NICE National Institute for Health and Care Excellence

NICE guidance

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

Diagnostics guidance

Published: 19 December 2018

www.nice.org.uk/guidance/dg34

Contents

This guidance replaces DG10, MIB44 and MIB27.

1 Recommendations

1.1 EndoPredict (EPclin score), Oncotype DX Breast Recurrence Score and Prosigna are recommended as options for guiding adjuvant chemotherapy decisions for people with oestrogen receptor (ER)-positive, human epidermal growth factor receptor 2 (HER2)-negative and lymph node (LN)-negative (including micrometastatic disease; see section 5.4) early breast cancer, only if:

- they have an intermediate risk of distant recurrence using a validated tool such as PREDICT or the Nottingham Prognostic Index
- information provided by the test would help them choose, with their clinician, whether or not to have adjuvant chemotherapy taking into account their preference
- the companies provide the tests to the NHS with the discounts agreed in the access proposals and
- clinicians and companies make timely, complete and linkable record-level test data available to the National Cancer Registration and Analysis Service as described in the data collection arrangements agreed with NICE (see section 5.29).

1.2 MammaPrint is not recommended for guiding adjuvant chemotherapy decisions for people with ER-positive, HER2-negative and LN-negative early breast cancer because it is not cost effective.

1.3 IHC4+C is not recommended for guiding adjuvant chemotherapy decisions for people with ER-positive, HER2-negative and LN-negative early breast cancer because the analytical validity of the test is uncertain.

Why the committee made these recommendations

People with early and locally advanced breast cancer may need further treatment (adjuvant treatment) after they have surgery. Tools such as PREDICT, which is used by many NHS trusts, provide prognostic information to help guide the selection of adjuvant treatment. Additional information from tumour profiling tests may be helpful for people whose cancer has an intermediate risk of distant recurrence when the decision to offer chemotherapy is unclear.

Evidence suggests that EndoPredict (EPclin score), Oncotype DX Breast Recurrence Score, MammaPrint and Prosigna can predict the risk of distant recurrence in people who have ER-positive, HER2-negative, early breast cancer. This evidence is strongest in the group with LN-negative disease, which is likely to include people with micrometastatic disease. Also, Oncotype DX Breast Recurrence Score may be able to predict who will respond to chemotherapy, but the evidence for this is uncertain.

There are uncertainties in the economic modelling, particularly around the pre- and post-test chemotherapy decisions and the effect of adjuvant chemotherapy on distant recurrence. Also, there are no data available to compare the tumour profiling tests with PREDICT, or to define the clinical risk groups using PREDICT.

Using the access proposal test costs for EndoPredict (EPclin score), Oncotype DX Breast Recurrence Score and Prosigna, all 3 tests would provide value for money in people with LN-negative disease and an intermediate risk of distant recurrence. But because of the uncertainty about their effects on clinical decision making and clinical outcomes, the tests are recommended as options for guiding adjuvant chemotherapy decisions only if the criteria in section 1.1 are met, which includes collecting data on their use. Discussion about treatment options within the multidisciplinary team may be particularly helpful for people who have micrometastatic disease.

In people with an intermediate to high risk of distant recurrence, MammaPrint is less clinically effective and costs more than current practice, which does not use tumour profiling tests. It is therefore not recommended for use in the NHS.

IHC4+C appears to be cost effective, but there are concerns about its analytical validity, for example the reproducibility of test results. It is therefore not recommended for use in the NHS.

2 Clinical need and practice

The problem addressed

2.1 The tumour profiling tests EndoPredict, MammaPrint, Oncotype DX Breast Recurrence Score, Prosigna and IHC4+C provide information on the activity of genes in tumour samples from people with early breast cancer. The results provide a risk profile of a person's breast cancer, which can be used with other routinely assessed clinical risk factors, 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. Some tests also claim to predict relative treatment effects for chemotherapy. This information is intended to help decision making about adjuvant chemotherapy use.

2.2 It is also claimed that people with early breast cancer identified as having a low risk of distant recurrence by a tumour profiling test may not need to have adjuvant chemotherapy. For these people, unnecessary treatment and therefore the comorbidities and negative effects on quality of life associated with chemotherapy could be avoided. Also, for people with early breast cancer at low risk of disease recurrence based on clinical and pathological features, the tests could confirm whether their risk is correct. If reclassified as being at high risk of recurrence, these people may benefit from chemotherapy. People with breast cancer and clinicians may also have more confidence that the treatment they are having or recommending is appropriate.

2.3 This assessment evaluates the clinical and cost effectiveness of EndoPredict, MammaPrint, Oncotype DX Breast Recurrence Score, Prosigna and IHC4+C when used to guide adjuvant chemotherapy decisions. The population was people with oestrogen receptor (ER)-positive (or progesterone receptor-positive or both), human epidermal growth factor receptor 2 (HER2)-negative early breast cancer (stages 1 or 2) with 0 to 3 positive lymph nodes.

2.4 This is a full update of NICE's diagnostics guidance on gene expression

profiling and expanded immunohistochemistry tests for guiding adjuvant chemotherapy decisions in early breast cancer management: MammaPrint, Oncotype DX, IHC4 and Mammostrat (DG10), which was published in 2013. This recommended Oncotype DX as an option for guiding adjuvant chemotherapy decisions for people with ER-positive, HER2-negative and lymph node-negative early breast cancer if the person was assessed as being at intermediate risk and the company provided Oncotype DX to NHS organisations according to the confidential arrangement agreed with NICE. The guidance also encouraged data collection on the use of Oncotype DX in the NHS, and further research on MammaPrint, IHC4 and Mammostrat. Since publication of the original guidance, Mammostrat is no longer available. Also, a new test, EndoPredict, has become available and PAM50 has been further developed into the Prosigna test.

The condition

2.5 Breast cancer is the most common cancer and the third most common cause of UK cancer-related deaths. One in 8 women and 1 in 870 men will be diagnosed with breast cancer during their lifetime (Cancer Research UK 2016). In 2014, 46,085 women and 332 men were newly diagnosed with breast cancer in England (Office for National Statistics 2016). Most breast cancer develops in women who are over the age of 50 (Cancer Research UK 2016).

2.6 Breast cancer survival depends on the stage of the disease at diagnosis, the treatment received and the biology of the tumour. More than 90% of women diagnosed with early breast cancer survive for at least 5 years, and 78% survive for 10 years (Cancer Research UK 2016). In contrast, only 13% of those diagnosed with advanced disease survive for more than 5 years.

The diagnostics and care pathways

Diagnosis

2.7 Breast cancer may be diagnosed following an abnormal result in the NHS

breast cancer screening programme, or after referral for further investigation because of signs or symptoms that could be associated with breast cancer. The referral criteria are described in NICE's guideline on suspected cancer.

2.8 When cancer cells have been detected in a biopsy sample, further tests are done to provide more information on the characteristics of the tumour. The results of these tests are used to categorise breast cancer into molecular subtypes and determine which types of treatment it is most likely to respond to. Recommendations on tumour testing are in NICE's guideline on early and locally advanced breast cancer. Tumour tests can include hormone receptor and HER2 tests. Although not routinely done, some laboratories may also test for Ki67, a marker of cell proliferation.

Care

2.9 NICE's guideline on early and locally advanced breast cancer describes the care pathway. Surgery is often the initial treatment. Neoadjuvant treatment may be used before surgery, to reduce the size of the tumour and enable breast-conserving surgery.

2.10 After surgery, further treatment (adjuvant treatment) may be needed and this can include radiotherapy, chemotherapy, hormone therapy, biological therapy or a combination of these. The decision to offer adjuvant therapy, and the treatments to use, is made taking into account the clinical history, the stage and grade of disease, the likely course of the disease (prognosis), the molecular characteristics of the tumour and the person's preferences.

2.11 A variety of tools are available that can help to predict the likelihood of breast cancer recurrence based on clinical and pathological features. These may be used to provide prognostic information for patients and to guide the selection of adjuvant therapy. Expert advice suggests that the PREDICT tool version 2.0, an online prognostic and treatment benefit calculator, is the most widely used tool in the NHS in England to calculate risk of recurrence. Adjuvant! Online is not currently available online because it is being updated. It is not certain when it will be reinstated,

and the website directs people to the PREDICT tool. Adjuvant! Online is described in the supplementary appendix of Cardoso et al. 2016. PREDICT is recommended in NICE's guideline on early and locally advanced breast cancer.

3 The diagnostic tests

3.1 The assessment compared 5 intervention tests with 1 comparator.

The interventions

EndoPredict (Myriad Genetics)

3.2 EndoPredict is a CE-marked assay that is designed to predict the likelihood of metastases developing within 10 years of an initial breast cancer diagnosis. The test is for pre and postmenopausal people with early breast cancer with oestrogen receptor (ER)-positive, human epidermal growth factor 2 (HER2)-negative and lymph node (LN)-negative or LN-positive disease (up to 3 positive nodes).

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

3.4 EndoPredict needs RNA extracted from a formalin-fixed, paraffin-embedded (FFPE) breast cancer tissue sample. The test can be done in a local laboratory or the Myriad Genetics pathology laboratory in Germany. It takes approximately 2 days from receipt of the tissue sample to get the results from a local laboratory, although the turnaround time may be longer if samples are tested in full batches. The turnaround time is also longer if samples are sent to Germany.

3.5 The test involves a reverse transcription-quantitative polymerase chain reaction. Online evaluation software calculates an EP score and an EPclin score. An EP score of 0 to less than 5 indicates low risk of distant disease recurrence in the next 10 years. An EP score of 5 to 15 indicates high risk of distant disease recurrence in the next 10 years.

3.6 The EPclin score estimates the probability of metastases developing within 10 years (assuming 5 years of endocrine therapy). It is calculated by adding clinical data about tumour size and nodal status to the EP

score. An EPclin score of less than 3.3 indicates low risk (less than 10%) of metastases in the next 10 years. An EPclin score of 3.3 or more indicates high risk of metastases in the next 10 years.

3.7 During consultation on the first diagnostics consultation document, NICE accepted an access proposal from the company in line with the Diagnostics Assessment Programme's interim addendum on access proposals. This provides a simple discount to the list price of EndoPredict, with the discount applied at the point of purchase or invoice. The level of the discount is commercial in confidence.

MammaPrint (Agendia)

3.8 MammaPrint is a CE-marked assay that is designed to assess the risk of distant recurrence within 5 and 10 years and whether a person would benefit from chemotherapy. The test is for pre and postmenopausal people with stage 1 or 2 breast cancer, with a tumour size of 5 cm or less, and LN-negative or LN-positive disease (up to 3 positive nodes). The test can be used irrespective of ER and HER2 status.

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

3.10 The MammaPrint test needs RNA extracted from an FFPE breast cancer tissue sample. The test is offered as an off-site service. In Europe, samples are analysed at the Agendia laboratory in the Netherlands. Results are available within 10 days of submitting the sample.

3.11 The test is based on diagnostic microarray. Software is used to calculate the MammaPrint result on a scale of −1 to +1. The score indicates the risk of developing distant metastases over the next 10 years without any adjuvant endocrine therapy or chemotherapy. A MammaPrint result of 0 or less indicates high risk of metastases in the next 10 years and a result of more than 0 indicates low risk (10% or less) of metastases in the next 10 years.

Oncotype DX Breast Recurrence Score (Genomic Health)

3.12 Oncotype DX Breast Recurrence Score (hereafter referred to as Oncotype DX) is designed to quantify the 10-year risk of distant recurrence and predict relative treatment effects for chemotherapy. The test also reports the underlying tumour biology: ER, progesterone receptor and HER2 status. The test is for pre and postmenopausal people with stage 1 or 2 breast cancer and ER-positive, HER2-negative, LN-negative or LN-positive disease (up to 3 positive nodes). The assay does not have a CE mark because it is provided as a service by Genomic Health in an accredited laboratory in the US.

3.13 Oncotype DX quantifies the expression of 21 genes: 16 cancer-related genes correlated with distant recurrence-free survival, and 5 reference (normalisation) genes.

3.14 The Oncotype DX test needs RNA extracted from a FFPE breast cancer tissue sample. Samples are processed centrally at a Genomic Health laboratory in the US. Results are usually available 7 to 10 days after the sample is received.

3.15 The test is based on a reverse transcription-quantitative polymerase chain reaction. It gives a Recurrence Score result of between 0 and 100, which is used to quantify the 10-year risk of distant recurrence, assuming 5 years of endocrine therapy. A result below 18 indicates low risk of distant recurrence and claims to predict little to no effect of chemotherapy on patient outcomes. A result between 18 and 30 indicates intermediate risk of recurrence and claims to predict no substantial effect of chemotherapy on patient outcomes. A result of 31 or more indicates high risk of recurrence and claims to predict a large effect of chemotherapy on patient outcomes. NICE is aware that the cut-offs may change in light of the results from the TAILORx study.

3.16 The breast Recurrence Score result can be combined with clinical and pathological factors using the recurrence score-pathology-clinical (RSPC) calculator. However, this calculator has not been validated in a cohort independent of that used to derive Oncotype DX.

3.17 During consultation on the first diagnostics consultation document NICE accepted the company's commitment to maintain the current confidential discount, which is in line with the Diagnostics Assessment Programme's interim addendum on access proposals. This provides a simple discount to the list price of Oncotype DX, with the discount applied at the point of purchase or invoice. The level of the discount is commercial in confidence.

Prosigna (NanoString Technologies)

3.18 Prosigna is a CE-marked assay designed to provide information on breast cancer subtype and to predict distant recurrence-free survival at 10 years. The test is for postmenopausal people with early breast cancer that is ER-positive, HER2-negative and LN-negative or LN-positive (up to 3 positive nodes).

3.19 Prosigna measures the expression of 50 genes used for intrinsic subtype classification, 8 housekeeping genes used for signal normalisation, 6 positive controls, and 8 negative controls.

3.20 The test needs RNA extracted from a FFPE breast tumour tissue sample. It is based on direct mRNA counting using fluorescent probes and an nCounter Digital Analyser.

3.21 Prosigna classifies the risk of distant recurrence within 10 years, assuming 5 years of endocrine therapy, based on the PAM50 gene signature, breast cancer subtype, tumour size, nodal status and proliferation score. The proliferation score is determined by evaluating multiple genes associated with the proliferation pathway. The test gives a score between 0 and 100. Based on this score and the nodal status, samples are classified into risk categories:

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

3.22 During consultation on the first diagnostics consultation document, NICE

accepted an access proposal from the company in line with the Diagnostics Assessment Programme's interim addendum on access proposals. This provides a simple discount to the list price of Prosigna, with the discount applied at the point of purchase or invoice. The level of the discount is commercial in confidence.

IHC4 and IHC4+C

3.23 The IHC4 test is a laboratory developed test that combines the results of 4 immunohistochemistry (IHC) measurements. The IHC4+C test combines the results of the 4 IHC4 tests with clinical and pathological features such as age, nodal status, tumour size, and grade. Both versions are designed to quantify the 10-year risk of distant disease recurrence, assuming 5 years of endocrine therapy. The test is for postmenopausal people with early breast cancer that is ER-positive and LN-negative or LN-positive (up to 3 positive nodes).

3.24 The IHC4+C test needs an FFPE breast tumour tissue sample. The 4 immunohistochemistry tests are: ER, progesterone receptor (PR), HER2 and the proliferation marker Ki67. ER and HER2 markers are commonly measured in NHS laboratories, but PR and Ki67 markers are not.

3.25 The IHC4+C test is in clinical use at 1 NHS centre (the Royal Marsden NHS Foundation Trust), which carries out the test with an average turnaround time of 1 week. The test could be run in local NHS laboratories providing that training and quality assurance programmes for the individual assays were in place.

3.26 The IHC4+C test uses a published algorithm to calculate a risk score for distant recurrence based on the results of the 4 assays and clinical factors. A calculator is available for use on request. A score of less than 10% is categorised as low risk for distant recurrence at 10 years. A score of more than 10% but less than 20% is intermediate risk, and a score of 20% or more is high risk for distant recurrence at 10 years.

The comparator

3.27 The comparator is decision making for adjuvant chemotherapy

prescribing, based on clinical and pathological features or the results of tools used to assess risk without the tumour profiling tests. Features may include the stage and grade of the disease, nodal status, ER or PR status, HER2 status and any previous treatment (for example, neoadjuvant therapy). Risk assessment tools include PREDICT, the Nottingham Prognostic Index (NPI) and Adjuvant! Online.

3.28 These risk assessment tools can be used to define the level of clinical risk. For example, in LN-negative disease a NPI of 3.4 or less is classed as low risk, and a NPI of more than 3.4 and up to or equal to 5.4 is classed as intermediate risk. If using the PREDICT tool, an absolute 10-year survival benefit from chemotherapy of less than 3% is classed as low risk; between 3% and 5% is classed as intermediate risk; and more than 5% is classed as high risk.

4 Evidence

The diagnostics advisory committee (section 7) considered evidence on EndoPredict (EP score and EPclin score), MammaPrint, Oncotype DX (with and without the recurrence score-pathology-clinical [RSPC] calculator), Prosigna and IHC4 or IHC4+C from several sources. Full details of all the evidence are in the committee papers.

Clinical effectiveness

4.1 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 relative treatment effect – the ability of the test to predict which patients have disease that will respond to chemotherapy. It can be assessed by considering whether the relative treatment effect of chemotherapy or 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.

4.2 A total of 153 references were included in the review. Studies assessing prognostic ability and prediction of relative treatment effect were quality assessed using relevant criteria 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).

Prognostic ability

4.3 Studies providing information on prognostic ability were retrospective analyses of RCT data or routinely collected data. Most of the 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, because patients with smaller tumours (who may be systematically different to those with large tumours) are likely to be under-represented. In many studies patients had chemotherapy, which could affect event rates and therefore potentially reduce the apparent prognostic performance of a test. In other studies, patients who had chemotherapy were excluded from analyses, which may also lead to spectrum bias. Therefore studies in which all patients had endocrine monotherapy were preferable.

4.4 Results for prognostic ability were generally presented as unadjusted or adjusted analyses. Unadjusted analyses look at differences in the event rates among low, intermediate and high-risk groups without adjusting for clinical and pathological variables. Adjusted analyses show whether the test has prognostic value over clinical and pathological variables.

Distribution of patients across risk categories

4.5 Among studies of patients with lymph node (LN)-negative disease who had endocrine monotherapy, in each group around 70% to 80% had disease that was categorised as low or low/intermediate risk across all tests (11 studies). Most MammaPrint studies had mixed endocrine and chemotherapy use, mixed hormone receptor status with or without mixed human epidermal growth factor receptor 2 (HER2) status, so results may not be comparable with results from other tests. In these studies 20% to 61% of patients had disease that was categorised as low risk (6 studies). Most IHC4 or IHC4+C studies used quartiles or tertiles to define risk groups. These studies do not provide useful information on the distribution of patients across risk categories.

4.6 The proportion of patients with low and intermediate risk was generally much lower in groups with LN-positive disease than in groups with LN-negative disease who had endocrine monotherapy (7 LN-positive studies). For Oncotype DX, however, the proportion of patients with low and intermediate risk was only slightly lower in the LN-negative group than in the LN-positive group. Studies of MammaPrint in patients with LN-positive disease were all done in groups with mixed hormone

receptor status and mixed or unknown HER2 status, so results may not be comparable with results from other tests. In these studies 38% to 41% of patients had disease that was categorised as low risk (2 studies).

Oncotype DX

4.7 There were 11 data sets that provided information on the prognostic ability of Oncotype DX: 7 reanalyses of RCT data and 4 retrospective studies of routinely collected data. All studies were validation studies, and in 4 RCTs patients had endocrine monotherapy. Three of the studies were done in East Asia and it is uncertain whether the commercial version of Oncotype DX was used in these studies. Also, they may not be generalisable to England because usual clinical practice may differ between countries enough to affect prognostic outcomes. In addition, it is possible that people of different ethnicities have different underlying risk profiles and natural history of disease.

4.8 Unadjusted analyses indicated that Oncotype DX had prognostic accuracy (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 group with LN-positive disease.

4.9 In adjusted analyses, Oncotype DX provided statistically significant additional prognostic information over most commonly used clinical and pathological variables (age, grade, size, nodal status), regardless of lymph node status. A bespoke analysis of TransATAC study data also showed that Oncotype DX provided additional prognostic information over clinical and pathological tools to assess risk.

MammaPrint

4.10 There were 10 data sets that provided information on the prognostic ability of MammaPrint: 1 reanalysis of RCT data and 9 retrospective studies of routinely collected data. In addition, a further 4 studies pooled data on specific patients from the same 10 data sets. All studies were validation studies, and in 2 studies patients had endocrine monotherapy.

Most studies included some patients who were out of scope (with HER2-positive or hormone receptor-negative disease or both).

4.11 In 6 of 7 unadjusted analyses, MammaPrint had prognostic accuracy (there were statistically significant differences between low-risk and high-risk groups) for 10 year distant recurrence-free survival or interval, regardless of LN status.

4.12 In adjusted analyses, a pooled analysis of patients with LN-negative and LN-positive disease showed that MammaPrint had statistically significant prognostic accuracy for 10-year distant recurrence-free survival after adjusting for clinical and pathological variables. In patients with LN-negative disease, MammaPrint had statistically significant prognostic accuracy for 10-year distant recurrence-free interval when adjusted for Adjuvant! Online, Nottingham Prognostic Index (NPI) or clinical and pathological variables. In patients with LN-positive disease, MammaPrint had borderline statistically significant prognostic accuracy for 10-year distant metastasis-free survival when adjusted for clinical and pathological variables.

Prosigna

4.13 There were 8 data sets that provided information on the prognostic ability of Prosigna: 6 reanalyses of RCT data and 3 retrospective analyses of 2 prospective cohort studies. All studies were validation studies, and in 5 studies patients had endocrine monotherapy. Some studies included some patients who were out of scope (with HER2-positive or hormone receptor-negative disease or both).

4.14 Prosigna had statistically significant prognostic accuracy for 10-year distant recurrence-free survival and interval in all unadjusted analyses of patients with LN-negative and LN-positive disease.

4.15 In analyses adjusted for clinical and pathological variables or tools, Prosigna had prognostic accuracy for 10-year distant metastasis-free survival and distant recurrence-free survival. In patients with LN-negative disease the results were statistically significant. In patients with LN-positive disease the results were statistically or borderline significant.

EndoPredict

4.16 There were 3 data sets that provided information on the prognostic ability of EndoPredict; all were reanalyses of RCT data. All studies were validation studies, and in 2 of the 3 studies patients had endocrine monotherapy.

4.17 In unadjusted analyses, EndoPredict (EPclin) had statistically significant prognostic accuracy for 10-year distant recurrence-free survival and interval in patients with LN-negative and LN-positive disease.

4.18 Adjusted analyses of TransATAC data show that EndoPredict (EPclin) had statistically significant increases in likelihood ratio for 10-year distant recurrence-free interval over clinical and pathological variables or tools, regardless of LN status.

IHC4 and IHC4+C

4.19 There were 12 data sets that provided information on the prognostic ability of IHC4 and IHC4+C: 6 reanalyses of RCT data and 6 reanalyses of routinely collected data. Most of the data related to the IHC4 score alone, without including clinical factors. One of the studies was based on the derivation cohort for IHC4, and therefore may have overestimated prognostic ability. The remaining studies were validation studies. Patients had endocrine monotherapy in only 2 studies, 1 of which was the derivation cohort study.

4.20 In unadjusted analyses, IHC4 had statistically significantly better prognostic performance in groups with high risk than in groups with low risk (defined by quartiles or tertiles), regardless of lymph node status. However, no studies reported survival or recurrence outcomes by risk group. Also, many used laboratory methods that differed from the derivation study methodology. In adjusted analyses, IHC4 had additional prognostic value over clinical and pathological factors in 3 studies, but patients had endocrine monotherapy in only 1 of these studies.

4.21 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 the NPI in LN-negative, but not LN-positive, disease.

Prediction of relative treatment effect

4.22 In addition to estimating the risk of recurrence, the ability of Oncotype DX and MammaPrint to predict which patients have disease that will respond to chemotherapy was explored in 7 data sets. The external assessment group (EAG) reviewed evidence in support of this.

Oncotype DX

4.23 In 5 data sets (2 reanalyses of RCT data and 3 observational studies) reported across 11 published references and 1 confidential manuscript, analyses assessed the ability of Oncotype DX to predict relative treatment effects for chemotherapy.

4.24 The 2 reanalyses of RCTs suggest that Oncotype DX may predict differences in relative treatment effects for chemotherapy. Hazard ratios for disease-free survival for patients having chemotherapy compared with those having no chemotherapy suggested that the greatest relative treatment effect was for patients in the Oncotype DX high-risk category. Unadjusted interaction tests between Oncotype DX risk group and relative treatment effects were mainly statistically significant. Adjusted interaction tests were statistically significant in an analysis of patients with HER2-negative, LN-negative disease, but in patients with LN-positive disease the interaction test was not significant when hormone receptor status was adjusted for. However, the data for the population with LN-negative disease came from the derivation cohort for Oncotype DX and may overestimate predictive performance.

4.25 Results from the 3 observational studies were mixed and at high risk from confounding. One reported a statistically significant interaction test but this was only adjusted for a limited number of factors. Two others reported hazard ratios for chemotherapy compared with no chemotherapy; 1 study in patients with intermediate Recurrence Score results, and another in patients with high Recurrence Score results. Both of these studies reported statistically non-significant results.

4.26 The RSPC algorithm incorporates Oncotype DX plus age, tumour size and grade. There was a non-significant interaction test result between relative chemotherapy treatment effects and RSPC risk group.

MammaPrint

4.27 Two studies reported the ability of MammaPrint to predict the relative treatment effects for chemotherapy. In a pooled analysis including patients with LN-negative and LN-positive disease, the effect of chemotherapy compared with no chemotherapy was statistically significant in the MammaPrint high-risk group but not in the low-risk group in unadjusted and adjusted analyses. Further, the interaction test for chemotherapy treatment and risk group was non-significant. In a pooled analysis of patients with LN-positive disease, there was a non-significant interaction between chemotherapy treatment and risk group.

Clinical utility

4.28 The EAG noted that the best evidence for clinical utility was an RCT of treatment guided by the test compared with treatment guided by the comparator. There were no clinical utility data available for EndoPredict, Prosigna or IHC4+C.

Oncotype DX

4.29 Five data sets, reported across 9 published references and 1 confidential manuscript, reported evidence on the clinical utility of Oncotype DX. These studies included the low-risk group from TAILORx. One further study did not meet the inclusion criteria (because of insufficient follow-up length), but presented subgroup data according to age, lymph node status and ethnicity, and was therefore discussed by the EAG. Studies generally reported different outcomes, making comparisons across studies difficult. All studies were judged to be of poor quality using the Cochrane risk of bias tool for RCTs.

4.30 In patients with LN-negative disease, using the test in clinical practice appeared to result in low rates of chemotherapy in patients with low risk (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 increased with increasing risk category, and were generally higher in patients with LN-positive disease. 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 because there were no comparator groups (patients who had treatment without Oncotype DX testing).

4.31 One study (TAILORx; Sparano et al. 2018) reporting evidence on clinical utility was published after completion of the diagnostics assessment report. This study was a prospective, partially randomised study in which patients with an Oncotype DX Recurrence Score result of 0 to 10 had endocrine therapy, patients with Recurrence Score results of 26 and above had endocrine therapy plus chemotherapy, and those with Recurrence Score results of 11 to 25 were randomised to have either endocrine therapy alone, or endocrine therapy plus chemotherapy. The cut-offs in this study were different to the cut-offs recommended by the company (less than 18, 18 to 30 and greater than 30; see section 3.15). The 2018 publication focused on the results from patients in the intermediate-risk group who were randomised to treatment. It reported that across all patients with Recurrence Score results of 11 to 25, there were no clinically relevant or statistically significant differences between those who had endocrine therapy alone and those who had chemotherapy plus endocrine therapy. Results for the primary end point of 9-year invasive disease-free survival were 84.3% with chemotherapy and 83.3% without chemotherapy; an absolute difference of 1.0% (hazard ratio [HR] 1.08, 95% confidence interval [CI] 0.94 to 1.24, p=0.26). The upper confidence interval was within the pre-specified non-inferiority margin (HR 1.322). Results for freedom from distant recurrence at 9 years were 95% with chemotherapy and 94.5% without chemotherapy; an absolute difference of 0.5% (HR 1.10, 95% CI 0.85 to 1.41, p=0.48). However, exploratory subgroup analyses suggested that chemotherapy may have an effect in some subgroups, such as those with Recurrence Score results of 21 to 25 and possibly Recurrence Score results of 16 to 20, particularly in people aged 50 or under. The EAG noted that no analysis was available for the subgroup of patients with Recurrence Score results of 11 to 25 and a modified Adjuvant! Online high risk score.

MammaPrint

4.32 Two studies reported evidence relating to the clinical utility of MammaPrint. MINDACT was a prospective, partially randomised study in which clinical risk was determined using a modified version of Adjuvant! Online. Patients with risk scores that disagreed from MammaPrint and modified Adjuvant! Online were randomised to chemotherapy or no chemotherapy. Of patients included in the study, 88% had HR-positive disease and 90% HER2-negative disease, therefore some patients were outside of the scope for this assessment. For the group who were high risk with modified Adjuvant! Online and low risk with MammaPrint, 5-year distant metastasis-free survival was 95.9% with chemotherapy and 94.4% without chemotherapy, a non-statistically significant absolute difference of 1.5% (adjusted hazard ratio for distant metastasis or death with chemotherapy compared with no chemotherapy, 0.78; 95% CI 0.50 to 1.21; p=0.27). For the group who were low risk with modified Adjuvant! Online and high risk with MammaPrint, 5-year distant metastasis-free survival was 95.8% with chemotherapy and 95.0% without chemotherapy, a non-statistically significant absolute difference of 0.8% (adjusted hazard ratio for distant metastasis or death with chemotherapy compared with no chemotherapy, 1.17; 95% CI 0.59 to 2.28; p=0.66). The EAG judged MINDACT to be at low risk of bias in terms of randomisation, allocation concealment and reporting. However, no details of blinding were reported.

4.33 Results from the RASTER study suggested 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 LN-negative disease was 97.0% for patients with low risk (15% had chemotherapy) and 91.7% for patients with high risk (81% had chemotherapy). In addition, MammaPrint provided additional prognostic information over Adjuvant! Online and the NPI, but not over the NHS PREDICT tool. The EAG judged RASTER to be at high risk of bias using the Cochrane risk of bias tool for RCTs.

Comparison of the tests with each other

4.34 There were 6 studies that compared more than 1 test: 4 reanalyses of

RCTs and 2 observational studies. Evidence shows that generally when a test placed more patients in a low-risk category than another test, the event-free survival in the low-risk group was reduced. Also, the tests generally performed differently in patients with LN-negative and LN-positive disease.

4.35 Thirteen studies reported data from microarray analyses on more than 1 test, however, these studies had methodological limitations. The comparability of test algorithms applied to microarray data with the commercial assays was unknown, so the generalisability of findings from microarray studies to the decision problem was uncertain. All the studies reported data on Oncotype DX and MammaPrint, and 2 also reported data on EndoPredict. No studies reported data on Prosigna or IHC4+C. The microarray studies generally supported the conclusions from studies using the commercial versions of the assays in suggesting that Oncotype DX, MammaPrint and EndoPredict can discriminate between patients with high and low risk regardless of LN status. In terms of additional prognostic performance of the tests over clinical and pathological variables, EndoPredict appeared to have the greatest benefit, followed by Oncotype DX and then MammaPrint. However, because of the methodological limitations, the EAG judged that these studies did not provide conclusive evidence of the superiority of 1 test over others.

4.36 The OPTIMA Prelim study, a UK-based feasibility phase of an RCT, analysed concordance between different tests. The study included Oncotype DX, MammaPrint, Prosigna and IHC4 plus 2 other tests. Out of the 4 in-scope tests, MammaPrint assigned the most patients to the low-risk category, but unlike the other 3 tests it does not have an intermediate category. When the low and intermediate categories were treated as 1 category for the 3 tests that have 3 risk groups, Oncotype DX assigned the most patients to this category, and MammaPrint the least. Kappa statistics indicated modest agreement between tests, ranging from 0.33 to 0.53. Also, across 5 tests in the study, only 39% of tumours were uniformly classified as either low/intermediate risk or high risk by all 5 tests. Of these, 31% were classified as low/intermediate risk by all tests and 8% were high risk by all tests. The study authors concluded that although the tests assigned similar proportions of patients to low/intermediate-risk and high-risk categories, test results for

an individual patient could differ markedly depending on which test was used.

Decision impact

4.37 The review of decision impact focused on studies done in the UK or the rest of Europe:

- 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.

4.38 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.

4.39 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. 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 a reduction of 31% to an increase of 8% for MammaPrint.

Anxiety and health-related quality of life

4.40 There were 6 studies that reported outcomes relating to anxiety (including worry and distress) and health-related quality of life. The lack 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 anxiety in some patients in some contexts, but generally there was little effect on health-related quality of life.

Cost effectiveness

Review of economic evidence

4.41 The EAG reviewed existing studies investigating the cost effectiveness of tumour profiling tests to guide treatment decisions in people with early breast cancer, and also did a detailed critique of the economic models and analyses provided by Agendia (MammaPrint), Genomic Health (Oncotype DX), and the chief investigator of a UK decision impact study (EndoPredict).

4.42 From the review, 26 studies were identified that had been published since the original assessment for diagnostics guidance 10. The models reported in the studies assessed the cost effectiveness of tumour profiling tests across different countries including the UK, the US, Canada, Mexico, Japan, Austria, Germany, France and the Netherlands. Most studies compared Oncotype DX (18 studies), MammaPrint (8 studies) or EndoPredict (1 study) with comparators such as Adjuvant! Online, the St Gallen guidelines, standard practice or other conventional diagnostic tools. There was variation between the analyses in the populations evaluated, the disease type and other patient characteristics.

4.43 There was a high level of consistency in the general modelling approach and structure, and several studies were based on a previously published model. Most 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 relative

treatment effects for chemotherapy.

Economic evaluation

4.44 None of the models identified in the literature review included all of the tests identified in the scope. Therefore, 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 without the use of the tumour profiling tests. The model used a lifetime time 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 year. Unit costs were valued at 2015/16 prices. The main source of evidence used to inform the analyses of Oncotype DX, Prosigna, IHC4+C and EndoPredict was a bespoke analysis of TransATAC provided by the study investigators. This was limited to UK data on patients with hormone receptor-positive, HER2-negative disease with 0 to 3 positive lymph nodes to match the scope for this assessment. Because this study did not include MammaPrint, MINDACT was used as the basis for evaluating the cost effectiveness of MammaPrint. PREDICT scores were not available in either data set, and so this tool could not be considered as a comparator or used to determine different risk subgroups. Therefore, the comparator for Oncotype DX, Prosigna, IHC4+C and EndoPredict was current practice (various tools and algorithms), and the comparator for MammaPrint was a modified version of Adjuvant! Online.

Model structure

4.45 The hybrid decision tree–Markov model was based on the model previously developed by Ward et al. (2013). The decision tree component of the model classified patients in the current practice group (no test) and the tumour profiling test group as high, intermediate and low risk. For EndoPredict and MammaPrint, the intermediate-risk category was excluded because the test provides results in terms of high and low risk only. In 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 EndoPredict and MammaPrint, 4 groups were used because there was no intermediate-risk category. Each group was 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 had chemotherapy.

4.46 Each Markov node included 4 health states: distant recurrence-free; distant recurrence; long-term adverse events (acute myeloid leukaemia [AML]); and dead. Patients entered the model in the distant 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 having adjuvant chemotherapy. The treatment effect for 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 had adjuvant chemotherapy.

Model inputs

4.47 The risk classification probabilities used in the model for Oncotype DX, Prosigna, IHC4+C and EndoPredict were from the bespoke data analysis of TransATAC, which only included postmenopausal women. For MammaPrint, they were from MINDACT.

4.48 The probability of developing distant metastases in each group and risk category was based on 10-year recurrence-free interval data from the bespoke data analysis of TransATAC for Oncotype DX, Prosigna, IHC4+C and EndoPredict. For MammaPrint the probability of developing distant metastases was based on an adjusted analysis of 5-year distant metastasis-free survival data from MINDACT. The model assumed that the risk of distant metastases between 10 and 15 years was halved, and after 15 years was 0.

4.49 The probability of having chemotherapy in the current practice group and in the tumour profiling test groups was taken from the sources in table 1.

Table 1 Source for post-test probability of having chemotherapy

Current practice group

Population	Source	Proportion of patients having chemotherapy
LN-negative, NPI≤3.4	NCRAS data set	0.07
LN-negative, NPI>3.4	Genomic Health access scheme data set	0.43
LN-positive (1–3 nodes)	NCRAS data set	0.63
Overall population (MammaPrint)	Expert opinion	0.47

Abbreviations: LN, lymph node; NCRAS, National Cancer Registration and Analysis Service; NPI, Nottingham Prognostic Index; UKBCG, UK breast cancer group.

The Genomic Health access scheme data set is based on the access scheme operated by NHS England and is a result of the research recommendation from NICE's original diagnostics guidance 10.

For the proportion of patients having chemotherapy, the low, intermediate and high risks are combined.

3-level tests (Oncotype DX, Prosigna and IHC4+C)

Population	Source	Proportion of patients having chemotherapy (low risk)	Proportion of patients having chemotherapy (intermediate risk)	Proportion of patients having chemotherapy (high risk)
LN-negative, NPI≤3.4	UKBCG survey data	0.00	0.20	0.77
LN-negative, NPI>3.4	Genomic Health access scheme data set	0.01	0.33	0.89

Population	Source	Proportion of patients having chemotherapy (low risk)	Proportion of patients having chemotherapy (intermediate risk)	Proportion of patients having chemotherapy (high risk)
LN-positive (1–3 nodes)	Loncaster et al. (2017) node-positive estimates	0.08	0.63	0.83

Abbreviations: LN, lymph node; NPI, Nottingham Prognostic Index; UKBCG, UK breast cancer group.

2-level tests (EndoPredict and MammaPrint

Population	Source	Proportion of patients having chemotherapy (low risk)	Proportion of patients having chemotherapy (intermediate risk)	Proportion of patients having chemotherapy (high risk)
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

4.50 In the base-case analysis, the relative treatment effect for 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 a 10-year relative risk of distant recurrence was estimated as 0.76 for chemotherapy compared with no chemotherapy (Early breast cancer trialists' collaborative group 2012), and was assumed to apply to the groups with LN-negative and LN-positive disease. For MammaPrint the 10-year relative risk of distant recurrence was estimated to be 0.77 (MINDACT) for chemotherapy compared with no chemotherapy.

4.51 In sensitivity analyses the effect of assuming that Oncotype DX could predict relative treatment effects for chemotherapy was explored, based on the B20 study by Paik et al. (2006) and the SWOG-8814 study by Albain et al. (2010). For the group with LN-negative disease, the 10-year relative risks of distant recurrence with chemotherapy compared with no chemotherapy were 1.31, 0.61 and 0.26 for the low, intermediate and high-risk categories respectively. For the group with LN-positive disease, the 10-year relative risks of relapse with chemotherapy compared with no chemotherapy were 1.02, 0.72 and 0.59 respectively. It is possible that the no-chemotherapy arm of B20 may have overestimated the difference in response rates between low and high-risk patients, because this arm was the derivation set for Oncotype DX. Therefore, additional sensitivity analyses in the group with LN-negative disease explored the impact of varying the relative chemotherapy treatment effect between risk groups on the incremental cost-effectiveness ratios (ICERs). Hazard ratios were based on naive indirect comparisons of the chemotherapy arms from the B20 study and the no-chemotherapy arms from the B14 study (estimated hazard ratios for treatment effects with chemotherapy compared with no chemotherapy were 0.64, 0.75 and 0.35 for the low, intermediate and high-risk categories respectively), and the chemotherapy arms of the B20 study and the no-chemotherapy arms of the TransATAC study (hazard ratios for treatment effects with chemotherapy compared with no chemotherapy were 0.86, 0.88 and 0.49 for the low, intermediate and high-risk categories respectively).

4.52 Survival following distant recurrence was based on a median of 40.1 months from Thomas et al. (2009). 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 following distant metastases was assumed to be the same across the different subgroups and across each test risk group.

4.53 The model assumed that 10.5% of patients entering the distant recurrence health state had previously had local recurrence, based on de Bock et al. (2009). The 6-month probability of developing 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 gave a 6-month probability of death

following AML of 0.53. Additional sensitivity analyses explored the effect of including congestive heart failure (average net lifetime QALY loss of 0.0385 and average net lifetime cost saving of £2 from Hall et al. 2017, using an excess congestive heart failure risk relative to that of the general population), permanent hair loss (disutility of 0.04495 from Nafees et al. 2008 applied to 15% of all patients having chemotherapy) and peripheral neuropathy (disutility of 0.02 from Shiroiwa et al. 2009 applied to 12% of all patients having chemotherapy) in the model.

Costs

4.54 The costs of the tumour profiling tests were based on company prices (see table 2).

Table 2 Test prices

Test	List price	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 doing the test in an NHS laboratory, which includes the laboratory costs (£240), the Prosigna kit (£1,650) and the nCounter system (£194,600) and is based on 2,500 samples per lifetime of the nCounter system). Commercial-in-confidence discounted test costs were used in scenario analyses to account for the access proposal.
EndoPredict	£1,500	Tests carried out in Myriad's laboratory in Munich. Commercial-in-confidence discounted test costs were used in scenario analyses to account for the access proposal.
IHC4	£203	The cost was based on 2014 prices. The total cost of the test (£198) was uplifted to current prices using the hospital and community health services indices.
MammaPrint	£2,326	Converted from Euros to UK pounds sterling, assuming an exchange rate of 1 British pound to 1.15 Euros.

4.55 The costs associated with adjuvant chemotherapy were from a previous costing analysis of the OPTIMA Prelim trial (Hall et al. 2017). The weighted mean cost of adjuvant chemotherapy acquisition, delivery and toxicity was estimated to be £3,145 per course.

4.56 All surviving patients had endocrine therapy for a period of between 5 and 8 years. Costs of endocrine therapy were taken from the British national formulary (2017). In addition, 30% of women with early breast cancer had 4 mg of bisphosphonates (zoledronic acid) by intravenous infusion every 6 months for up to 3 years, at a cost of £58.50, excluding administration.

4.57 All patients had 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.

4.58 Costs associated with treating local recurrence were taken from Karnon et al. (2007) and uplifted to current prices (£13,913). This was applied as a once-only cost to 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

4.59 Health utilities were taken from published studies (see table 3).

Table 3 Health utilities applied in the base case

Health state / event	Duration applied in model	Mean	Standard error	Source
Recurrence-free	Indefinite	0.824	0.002	Lidgren et al. 2007

Health state / event	Duration applied in model	Mean	Standard error	Source
Disutility distant metastases	Indefinite	0.14	0.11	Calculated from Lidgren et al. 2007
Local recurrence	Once-only QALY loss applied on transition to distant recurrence state	−0.108	0.04 (assumed)	Campbell et al. 2011
Chemotherapy AEs	Once-only QALY loss applied in first cycle	−0.038	0.004	Campbell et al. 2011
AML	Indefinite	0.26	0.04 (assumed)	Younis et al. 2008

Abbreviations: AEs, adverse events; AML, acute myeloid leukaemia; QALY, quality-adjusted life year.

Base-case results

4.60 The following key assumptions were applied in the base-case analysis:

- Clinicians interpreted each of the 3-level tests in the same way (for example, an Oncotype DX high-risk Recurrence Score result would lead to the same chemotherapy decision as a Prosigna high-risk score).
- Clinicians interpreted 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 treatment effect for adjuvant chemotherapy was the same across all risk score categories for all tests.
- The prognosis of patients with AML and the costs and QALYs accrued within the AML state were independent of whether they had previously developed distant metastases.
- A disutility associated with adjuvant chemotherapy was applied once during the first model cycle only (while the patient is taking the regimen).

- Costs associated with endocrine therapy, bisphosphonates, follow-up appointments and mammograms were assumed to differ according to time since model entry.
- The model assumed that people entered at an age of around 60 years.

4.61 In the subgroup with LN-negative disease and a NPI of 3.4 or less, compared with current practice, the probabilistic model gave ICERs of:

- £147,419 per QALY gained (EndoPredict)
- £122,725 per QALY gained (Oncotype DX)
- £91,028 per QALY gained (Prosigna)
- £2,654 per QALY gained (IHC4+C).

4.62 In the subgroup with LN-negative disease and a NPI of more than 3.4, compared with current practice, the probabilistic model gave ICERs of:

- £46,788 per QALY gained (EndoPredict)
- £26,058 per QALY gained (Prosigna)
- Oncotype DX was dominated by current practice (that is, it was more expensive and less effective)
- IHC4+C was dominant over current practice (that is, it was less expensive and more effective).

4.63 In the population with LN-positive disease, compared with current practice, the probabilistic model gave ICERs of:

- £28,731 per QALY gained (Prosigna)
- £21,458 per QALY gained (EndoPredict)
- Oncotype DX was dominated by current practice
- IHC4+C was dominant over current practice.

4.64 In the overall MINDACT population, MammaPrint compared with 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.

4.65 The risk classification probabilities and the probability of having chemotherapy were combined in the model to estimate chemotherapy use with and without tumour profiling. The modelled chemotherapy use in the base case is shown in table 4.

Table 4A Modelled chemotherapy use with and without tumour profiling: Oncotype DX

Test, subgroup compared with current practice	Chemotherapy use with tumour profiling	Chemotherapy use with no tumour profiling	Net change
LN0 NPI≤3.4	0.076	0.072	0.004
LN0 NPI>3.4	0.273	0.430	−0.157
LN+ (1–3 nodes)	0.337	0.627	−0.290

Abbreviations: LN0, lymph node negative; LN+, lymph node positive, mAOL, modified Adjuvant! Online; NPI, Nottingham Prognostic Index.

Table 4B Modelled chemotherapy use with and without tumour profiling: IHC4+C

Test, subgroup compared with current practice	Chemotherapy use with tumour profiling	Chemotherapy use with no tumour profiling	Net change
LN0 NPI≤3.4	0.030	0.072	−0.042
LN0 NPI>3.4	0.355	0.430	−0.075
LN+ (1–3 nodes)	0.554	0.627	−0.073

Table 4C Modelled chemotherapy use with and without tumour profiling: Prosigna

Test, subgroup compared with current practice	Chemotherapy use with tumour profiling	Chemotherapy use with no tumour profiling	Net change
LN0 NPI≤3.4	0.075	0.072	0.003
LN0 NPI>3.4	0.435	0.430	0.005

Test, subgroup compared with current practice	Chemotherapy use with tumour profiling	Chemotherapy use with no tumour profiling	Net change
LN+ (1–3 nodes)	0.709	0.627	0.082

Table 4D Modelled chemotherapy use with and without tumour profiling: EndoPredict

Test, subgroup compared with current practice	Chemotherapy use with tumour profiling	Chemotherapy use with no tumour profiling	Net change
LN0 NPI≤3.4	0.140	0.072	0.068
LN0 NPI>3.4	0.438	0.430	0.008
LN+ (1–3 nodes)	0.603	0.627	−0.024

Table 4E Modelled chemotherapy use with and without tumour profiling: MamaPrint

Test, subgroup compared with current practice	Chemotherapy use with tumour profiling	Chemotherapy use with no tumour profiling	Net change
MINDACT overall population	0.319	0.466	−0.148
mAOL high risk	0.445	0.772	−0.327
mAOL low risk	0.191	0.159	0.033

Probabilistic sensitivity analyses

4.66 The cost-effectiveness planes from the probabilistic sensitivity analyses showed considerable uncertainty in the cost-effectiveness estimates.

4.67 In the subgroup with LN-negative disease and a NPI of 3.4 or less, the only test with a non-zero probability of producing more net benefit than current practice at maximum acceptable ICERs of £20,000 and £30,000 per QALY gained was IHC4+C.

4.68 In the subgroup with LN-negative disease and a NPI of more than 3.4, at a maximum acceptable ICER of £20,000 per QALY gained, IHC4+C had a probability of 0.69 of being cost effective compared with current practice. For EndoPredict, Oncotype DX and Prosigna, the probability

that the test was cost effective compared with current practice at this threshold was 0.24 or less. In the same subgroup, at a maximum acceptable ICER of £30,000 per QALY gained, IHC4+C had a probability of 0.67 and Prosigna had a probability of 0.60 of being cost effective compared with current practice. Oncotype DX had a probability of 0.04 and EndoPredict had a probability of 0.26 of being cost effective compared with current practice.

4.69 In the subgroup with LN-positive disease, IHC4+C had probabilities of 0.95 and 0.94 of being cost effective compared with current practice at maximum acceptable ICERs of £20,000 and £30,000 per QALY gained respectively. In the same subgroup, the probabilities of EndoPredict producing more net benefit than current practice were 0.44 and 0.73, at maximum acceptable ICERs of £20,000 and £30,000 per QALY gained respectively. For Prosigna the probabilities were 0.24 and 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).

4.70 In the overall MINDACT population and in the subgroups, the probability that MammaPrint would be cost effective compared with current practice at maximum acceptable ICERs of £20,000 and £30,000 per QALY gained was approximately 0.

Deterministic sensitivity analyses

4.71 The EAG did deterministic sensitivity analyses, testing a wide range of plausible values of key parameters.

4.72 Deterministic sensitivity analysis results for Oncotype DX compared with current practice were:

- Subgroup with LN-negative disease and a NPI of 3.4 or less: ICERs remained over £34,000 per QALY gained across all analyses.

- Subgroup with LN-negative disease and a NPI of more than 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 relative treatment effects for chemotherapy. In this analysis, Oncotype DX dominated current practice.
- Population with LN-positive disease: Oncotype DX remained dominated across most analyses. The exceptions were when Oncotype DX was assumed to predict relative treatment effects for chemotherapy (it was dominant), and when the cost of chemotherapy was doubled (£3,700 saved per QALY lost).

4.73 Deterministic sensitivity analysis results for IHC4+C compared with current practice were:

- Subgroup with LN-negative disease and a NPI of 3.4 or less: ICERs 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). Also, IHC4+C dominated current practice when the cost of chemotherapy was doubled.
- Subgroup with LN-negative disease and a NPI of more than 3.4: IHC4+C dominated current practice or had an ICER below £6,000 per QALY gained across all scenarios.
- Population with LN-positive disease: IHC4+C dominated current practice across all but 1 scenario. When the probability of having chemotherapy was based on the UK breast cancer group (UKBCG) survey the ICER was £1,929 per QALY gained.

4.74 Deterministic sensitivity analysis results for Prosigna compared with current practice were:

- Subgroup with LN-negative disease and a NPI of 3.4 or less: ICERs were greater than £71,000 per QALY gained across all analyses.
- Subgroup with LN-negative disease and a NPI of more than 3.4: ICERs were below £34,000 per QALY gained across all analyses.
- Population with LN-positive disease: ICERs were below £38,000 per QALY gained across all analyses.

4.75 Deterministic sensitivity analysis results for EndoPredict compared with current practice were:

- Subgroup with LN-negative disease and a NPI of 3.4 or less: ICERs remained greater than £91,000 per QALY gained across all analyses.
- Subgroup with LN-negative disease and a NPI of more than 3.4: ICERs 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 having chemotherapy (£25,250 per QALY gained), and when Cusumano et al. (2014) was used to inform the probability of having chemotherapy based on the EndoPredict test result (£26,689 per QALY gained).
- Population with LN-positive disease: ICERs remained below £30,000 per QALY gained across all scenarios.

4.76 Deterministic sensitivity analysis results for MammaPrint compared with current practice were:

- Overall MINDACT population: ICERs were estimated to be greater than £76,000 per QALY gained across all scenarios.
- Modified Adjuvant! Online high-risk subgroup: MammaPrint was dominated by current practice across almost all scenarios.
- Modified Adjuvant! Online low-risk subgroup: ICERs were greater than £161,000 per QALY gained across all analyses.

4.77 After consultation, the EAG did more deterministic sensitivity analyses varying the estimated relative risk of distant recurrence associated with chemotherapy, which was assumed to be 0.76 in the base case. Results showed that as the relative risk moved from 0.6 to 0.9, the tests became less cost effective.

5 Committee discussion

5.1 The committee discussed current practice for making adjuvant chemotherapy prescribing decisions. The clinical experts explained that NHS clinical practice has changed since NICE's diagnostics guidance 10 was published in 2013. The PREDICT tool is now used by many NHS trusts rather than the Nottingham Prognostic Index (NPI). Adjuvant! Online is not currently available. The committee also heard that Oncotype DX is currently used in NHS clinical practice and may be used for a broader group than the population defined in the original diagnostics guidance 10, that is, people with oestrogen receptor (ER)-positive, human epidermal growth factor receptor 2 (HER2)-negative and lymph node (LN)-negative early breast cancer who are assessed as being at intermediate risk using existing risk assessment tools.

5.2 The committee discussed the potential benefits of the tumour profiling tests for people with early breast cancer who are deciding whether to have adjuvant chemotherapy. It acknowledged that chemotherapy is an unpleasant treatment associated with short-term physical, emotional and financial effects, and also long-term consequences such as infertility and increased risk of cardiomyopathy and leukaemia. The committee heard that there is potential benefit for people with cancer identified as being at low clinical risk, when test results suggest a high risk of distant recurrence. These people could potentially benefit from chemotherapy. It also heard that there is potential benefit for people with cancer categorised as high clinical risk, when test results suggest a low risk of distant recurrence. The committee heard that these people could decide not to have chemotherapy, therefore avoiding toxic side effects and effects on fertility. They could potentially resume normal daily activities earlier, although some may wish to have chemotherapy regardless of the test result. However, the committee noted that the claimed benefits of the tests depend on them having sufficient accuracy and discrimination to correctly classify risk and provide valid clinical information. The clinical experts explained that the additional clinical information provided by the tests may help people discuss further treatment options. This information is particularly helpful for people with cancers identified as intermediate clinical risk when the decision to offer chemotherapy is unclear. However,

the final decision to recommend a course of adjuvant chemotherapy would always take into account the person's circumstances and preferences.

Clinical effectiveness

5.3 The committee considered the prognostic ability of the tumour profiling tests. It noted that for people with LN-negative disease, all the tests had statistically significant prognostic accuracy over clinical and pathological features or risk assessment tools such as the NPI. It also noted that for people with LN-positive disease, results for prognostic ability were more variable but all tests except IHC4+C showed statistically significant or borderline statistically significant prognostic ability over clinical and pathological features or risk assessment tools. The external assessment group (EAG) explained that there were concerns about bias in all studies reporting prognostic ability. This was because in many of the studies some or all patients had chemotherapy or patients who had not had chemotherapy were selected for analyses. Also, most studies excluded tumour samples with insufficient tissue or missing clinical and pathological data, and some studies included patients who had hormone receptor-negative or HER2-positive disease. The committee concluded that despite the potential spectrum bias, the evidence suggested that all the tumour profiling tests have the ability to predict the risk of distant recurrence in the population included in the assessment. It also concluded that the evidence was weaker in the group with LN-positive disease than in the group with LN-negative disease.

5.4 The committee considered the evidence on micrometastases. The EAG explained that 2 studies with Oncotype DX reported subgroup data on people with micrometastases, but no studies with the other tests reported such data. The EAG noted that in patients with micrometastases and a Recurrence Score result of less than 18, outcomes were more similar to those in patients with LN-negative disease than in those with LN-positive disease. However, in patients with micrometastases and a Recurrence Score result of more than 30, outcomes were more similar to those in patients with LN-positive disease. Results were variable in patients with Recurrence Score results between 18 and 30. The EAG noted that the data were uncertain

because of the high risk of confounding. The clinical experts explained that micrometastatic disease is classified as LN-positive disease but treated as LN-negative disease for clinical and shared decision making. In clinical practice some centres send samples from patients with micrometastases for Oncotype DX testing, but others do not. A clinical expert also explained that for patients with micrometastatic disease who have reasons to avoid chemotherapy, such as being older or having comorbidities, tumour profiling tests would be helpful. The committee noted that the ongoing OPTIMA study in LN-positive disease excludes patients with micrometastases, unless the tumour size is 20 mm or more. The EAG reviewed whether all studies in the diagnostics assessment report included or excluded patients with micrometastatic disease. It found that in TransATAC micrometastases were not assessed and therefore the disease was treated as LN-negative, but other studies did not report whether patients with micrometastatic disease were included or not. A company representative explained that in the MINDACT study, micrometastases measuring 0.2 mm to 2 mm were classified as LN-positive and isolated tumour cells were classified as LN-negative. The clinical experts judged that, on balance, patients with micrometastases were likely to have been included in the studies as having LN-negative disease. The committee concluded that tumour profiling tests should be available as an option for people fulfilling the recommendation requirements and who have micrometastatic disease. Discussion within the multidisciplinary team may be particularly helpful for this group.

5.5 The committee considered the evidence on whether the tumour profiling tests can predict relative treatment effects associated with chemotherapy. The clinical experts stated that it is likely some patients could have a greater relative treatment effect from chemotherapy than others, for example, patients with hormone receptor-positive cancer that is not sensitive to endocrine therapy, but evidence is not available to support this. The EAG explained that the only evidence available to show a relative treatment effect for chemotherapy across different risk groups was for Oncotype DX, and the evidence included in the diagnostics assessment report was weak because it was at high risk of bias from potential confounding. The results of interaction tests (which show whether the tumour profiling test was able to predict a different treatment effect by risk group) in the adjusted analysis in the B20 study

by Paik et al. (2006; LN-negative disease) remained statistically significant when adjusting simultaneously for clinical and pathological variables. However, the EAG also explained that the difference in relative treatment effects for chemotherapy in the B20 study may be overestimated because this was the Oncotype DX derivation data set. In the SWOG-8814 study by Albain et al. (2010; LN-positive disease) the results of the interaction tests remained statistically significant when adjusting for some individual clinical and pathological variables, but there was no analysis that adjusted for these simultaneously, and the test was non-significant when adjusted for Allred-quantified ER status. The clinical experts explained that hormone receptor status may also predict relative treatment effects for chemotherapy. The committee considered that if all known clinical and pathological variables were included in the analyses of SWOG-8814 then it was likely that the results of the interaction test would no longer be statistically significant. This suggested highly uncertain relative treatment effects for chemotherapy according to the results of the tumour profiling tests for this group with LN-positive disease. The committee concluded that the evidence on the extent to which tumour profiling tests are able to predict relative treatment effects for chemotherapy is highly uncertain, but there may be some differences between Oncotype DX risk groups. The committee noted that no data were available to assess a difference in relative treatment effects for chemotherapy for EndoPredict, IHC4+C and Prosigna risk groups. However, it considered that it would be unethical to do a randomised controlled trial looking at the benefit of chemotherapy compared with endocrine therapy in patients with a clinically low or high risk of distant recurrence. It also noted that data on MammaPrint suggested no difference in relative treatment effects for chemotherapy.

5.6 The committee considered the evidence on clinical utility, that is, data from studies which assessed the ability of the tumour profiling tests to affect patient outcomes. It discussed the recently published results from TAILORx on Oncotype DX (see section 4.31), which showed that across all patients with Recurrence Score results of 11 to 25, there were no clinically relevant or statistically significant differences between those who had endocrine therapy alone and those who had chemotherapy plus endocrine therapy. The EAG noted that some subgroups, such as those with Recurrence Score results of 21 to 25 and those aged 50 or under,

had results with confidence intervals above the non-inferiority margin, which suggested that there could be a clinically relevant difference in these subgroups. The EAG noted that patients included in TAILORx had hormone receptor-positive, HER2-negative and LN-negative disease, and met National Comprehensive Cancer Network guidelines for recommendation or consideration of chemotherapy. In the group with Recurrence Score results of 11 to 25, 73% to 74% were clinically low risk according to modified Adjuvant! Online. The committee acknowledged that the results from TAILORx may not be generalisable to clinical practice in the UK because the population who had chemotherapy in the study would not be routinely offered chemotherapy in an NHS pathway. It noted that, to fully understand the implications of the study for UK practice, a subgroup analysis of the TAILORx data would be needed investigating the performance of the test in predicting chemotherapy benefit for patients eligible for chemotherapy in the UK. The EAG explained that this had been requested but was not made available. The committee concluded that in principle TAILORx is an important piece of evidence showing the effectiveness of gene profiling to guide adjuvant chemotherapy decisions in breast cancer. But it is uncertain how applicable it is to people with breast cancer in the UK who are considering adjuvant chemotherapy treatment.

5.7 The committee considered the evidence on clinical utility for the other tests. It noted that the only other test with evidence from randomised controlled trials was MammaPrint (the MINDACT study). The committee noted that none of the other tumour profiling tests (EndoPredict, IHC4+C and Prosigna) had similar evidence of clinical utility, but it was aware that this evidence was being collected for Prosigna (see section 5.27). The committee noted that MINDACT (see section 4.32) was a well-designed study. The results suggested that patients with high clinical risk and MammaPrint low-risk scores can forgo chemotherapy without a statistically significant increase in the 5-year risk of distant recurrence. However, a clinical expert explained that the risk of recurrence often continues beyond 5 years and noted that the MINDACT authors (Cardoso et al. 2016) stated that long-term follow-up and outcome data will be essential. These data are being collected and a 10-year follow-up analysis is planned. The committee concluded that none of these tests had strong enough evidence to demonstrate an effect on subsequent

patient outcomes.

5.8 The committee was encouraged by the availability of the data set provided in confidence to NICE by Genomic Health. The data set was based on the access scheme operated by NHS England, which provided real world evidence on the use of adjuvant chemotherapy in the NHS following testing with Oncotype DX for the population included in the scope for this assessment. The committee noted that the total number of patients in the data set appeared to be much larger than the number of patients with complete data in the population of interest, and that the advice from clinical experts (see section 5.1) was that the test had been used on a wider group of patients in practice. The committee also noted that the publication on TAILORx (Sparano et al. 2018) may influence chemotherapy decision making in people with a Recurrence Score result of 11 to 25, and therefore the data set may not represent clinical decision making in this group. The committee concluded that the access scheme data set was an important piece of real world evidence for use in the economic model, but that more complete data could have been collected and reported, and that it will be important to continue the data collection to capture the influence of TAILORx. It also concluded that future data collection should be done as part of a national database, rather than by individual companies, to increase transparency and enable it to be linked to outcome data (see section 5.29).

5.9 The committee discussed the analytical validity of IHC4+C. The EAG explained that the evidence has developed since diagnostics guidance 10 was published. The committee noted that the data showed good correlation between different centres when scoring and staining were assessed separately for measurement of the Ki67 marker, which had been achieved with training. But it also noted that when studies looked at staining and scoring combined, the correlation between centres decreased substantially. A clinical expert noted that different antibody clones are available for testing Ki67, ER and progesterone receptor (PR) status. Different studies used different antibody clones which means that the studies are not directly comparable. The committee heard that different methods of assessing ER and PR receptors may be needed for IHC4+C compared with those already used routinely, which may introduce additional complexity. The committee

concluded that because of these issues, the reproducibility of IHC4+C was poor. It also concluded that if this test were to be developed further, the antibody clones used in the assays for ER, PR and Ki67 should be specified, and there would need to be substantial investment in staff training and quality assurance.

Cost effectiveness

5.10 The committee discussed the assumptions and inputs used in the model, and considered the extensive stakeholder comments on the model and the EAG responses to these comments. It noted that a specific analysis of the TransATAC data was used for risk classification probabilities and for distant recurrence rates based on test result for Oncotype DX, EndoPredict (EPclin), Prosigna and IHC4+C. The results from this specific analysis of the data set have now been published (Sestak et al. 2018). The EAG explained that this data source was chosen because it included data on 4 of the 5 tests of interest and was specific to the population included in the scope (patients with hormone receptor-positive, HER2-negative disease). The committee heard that although the TransATAC data were slightly older and some patients were not candidates for chemotherapy, the patient characteristics matched well with the more recent MINDACT study. The alternative would be to use different data sources for each test, which would have introduced additional uncertainty and complexity. Also, the group with LN-negative disease could not have been split according to level of clinical risk. The EAG described the limitations of using data from the TAILORx study (Sparano et al. 2018) for the health economic analysis. It also explained that the distant recurrence-free rates from the TransATAC analysis used in the model were consistent with results from other studies (B14, B20, TAILORx, MD Anderson, Clalit, Memorial Sloan Kettering, SEER and WSG PlanB) both when grouped separately by clinical risk and when all clinical risk groups were pooled together. The committee concluded that the TransATAC analysis had some limitations, but was the best available data for use in the model.

5.11 The committee considered the data on pre- and post-test chemotherapy decisions used in the model. The EAG explained that for 3-level tests (tests with low, intermediate and high-risk categories [IHC4+C, Oncotype

DX, Prosigna]), data on pre- and post-test chemotherapy decisions for the group with LN-negative disease and a NPI of more than 3.4 were taken from the Genomic Health access scheme data set (see section 5.8). For other clinical risk subgroups with the 3-level tests, and for all clinical risk subgroups with 2-level tests (tests with low and high-risk categories; EndoPredict, MammaPrint), data on pre-test chemotherapy decisions were taken from different sources to data on post-test chemotherapy decisions. There were also very limited UK data for these groups. The committee considered the modelled impact of these data on chemotherapy use, and noted that although clinical and patient experts thought that the main benefit of the tests was in avoiding unnecessary chemotherapy, most tests were estimated to increase chemotherapy use at least in some subgroups (see section 4.49). The committee concluded that there was much more uncertainty around chemotherapy decision making for the 2-level tests, and for the subgroups who were not included in the original NICE recommendation on tumour profiling tests (LN-negative disease and a NPI of 3.4 or less, and LN-positive disease).

5.12 The committee considered how adjuvant chemotherapy treatment effects had been applied in the economic model, particularly the relative treatment effects of chemotherapy between the risk groups predicted by the tumour profiling tests. It noted its earlier conclusion that the evidence on whether tumour profiling tests can predict relative treatment effects for chemotherapy is highly uncertain, but that there may be some differences between Oncotype DX risk groups (see section 5.5). It agreed that for EndoPredict, IHC4+C and Prosigna, no evidence was available to show a difference in relative treatment effects of chemotherapy across risk groups, and that data on MammaPrint suggested no difference in relative treatment effects. Therefore for these tests it was appropriate to assume the same relative risk of distant recurrence across all test risk categories. The EAG noted that a relative risk of distant recurrence for chemotherapy compared with no chemotherapy of 0.76 estimated from data reported in a large meta-analysis by the Early Breast Cancer Trialists' Collaborative Group was used in the base case, and that this value had been varied between 0.6 and 0.9 in sensitivity analyses. The committee acknowledged that the ICERs were sensitive to this assumption, increasing as the relative risk moved from 0.6 to 0.9. It concluded that, although the true treatment effect is unknown, the

relative risk was unlikely to be 0.9 or more.

5.13 The committee considered stakeholder comments submitted during the first consultation suggesting that Oncotype DX has the ability to predict which patients have disease that will respond to chemotherapy. The EAG noted that in response to the comments it had done additional exploratory analyses for Oncotype DX to show the impact on the incremental cost-effectiveness ratios (ICERs) if a smaller relative treatment effect than that taken from the B20 study (Paik et al. 2006) was applied in the model in the group with LN-negative disease and a NPI of more than 3.4 (see section 4.51). The EAG noted that the hazard ratios used in these analyses were from comparisons of independent arms of trials and were therefore very uncertain. The EAG also said that using hazard ratios calculated from the B20 and the B14 (Paik et al. 2004) studies resulted in an ICER of around £24,000 per quality-adjusted life year (QALY) gained for Oncotype DX compared with current practice. Using hazard ratios calculated from the B20 and TransATAC studies resulted in an ICER of around £8,000 per QALY gained. Based on the results of the Sparano et al. 2018 publication on TAILORx, the EAG repeated the analysis incorporating an additional assumption of 0 chemotherapy benefit for patients in the Oncotype DX low Recurrence Score result category. It noted that this analysis was 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. When this assumption was included in the analysis using B20, the analysis using B20 and B14, and the analysis using B20 and TransATAC, the ICERs were below £4,000 per QALY gained. The committee concluded that although these analyses were associated with considerable uncertainty, they gave an indication of Oncotype DX's likely cost effectiveness if the relative treatment effects for chemotherapy did differ between Oncotype DX risk groups, but not to the extent reported in the Paik et al. (2006) study.

5.14 The committee considered stakeholder comments submitted during the first consultation suggesting that chemotherapy adverse events had not been adequately captured in the economic model; in particular, congestive heart failure, permanent hair loss and peripheral neuropathy. The EAG noted that in response to the comments it had done additional

exploratory analyses to include these adverse events in the model. Congestive heart failure was added into the model by incorporating estimated lifetime QALY losses and costs taken from an alternative model (Hall et al. 2017). Hair loss and peripheral neuropathy were incorporated using a disutility applied to a proportion of the population for the lifetime of the model. The EAG highlighted the considerable limitations of these analyses, and noted that for tests that increased chemotherapy use in some subgroups, the ICERs became less favourable. The committee noted that including additional adverse events in the model did reduce some of the ICERs, but not enough to change the conclusions. It also noted a further EAG analysis, which suggested that for tests that reduced chemotherapy use but were not cost effective, the QALY gain from avoiding adverse events would have to be in the range of 1.1 to 1.3 to result in cost-effective ICERs. The committee concluded that it was important to consider potential adverse events that could be caused by chemotherapy. However, the reduction in adverse events from reduced chemotherapy use, although beneficial for patients, was unlikely to affect its conclusions on the cost effectiveness of the tumour profiling tests based on the EAG's analysis.

5.15 The committee considered other assumptions used in the model such as the cost of chemotherapy and how the risk of distant recurrence was applied over time. The EAG explained that there was some uncertainty around these inputs, but all had been tested in sensitivity analyses. The committee concluded that the assumptions and inputs used in the model were reasonable, but they were associated with considerable uncertainty because of the limitations in the data that underpinned them.

5.16 The committee noted its discussion on current practice (see section 5.1) and considered the absence of comparisons of the tumour profiling tests with the PREDICT tool. The EAG explained that in the model it was not possible to compare the tumour profiling tests with PREDICT, or to define the clinical risk groups using PREDICT, because relevant data were not available. The committee noted that the comparisons in the model did not fully reflect current NHS clinical practice, which led to uncaptured uncertainty in the model results. The committee concluded that research on tumour profiling tests should include comparisons with PREDICT (see section 5.26) so that the cost effectiveness of the tests relative to

current practice can be fully assessed in future.

5.17 The committee considered the subgroups that were included in the model, that is, people with LN-negative disease and a NPI of 3.4 or less, people with LN-negative disease and a NPI of more then 3.4, and people with LN-positive disease. It noted its earlier conclusion that the evidence suggested that all the tumour profiling tests have the ability to predict risk of distant recurrence (prognosis), but this ability was less certain in the group with LN-positive disease (see section 5.3). The committee also recalled that the test results were particularly helpful for people with cancers identified as intermediate clinical risk when the decision to offer chemotherapy is unclear (see section 5.2). The clinical experts explained that tumour profiling tests were also helpful for people with LN-positive cancer who have comorbidities and therefore an additional reason to want to avoid chemotherapy. The EAG noted that this subgroup of the LN-positive population could not be modelled because of a lack of data. In addition, the committee noted that the EAG's systematic review had highlighted substantial lack of agreement between the tests in risk categorising the group with LN-positive disease. The committee decided to consider the ICERs in the group with LN-negative disease only, but noted that further studies would be helpful to assess the clinical effectiveness of the tests in the group with LN-positive disease (see section 5.27).

5.18 The committee considered the results from the model. It noted that the differences in the QALYs were small, and that the ICERs for all tumour profiling tests were highly uncertain because of the available clinical data and the assumptions used in the modelling (see section 5.10 to section 5.15). It also noted that the base-case ICERs for many of the tumour profiling tests were higher than those normally considered to be cost effective. However, it heard that access proposals had been made by Myriad Genetics (for EndoPredict) and NanoString Technologies (for Prosigna). Genomic Health confirmed that the confidential discount for Oncotype DX would continue in the NHS. The committee concluded that the availability of the access proposals for EndoPredict and Prosigna may reduce the ICERs to a range that could be considered plausibly cost effective despite the clinical uncertainties.

5.19 The committee considered the EndoPredict and Prosigna access proposals. Compared with current practice, the ICERs for EndoPredict and Prosigna in the group with LN-negative disease and a NPI of 3.4 or less were still higher than those normally considered to be a cost-effective use of NHS resources. In the group with LN-negative disease and a NPI of more than 3.4, Prosigna compared with current practice had an ICER of less than £20,000 per QALY gained, and therefore could be considered cost effective. In the same group, EndoPredict compared with current practice had ICERs between £20,000 and £30,000 per QALY gained, which varied depending on whether the testing was done at a local or a central laboratory. The committee noted that local testing was more cost effective than central testing, and that testing became more cost effective as test throughput increased. It also recalled its conclusion that the data on post-chemotherapy decisions were more uncertain for 2-level tests than for 3-level tests (see section 5.11), and noted that the EAG's sensitivity analyses using plausible alternative sources for post-chemotherapy decisions resulted in ICERs that were lower than £20,000 per QALY gained. The committee noted that in sensitivity analyses, when the relative risk of distant recurrence for chemotherapy compared with no chemotherapy was changed to 0.9 from the base-case value of 0.76, the ICERs increased for both EndoPredict and Prosigna to more than £30,000 per QALY gained. It considered that a relative risk of 0.9 or more across all genomic risk groups was unlikely, but accepted the uncertainty around this parameter (see section 5.12). The committee decided that although there is uncertainty around the ICERs for EndoPredict compared with current practice, sensitivity analyses suggested that the ICER will be around £20,000 per QALY gained, and therefore it could be considered cost effective. The committee concluded that EndoPredict (EPclin) and Prosigna, when provided at the costs stated in the access proposals, were likely to be cost effective in the group with LN-negative disease and a NPI of more than 3.4, but evidence on clinical outcomes will be important to confirm this (see section 5.29).

5.20 The committee considered the ICERs for Oncotype DX compared with current practice. It heard that the proposed confidential test cost for Oncotype DX was the same as in current NHS practice, and that this cost had been used in the EAG's economic model. It noted that compared with current practice, the ICERs for Oncotype DX in the group with LN-

negative disease and a NPI of 3.4 or less were higher than those normally considered to be a cost-effective use of NHS resources. In the group with LN-negative disease and a NPI of more than 3.4, the committee noted that in the base-case analyses Oncotype DX was dominated by the comparator. The committee recalled its earlier conclusions; Oncotype DX may be able to predict relative treatment effects for chemotherapy, and the ICERs for Oncotype DX compared with current practice when some relative treatment effect across different risk groups was applied in the model were most likely to be between £2,000 and £25,000 per QALY gained (see section 5.5 and section 5.13). However, it noted that this was very uncertain. The committee concluded that Oncotype DX, when provided at the test cost stated in the access proposal, was likely to be cost effective in the group with LN-negative disease and a NPI of more than 3.4, but evidence on clinical outcomes will be important to confirm this (see section 5.29).

5.21 The committee considered how EndoPredict, Oncotype DX and Prosigna compare with each other. It noted that only pairwise ICERs of each tumour profiling test compared with no testing had been presented, rather than a fully incremental analysis. The EAG explained that a fully incremental analysis could not be done because there was no clinical evidence which directly compared the tests. The committee noted that since the publication of TAILORx (Sparano et al. 2018) evidence on clinical utility was strongest for Oncotype DX. It also noted that it was not possible to determine which test was the most cost-effective use of NHS resources, and that it may not be the test with the lowest acquisition price.

5.22 The committee considered the ICERs for MammaPrint compared with modified Adjuvant! Online. It noted that in the base-case analyses, MammaPrint was dominated by the comparator in the modified Adjuvant! Online high-risk subgroup. In the modified Adjuvant! Online low-risk subgroup, the ICERs were much higher than those normally considered to be cost effective. The committee concluded that MammaPrint would not be a cost-effective use of NHS resources.

5.23 The committee considered the ICERs for IHC4+C compared with current practice. It noted that the ICERs were low or that IHC4+C dominated

current practice in all subgroups. The committee felt that the test cost had been underestimated because it did not include any costs for training or for setting up a quality assurance programme. But even if these costs were included, IHC4+C may still be cost effective. However, the committee noted its earlier conclusion on the analytical validity of IHC4+C (see section 5.9) and concluded that it could not be recommended for use in the NHS until issues around reproducibility and implementation had been resolved.

5.24 The committee noted that the model for EndoPredict, IHC4+C, Oncotype DX and Prosigna related only to a postmenopausal population because TransATAC was used as the data source for these tests. It considered whether the model results could also apply to a premenopausal population. A clinical expert explained that the biology of a cancer and its molecular subtype, for example hormone receptor status and HER2 status, is more influential in determining the risk of distant recurrence than menopausal status. Therefore the committee concluded that the model results apply to premenopausal and postmenopausal populations, but noted that clinicians wishing to use a tumour profiling test should first check which populations the test is indicated for (see section 3).

5.25 The committee discussed the generalisability of the data to men. It acknowledged that men make up a small proportion of people with breast cancer. The committee noted that all the clinical and economic evidence was based on trials with women, but that the general subtypes of breast cancer are identical in men and women, and in clinical practice men would have treatment in the same way as women. The committee concluded that the recommendations in this guidance should also apply to men.

Research considerations

5.26 The committee noted that there are several ongoing studies which will provide evidence of long-term patient outcomes: further data collection from the MINDACT study on MammaPrint and the OPTIMA trial on Prosigna. The committee concluded that these studies are relevant to this assessment and data from them may be important when the guidance is considered for updating in the future. It also recalled its

earlier conclusion that a subgroup analysis of TAILORx would be welcomed (see section 5.6). But it noted that not all studies would provide UK-specific data and comparisons with the PREDICT tool, which would be important for future updates to fully assess the cost effectiveness of the tests compared with current practice.

5.27 The committee also recalled its previous conclusion on the potential utility of the tests in the group with LN-positive disease (see section 5.17), particularly for people who have comorbidities and who may be particularly affected by the side effects of adjuvant chemotherapy. It noted that further research in this group would be welcome and heard from clinical experts that the ongoing OPTIMA trial may help to reduce some of the uncertainties identified during this assessment.

5.28 The committee considered consultation comments from the Cancer and Society in the 21st century research team about their qualitative research on women's experiences of gene expression profiling for chemotherapy decision making, and noted the importance of this work.

Data collection arrangements

5.29 The committee recalled its previous conclusions on the uncertainties associated with both the clinical and cost effectiveness of EndoPredict (EPclin), Oncotype DX and Prosigna. It had identified clinical uncertainties associated with the effect of the technologies on patient outcomes (see section 5.6 and 5.7) and also on clinical decision making (see section 5.8 and section 5.11). These limitations meant that the estimated cost effectiveness of the technologies in the NHS was highly uncertain (see section 5.18 to section 5.20). On balance the committee concluded that EndoPredict (EPclin), Oncotype DX and Prosigna, when provided at the test cost stated in the access proposal, were likely to be cost effective in the group with LN-negative disease and a NPI of more than 3.4, but evidence on clinical outcomes will be needed to confirm this in the NHS. Further, it considered that this should be addressed through data collection using the National Cancer Registration and Analysis Service which would provide data on NHS use. It also believed that it is necessary that data is collected as part of a national database,

rather than by individual companies, to increase transparency, enable the data to be linked to clinical outcomes and ensure evidence is available that can be considered in future updates of this guidance. It therefore decided that its recommendations for EndoPredict (EPclin), Oncotype DX and Prosigna are conditional on data collection arrangements agreed with NICE being put in place. It is anticipated that arrangements will be made to collect timely and complete record-level test data, which can be submitted to the National Cancer Registration and Analysis Service, with the aim of linking test data to chemotherapy use, recurrence and survival outcomes.

6 Implementation

6.1 NICE intends to develop tools, in association with relevant stakeholders, to help organisations put this guidance into practice.

6.2 Genomic Health, Myriad Genetics and NanoString Technologies have all offered their tumour profiling tests to the NHS under access proposals that make each test available to the NHS at a revised price. The proposal prices are commercial in confidence. It is the responsibility of the companies to communicate details of their proposal to the relevant NHS organisations.

7 Diagnostics advisory committee members and NICE project team

Diagnostics advisory committee

The diagnostics advisory committee is an independent committee consisting of 22 standing members and additional specialist members. A list of the committee members who participated in this assessment appears below.

Standing committee members

Dr Mark Kroese
Chair, diagnostics advisory committee

Ms Elizabeth Adair
Quality director, ViaPath

Mr John Bagshaw
In-vitro diagnostics consultant

Professor Enitan Carrol
Chair in paediatric infection, University of Liverpool

Dr Owen Driskell
Lead for laboratory medicine, National Institute for Health Research (NIHR) Clinical Research Network West Midlands

Dr Steve Edwards
Head of health technology assessment, BMJ Evidence Centre

Dr Simon Fleming
Consultant in clinical biochemistry and metabolic medicine, Royal Cornwall Hospital

Dr James Gray
Consultant microbiologist, Birmingham Children's Hospital

Professor Steve Halligan
Professor of radiology, University College London

Mr John Hitchman
Lay member

Professor Chris Hyde
Professor of public health and clinical epidemiology, Exeter Test Group, University of Exeter Medical School

Mr Patrick McGinley
Head of costing and service line reporting, Maidstone and Tunbridge Wells NHS Trust

Dr Michael Messenger
Deputy director and scientific manager NIHR Diagnostic Evidence Co-operative, Leeds

Mrs Alexandria Moseley
Lay member

Dr Peter Naylor
GP, Wirral

Dr Dermot Neely
Consultant in clinical biochemistry and metabolic medicine, Newcastle upon Tyne NHS Trust

Dr Shelley Rahman Haley
Consultant cardiologist, Royal Brompton and Harefield NHS Foundation Trust

Dr Simon Richards
VP regulatory affairs, EME, Alere Inc

Professor Mark Sculpher
Professor of health economics, Centre for Health Economics, University of York

Professor Matt Stevenson
Professor of health technology assessment, School of Health and Related Research, University of Sheffield

Dr Nick Summerton
GP, East Yorkshire

Professor Anthony Wierzbicki
Consultant in metabolic medicine/chemical pathology, St Thomas' Hospital

Specialist committee members

Miss Maria Bramley
Consultant oncoplastic breast surgeon, Pennine Acute NHS Trust c/o Royal Oldham Hospital

Dr John Graham
Consultant oncologist, Taunton & Somerset NHS Foundation Trust

Linda Pepper
Lay specialist committee member

Dr Deirdre Ryan
Consultant cellular pathologist, Barts Health NHS Trust

Dr Britta Stordal
Senior lecturer, Middlesex University

Ursula Van Mann
Lay specialist committee member

Professor Andrew Wardley
Professor of breast medical oncology, The Christie NHS Foundation Trust

NICE project team

Each diagnostics assessment is assigned to a team consisting of a technical analyst (who acts as the topic lead), a technical adviser and a project manager.

Frances Nixon
Topic lead

Rebecca Albrow
Technical adviser

Donna Barnes
Project manager

ISBN: 978-1-4731-3128-6

Accreditation