NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

DIAGNOSTICS ASSESSMENT PROGRAMME

Diagnostics consultation document

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

The National Institute for Health and Care Excellence (NICE) is producing guidance on using tumour profiling tests (EndoPredict, MammaPrint, Oncotype DX Breast Recurrence Score, Prosigna and IHC4+C) to guide adjuvant chemotherapy decisions in people with breast cancer in the NHS in England. The diagnostics advisory committee has considered the evidence base and the views of clinical and patient experts.

This document has been prepared for public consultation. It summarises the evidence and views that have been considered, and sets out the draft recommendations made by the committee. NICE invites comments from registered stakeholders, healthcare professionals and the public. This document should be read along with the evidence base (the diagnostics assessment report and the diagnostics assessment report addendum).

The advisory committee is interested in receiving comments on the following:

- Has all of the relevant evidence been taken into account?
- Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?
- Are the provisional recommendations sound, and a suitable basis for guidance to the NHS?

Equality issues

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others. Please let us know if you think that the preliminary recommendations may need changing in order to meet these aims. In particular, please tell us if the preliminary recommendations:

- could have a different effect on people protected by the equality legislation than on the wider population, for example by making it more difficult in practice for a specific group to access the technology
- could have any adverse effect on people with a particular disability or disabilities.

Please provide any relevant information or data you have regarding such effects and how they could be avoided or reduced.

Note that this document is not NICE's final guidance on tumour profiling tests to guide adjuvant chemotherapy decisions in people with breast cancer. The recommendations in section 1 may change after consultation.

After consultation, the committee will meet again to consider the evidence, this document and comments from the consultation. After considering these comments, the committee will prepare its final recommendations, which will be the basis for NICE's guidance on the use of the technology in the NHS in England.

For further details, see the Diagnostics Assessment Programme manual.

Key dates:

Closing date for comments: 31 January 2018

Second diagnostics advisory committee meeting: 8 February 2018

1 Draft recommendations

1.1 There is not enough evidence to recommend the routine adoption of EndoPredict, MammaPrint, Oncotype DX Breast Recurrence Score, Prosigna and IHC4+C to guide adjuvant chemotherapy decisions for people with hormone receptor-positive, human epidermal growth factor receptor 2 (HER2)-negative early breast cancer with 0 to 3 positive lymph nodes. In particular, more evidence is needed to prove that these tests have a positive effect on patient outcomes. Their cost effectiveness compared with current practice is highly uncertain.

1.2 Further research is recommended on the effect of EndoPredict, MammaPrint, Oncotype DX Breast Recurrence Score and Prosigna

on long-term patient outcomes such as distant recurrence, and on pre- and post-test adjuvant chemotherapy decisions compared with the PREDICT tool (see sections 5.16 and 6.1).

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 the relative benefit of chemotherapy. This information is intended to help decision-making about adjuvant chemotherapy use.

2.2 It is also claimed that the tumour profiling tests may improve the identification of early breast cancer that may not benefit from adjuvant chemotherapy because there is a low risk of disease recurrence. For these people unnecessary treatment could be avoided, and therefore the comorbidities and negative effects of chemotherapy on quality of life. 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, those people may benefit from chemotherapy. People with breast cancer and clinicians may also benefit from improved 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 [PR] or both), human epidermal growth factor receptor 2 (HER2)-negative early breast cancer (stages I or II) with 0 to 3 positive lymph nodes.

2.4 This is a full update of NICE's diagnostics guidance 10 on gene expression profiling and expanded immunohistochemistry tests for guiding adjuvant chemotherapy decisions in early breast cancer management: MammaPrint, Oncotype DX, IHC4 and Mammostrat, which was published in 2013. This recommended Oncotype DX as an option for guiding adjuvant chemotherapy decisions for people with ER-positive, lymph node-negative and HER2-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 and a new test, EndoPredict, has become available.

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 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.

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 women 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 to get the results from a local laboratory, and 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.

MammaPrint (Agendia)

3.7 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 women with stage I or II 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.

3.8 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.

3.9 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.

3.10 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)

3.11 Oncotype DX Breast Recurrence Score (hereafter referred to as Oncotype DX) is designed to quantify the 10-year risk of distant recurrence and predict the likelihood of chemotherapy benefit. The test also reports the underlying tumour biology: ER, progesterone receptor (PR) and HER2 status. The test is for pre- and postmenopausal women with stage I or II 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 done by Genomic Health.

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

3.13 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.

3.14 The test is based on a reverse transcription-quantitative polymerase chain reaction. It gives a recurrence score of between 0 and 100, which is used to quantify the 10-year risk of distant recurrence, assuming 5 years of endocrine therapy. A score below 18 indicates low risk of distant recurrence and claims to predict little to no chemotherapy benefit. A score between 18 and 30 indicates intermediate risk of recurrence and claims to predict no substantial chemotherapy benefit. A score of 31 or more indicates high risk of recurrence and claims to predict a large benefit from chemotherapy.

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

Prosigna (NanoString Technologies)

3.16 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 women with early breast cancer that is ER-positive, HER2-negative and LN-negative or LN-positive (up to 3 positive nodes).

3.17 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.

3.18 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.

3.19 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).

IHC4 and IHC4+C

3.20 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 women with early breast cancer that is ER-positive and LN-negative or LN-positive (up to 3 positive nodes).

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

3.22 The IHC4+C test is used in the Royal Marsden Breast Cancer Unit, but the test could be run in local NHS laboratories if appropriate training and quality assurance programmes for the individual assays are in place. At the Royal Marsden NHS foundation trust, the average turnaround time is 1 week.

3.23 The IHC4+C 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

3.24 The comparator is current decision-making for adjuvant chemotherapy prescribing, which is based on clinical and pathological features or the results of tools used to assess risk. Features may include the stage 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. However, Adjuvant! Online is currently unavailable because it is being updated. It is not certain when it will be reinstated, and the website directs people to the PREDICT tool.

4 Evidence

The diagnostics advisory committee (section 9) considered evidence on EndoPredict, MammaPrint, Oncotype DX, 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 chemotherapy benefit – 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 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 chemotherapy benefit were quality assessed using relevant criteria selected from the draft prediction model study risk of bias assessment tool (PROBAST). Clinical utility studies were quality assessed using the Cochrane risk of bias tool for randomised controlled trials (RCTs).

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 studies patients had endocrine monotherapy. Three of the studies were done in East Asia and may not be generalisable to England because usual clinical practice may differ between countries enough to affect prognostic outcomes. Also, it is possible that people of different ethnicities have different underlying risk profiles and natural history of disease.

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. However, the details were academic in confidence.

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 5 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 or Nottingham Prognostic Index (NPI). In patients with LN-positive disease, MammaPrint had borderline statistically significant prognostic accuracy for 10-year distant recurrence-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 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 Results from the bespoke analysis of TransATAC, which reported adjusted analyses on the EPclin score part of EndoPredict were academic in confidence. Two studies reported adjusted analyses on the EP score part of EndoPredict, showing that it provided statistically significant additional information over clinical and pathological variables 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 NPI in LN-negative, but not LN-positive, disease.

Prediction of chemotherapy benefit

4.22 Oncotype DX and MammaPrint claim to be able to identify patients who will benefit from chemotherapy. The external assessment group (EAG) reviewed evidence in support of this claim.

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 chemotherapy benefit.

4.24 The 2 reanalyses of RCTs suggest that Oncotype DX may predict chemotherapy benefit. Hazard ratios for disease-free survival for patients having chemotherapy compared with those having no chemotherapy suggested that the greatest relative benefit was for patients in the Oncotype DX high-risk category. Unadjusted interaction tests between Oncotype DX risk group and chemotherapy benefit were mainly statistically significant, but

adjusted interaction tests were not always statistically significant. Therefore the EAG concluded that the significant results could be because potentially important covariates were omitted from the statistical model.

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 risk, and another in patients with high risk. Both of these studies reported statistically non-significant results.

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

MammaPrint

4.27 Two studies reported the ability of MammaPrint to predict the benefit of 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 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. 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. The EAG noted that the best evidence for clinical utility is an RCT of treatment guided by the test compared with treatment guided by the comparator, and that this type of evidence is not currently available for Oncotype DX. All studies reporting on the clinical utility of Oncotype DX are 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).

MammaPrint

4.31 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 discordant risk scores 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 modified Adjuvant! Online high clinical risk, MammaPrint low-risk group, 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 modified Adjuvant! Online low clinical risk, MammaPrint high-risk group, 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.

4.32 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 NPI, but not over the NHS PREDICT tool.

RASTER was judged to be at high risk of bias by the EAG using the Cochrane risk of bias tool for RCTs.

Comparison of the tests with each other

4.33 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.

4.34 Thirteen studies reported data from microarray analyses on more than 1 test, however, these studies have methodological limitations. The comparability of test algorithms applied to microarray data with the commercial assays is unknown, so the generalisability of findings from microarray studies to the decision problem is uncertain. All the studies reported data on Oncotype DX and MammaPrint, and 2 also reported data on EndoPredict. The microarray studies generally support 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.

4.35 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

4.36 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.

4.37 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.

4.38 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

4.39 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

4.40 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).

4.41 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.

4.42 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 the benefit of chemotherapy.

Economic evaluation

4.43 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. 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. 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 dataset, 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

4.44 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.

4.45 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 benefit of adjuvant chemotherapy was modelled using a relative risk reduction for distant recurrence within each risk classification group. The benefit of the test was therefore captured in the model by changing the probability that patients with each test risk classification had adjuvant chemotherapy.

Model inputs

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

4.47 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 zero.

4.48 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. The NHS England access scheme dataset is owned by Genomic Health and is a result of the research recommendation from NICE's original diagnostics guidance 10.

Table 1 Source for probability of having chemotherapy

Population	Source	Proportion of patients having chemotherapy		
		Low risk	Intermediate risk	High risk
Current practice group				
LN-negative, NPI≤3.4	NCRAS dataset	0.07		
LN-negative, NPI>3.4	NHS England access scheme dataset	0.43		
LN-positive (1–3 nodes)	NCRAS dataset	0.63		
Overall population (MammaPrint)	Expert opinion	0.47		
3-level tests (Oncotype DX, Prosigna and IHC4+C)				
LN-negative, NPI≤3.4	UKBCG survey data	0.00	0.20	0.77
LN-negative, NPI>3.4	NHS England access scheme dataset	0.01	0.33	0.89
LN-positive (1–3 nodes)	Loncaster et al. (2017) node-positive estimates	0.08	0.63	0.83
2-level tests (EndoPredict and MammaPrint)				
EndoPredict: all 3 subgroups	Bloomfield et al. (2017) study	0.07	–	0.77
MammaPrint: all subgroups	Bloomfield et al. (2017) study	0.07	–	0.77
Abbreviations: LN, lymph node; NCRAS, national cancer registration and analysis service; NPI, Nottingham Prognostic Index; UKBCG, UK breast cancer group				

4.49 In the base-case analysis, the benefit of chemotherapy was assumed to be the same across all test risk groups, that is, all tests were assumed to be associated with prognostic benefit only. For Oncotype DX, Prosigna, IHC4+C and EndoPredict 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. Sensitivity analyses explored the relative risks of distant recurrence in the modified Adjuvant! Online low- and high-risk subgroups, which were estimated to be 0.84 and 0.74, respectively.

4.50 In sensitivity analyses, the effect of assuming that Oncotype DX could predict the benefit of chemotherapy was explored, based on the studies by Paik et al. (2006) and Albain et al. (2010). For the group with LN-negative disease, the 10-year relative risks of relapse 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.

4.51 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.

4.52 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.

Costs

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

Table 2 Test costs

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

4.54 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.

4.55 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.

4.56 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.

4.57 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

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

Table 3 Health utilities applied in the model

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

4.59 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 score 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 benefit of 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 the patient had previously developed distant metastases.
- A disutility associated with adjuvant chemotherapy was applied once during the first model cycle only (while the patient is receiving the regimen).
- Costs associated with endocrine therapy, bisphosphonates, follow-up appointments and mammograms were assumed to differ according to time since model entry.
- The model assumed that people entered at an age of around 60 years.

4.60 In the subgroup with LN-negative disease and an NPI of 3.4 or less, for tumour profiling tests compared with current practice, the probabilistic model gave incremental cost-effectiveness ratios (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).

4.61 In the subgroup with LN-negative disease and an NPI of more than 3.4, for tumour profiling tests 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)
- ICH4+C was dominant over current practice (that is, it was less expensive and more effective).

4.62 In the population with LN-positive disease, for tumour profiling tests 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
- ICH4+C was dominant over current practice.

4.63 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.

Probabilistic sensitivity analyses

4.64 The cost-effectiveness planes from the probabilistic sensitivity analyses showed considerable variability in cost-effectiveness estimates.

4.65 In the subgroup with LN-negative disease and an 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.

4.66 In the subgroup with LN-negative disease and an 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 all other tests, 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.

4.67 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 at the same maximum acceptable ICERs, the probability of EndoPredict producing more net benefit than current practice ranged from 0.44 to 0.73. For Prosigna the range was 0.24 to 0.55. In this subgroup Oncotype DX had very low probabilities of producing more net benefit than current practice at the same maximum acceptable ICERs (0.01 or lower).

4.68 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 zero.

Deterministic sensitivity analyses

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

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

- Subgroup with LN-negative disease and an NPI of 3.4 or less: ICERs remained over £34,000 per QALY gained across all analyses.
- Subgroup with LN-negative disease and an 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 chemotherapy benefit. 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 chemotherapy benefit (it was dominant), and when the cost of chemotherapy was doubled (£3,700 saved per QALY lost).

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

- Subgroup with LN-negative disease and an 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 an 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.

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

- Subgroup with LN-negative disease and an NPI of 3.4 or less: ICERs were greater than £71,000 per QALY gained across all analyses.
- Subgroup with LN-negative disease and an 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.

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

- Subgroup with LN-negative disease and an NPI of 3.4 or less: ICERs remained greater than £91,000 per QALY gained across all analyses.
- Subgroup with LN-negative disease and an 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 conditional 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.

4.74 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.

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. Also, the PREDICT tool is now used rather than Adjuvant! Online or the Nottingham Prognostic Index (NPI). 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, lymph node (LN)-negative and human epidermal growth factor receptor 2 (HER2)-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 that the tumour profiling tests may have for people with early breast cancer who are deciding whether to have adjuvant chemotherapy. It heard that there is potential benefit for cancers identified as being at low clinical risk, when test results suggest these have a high risk of distant recurrence. These cancers would therefore benefit from chemotherapy. It also heard that there is potential benefit for cancers categorised as high clinical risk, when test results suggest

a low risk of distant recurrence. The committee heard that some people with these cancers could decide not to have chemotherapy, therefore avoiding toxic side effects and effects on fertility. They could potentially resume normal daily activities earlier. Alternatively, others may wish to have chemotherapy regardless of the test result. Also, the clinical experts explained that the tests may mean that additional information can be provided to help people discuss further treatment options. 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 tests had statistically significant prognostic accuracy over clinical and pathological features or risk assessment tools such as the NPI (see section 4). It also noted that for people with LN-positive disease, results on prognostic ability were more variable but all tests except IHC4+C showed statistically 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 patient spectrum 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, and some studies included some patients who had hormone receptor-negative or HER2-positive disease. The committee concluded that despite the potential spectrum bias, evidence suggests that all the tumour profiling tests have the ability to predict the risk of distant recurrence in the population included in the assessment.

5.4 The committee considered the evidence on whether the tumour profiling tests can predict the benefit of chemotherapy. The committee heard from clinical experts that it is likely some patients could have a greater relative benefit from chemotherapy than others, for example, patients with hormone receptor-positive cancer that is insensitive to endocrine therapy, but that evidence is not available to support this. The EAG explained that the evidence on differential chemotherapy benefit according to tumour profiling test results is weak because it is at high risk of bias from potential confounding. The interaction tests (which show whether the tumour profiling test is able to predict a differential relative treatment effect by risk group) in the adjusted analysis in the Paik et al. (2006) study and the Albain et al. (2010) study were statistically significant, but did not adjust for all relevant clinical and pathological variables, in particular hormone receptor status. The committee heard from clinical experts that hormone receptor status may also predict chemotherapy benefit. The committee considered that if all known clinical and pathological variables were included in the analyses then it was likely that the interaction test would no longer be statistically significant, suggesting no differential chemotherapy benefit of the tumour profiling tests alone. The committee concluded that the evidence does not support the assumption that tumour profiling tests can predict chemotherapy benefit.

5.5 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 noted that the only test with evidence from randomised controlled trials (RCTs) in which patients were randomised to treatment guided by either test result or usual clinical practice was MammaPrint. The committee noted that MINDACT was a well-designed study, which 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 noted that none of the other tumour profiling tests had similar evidence of clinical utility, but it was aware that this evidence was being collected for Oncotype DX and Prosigna (see section 5.16). The committee concluded that none of the tests had strong enough evidence to demonstrate an effect on subsequent patient outcomes.

5.6 The committee was encouraged by the availability of the dataset provided in confidence to NICE by Genomic Health. The dataset 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 dataset 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 has been used on a wider group of patients in practice. The committee concluded that the access scheme dataset was an important piece of real world evidence for use in the economic model, but that more complete data could potentially have been collected and reported.

5.7 The committee discussed the analytical validity of IHC4+C. The EAG explained that evidence has developed since diagnostics guidance 10 was published. The committee noted that the data showed good correlation between different centres on scoring and staining when 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 decreased substantially. The committee concluded that because of these issues with Ki67, the reproducibility of IHC4+C is poor. It also heard that different methods of assessing ER and PR receptors may be needed for the IHC4+C method compared with those already used routinely, which may introduce additional complexity. The committee concluded that if this test were to be developed further there would need to be substantial investments in staff training and a quality assurance scheme would need to be set up.

Cost effectiveness

5.8 The committee discussed the assumptions and inputs used in the model, and considered the stakeholder comments on the model and EAG responses to these comments. It noted that a bespoke analysis of the TransATAC data was used for risk classification probabilities and for distant recurrence rates conditional on test result for Oncotype DX, EndoPredict, Prosigna and IHC4+C. 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 are 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 committee agreed that the bespoke TransATAC analysis was the best available data for use in the model. The committee also noted their earlier conclusions that current evidence does not support the assumption that tumour

profiling tests can predict chemotherapy benefit (see section 5.4). It agreed that in the base case it was appropriate to assume the same relative risk of distant recurrence across all test risk categories (0.76). The committee considered other assumptions used in the model such as the cost of chemotherapy, the fixed benefit of chemotherapy, and the probability of having chemotherapy. 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.

5.9 The committee noted that the baseline probability of having adjuvant chemotherapy and the probability of having chemotherapy dependent on test result were key inputs driving the differences between the results of the original model for diagnostics guidance 10 and the results of the updated model for the intermediate clinical risk group in this assessment. It also noted that the data sources for these inputs in the original model were a published abstract with few details available about the methods (Holt et al. 2013) and data from the English Cancer Registry, and that the updated EAG model used data from Genomic Health's access scheme dataset. The committee concluded that this dataset is most likely to reflect chemotherapy use in NHS clinical practice and is therefore a more suitable data source for the model.

5.10 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 do not reflect current NHS clinical practice, which leads to uncaptured uncertainty in the model results. The committee concluded that future research on tumour profiling tests should include comparisons with PREDICT (see section 6.1).

5.11 The committee considered the incremental quality-adjusted life year (QALY) results from the model. It noted that the differences in the QALYs were small and that in the base-case analyses Oncotype DX and MammaPrint both had a QALY loss in some subgroups compared with current practice or modified Adjuvant! Online respectively. This led to these tests being dominated by the comparator. It also noted that in deterministic sensitivity analyses, Oncotype DX and MammaPrint sometimes had a QALY gain compared with the comparator. The committee concluded that the QALYs derived from the model are uncertain.

5.12 The committee considered the incremental cost-effectiveness ratios (ICERs) resulting from the model. It noted that the ICERs for EndoPredict in LN-positive disease, and for Prosigna in LN-negative intermediate-risk disease and LN-positive disease, fell between £20,000 and £30,000 per QALY gained. However, the committee considered these ICERs to be highly uncertain because of the available clinical data. It also noted that the ICERs for IHC4+C were low or dominating in all subgroups. But the committee noted its earlier conclusion on the analytical validity of IHC4+C (see section 5.7) and felt that the test cost had been underestimated because it did not include any costs for training or for setting up a quality assurance scheme. The committee concluded that the cost effectiveness of all tumour profiling tests was highly uncertain.

5.13 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. The committee heard from a clinical expert that the biology of a cancer and its molecular subtype, for example the 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.

5.14 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 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.

5.15 The committee considered its earlier conclusions that none of the tests had strong enough evidence of a positive effect on patient outcomes (see section 5.5) and that their cost effectiveness compared with current practice was uncertain (see sections 5.10 and 5.12). The committee concluded that none of the tumour profiling tests should be recommended for routine use in the NHS.

Research considerations

5.16 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, the TAILORx trial on Oncotype DX 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. But it noted that not all studies would provide UK specific data and comparisons with the PREDICT tool (see section 6).

6 Draft recommendations for further research

6.1 Further research is recommended comparing the tumour profiling tests (EndoPredict, Oncotype DX, MammaPrint and Prosigna) with the PREDICT tool. The results should record both pre- and post-test adjuvant chemotherapy decisions.

7 Implementation

NICE will support this guidance through a range of activities to promote the recommendations for further research. The research proposed will be considered by the NICE Medical Technologies Evaluation Programme research facilitation team for the development of specific research study protocols as appropriate. NICE will also incorporate the research recommendations in section 6 into its guidance research recommendations database (available on the NICE website) and highlight these recommendations to public research bodies.

8 Review

NICE reviews the evidence 3 years after publication to ensure that any relevant new evidence is identified. However, NICE may review and update the guidance at any time if significant new evidence becomes available.

Mark Kroese
Chair, diagnostics advisory committee
January 2018

9 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

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, Peninsula Technology Assessment Group (PenTAG)

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

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 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:

