NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE

Diagnostics consultation document

Gene expression profiling and expanded immunohistochemistry tests to guide the use of adjuvant chemotherapy in early breast cancer management: MammaPrint, Oncotype DX, IHC4 and Mammostrat

The National Institute for Health and Clinical Excellence (NICE) is producing guidance on using gene expression profiling and expanded immunohistochemistry tests in the NHS in England. The Diagnostics Advisory Committee has considered the evidence submitted and the views of expert advisers.

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), which is available from http://guidance.nice.org.uk/DT/InDevelopment.

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?
- Are there any aspects of the recommendations that need particular consideration to ensure we avoid unlawful discrimination against any group of people on the grounds of race, gender, disability, religion or belief, sexual orientation, age, gender reassignment, pregnancy and maternity?

Note that this document is not NICE's final guidance on these technologies. The recommendations in section 1 may change after consultation. This is the third diagnostics consultation document on this

topic.

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 process guide' (available at www.nice.org.uk/aboutnice/howwework/developingnnicediagnostictechnologiesguidance).

Key dates:

Closing date for comments: 11th March 2013

Third Diagnostics Advisory Committee meeting: 10th April 2013

1 Provisional recommendations

1.1 Oncotype DX is recommended in people with oestrogen receptor positive (ER+), lymph node negative (LN−) and human epidermal growth factor receptor 2 negative (HER2−) early breast cancer to guide chemotherapy decisions if:

- the person is assessed as being at intermediate risk, **and**
- where the decision to prescribe chemotherapy remains unclear, so that information on the biological features of the cancer provided by Oncotype DX is likely to help in predicting the course of the disease, **and**
- the manufacturer provides it to NHS organisations at the price offered through the confidential arrangement agreed with NICE.

1.2 NICE encourages further data collection on the use of Oncotype DX in the NHS (see section 7).

The analysis leading to recommendation 1.1 was based on intermediate risk of distant recurrence being defined as a Nottingham Prognostic Index (NPI) score above 3.4. It is anticipated that an NPI score can be simply calculated from information that is routinely collected about people with breast cancer. Other decision-making tools or protocols (for example Adjuvant! Online) are also currently used in the NHS and these may be used to identify people at intermediate risk.

1.3 MammaPrint, IHC4 and Mammostrat are only recommended for use in research in people with ER+, LN− and HER2− early breast cancer, to collect evidence about potentially important clinical outcomes and to determine the ability of the tests to predict the benefit of chemotherapy. The tests are not recommended for general use in these people because of uncertainty about their overall clinical benefit and consequently their cost effectiveness.

2 The technologies

2.1 Four tests available to the NHS were evaluated. Two are based on gene expression profiling: MammaPrint (Agendia) and Oncotype DX (Genomic Health). Two are based on immunohistochemistry (also known as protein expression profiling): IHC4 (academic sponsor – Royal Marsden Hospital and Queen Mary University, London) and Mammostrat (Clarient). These tests measure multiple markers within the tumour that may indicate how the tumour is likely to develop. Additional details are provided in section 4.

3 Clinical need and practice

The problem addressed

3.1 Gene expression profiling and immunohistochemistry tests aim to improve the targeting of chemotherapy in breast cancer by more accurately identifying patients who will gain the most benefit. This rationale is based on the knowledge that certain biological features of cancers may indicate an increased likelihood of rapid growth and metastasis (in particular, distant recurrence). Distant recurrence is the return of detectable cancer in another part of the body. The tests may also identify, in some instances, which patients are most likely to benefit from chemotherapy. Some tools or tests provide mainly prognostic information (such as the NPI and Adjuvant! Online). Others may or may not be able to predict the extent to which the patient could benefit from chemotherapy (such as Oncotype DX, MammaPrint, Mammostrat and IHC4). Breast cancer patients face significant emotional and psychological strain when considering chemotherapy. It can be particularly distressing for patients in whom the decision to undergo chemotherapy is unclear using currently available tools (especially people with an intermediate risk of distant recurrence). Tools or tests that help people decide whether or not to have chemotherapy are likely to be greatly appreciated by patients. The aim of this evaluation is to determine whether using gene expression profiling and expanded immunohistochemistry tests (MammaPrint, Oncotype DX, IHC4 and Mammostrat) in conjunction with current decision-making protocols (including tools such as the NPI and Adjuvant! Online) to guide the use of adjuvant chemotherapy, cost-effectively improves health outcomes and quality of life of people with early stage breast cancer, compared with current decision-making protocols alone.

The condition

Epidemiology and incidence

3.2 Breast cancer is the most commonly diagnosed cancer in women in England and Wales, but it can affect both men and women. In 2010 there were approximately 49,600 new cases in women and 400 in men. For both sexes, incidence varies with age. Just over 80% of cases occur in women aged 50 years and over. In England and Wales, 2006–2008 data demonstrate highest incidence rates for women in the 60- to 70-year age range.

3.3 Incidence also varies with family origin. In England, people of Asian, Chinese and black family origin and those with mixed heritage have a lower incidence than those of white family origin. Incidences are 0.65, 0.75, 0.49 and 0.58 that of those of white family origin respectively.

3.4 Breast cancer is the second largest cause of cancer-related death in women after lung cancer, with an age-standardised mortality rate of 24 per 100,000 women. In 2010 this constituted 10,328 deaths for women in England and Wales.

Prognosis

3.5 Overall, 5-year age-standardised survival rates for breast cancer are around 80%. Breast cancer survival rates have improved over the last 2 decades and now almost 2 out of 3 women with breast cancer survive beyond 20 years. Survival varies with age, stage of disease, family origin, socioeconomic status and tumour characteristics.

3.6 Clinicians currently estimate prognosis using tools such as the NPI (see paragraph 4.10) or Adjuvant! Online (see paragraph 4.11).

The NPI takes into account grade as well as size and spread of the tumour, whereas Adjuvant! Online uses age of the patient, tumour size, nodal involvement, hormonal receptor status, histological grade and comorbidities to predict disease course and treatment options. Better prognosis is associated with small tumour size, younger age, LN−, ER+ and progesterone receptor positive (PR+) status. HER2 over-expression (also known as HER2+) is associated with a poor prognosis. A tool called PREDICT, which is based on cancer registry data for women treated in England (East Anglia) and includes HER2 and Ki-67 status, has recently become available to the NHS.

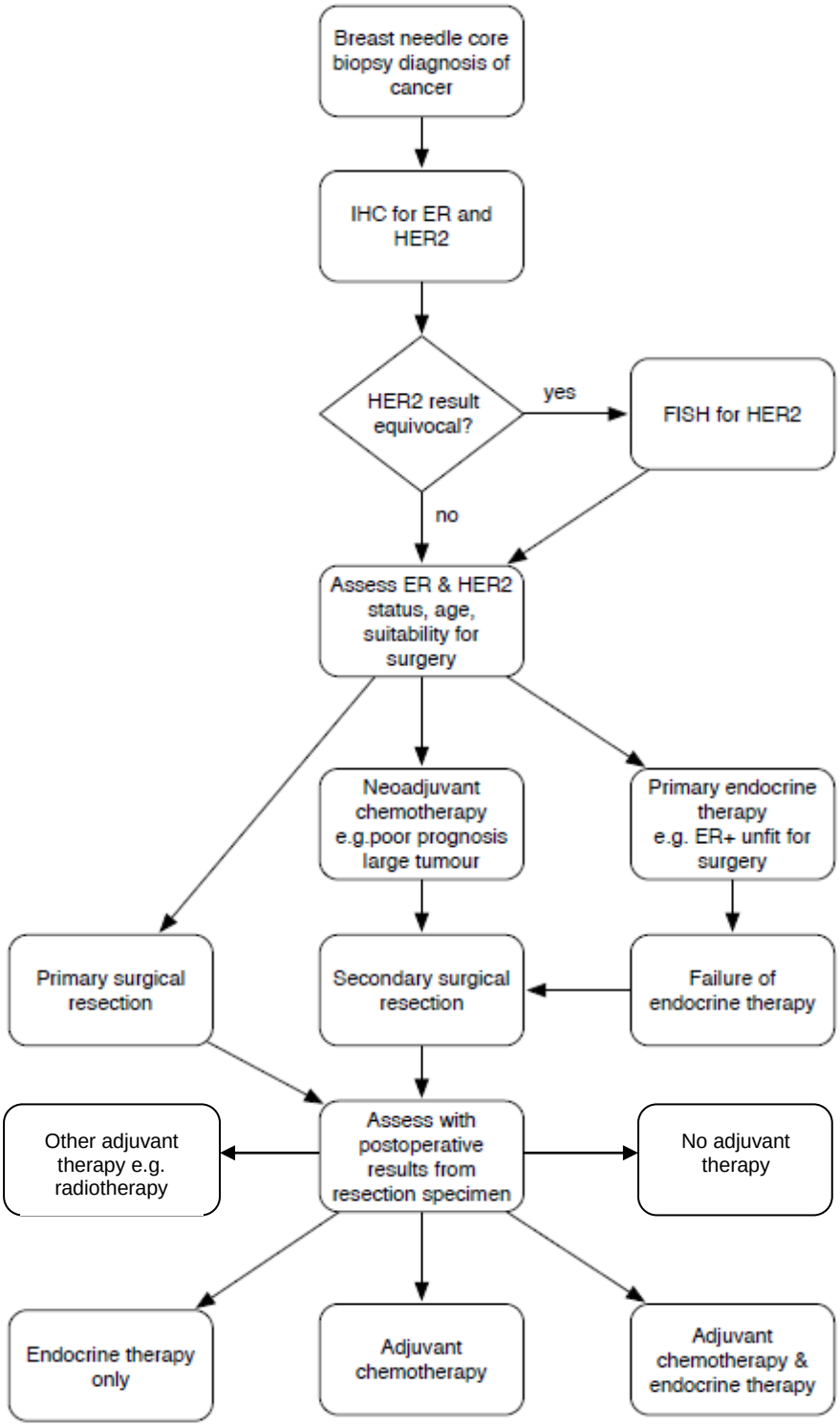
3.7 Some patients considered to have a 'good' prognosis using current tools may still have recurrence after curative surgery and adjuvant therapy. Some patients considered to have a 'poor' prognosis may never develop metastatic disease. It is therefore challenging to decide whether to treat early stage breast cancer with adjuvant chemotherapy.

3.8 The decision whether to offer adjuvant chemotherapy is uncertain in people with ER+, LN− and HER2− early breast cancer. The External Assessment Group and clinical specialists who were consulted advised that the tests being evaluated would most likely be of benefit to the NHS in this patient group. Moreover, the evidence base was most robust for this population. Therefore the economic analysis for this evaluation focused on people with ER+, LN− and HER2− early breast cancer.

The diagnostic and care pathways

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

Figure 1 Diagnosis and management pathway in breast cancer



FISH: fluorescence in situ hybridisation

3.10 For the purposes of this assessment, chemotherapy is defined as the use of cytotoxic drugs with the intention of preventing cancer recurrence and does not include other forms of systemic therapy such as endocrine treatments or targeted biological therapy. Generally, chemotherapy regimens containing anthracyclines are used after cancer surgery (in the adjuvant setting).

Current guidelines

3.11 NICE cancer service guidance 'Improving outcomes in breast cancer' (2002) recommends that women at intermediate or high risk of recurrence who have not had neoadjuvant chemotherapy should normally be offered multi-agent chemotherapy, which includes anthracyclines.

3.12 'Early and locally advanced breast cancer: diagnosis and treatment' (NICE clinical guideline 80 [2009]) recommends that adjuvant therapy should be considered for all patients with early invasive breast cancer after surgery, based on assessment of the prognostic and predictive factors, and the potential benefits and side effects of the treatment. These guidelines do not refer to the use of gene expression profiling and expanded immunohistochemistry tests to aid decision making. NICE clinical guideline 80 recommends that decisions should be made following discussion of these predictive and prognostic factors with the patient and that Adjuvant! Online should be considered to support estimations of individual prognosis and the absolute benefit of adjuvant treatment. The NPI is also commonly used locally to aid decisions about chemotherapy for patients with early stage breast cancer and is discussed in NICE clinical guideline 80.

3.13 In the UK, local guidance based on the NPI and Adjuvant! Online has been developed to help clinicians decide about the benefits of adjuvant chemotherapy for a particular patient. However, it has been suggested that these tools may be imperfect and different local approaches to the use and interpretation of these tools leads to a proportion of people with early stage breast cancer being over- or under-treated. This may result in unnecessary use of expensive chemotherapy with its associated adverse effects for people who derive little or no benefit. In addition, there may be avoidable deaths in people who would have benefitted from chemotherapy had it been offered.

4 The diagnostic tests

The individual tests: MammaPrint, Oncotype DX, IHC4, Mammostrat

4.1 Gene expression profiling and immunohistochemistry tests typically report 1 or 2 types of information – breast cancer subtype and risk of recurrence. Tests developed to provide information on subtypes can be used either before surgery to inform decisions on neoadjuvant therapy or after primary surgery (for removal of the tumour, which may also be used for further assessment of the tumour characteristics) to inform decisions on adjuvant chemotherapy (see figure 1). Tests predicting the risk of recurrence in a specific population are typically used after surgery, in conjunction with other information such as tumour size and grade, to guide the use of adjuvant chemotherapy. Such tests are typically indicated for women with ER+ and LN− (and sometimes LN+ if the number of nodes is small) breast cancer in whom there is significant uncertainty about the value of chemotherapy. The

current evaluation addresses the use of MammaPrint, Oncotype DX, IHC4 and Mammostrat after primary surgery to inform decisions on the use of adjuvant chemotherapy.

4.2 Three tests (MammaPrint, Oncotype DX and Mammostrat) require that samples are sent to a central laboratory for processing following surgery, with an estimated shipping and processing time of up to 7–10 days. IHC4 is processed in a local laboratory with estimated turnaround times of less than 1 week.

Gene expression profiling

4.3 Some gene expression profiling tests work by identifying and quantifying mRNA transcripts in a specific tissue sample. Because only a fraction of the genes encoded in the genome of a cell are transcribed into mRNA, gene expression profiling provides information about the activity of genes that give rise to these mRNA transcripts. Other gene expression profiling tests work by measuring levels of cDNA, which is synthesised from mRNA. There are a range of different techniques for measuring mRNA levels in breast cancer tumour samples, including real-time reverse transcription polymerase chain reaction (RT-PCR) and DNA microarrays.

4.4 Different tests use different protocols for preparing the samples (for example, formalin fixation, paraffin embedding, snap freezing and fresh samples) and different methods for preparing the RNA. Furthermore, there are different algorithms for combining the raw data into a summary profile. All of these factors can affect the reproducibility and reliability of gene expression profiling tests.

4.5 The 2 gene expression profiling tests included in this evaluation are described below:

- MammaPrint is based on microarray technology and uses an expression profile of 70 genes. MammaPrint is intended as a prognostic test for women of all ages, with LN− and LN+ (up to 3 nodes positive) breast cancer with a tumour size of 5 cm or less. MammaPrint is used to estimate the risk of distant recurrence of early breast cancer. It stratifies patients into 2 distinct groups — low risk (good prognosis) or high risk (poor prognosis) of distant recurrence. MammaPrint has been cleared by the FDA as an In Vitro Diagnostic Multivariate Index Assay. The test uses fresh or formalin-fixed paraffin-embedded samples that are processed centrally at laboratories run by the manufacturer in the USA or The Netherlands.
- Oncotype DX quantifies the expression of 21 genes in breast cancer tissue by RT-PCR. It predicts the likelihood of recurrence in women of all ages with newly diagnosed stage I or II, ER+ LN− or LN+ (up to 3 nodes positive) breast cancer treated with tamoxifen. The test assigns the breast cancer a continuous recurrence score (RS) and a risk category – low (RS<18), intermediate (18≤RS≤30) or high (RS≥31). The test also reports ER, PR and HER2 status. The test uses formalin-fixed paraffin-embedded samples that are processed centrally at a laboratory run by the manufacturer in the USA.

Immunohistochemistry (protein expression profiling)

4.6 Immunohistochemistry tests measure protein levels in the tumour sample rather than RNA or cDNA. Some of these tests offer the advantage of using existing immunohistochemical markers (such as ER and HER2) which are routinely tested in UK pathology

departments. The term ‘expanded’ has been used to describe the immunohistochemistry tests evaluated in this assessment that are used in addition to standard immunohistochemistry testing (such as ER and HER2) for early invasive breast cancer.

4.7 The expanded immunohistochemistry tests included in this evaluation are described below:

- IHC4 measures the levels of 4 key proteins (ER, PR, HER2 and Ki-67) in addition to classical clinical and pathological variables (for example, age, nodal status, tumour size and grade) and calculates a risk score for distant recurrence using an algorithm. Quantitative assessments of ER, PR, and Ki-67 are needed for the IHC4 test. An online calculator for IHC4 is in development. The test uses formalin-fixed paraffin-embedded samples that can be processed in local NHS laboratories.
- Mammostrat uses 5 immunohistochemical markers (SLC7A5, HTF9C, P53, NDRG1 and CEACAM5) to stratify patients into risk groups to inform treatment decisions. These markers are independent of one another and do not directly measure either proliferation or hormone receptor status. The test calculates the relative risk of recurrence by using a weighted algorithm that is interpreted in the context of published clinical studies of appropriate patient populations. Patients are classified into 3 risk categories: prognostic index ≤0 defined as the ’low risk’ group; prognostic index >0 and ≤0.7 defined as the ’moderate-risk’ group; prognostic index >0.7 defined as the ’high risk’ group. The test uses formalin-fixed paraffin-embedded samples that are processed centrally at a laboratory run by the manufacturer in the USA.

The comparator

4.8 The comparator is standard practice in England. Although this varies between hospitals, Adjuvant! Online and/or the NPI are often used to guide decisions on which patients with early breast cancer should be offered adjuvant chemotherapy. The economic analysis used cancer registry data on levels of chemotherapy prescribing to reflect standard practice in England and, therefore, is likely to incorporate the impact on the decision to use chemotherapy based on a range of different decision tools currently used in the NHS.

4.9 Further information on the importance of individual molecular markers (for example, ER and HER2, which are routinely assessed for early breast cancer) in the decision to offer adjuvant chemotherapy (and other therapies such as endocrine therapy) has led to varying local practice. Although some hospitals use Adjuvant! Online and the NPI in their original forms, others use adaptations of these tools. Adjuvant! Online is often used in conjunction with the HER2 score. Management algorithms based on the combined use of the NPI and molecular markers such as ER and HER2 are also used.

Nottingham Prognostic Index

4.10 The NPI is a composite prognostic parameter involving both time-dependent factors and aspects of tumour aggressiveness. The NPI score is based on a mix of grade, lymph node involvement and tumour size. The score is calculated by adding numerical grade (1, 2 or 3), lymph node score (negative=1, 1 to 3 nodes=2, >3 nodes=3) and 0.2 times tumour size in centimetres. Patients can be divided into 3 prognostic groups (other subdivisions are also possible, for example 5 prognostic groups) on the basis of the NPI

score: a good prognostic group (NPI≤3.4), a moderate prognostic group (3.4<NPI≤5.4), and a poor prognostic group (NPI>5.4).

Adjuvant! Online

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

5 Outcomes

5.1 The Diagnostics Advisory Committee (appendix A) considered evidence from a number of sources (appendix B).

How outcomes were assessed

5.2 The assessment consisted of a systematic review of the evidence on test performance and clinical-effectiveness data for the 4 tests included in the evaluation. The outcome measures included:

- analytical validity, defined as the ability of the test to accurately and reliably measure the expression of mRNA or proteins by breast cancer tumour cells (that is, repeatability and reproducibility)

- clinical validity, defined as prognostic ability or the degree to which the test can accurately predict the risk of an outcome (for example, the risk of distant metastases in 10 years)
- clinical utility, defined as the ability of the test to improve clinical outcomes such as overall survival. This includes direct harms arising from the test, reclassification of risk compared with existing tools, its impact on clinical decision-making and the ability of the test to predict benefit from chemotherapy. Within the context of clinical utility, the predictive ability of a test refers to the capability of the test to accurately predict patients who will benefit most, in relative terms, from chemotherapy, that is, whether patients classified as high risk benefit more in relative terms than patients classified as low risk.

5.3 In the base-case economic analysis, the External Assessment Group used the available data on the clinical validity and clinical utility of the tests to populate the model. The risk of distant recurrence (prognosis) was computed from data on the clinical validity of the tests. The reclassification of risk by the new tests (presented as 2 NPI subgroups), the impact of the test results on clinical decision-making (the proportion of patients receiving adjuvant chemotherapy) and the predicted benefit of chemotherapy by risk group (reduction in the risk of distant recurrence) were based on data on the clinical utility of the tests. In all cases, the systematic review showed that data on the clinical validity of the tests were more robust than data on their clinical utility. Therefore the External Assessment Group used sensitivity analysis to explore alternative scenarios with different assumptions of the clinical utility of the tests and, in some cases alternative assumptions of clinical validity.

5.4 For 2 of the 4 tests (Oncotype DX and MammaPrint), the current systematic review updated an existing systematic review of gene expression profiling tests for breast cancer. Two previous systematic reviews (one an update of the other) reviewed the literature relating to both Oncotype DX and MammaPrint. Marchionni et al. (2008) is an exhaustive literature review of various electronic databases covering biomedical literature between 1990 and 2006. In 2010, Smartt updated this systematic review and included all relevant evidence that became available between 2007 and 2009.

5.5 The External Assessment Group undertook a systematic review of the evidence on cost effectiveness for the 4 tests. Genomic Health (Oncotype DX) submitted an economic model and Clarient submitted a report detailing an economic analysis of Mammostrat. The External Assessment Group constructed a de novo economic model. The outcomes of interest for the economic evaluation were the morbidity and mortality associated with invasive breast cancer and its treatment. These included survival and health-related quality of life, including the impact of adverse events associated with chemotherapy. The de novo economic model followed a linked evidence approach in which intermediate outcomes (results of the tests) were linked to treatment outcomes and hence quality-adjusted life year (QALY) gains. Costs and QALYs were assigned to each of the 4 tests and the comparator.

5.6 The population identified in the scope for this evaluation included assessment of the gene expression profiling and expanded immunohistochemistry tests in men with breast cancer if data were available. No such data that would allow the evaluation of these technologies in men were identified in the systematic review.

Clinical effectiveness

5.7 The terms analytical validity, clinical validity and clinical utility, used in this section, are defined in paragraph 5.2.

5.8 Much of the clinical evidence was related to the Oncotype DX and MammaPrint tests because these tests are much further along the validation pathway than IHC4 and Mammostrat. The highest quality evidence was reported for Oncotype DX, although limitations or gaps in the clinical data were identified for all tests. Most studies, for all tests, were retrospective in design, analysing archived tumour samples from a cohort of patients with documented information on patient characteristics and outcomes. Retrospective analyses are associated with increased bias compared with prospective randomised controlled trials. Some of the studies involved a prospective analysis of retrospective archived material from a previous randomised controlled trial. Potential issues still remain, including the effects of confounding and the possible incompleteness of some biological specimens.

5.9 Study populations were generally heterogeneous, although most of the evidence on Oncotype DX came from ER+, LN− populations. Some studies included a small number of participants. Studies including larger sample sizes, in excess of 1000 samples, were available for Oncotype DX, Mammostrat and IHC4. Follow-up was short or not reported in a number of studies. Five studies were specific to a UK population, including 3 for Oncotype DX, 1 for IHC4 and 1 for Mammostrat.

MammaPrint

5.10 A range of studies provided evidence on the prognostic ability of MammaPrint in heterogeneous populations. However, the previous

systematic reviews indicated that evidence relating to the clinical validity of MammaPrint was not always conclusive nor supported the prognostic value of the test. Four studies suggested that the test could predict prognosis, 1 study failed to verify the prognostic utility of the test and in another the methods and results were at variance with other studies. In terms of clinical utility, the previous reviews identified 1 prospective study demonstrating that MammaPrint had an impact on clinical decision-making. The addition of MammaPrint to the standard Dutch clinical assessment of risk (modified by patient preference) in a cohort of 427 patients increased the number of patients receiving adjuvant systemic therapy by 20 (5%). However, follow-up was not long enough to provide evidence of its effect on clinical endpoints such as distant metastasis-free survival or its utility in predicting treatment benefit. The previous systematic reviews recommended that further evidence from randomised controlled trials was needed in addition to robust evidence on the prediction of benefit from chemotherapy.

5.11 The External Assessment Group identified 7 additional, non-UK-based, studies of MammaPrint. Of these 7 studies, 4 on the clinical validity of MammaPrint demonstrated that the MammaPrint score is a strong independent prognostic factor, and may provide additional value to standard clinicopathological measures. A mix of evidence exists for outcomes at 5 and 10 years. The population in all these studies was relatively small. One of the studies was of a Japanese population, and follow-up was limited to only 5 years in 2 of the studies. For example, Mook et al. (2010) showed that in 148 women the distant metastasis-free survival at 5 years was 93% in the low-risk group and 72% in the high-risk group with an associated hazard ratio (HR) of 4.6 (95% confidence interval [CI] 1.8 to 12.0, p=0.001). The External Assessment Group did not

identify any prospective studies of the impact of MammaPrint on long-term outcomes such as overall survival. Six studies with data on the clinical utility of MammaPrint were identified by the External Assessment Group. Five studies reported use of MammaPrint to reclassify patients into high- and low-risk groups and compared this with the risk assigned according to current local guidance. They reported a high level of discordance between MammaPrint and current classification, although these studies did not demonstrate how this would impact on treatment decisions. For example, Bueno-de-Mesquita et al. (2009) compared MammaPrint risk categories and risk assessment based on Adjuvant! Online, St Gallen guidelines, NPI, and Dutch Institute for Healthcare Improvement (CBO) guidelines (2004). Discordance between MammaPrint and the other risk assessment measures was 38%, 41%, 26%, and 30% respectively. One study reported that the use of MammaPrint would result in altered treatment advice for 40% of patients based on the assumption that all patients classified as high risk would receive chemotherapy and patients classified as low risk would not receive chemotherapy. Because the study was retrospective, altered treatment advice assumed in the analysis represented potential changes and not actual changes from using the test in clinical practice.

5.12 A study of the benefit of chemotherapy according to risk group stratification by MammaPrint was identified (Knauer et al. 2010) but was omitted from the systematic review because it was based on a pooled analysis of 6 primary studies. The External Assessment Group did not include the pooled analysis in the systematic review to avoid double counting of studies already included in the review. In addition, it did not consider the findings of the pooled analysis to be robust because the authors did not reanalyse the tumour

samples, it is unclear how individual patient data were combined, and there were potential issues with the statistical analyses performed (for example, although median follow-up was 7.1 years the data were arbitrarily truncated at 5 years).

5.13 Robust evidence of clinical utility is not available for MammaPrint so it is not yet clear whether using the test will improve the use of adjuvant chemotherapy in the management of breast cancer in the UK. In summary, most studies of MammaPrint were retrospective in design, used small sample sizes and had heterogeneous patient populations. For example, some studies included only pre-menopausal women, which may overestimate the benefit of MammaPrint in the early breast cancer population as a whole. Moreover, no studies were conducted in the UK. The evidence for MammaPrint is based on the use of the test with fresh samples. It is not clear whether this evidence would apply if the test were used on formalin-fixed paraffin-embedded samples. Overall, the External Assessment Group considered that further robust evidence on the clinical validity and clinical utility of the test would be helpful.

Oncotype DX

5.14 Oncotype DX was reported to be furthest along the validation pathway by previous systematic reviews. In terms of clinical validity, these reviews reported evidence that the Oncotype DX recurrence score was significantly correlated with disease-free survival and overall survival. Furthermore, the recurrence score was shown to be a better predictor of distant recurrence at 10 years than traditional clinicopathological predictors. The evidence on clinical utility was limited. One study (Paik et al. 2006) demonstrated a significantly increased benefit from the use of chemotherapy in the Oncotype DX high-risk group compared with

the low-risk group, although the review highlighted that the study may have potential flaws. The study indicated that this benefit difference was caused by the better prognosis without chemotherapy (and hence the reduced absolute benefit these patients would receive) and the decreased relative benefit of chemotherapy in the lower-risk groups. The specific cancers in the low-risk groups were less likely to respond to chemotherapy, independent of actual survival probability. Key gaps were identified in the evidence base related to the extent to which the test added to the management of breast cancer and the proportion of patients who would benefit from the test. The previous systematic reviews indicated that prospective confirmation of the clinical utility of Oncotype DX was needed.

5.15 The External Assessment Group identified 12 additional studies of Oncotype DX. Further larger studies now support the prognostic ability of Oncotype DX. One large-scale UK study in post-menopausal women with ER+, LN− early breast cancer found that an increase in risk score was significantly associated with an increased risk of distant recurrence. This study demonstrated that a 50-point increase in the recurrence score (RS) in all LN− patients (for example, RS=55 compared with RS=5) was significantly associated with an increased risk of distant recurrence (HR 3.92; (95% CI 2.08 to 7.39; p<0.001) when adjusted for the effects of tumour size, grade (assessed locally), age and treatment. Furthermore, the evidence base has been extended to include the LN+ population. The External Assessment Group did not identify any prospective studies of the impact of Oncotype DX on long-term outcomes such as overall survival. Four studies were identified that presented further evidence on the impact of Oncotype DX on clinical decision-making. These indicated that the use of Oncotype

DX leads to changes in treatment decisions for between 32% and 38% of patients. However, only 1 of these studies was performed in the UK and limitations in relation to the generalisability of the study were identified. In addition, the study only included a small sample of patients, and although only interim results on 106 patients were available for the systematic review, the dataset for 142 patients was available for and used in the External Assessment Group's cost-effectiveness analysis. Four recent publications provided evidence that Oncotype DX predicts benefit from chemotherapy. However, only 1 study on an LN+ population (Albain et al. 2010) presented new data. The other 3 publications (Tang et al. 2011, Mamounas et al. 2010 and Tang et al. 2010) reported the same trial data as Paik et al. (2006). The first evidence of improvements in quality of life and reduced patient anxiety as a result of using Oncotype DX have been reported, but the studies had small sample sizes (for example, Lo et al. 2010 included 89 patients). In summary, Oncotype DX is considered to have the most robust evidence base of the tests reviewed in this guidance, with data on the analytical validity, clinical validity and clinical utility of the test. The studies varied considerably in their size, design and patient populations. Many of the Oncotype DX studies were small and retrospective. A small number of studies were conducted in the UK. The External Assessment Group considered that further robust evidence on the clinical utility of the test would be helpful.

IHC4

5.16 No studies on analytical validity of IHC4 (based on ER, PR, HER2 and Ki-67 in addition to classical clinical and pathological variables combined using an algorithm) were identified. Of the 4 individual tests that make up IHC4 (ER and HER2), 2 are commonly

measured in the NHS. However, the quantitative assessment of ER needed for IHC4 calculations is not routinely performed. Outstanding issues around the reproducibility of detecting Ki-67 also exist. This is noteworthy because the test is designed for local use and different local processing methods may potentially lead to different results. The External Assessment Group identified 1 study on the clinical validity of IHC4 (Cuzick, 2011), which reports that the IHC4 score is a highly significant predictor of distant recurrence. The authors validated the test in a cohort of 786 patients with ER+ cancer treated in the UK, and demonstrated that the IHC4 score was highly significantly predictive of outcome, with a hazard ratio of 4.8 (95% CI 2.2 to 10.2) for a change from the 25th to 75th percentile in a univariate analysis. This study also reported evidence comparing IHC4 against Oncotype DX. The study was rated as high quality. The External Assessment Group did not identify any prospective studies of the impact of IHC4 on long-term outcomes such as overall survival. It did not identify any published evidence on the clinical utility of IHC4 in terms of its ability to change treatment decisions or its ability to predict chemotherapy benefit. In summary, the External Assessment Group considered that the evidence base for IHC4 is currently limited to clinical validity (prognostic ability), although this evidence is considered to be relatively robust, and further evidence would be helpful on analytical validity and clinical utility.

Mammostrat

5.17 The External Assessment Group did not identify any studies on the analytical validity of Mammostrat, although some limited evidence on analytical validity was reported in a study of clinical validity. Three studies were identified that provided data to support the use

of Mammostrat as an independent prognostic tool for women with ER+, tamoxifen-treated breast cancer. Although the evidence base for Mammostrat is at present relatively limited, these studies included a large sample size, appeared to be of reasonable quality, and 1 study provided data from a UK setting. The External Assessment Group did not identify any prospective studies of the impact of Mammostrat on long-term outcomes such as overall survival. In addition, clinical utility data on Mammostrat (from 1 study) suggest that the low- and high-risk groups benefit from chemotherapy, but not the intermediate-risk group. There was no published evidence on reclassification of risk groups compared with conventional means of risk classification, and no evidence on the impact of the test on clinical decision-making. Overall, the External Assessment Group considered that further evidence of analytical validity and clinical utility would be helpful.

Economic analysis

5.18 Four studies were identified by the systematic review of cost-effectiveness evidence (2 for MammaPrint and 2 for Oncotype DX). None were conducted in England. Genomic Health and Clarient also submitted economic analyses on the cost effectiveness of Oncotype DX and Mammostrat in England respectively. Several issues were highlighted in the critique of these analyses, which needed further consideration. These included assumptions about: the baseline level of chemotherapy in clinical practice; the risk of distant recurrence; patients who would be offered the test; the proportion of patients who would be offered chemotherapy after reclassification with the new test; the cost of chemotherapy and therapy used to prevent or treat associated adverse events.

5.19 The External Assessment Group constructed a de novo economic model to specifically address the decision problem for this evaluation and to estimate the cost effectiveness of the 4 tests in England.

5.20 The population assessed in the economic model was women with ER+, LN–, HER2– early breast cancer up to 75 years old at diagnosis. One analysis assumed that all women in the group received the new tests. However, the External Assessment Group's clinical specialists suggested that the new tests may be targeted at a subgroup of this population – those at intermediate risk of distant recurrence for whom the decision about whether or not to give chemotherapy is most uncertain. A subgroup analysis was performed that assumed that the new test was given only to women with an NPI score above 3.4 (used as a proxy for those women at intermediate risk of distant recurrence), based on the assumption that most women at low risk (with an NPI score below 3.4) would not be considered for chemotherapy and that there would not be many women at high risk (with an NPI score of above 5.4) within the population considered.

Clinical outcomes

5.21 Modelling was used to estimate clinical outcomes. All women in the model were assumed to be treated with endocrine therapy. A state transition model was used to simulate breast cancer outcomes for patients treated with endocrine therapy alone or with the addition of chemotherapy. Outcomes associated with breast cancer were simulated using multiple health states including recurrence-free survival, recurrence (distant and local), adverse events from chemotherapy, and death.

Costs

5.22 The costs included in the economic model were the costs of the different tests, treatment costs (endocrine therapy and chemotherapy), costs of short-term and long-term adverse events associated with chemotherapy (including the secondary prevention of short-term adverse events), costs associated with managing distant recurrence, local recurrence and terminal care.

5.23 The cost of the MammaPrint test is £2675 (this cost was used in the economic model). The Oncotype DX test costs £2580 (this cost was used in the original economic analysis, but a revised cost was used for the economic analysis conducted for the confidential revised price). IHC4 was estimated to cost £100–£200 (£150 was used in the economic model) for quantitative analysis of ER (which may need additional time compared with traditional assessment of ER status), and assessment of PR and Ki-67 (which are not routinely collected) and running the algorithm (it was assumed that HER2 would be measured as part of standard practice). The Mammostrat test has an indicative cost between £1120 and £1620 (£1135 was used in the economic model).

Cost effectiveness

5.24 The primary analysis compared current clinical practice with treatment guided by Oncotype DX and IHC4. The systematic review of the evidence indicated most evidence for Oncotype DX compared with the other tests, and that the evidence base for Oncotype DX, in particular in relation to prognostic ability, was reasonably sound. There was less evidence for IHC4, but there was evidence relating to the performance of IHC4 compared with Oncotype DX. This evidence, with some additional assumptions when compared with the analysis of Oncotype DX, was used to

model the cost effectiveness of IHC4. Additional assumptions include the reproducibility of the test and the use of risk groups as opposed to a continuous risk score. There was no evidence on the ability of IHC4 to predict benefit from chemotherapy; in the IHC4 analysis the predicted benefit of chemotherapy was applied according to the Oncotype DX risk classification.

5.25 In addition to the primary economic analysis, further economic analyses were undertaken for Mammostrat and MammaPrint. These additional analyses were deemed exploratory by the External Assessment Group because there are significant limitations in the evidence base and the generalisability of the data to practice in England. Only data from studies conducted in the USA (Mammostrat) and The Netherlands (MammaPrint) were available to estimate the reclassification of patients using the new test. There was concern whether these data are generalisable to England and there was also uncertainty in the data. In addition, the External Assessment Group reported considerable uncertainty in the data used to estimate the predicted benefit of chemotherapy by MammaPrint risk groups.

5.26 All analyses assumed that the new tests were used in addition to current practice (for IHC4 it was assumed that quantitative analysis of ER (which may need additional time compared with traditional assessment of ER status), PR and Ki-67 is carried out in addition to current practice and data combined in an algorithm). Full details of the results can be found in section 5.6 of the diagnostics assessment report. A brief summary of the key results (including incremental cost-effectiveness ratios [ICERs]) of the base-case analysis and sensitivity analyses is presented below. Results of the 3 analyses (Oncotype DX and IHC4; Mammostrat; MammaPrint)

cannot be directly compared because the data came from different studies with different patient characteristics and methodologies, and the basis for each model therefore varies significantly.

5.27 The base-case analysis modelled a hypothetical cohort of 1000 women over a lifetime horizon (100 years was used as the upper age limit). Two analyses are presented. In the first the tests were used for all women with ER+, LN−, HER2− early breast cancer aged up to and including 75 years. In the second the tests were used only for women with ER+, LN−, HER2− early breast cancer up to and including 75 years with an NPI score above 3.4 (used as a proxy for those women at intermediate risk of distant recurrence). Results are presented on a per patient basis and any differences in expected values are a result of rounding error.

Oncotype DX and IHC4

5.28 **Tests used for all women with ER+, LN−, HER2− early breast cancer.** In the primary economic analysis comparing Oncotype DX and IHC4 with current practice, the proportions of patients receiving chemotherapy were 19.11%, 9.57% and 14.42% respectively. The model predicted that there would be 64, 71 and 76 distant recurrences when using Oncotype DX, IHC4 or current practice respectively. Total costs and QALYs, assuming predictive benefits (that is, benefits from identifying who will benefit most from chemotherapy) based on Paik et al., are summarised in table 1.

Table 1 Per-patient costs, QALYs and ICERs in the primary economic analysis (Oncotype DX and IHC4 compared with current practice) assuming the tests to be used for all women with ER+, LN−, HER2− early breast cancer in England. The model assumes predictive benefit based on Paik et al.

	Mean cost (£)	Mean QALYs	ICER – compared with current practice
Oncotype DX	£9094	13.54	£26,940[a]
IHC4	£6340	13.49	Dominant
Current practice	£6519	13.44	

[a]Rounding error contributes to the difference from expected value

ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year.

5.29 Compared with current practice, Oncotype DX was associated with an incremental cost of £2575 and incremental QALYs of 0.1 yielding an ICER of £26,940 per QALY gained. IHC4 was £179 cheaper than current practice (cost saving), with incremental QALYs of 0.05 and was found to be the dominant test (yielding greater QALYs at lower cost). Oncotype DX, IHC4 and current practice were also compared using incremental analysis; that is, the least effective strategy was compared with the next least effective strategy that was neither dominated nor extendedly dominated. The cost-effectiveness acceptability curve showed that the probability of IHC4 being cost effective (when compared with current practice) was almost 100% if the maximum acceptable ICER was £20,000 per QALY gained. At the same maximum acceptable ICER, the probability of Oncotype DX being cost effective, when compared with current practice only, was 12.4%.

5.30 **Tests used for women with ER+, LN−, HER2− early breast cancer and an NPI score above 3.4**. In the primary economic analysis the proportion of these patients predicted to receive

chemotherapy was 34.72%, 26.31% and 33.60% with Oncotype DX, IHC4 and current practice respectively. The model predicted that there would be 117, 129 and 144 distant recurrences when using Oncotype DX, IHC4 or current practice respectively. Total costs and QALYs, assuming predictive benefits based on Paik et al., are summarised in table 2.

Table 2 Per patient costs, QALYs and ICERs in the primary economic analysis (Oncotype DX and IHC4 compared with current practice) assuming the tests to be used for women with ER+, LN−, HER2− early breast cancer and an NPI score above 3.4 in England. The model assumes predictive benefit based on Paik et al.

	Mean cost (£)	Mean QALYs	ICER – compared with current practice
Oncotype DX	£10,911	13.06	£9007[a]
IHC4	£8318	12.97	Dominant
Current practice	£8816	12.83	

[a]**Rounding error contributes to the difference from expected value**

ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year.

5.31 Compared with current practice, Oncotype DX was associated with an incremental cost of £2095 and incremental QALYs of 0.23, which resulted in an ICER of £9007 per QALY gained. IHC4 was £498 cheaper than current practice (cost saving), with incremental QALYs of 0.14 and so was the dominant test. Oncotype DX, IHC4 and current practice were also compared using incremental analysis; that is, the least effective strategy was compared with the next least effective strategy that was neither dominated nor extendedly dominated. The cost-effectiveness acceptability curve

showed that the probability of IHC4 being cost effective (when compared with current practice) was almost 100% if the maximum acceptable ICER was £20,000 per QALY gained. At the same threshold, the probability of Oncotype DX being cost effective, when compared with current practice only, was 91.6%.

5.32 **Sensitivity analyses (univariate)**. A range of univariate sensitivity analyses were undertaken to explore the impact of varying the main model parameters. Analyses of varying the assumptions underlying the structure of the model were also performed. The ICERs for Oncotype DX compared with current clinical practice, for all women with ER+, LN−, HER2− early breast cancer and those with an NPI above 3.4, were sensitive (defined as changes in the ICER by 10% or more) to some of the assumptions made in the model. These included the time horizon modelled, the starting age of the cohort, the risk of recurrence, the proportion of patients receiving chemotherapy after reclassification with the new test, the benefit of chemotherapy in the different risk groups and the distribution of patients by NPI score. For example, the ICERs for Oncotype DX (compared with current practice) when offered to all women with ER+, LN−, HER2− early breast cancer were £91,300 (assuming 30% relative risk reduction from chemotherapy for all patients) per QALY gained and £64,900 (assuming 40% relative risk reduction from chemotherapy for all patients) per QALY gained. The ICERs for IHC4 compared with current clinical practice, for all women with ER+, LN−, HER2− early breast cancer and those with an NPI score above 3.4, were sensitive to more assumptions (such as the time spent in the distant recurrence health state, the proportion of patients receiving chemotherapy under current practice and the cost of chemotherapy), but IHC4 remained dominant compared with current practice (that is, it provided more

QALYs at a lower cost) except when the cost of IHC4 was set at £400 (the resulting ICER was £1557 per QALY gained).

5.33 Following the first consultation, the manufacturer of Oncotype DX submitted a proposal to make it easier for the NHS to access the technology. The proposal makes Oncotype DX available at a revised price. The proposed price is commercial in confidence. The proposal is made for patients at an intermediate risk of distant recurrence, defined as an NPI score above 3.4 in this guidance. An External Assessment Group analysis of the proposal, using the proposal price and the assumption that Oncotype DX is validated as a prognostic tool but does not predict the benefit patients will get from chemotherapy, yielded an ICER of £22,600 per QALY gained for patients with an NPI score above 3.4.

Mammostrat (exploratory analysis)

5.34 **Test used for all women with ER+, LN−, HER2− early breast cancer.** The proportion of patients receiving chemotherapy increased with the use of Mammostrat when compared with current practice (21.16% and 14.42% respectively). Current practice was associated with a mean cost of £7699 and mean QALYs of 12.86. Mammostrat was associated with a mean cost of £9040 and mean QALYs of 12.91. The ICER for Mammostrat was estimated to be £26,598 per QALY gained. The cost-effectiveness acceptability curve showed that the probability of Mammostrat being cost effective if the maximum acceptable ICER was £20,000 is 36.0%.

5.35 **Test used for women with ER+, LN−, HER2− early breast cancer and an NPI score above 3.4**. The proportion of patients receiving chemotherapy increased slightly with the use of Mammostrat when compared with current practice (34.27% and

33.60% respectively). Current practice was associated with a mean cost of £9717 and mean QALYs of 12.34. Mammostrat was associated with a mean cost of £10,985 and mean QALYs of 12.29. Mammostrat was shown to be dominated by current practice. The cost-effectiveness acceptability curve showed that the probability of Mammostrat being cost effective if the maximum acceptable ICER was £20,000 is 18.0%.

5.36 **Sensitivity analyses (univariate).** A range of univariate sensitivity analyses were undertaken to explore the impact of varying model parameters. When offering the test to all women with ER+, LN−, HER2− early breast cancer the ICER was very sensitive to the proportion of patients who would receive chemotherapy based on the test result. The ICER ranged between £18,879 per QALY gained to being dominated, when using the confidence intervals from the Ross et al. (2008) study for the predicted benefit of chemotherapy in terms of the reduction of distant recurrence. The ICER was not sensitive to the assumptions about utility values, management costs and the time spent in the recurrence health state. Mammostrat remained dominated under the assumptions examined in the sensitivity analysis when the test was offered to women with ER+, LN−, HER2− early breast cancer and an NPI score above 3.4.

MammaPrint (exploratory analysis)

5.37 **Test used for all women with ER+, LN−, HER2− early breast cancer.** The proportion of patients receiving chemotherapy increased with the use of MammaPrint when compared with current practice (44.18% and 14.42% respectively). Current practice was associated with mean costs of between £6408 and £6629, and mean QALYs of between 13.39 and 13.49. MammaPrint was

associated with mean costs of between £10,017 and £10,748 and mean QALYs of between 13.47 and 13.78. The ICER for MammaPrint was estimated to be between £12,240 and £53,058 per QALY gained. The ICER is given as a range because of the lack of UK data, the reliance on data mainly from pre-menopausal populations and uncertainty around the data for the predicted benefit of chemotherapy by MammaPrint risk group.

5.38 **Test used for women with ER+, LN−, HER2− early breast cancer and an NPI score above 3.4**. The proportion of patients receiving chemotherapy increased with the use of MammaPrint when compared with current practice (90.31% and 33.60% respectively). Current practice was associated with mean costs of between £8281 and £8872 and mean QALYs of between 12.81 and 13.07. MammaPrint was associated with mean costs of between £12,278 and £14,014 and mean QALYs of between 12.99 and 13.73. The ICER for MammaPrint was estimated to be between £6053 and £29,569 per QALY gained. The ICER is given as a range because of the lack of UK data, the reliance on data mainly from pre-menopausal populations and uncertainty around the data for the predicted benefit of chemotherapy.

5.39 **Sensitivity analyses (univariate and multivariate)**. Given the uncertainty in the base-case analysis a limited number of sensitivity analyses were undertaken. Univariate analyses included: assuming no additional cost to the NHS for the use of fresh tissue samples and that 5% of patients classified as good prognosis and 95% of patients classified as poor prognosis received chemotherapy. A multivariate sensitivity analysis explored different values for the benefit of chemotherapy in terms of reduction in the risk of distant

recurrence, assuming MammaPrint was used in all women with ER+, LN−, HER2− early breast cancer.

6 Considerations

6.1 The Diagnostics Advisory Committee discussed the focus of the evaluation and the evidence available for the 4 tests. It noted that gene expression profiling and immunohistochemistry tests other than those included in this evaluation are being developed. The Committee also noted that at present, the level and quality of the available evidence varies for the 4 tests. In particular, evidence on the tests' ability to guide clinical decisions on the use of chemotherapy in England and to predict response to chemotherapy in women with early breast cancer was limited. The External Assessment Group's economic model for women with ER+, LN− HER2− early breast cancer was used by the Committee when considering the likely cost effectiveness of the 4 tests. The Committee considered that the most appropriate use of these tests is in women for whom the decision to offer chemotherapy is uncertain, that is, women at intermediate risk of distant recurrence. It therefore considered that the subgroup analyses of women with an NPI score above 3.4 were the most relevant, based on the likelihood that there would not be many women with an NPI score of above 5.4 within the target population.

6.2 The Committee acknowledged the emotional and psychological strain for patients with breast cancer when considering therapy, in particular, chemotherapy and its associated adverse events. The Committee noted that this is likely to be significant in patients for whom the decision about whether or not to have chemotherapy is difficult after receiving the results of current tools used in the NHS

(especially patients deemed to be at intermediate risk). The Committee also noted that tools used by the NHS to assess the suitability of patients with breast cancer for adjuvant chemotherapy vary across England. The Committee concluded that any tests that can help to alleviate patient strain and promote consistency of practice within the NHS are likely to be appreciated by patients and clinicians alike.

6.3 The Committee discussed the generalisability of the data to men. The Committee acknowledged that breast cancer is not only observed in women and that men make up a small proportion of patients with breast cancer. The Committee noted that all the clinical and economic evidence had been based on trials with women; however, experts on the Committee stated that even though there are some subtle gender-specific differences in the pathobiology of breast cancer, the general subtypes are identical in men and women. Therefore, in clinical practice men would be treated in the same way as women. The Committee therefore concluded that the recommendations in this guidance should also apply to men.

6.4 The Committee discussed the evidence base for Oncotype DX and concluded that, in general, it was the most developed of the 4 tests in the evaluation. The Committee discussed the analytical validity of Oncotype DX. The Committee noted that no new evidence was identified in the External Assessment Group review, but that evidence was identified in the previous systematic review (Marchionni et al. [2008]) that showed reasonable within-laboratory replicability. The Committee also noted that the test is processed centrally by the manufacturer in the US and the laboratory is CLIA (Clinical Laboratory Improvement Amendments) certified. Given the

above, the Committee was satisfied with the analytical validity of the test. The Committee discussed the prognostic ability (clinical validity) of Oncotype DX. Experts on the Committee pointed out and the Committee agreed that the prognostic ability (the ability to predict the risk of distant recurrence) of Oncotype DX had been well validated. The Committee therefore concluded that the prognostic ability of Oncotype DX was supported by robust evidence in the early breast cancer population.

6.5 The Committee then discussed the clinical utility of Oncotype DX. It heard from the External Assessment Group that a key aspect of clinical utility is the ability of a test to accurately predict those patients who will benefit most from chemotherapy. The Committee therefore considered whether gains from chemotherapy could differ between patients with different prognoses (that is, patients in different risk groups). Experts on the Committee pointed to data from recent meta-analyses that showed proportional gains from chemotherapy were generally constant across clinical parameters such as tumour diameter and ER status (used to help determine a patient's prognosis). However, these constant proportional gains meant that those with a good prognosis would receive less absolute benefit from chemotherapy than those with a poor prognosis. Furthermore, the possibility that chemotherapy might be more effective both proportionally and absolutely in patients identified by Oncotype DX, given that the test provides information about the biological features of the tumour, was discussed. The possibility that tumours with the genetic characteristics identified by Oncotype DX might be more susceptible to chemotherapy was also explored. The evidence on the predicted benefit of chemotherapy (reduction in the risk of distant recurrence) for women receiving chemotherapy in addition to endocrine therapy compared with

endocrine therapy alone was discussed. The Committee heard that data were available that suggest that Oncotype DX can predict the relative benefit of chemotherapy and that the effectiveness of chemotherapy varies depending on the classification of patients by the Oncotype DX test in LN− patients (Paik et al. 2006). These data indicated that women in lower risk groups benefit proportionally less from chemotherapy than those in higher risk groups (see section 5.14). The Committee considered that the Paik study was limited by its design, the sample sizes of individual risk groups, and the use of some results from the training dataset (tamoxifen-treated patients of the NSABP B-20 trial) in the study dataset, the applicability of the study population (a younger population that includes patients with HER2+ breast cancer) to the population considered in this guidance, and the fact that the treatments (endocrine therapy and chemotherapy) used are different to those currently used in the NHS. In addition, the Committee considered that the relative benefit from chemotherapy by risk group was unclear. The Committee concluded that the evidence implying a predicted differential relative benefit of chemotherapy according to Oncotype DX risk group in LN− patients (Paik et al. 2006) was not robust. The Committee also reviewed evidence implying a predicted differential relative benefit of chemotherapy according to Oncotype DX risk group in LN+ patients (Albain et al. 2010) and data from the neoadjuvant setting. The Committee concluded that these data were not robust enough to support the test's ability to predict the benefit of chemotherapy. In the absence of robust data the Committee concluded that equal benefit of chemotherapy should be assumed across all Oncotype DX risk groups. Therefore, although the Committee considered that adequate evidence supported the prognostic ability of Oncotype DX (that is, its ability

to predict the risk of distant recurrence, see section 6.4) it concluded that it was not confident in the ability of Oncotype DX to predict benefit from chemotherapy.

6.6 The Committee discussed the cost effectiveness of Oncotype DX based on the original price proposed by the manufacturer (list price). The Committee considered that the ICERs from the base-case analysis of Oncotype DX were not robust because of the assumption of a predicted differential relative benefit of chemotherapy according to Oncotype DX risk group. The Committee discussed the ICERs presented in the sensitivity analysis that assumed equal benefit of chemotherapy across all Oncotype DX risk groups (at a level of either 30% or 40% relative risk reduction from chemotherapy). The Committee noted the ICERs for Oncotype DX (compared with current practice) when offered to all women with ER+, LN−, HER2− early breast cancer were £91,300 (30% relative risk reduction from chemotherapy) per QALY gained and £64,900 (40% relative risk reduction from chemotherapy) per QALY gained. The Committee considered the ICERs to be too high to recommend Oncotype DX for use in the NHS for all women with ER+, LN−, HER2− early breast cancer. The Committee considered that the overall benefit of chemotherapy was likely to be closer to 27% relative risk reduction from chemotherapy across all Oncotype DX risk groups (EBCTCG overviews 2005, 2011). The Committee noted that there are potential differences in the population included in the EBCTCG review compared with the population in the economic analysis and that the outcome measures differed. Without data specific to the population under consideration, the Committee considered the EBCTCG figure to be the most appropriate for use at this time. The Committee concluded that the most plausible ICER, based on the

evidence presented, was likely to exceed £91,300 for all women with ER+, LN-, HER2- early breast cancer. Therefore, based on the original proposed price (list price), Oncotype DX would not be a cost-effective use of NHS resources in this group.

6.7 The Committee then considered a proposal, submitted by the manufacturer of Oncotype DX. The proposal makes Oncotype DX available to the NHS at a revised price. The proposed price is commercial in confidence. NICE advised the Committee, and the Committee agreed, that the access proposal appeared workable and efficient, and did not appear to constitute an excessive administrative burden on the NHS. The Committee went on to discuss the impact of the proposal on the cost effectiveness of Oncotype DX in patients with an NPI score above 3.4 when it was assumed that Oncotype DX was able to predict a patient's prognosis but not the benefit of chemotherapy (relative risk reduction of distant recurrence from chemotherapy). The Committee accepted an analysis performed by the External Assessment Group which showed that the ICER for Oncotype DX (compared with current practice) in this group of patients was £22,600 per QALY gained, assuming prognostic benefits of the test but no predictive effect. The Committee also noted the ICER could be significantly lower if Oncotype DX were shown to predict the benefit of chemotherapy by robust evidence from future research. The Committee noted that an NPI score above 3.4 was used in the analysis as a mechanism for identifying patients at intermediate risk, but also noted that other methods for determining the risk group were available and in use in the NHS. The Committee believed that the subgroup analysis of people with an NPI score above 3.4 was likely to represent a reasonable approximation for people at intermediate risk generally. Therefore, given the strength

of the evidence on the prognostic ability of the test (and evidence of analytical validity), the Committee concluded that Oncotype DX for use in people at intermediate risk of distant recurrence where the decision to prescribe chemotherapy remains unclear, would represent a cost-effective use of NHS resources when acquired at the confidential revised price offered by the manufacturer.

6.8 The Committee discussed the need for further robust evidence to demonstrate the ability of Oncotype DX to identify patients who will benefit most from chemotherapy (see section 6.5). The Committee considered that further information on the clinical utility of the test is warranted, including robust evidence on the impact of Oncotype DX on clinical decision-making in England (the Committee noted that a decision-impact study is underway in Bristol) and the ability of the test to predict the benefit of chemotherapy. The Committee noted there is an ongoing prospective trial (TAILORx) that will provide further information on the benefit of chemotherapy in women classified as intermediate risk by Oncotype DX. As the patient population included in TAILORx is from North America, the Committee encouraged the collection of data on Oncotype DX when used in the NHS in England (see section 7).

6.9 The Committee discussed the evidence available on the analytical validity of IHC4. It noted the test was at a comparatively early stage of development. In particular, it was noted that although there are data on the reliability and reproducibility of the measurement of ER, PR and HER2 markers, data were lacking on the reliability and reproducibility of the Ki-67 marker measurement. The Committee heard that ER, PR and HER2 have an established UK National External Quality Assessment Scheme (NEQAS) and were commonly measured in the NHS, and that a study was published

recently on the reproducibility of Ki-67 and a UK NEQAS was being investigated for the marker. The Committee noted however that quantitative assessments of ER, PR and Ki-67 (not routinely reported in the NHS) should be appropriately considered in the NEQAS if not already done so. The Committee considered that data are needed on the reproducibility and reliability (analytical validity) of the complete IHC4 test (an algorithm combining 4 markers and classical clinical and pathological variables). This is particularly important as the test is designed for local processing in NHS laboratories. An additional study on quality assurance was also considered by the Committee. This was a small preliminary study that did not materially change the results of the External Assessment Group analysis. The Committee concluded that the lack of data on analytical reliability meant that it was not possible to make a recommendation for general use of the IHC4 test at this time.

6.10 The Committee then discussed the clinical validity and clinical utility of IHC4. It noted that only 1 study was available on the clinical validity of the test. The Committee discussed the separate cohort of 786 patients used for external validation of the test in this study and concluded that it was not fully representative of the population of interest in this assessment because approximately 50% of patients did not receive 5 years of endocrine therapy. The Committee also noted that the External Assessment Group review did not identify any data on the clinical utility of IHC4. An additional study on how the test classifies patients by risk group compared with the NPI and Adjuvant! Online was also considered by the Committee. Although encouraging, this was a small preliminary study that did not provide an indication of how management decisions would actually change and did not materially change the results of the External

Assessment Group analysis. The Committee also noted the recent availability of further data on IHC4 in the large TEAM study. The Committee considered that the general uncertainty in the clinical effectiveness evidence for IHC4 limited the validity of the economic analysis. It concluded that robust data on how the test might be used in the NHS in England (a continuous risk score, defined risk groups or both) and the impact of the test on clinical decision-making are needed. The Committee also indicated that data on the benefit of chemotherapy according to IHC4 score (or defined risk groups) would be useful. Although IHC4 was found to dominate current practice in the base-case economic analysis when offered to all women with ER+, LN−, HER2− early breast cancer and in a subgroup of women with an NPI score above 3.4, in addition to most of the sensitivity analyses, the Committee concluded that the uncertainty in the estimates of the clinical effectiveness of the test was too great to recommend adoption at this time. The Committee considered that further evidence was needed before the test could be adopted for general use by the NHS. Given the estimated low cost of the test and the modelling results that showed it has the potential to dominate current practice, the Committee considered it prudent to recommend the use of IHC4 for research in the NHS to collect information on the analytical validity, and hence, clinical validity and clinical utility of the test (see section 7).

6.11 The Committee discussed the clinical evidence and the uncertainty in the estimates of the cost effectiveness of Mammostrat (because of uncertainty in the evidence underpinning the economic analysis). The Committee noted there were limited published data on the analytical validity of the test. It went on to discuss the clinical validity of the test, and considered the results of the economic analysis to be limited because the risk reclassification data

(provided in confidence) derived from a subset of women included in the study by Ring et al. 2006 (US cohort) were not sufficiently robust in demonstrating the ability of the test to predict the risk of distant recurrence in the subgroup of women with an NPI score above 3.4. The Committee noted the recent availability of data on the clinical validity of Mammostrat in the TEAM study. The Committee considered that this study provides additional supportive data for a large UK population on the prognostic ability of the test and expands the evidence base to patients treated with aromatase inhibitors, rather than tamoxifen. Although further supportive data are available on the clinical validity of the test, the Committee considered that the economic analysis was also limited because there was uncertainty in the clinical utility of the test, including how the test would affect clinical decision-making in England, and because of the discordant results (Ross et al. 2008) on the benefit of chemotherapy (only the low- and high-risk groups benefitted from chemotherapy but not the intermediate-risk group). The Committee considered that the uncertainty in the clinical effectiveness evidence for Mammostrat limited the validity of the economic analysis. Therefore, given the uncertainty in the clinical effectiveness of the test (in particular, the analytical validity and clinical utility), the Committee was unable to recommend the adoption of Mammostrat for general use in the NHS at this time and recommended the test for research use only. The Committee heard that there is an extensive ongoing research programme for this relatively new test.

6.12 The Committee discussed the clinical evidence and the uncertainty in the estimates of cost effectiveness of MammaPrint. It noted that no new evidence was identified in the External Assessment Group's review on the analytical validity of the test, but that

evidence had been identified in the previous systematic review by Marchionni et al. (2008). The Committee went on to consider the different sample types used by the test and concluded that it was not clear whether the evidence on the clinical effectiveness of the test when used on fresh samples would apply if the test is used on formalin-fixed paraffin-embedded samples. The Committee was aware that the manufacturer had submitted data to the FDA showing that the performance of MammaPrint in formalin-fixed paraffin-embedded samples is equivalent to that of fresh samples, but that this evidence had not been reviewed by the External Assessment Group or the Committee. The Committee discussed the clinical validity of the test and agreed with the External Assessment group that such evidence, although developing, is based on cohort studies of small sample sizes that have been conducted outside of England in predominantly pre-menopausal women. The Committee discussed the clinical utility of the test and noted that the risk reclassification data and the proportion of patients receiving chemotherapy were taken from studies of Dutch patients that included predominantly pre-menopausal women (younger women are more likely to be classified as having a poor prognosis using MammaPrint, which may overestimate the benefit of the test in the early breast cancer population as a whole) as well as post-menopausal women. Furthermore, the Committee agreed with the assessment of the External Assessment Group that the Knauer et al. study (2010) had considerable methodological limitations (see section 5.12) and therefore the Committee considered that clinical utility of MammaPrint (in particular, the benefit from chemotherapy) had not been robustly demonstrated. The Committee considered that the uncertainty in the clinical-effectiveness evidence for MammaPrint limited the validity of the

economic analysis. The Committee also considered additional evidence forwarded by the manufacturer, including a summary of cost-effectiveness results based on NPI scores from 2 patient series data and the RASTER study (updated analysis); however, the Committee agreed with the External Assessment Group that concluded this evidence did not materially change the results of the analysis. The Committee concluded that the uncertainty in the clinical effectiveness, in particular the clinical validity and clinical utility of the test, was too high to recommend the adoption of MammaPrint for general use in the NHS at this time and recommended the test for research only. The Committee noted that there is an ongoing prospective clinical trial (MINDACT) on the value of MammaPrint in predicting which patients would benefit from chemotherapy and that results from this trial may help to reduce the uncertainty about effectiveness.

6.13 The Committee expressed general concern over the lack of information on the impact of the use of gene expression profiling and expanded immunohistochemistry tests on clinical decision-making in England. The Committee requested further data on the ability of the tests to impact clinical decision-making. The applicable data forwarded by the manufacturers came from studies that were conducted outside England. The Committee agreed with the External Assessment group that a lower baseline level of chemotherapy prescribing in England increases the uncertainty in the generalisability of studies conducted outside England.

6.14 A potential equality issue was raised by the Committee, which was concerned about the lack of evidence on the use of the tests in women older than 75 years. The Committee accepted the evidence had been limited to women younger than 75 years; however, the

recommendations in section 1 do not restrict access to the tests based on age of the patient. The Committee also discussed potential equality issues concerning the use of the new tests in men. The Committee heard that given the relatively low number of men with breast cancer when compared with women, evidence on the performance of these tests in men was less likely to be generated. Experts on the Committee pointed out that breast cancer in men shared many characteristics with that seen in women and that both groups were treated similarly in clinical practice (see section 6.3). Therefore, the Committee felt it appropriate that the recommendations apply to both men and women.

7 Proposed recommendations for further research

7.1 The following research is recommended in the context of people with ER+, LN−, HER2− early breast cancer.

7.2 MammaPrint: research is recommended on the clinical validity of the test in people that are representative of the population in England. In particular, information on how the test reclassifies people when compared with current practice in England and their risk of distant recurrence would be useful. Research into the clinical utility of the test is also recommended; in particular, evidence of the impact of the test on clinical decision-making in England and robust data on its ability to predict the benefit of chemotherapy.

7.3 Oncotype DX: research is recommended on the clinical utility of the test, including robust evidence on the impact of Oncotype DX on clinical decision-making in England and its ability to predict the

benefit of chemotherapy. As part of the adoption of Oncotype DX by the NHS the Committee encourages the collection of clinical utility and any other useful data by the health system.

7.4 The Committee noted that the MINDACT study is being conducted on MammaPrint and the TAILORx study is being conducted on Oncotype DX. The Committee encourages the availability of data showing the risk of distant recurrence of patients in these trials using tools representative of current practice adopted by the NHS (for example, NPI and Adjuvant! Online). Researchers should be mindful of the evolving breast cancer practice in England. For example, the emergence of the PREDICT tool; although a new tool at present, this may be more widely used in the future.

7.5 IHC4: research into the analytical validity (reliability and reproducibility) of the complete IHC4 test is recommended (an algorithm combining 4 markers and classical clinical and pathological variables), particularly within the NHS and when performed in local laboratories. Studies to confirm the prognostic ability and to determine the impact of IHC4 on clinical decision-making in England and, ideally, to predict the benefit of chemotherapy are recommended.

7.6 Mammostrat: research on the analytical validity (reliability and reproducibility) of the test is recommended. Research on the clinical utility of the test is also recommended. In particular, evidence of how the test reclassifies people's risk when compared with current practice in England, evidence on the impact of Mammostrat on clinical decision-making in England, and its ability to predict the benefit of chemotherapy.

8 Implementation

8.1 NICE will support this guidance with a range of activities to promote the recommendations for further research. This will include incorporating the research recommendations in section 7 into the NICE guidance research recommendations database (available on the NICE website at www.nice.org.uk) and highlighting these recommendations to public research bodies. The research proposed will also be put forward to NICE's Medical Technologies Evaluation Programme research facilitation team for consideration of the development of specific research protocols.

8.2 The manufacturer has offered Oncotype DX to the NHS under a proposal (December 2012) that makes Oncotype DX available to the NHS at a revised price. The proposal price is commercial-in-confidence. It is the responsibility of the manufacturer to communicate details of the proposal to the relevant NHS organisations. Any enquiries from NHS organisations about the access proposal should be directed to [to be added by NICE at the time of publication of final guidance].

9 Related NICE guidance

See www.nice.org.uk for related guidance.

10 Review

NICE updates the literature search at least every 3 years to ensure that relevant new evidence is identified. NICE will contact product sponsors and other stakeholders about issues that may affect the value of the diagnostic technologies. NICE may review and update diagnostics guidance at any time if significant new evidence becomes available.

Professor Adrian Newland

Chair, Diagnostics Advisory Committee

February 2013

Appendix A: 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 Trevor Cole
Consultant Clinical and Cancer Geneticist, Birmingham Women's Hospital

Dr Paul Collinson
Consultant Chemical Pathologist, St George's Hospital

Dr Sue Crawford
General Practitioner (GP) Principal, Chillington Health Centre

Professor Ian Cree
Senior Clinical Advisor, NIHR Evaluation Trials and Studies Coordinating Centre, University of Southampton

Professor Erika Denton
National Clinical Director for Imaging, Department of Health, Honorary Professor of Radiology, University of East Anglia and Norfolk and Norwich University Hospital

Dr Simon Fleming
Consultant in Clinical Biochemistry and Metabolic Medicine, Royal Cornwall Hospital

Professor Lisa Hall
Professor of Analytical Biotechnology, University of Cambridge

Professor Chris Hyde

Professor of Public Health and Clinical Epidemiology, Peninsula Technology Assessment Group (PenTAG)

Professor Noor Kalsheker

Professor of Clinical Chemistry, University of Nottingham

Dr Mark Kroese

Vice Chair, Diagnostics Advisory Committee and Consultant in Public Health Medicine, PHG Foundation, Cambridge and UK Genetic Testing Network

Professor Adrian Newland

Chair, Diagnostics Advisory Committee

Dr Richard Nicholas

Consultant Neurologist, Honorary Senior Lecturer, Heatherwood and Wexham Park Hospitals

Ms Margaret Ogden

Lay representative

Mr Stuart Saw

Director of Finance, North East London and the City PCTs

Professor Mark Sculpher

Professor of Health Economics at the Centre for Health Economics, University of York

Dr Steve Thomas

Consultant Vascular and Cardiac Radiologist at Sheffield Teaching Hospitals Foundation Trust

Mr Paul Weinberger

CEO, Diasolve Ltd, London

Mr Christopher Wiltsher

Lay representative

Specialist Committee members

Professor Anthony Howell

Director of the Breakthrough Breast Cancer Research

Mr Simon Pain

Consultant Breast and Endocrine Surgeon, Department of General Surgery, Norfolk and Norwich University Hospital

Mrs Ursula Van Mann

Lay representative

Mrs Carole Farrell

Nurse Clinician, The Christie NHS Foundation Trust

Professor Louise Jones

Consultant Clinical Scientist, Barts and the London NHS Trust

NICE project team

Each diagnostics assessment is assigned to a team consisting of a Technical Analyst (who acts as the topic lead), Technical Advisers and a Project Manager.

Gurleen Jhuti

Topic Lead

Hanan Bell and Pall Jonsson

Technical Advisers

Jackson Lynn

Project Manager

Appendix B: Sources of evidence considered by the Committee

The diagnostics assessment report was prepared by the School of Health and Related Research (ScHARR).

- Ward S, Scope A, Rachid R et al. Gene expression profiling and expanded immunohistochemistry tests to guide the use of adjuvant chemotherapy in breast cancer management, October 2011.

Registered stakeholders

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

Manufacturers/sponsors:

The technologies under consideration

- Agendia Bvd (MammaPrint)
- Clarient (Mammostrat)
- Genomic Health (Oncotype DX)
- Royal Marsden Hospital and Queen Mary University London – academic sponsor (IHC4)

Comparator technologies

- None

Other

- ARUP Laboratories
- Ipsogen
- Nottingham University
- Randox Laboratories
- Roche Diagnostics

Professional/specialist and patient/carer groups:

- Breakthrough Breast Cancer
- Bupa
- Royal College of Physicians
- Southport and Ormskirk Hospital NHS Trust