 2017.
3. Sestak I, Buus R, Cuzick J, et al. Comparison of the performance of 6 prognostic signatures for estrogen receptor–positive breast cancer: A secondary analysis of a randomized clinical trial. *JAMA Oncology* 2018; 10.1001/jamaoncol.2017.5524.

4. 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. 5480.

5. 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.

6. 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.

7. 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.

8. 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.

9. 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.

10. 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.

11. 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.

12. 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.

13. 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.

14. 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.

15. 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.

16. 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.

17. 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.

18. 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.

19. 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.

20. 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.

21. 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.

22. 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.

23. 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.

24. 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.

25. 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.

26. Shabaruddin FH, Chen LC, Elliott RA, Payne K. A systematic review of utility values for chemotherapy-related adverse events. *PharmacoEconomics* 2013;31:277-88.

27. Nafees B, Stafford M, Gavriel S, Bhalla S, Watkins J. Health state utilities for non small cell lung cancer. *Health & Quality of Life Outcomes* 2008;6.

28. Shiroiwa T, Fukuda T, Tsutani K. Health utility scores of colorectal cancer based on societal preference in Japan. *Quality of Life Research* 2009;18:1095-103.

29. 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.

30. 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.

31. 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.