EndoPredict gene expression profiling assay for assessing risk of breast cancer recurrence

Medtech innovation briefing
Published: 10 November 2015
nice.org.uk/guidance/mib44

Summary

EndoPredict is a multi-gene assay that predicts the likelihood of women with oestrogen receptor (ER)-positive and human epidermal growth factor receptor 2 (HER2)-negative breast cancer developing metastases within 10 years of the initial diagnosis. The test combines measurements of gene expression (the EndoPredict [EP] score) with nodal status and tumour size to generate a comprehensive risk score (the EPclin score) which is used to identify tumour types that will not benefit from chemotherapy. The published clinical evidence summarised in this briefing comes from 3 analytical validation and 5 clinical validation studies in which the test was generally shown to be reproducible and to have prognostic power. In 1 impact evaluation study, EndoPredict results were reported to change treatment decisions. A cost-effectiveness analysis found that using EndoPredict in combination with non-UK clinical guidelines was less costly and more effective than clinical guideline risk stratification alone. One EndoPredict test costs £1000 if performed in a local laboratory or £1500 if it is sent to the distributor's laboratory.
### Product summary and likely place in therapy

- EndoPredict is a prognostic gene signature assay that uses gene expression to calculate the risk of women with breast cancer developing metastases within 10 years of diagnosis.

- Gene expression data (EP score) is combined with clinical and pathology information to produce a hybrid score (EPclin score) to guide treatment decisions for women with ER-positive and HER2-negative breast cancer. The test can be used to determine whether there is any added benefit in giving chemotherapy to women undergoing endocrine therapy.

### Effectiveness and safety

- The published evidence summarised in this briefing comes from 9 studies including a total of 3404 patients.

- Three of the studies examined analytical validity and found that the test was reproducible.

- Three studies retrospectively investigated the clinical validity of EndoPredict by assessing large prospective clinical trials. The studies found that the EndoPredict score was predictive of metastasis.

- One retrospective clinical validation study (n=553) found that the score was associated with the tumour response to chemotherapy. Another retrospective clinical validation study compared 9 breast cancer gene signature tests and found EndoPredict had significant prognostic power.

- One impact evaluation study (n=167) examined the performance of EndoPredict in clinical practice and found that the results changed clinicians' treatment decisions.
### Technical and patient factors

- The assay measures the gene expression profiles of 8 disease-relevant genes and 3 control or reference genes. The EndoPredict assay uses ribonucleic acid (RNA) that has been isolated from formalin-fixed, paraffin-embedded (FFPE) breast tumour tissue.
- The test must be carried out by experienced laboratory personnel using validated reagents and equipment following the manufacturer’s instructions. The test can be performed in a decentralised setting (local pathology laboratory) or at 1 of several specialist laboratories worldwide.
- It takes 2 days to receive the test results (including preparation time) if the local pathology laboratory is used. The results may be delayed by a few days if the sample is sent to one of the specialist laboratories.

### Cost and resource use

- Each EndoPredict test costs £1000 (including reagents and extraction) if done in a local laboratory or £1500 if it is sent to the distributor's laboratory.
- A cost-effectiveness study carried out in Germany found that the average life-long cost per patient treated according to clinical guidelines was less when combined with the EndoPredict score.
- The PCR module costs £40,000, but the manufacturer offers a free reagent rental agreement if the laboratory agrees to use EndoPredict for 3 to 5 years and returns the PCR module at the end of the contract term (or renews the contract).

### Introduction

Breast cancer is the most common cancer overall and the second most common cause of cancer death in women in the UK. In 2011, 49,936 women were diagnosed and 11,643 died from breast cancer (Cancer Research UK, 2014a). Breast cancer also occurs in men, although it is very rare. Approximately 350 men are diagnosed with breast cancer in the UK annually (Cancer Research UK, 2014b).

Breast cancer survival depends on the biology of the tumour, the stage of the disease at diagnosis and the treatment received. More than 90% of women diagnosed with early breast cancer survive for at least 5 years. In contrast, only 13% of those diagnosed with advanced disease survive for more than 5 years (Cancer Research UK, 2014a). Because there is limited awareness of the condition in men, they often present only when their symptoms become severe and the disease is at an advanced stage (NHS Choices, 2013). As a result, breast cancer prognosis is not as good in men as in women.
Breast cancer is diagnosed through a range of tests. These include clinical examination, imaging tests such as mammography, ultrasound, MRI and CT, and breast biopsy. Based on the results of these tests, breast cancer can be categorised into 4 stages (Cancer Research UK, 2014c):

- **Early breast cancer (stages 1 and 2).** In stage 1, the tumour, if present, is 2 cm or smaller in size and/or small clusters of cells (0.02–0.2 cm) are found in the lymph nodes close to the breast. Stage 2 usually means that the tumour is larger than in stage 1 (2–5 cm) and/or cancer cells are found in 1 to 3 lymph nodes in the armpit or in the lymph nodes near the breastbone.

- **Locally advanced breast cancer (stage 3).** This refers to a tumour which has not spread to another part of the body, but may be larger than 5 cm in diameter, growing into the skin or muscle of the chest, or present in 4 or more lymph nodes in the armpit.

- **Advanced breast cancer (stage 4).** This refers to cancer which has metastasised and spread to other parts of the body such as the bones, lungs, liver or brain.

Stages 1 to 3 can be split into subcategories based on the size of the tumour and whether the lymph nodes are affected. Sometimes clinicians use the term stage 0 to describe non-invasive breast cancer tumours such as ductal carcinoma in situ and lobular carcinoma in situ.

The main treatments used for breast cancer include surgery, radiotherapy, chemotherapy, hormone therapy and biological treatments (Cancer Research UK, 2014d). Oncologists use a number of tools (for example Adjuvant! Online or PREDICT) to make decisions about the best treatment or combination of treatments for each person and the risk of disease recurrence. These tools categorise prognosis as 'good' (low risk of metastases) to 'poor' (high risk of metastases). They can also indicate which people may benefit most from adjuvant treatment with endocrine therapy or chemotherapy. Risk is calculated using a number of factors, including the following:

- **Disease stage (size of the tumour, number of positive lymph nodes):** large tumours and positive lymph nodes are associated with a poorer prognosis. A diagnosis of ductal carcinoma in situ or lobular carcinoma in situ is also thought to increase the risk of invasive breast cancer (Li et al. 2014).

- **Grade of cancer cells (the extent to which the appearance of the cancer cells differs from that of normal cells):** a higher grade is associated with a poorer prognosis.

- **Hormone receptor (HR) expression and human epidermal growth factor 2 (HER2) status of the cancer cells:** HR-positive/HER2-negative cancer has a better prognosis.
A combination of the grade of the cancer cells and the expression of hormone receptor genes. The 4 most common molecular subtypes are:

- Luminal A, characterised by being oestrogen receptor (ER)-positive and low grade, and associated with good prognosis
- Luminal B, characterised by being ER-positive but often high grade and associated with worse prognosis than Luminal A
- Basal-like, characterised by being ER-negative, progesterone receptor (PR)-negative, HER2-negative, (also called triple-negative breast cancer) and associated with poor prognosis
- HER2-enriched, characterised by HER2-overexpression and associated with poor prognosis.

A reliable test for the prediction of long-term outcomes would help clinicians and patients make more informed decisions on treatment after surgery. There is, therefore, considerable interest in molecular genetic tests designed for this purpose.

Technology overview

This briefing describes the regulated use of the technology for the indication specified, in the setting described, and with any other specific equipment referred to. It is the responsibility of health care professionals to check the regulatory status of any intended use of the technology in other indications and settings.

About the technology

CE marking

EndoPredict comprises the EndoPredict kit and the EndoPredict Report Generator (EPRG) software, and was classified in November 2012 according to article 9 of the in vitro diagnostic directive 98/79/EC as 'Other product'. EndoPredict is manufactured by Sividon Diagnostics. In the UK, EndoPredict is distributed by Myriad Genetics.

Description

EndoPredict is an in vitro diagnostic tool that uses information from gene expression levels to calculate the risk of women with early-stage ER-positive/HER2-negative breast cancer developing
metastases within 10 years of diagnosis. The test is suitable for use in women without positive lymph nodes, as well as with up to 3 positive lymph nodes.

The test measures the expression profiles of 11 genes:

- Eight disease-relevant genes
  - three proliferation-associated genes (BIRC5, UBE2C, DHCR7)
  - five hormone receptor-associated genes (RBBP8, IL6ST, AZGP1, MGP, STC2)
- Three normalisation genes (CALM2, OAZ1 and RPL37A).

In addition, the test measures 1 DNA reference gene (HBB). This is used to check if there is DNA contamination in the sample which may affect the readings of the other genes.

The gene expression levels are analysed using reverse transcription-quantitative polymerase chain reaction (RT-qPCR), in which target messenger RNAs are reverse transcribed, amplified and simultaneously detected. In order to carry out the EndoPredict test the following components are required (Sividon Diagnostics, 2014):

- The EndoPredict kit. This includes:
  - EndoPredict UNO plates or EndoPredict DUO plates (including cap strips and quick start guide): these plates include polymerase chain reaction (PCR) primers and probes and are suitable for analysis of 1 or 2 patient samples respectively.
  - EndoPredict PICO qREF and EndoPredict PICO Qpcr-H$_2$O: these are the positive (RNA) and the negative (water) controls for the PCR. These are added in the UNO and DUO plates.
  - EndoPredict PICO qPCR-H$_2$O is also used in the preparation of the PCR master mix, which is added to the patient RNA samples before starting the PCR reaction.

- The VERSANT kPCR AD module (Siemens Healthcare Diagnostics Products GmbH): this carries out all of the steps involved in the test from RNA extraction to quantification of the PCR products. According to the manufacturer, EndoPredict has only been validated for use with this instrument.
- Other consumables that are necessary but are not included in the kit:
- The Tissue Preparation Reagents (Sividon Diagnostics GmbH) or the VERSANT Tissue Preparation Reagents (Siemens Healthcare Diagnostics Products): these are the reagents for RNA extraction.

- The EndoPredict TAQO (SuperScript III Platinum One-Step Quantitative RT-PCR System; Sividon Diagnostics GmbH/Thermo Fisher Scientific): this product includes the enzymes, PCR reaction mix and other components necessary for the reverse transcription and PCR process.

- EndoPredict QC: this is a kit (including quality control strips) for the optional quality control of the RNA (determination of the RNA yield and the quality of the DNA digestion).

- The EPRG web application: analysis software that calculates the EP and EPclin score. EPRG is freely available on 2 separate web servers. ([Server 1](#) and [Server 2](#)).

The EndoPredict test uses RNA samples extracted from formalin-fixed, paraffin-embedded (FFPE) breast cancer tissue. If the test cannot be performed locally, FFPE samples can be submitted to a Myriad-accredited pathology laboratory. The shipment must include an H&E (haematoxylin and eosin stain) slide.

The tissue section should meet the following requirements:

- It is recommended to isolate RNA from enough tissue to give a mean Ct of less than 27 for the 3 reference genes. The Ct value is defined as the number of cycles required for the signal to exceed background level. This is verified as a part of the quality control that is automatically carried out with every EndoPredict test. In most cases a 10 micrometre tissue section – either from a resected tumour block or from a core needle biopsy – will yield sufficient RNA for 1 EndoPredict test. Alternatively, FFPE sections scraped from slides can also be used. The manufacturer also provides a QC strip (included in the EndoPredict QC kit) so that a laboratory can check that the yield and quality of the RNA are sufficient for EndoPredict analysis. This optional QC process can be used before adding the reagents and initiating the PCR reaction.

- The tumour should make up 30% or more of the total tissue volume; only if the tumour content is less than 30% should adjacent (healthy) tissue be removed. This can be normally accomplished with a scalpel.

After the PCR has been completed the raw data are exported to the EPRG where data processing, including evaluation of the QC data and normalisation of gene expression values takes place.
Finally, the software calculates the risk classification and the metastasis risk for the patient over the next 10 years. The results consist of 2 scores:

- The EP score (also known as the 'molecular fingerprint') is represented numerically and graphically. The score refers to the expression of the 8 disease-relevant genes and is calculated using a mathematical formula that evaluates the results on a scale from 0 to 15. EP scores of less than 5 are designated as 'low risk' and EP scores of 5 or more as 'high risk' (Filipits et al. 2011).

- The EPclin score is calculated by adding clinical data about tumour size and nodal status to the EP score. From the EPclin score, the probability of metastasis formation within 10 years is estimated using an algorithm (Filipits et al. 2011), and presented graphically.

It takes approximately 2 days to receive the tests results (including preparation time) if the local pathology laboratory is used. The results may be delayed by a few days if the samples are sent to one of the specialty laboratories.

**Setting and intended use**

EndoPredict is intended for the long-term prognosis of women with early-stage breast cancer who meet the following criteria:

- the cancer is ER-positive/HER2-negative
- node-negative or node-positive with up to 3 positive lymph nodes.

EndoPredict is not currently indicated for men with breast cancer.

EndoPredict is intended for use in local pathology laboratories that use validated PCR systems and reagents. The test would be requested by a clinician involved in managing breast cancer, most likely an oncologist. The test is intended to be performed by professionals who are trained in general molecular biology techniques and have received EndoPredict-specific training from Sividon Diagnostics or Myriad Genetics (the distributor) on site. If no local pathology laboratory offers the test, the tumour sample can be sent to one of several accredited molecular pathology laboratories in the UK, Germany, Austria, Switzerland, Spain, Italy or France, or other laboratories in America, Asia and Australia. Myriad provides shipping instructions for the sample and arranges for the test to be carried out in one of the laboratories. The clinical data are combined with the EP score and the final results are sent to the hospital.
Current NHS options

Currently, the process for predicting the risk of disease recurrence and death in people with breast cancer involves the use of tools such as the Nottingham Predictive Index, the online prognosis algorithm Adjuvant! Online and PREDICT.

The Nottingham Predictive Index uses information about tumour grade, lymph node involvement and tumour size to predict prognosis. The Nottingham Predictive Index is most commonly used to divide people into 5 prognostic groups ranging from good to poor prognosis.

The Adjuvant! Online is a web-based risk assessment programme. Adjuvant! Online uses similar factors to the Nottingham Predictive Index but also includes patient age, hormone receptor status and comorbidity level. These variables are used to calculate the patient's estimated 10-year survival probabilities, risk of relapse and the expected benefit of adjuvant therapy.

PREDICT was launched to the NHS in September 2010. It is a tool similar to the Nottingham Predictive Index and Adjuvant! Online which is based on cancer registry data from women treated in the UK and includes HER2 and Ki-67 status (a histology marker of proliferation). The PREDICT tool is used to estimate breast cancer survival and the benefits of hormone therapy, chemotherapy and the biological therapy trastuzumab.

The NICE guideline on early and locally advanced breast cancer: diagnosis and treatment recommends using Adjuvant! Online to support estimates of risk of recurrence and calculate the absolute benefit of adjuvant treatment for patients with early invasive breast cancer. The guideline also recommends that adjuvant therapy should be considered for all patients with early invasive breast cancer after surgery. The decision about adjuvant therapy should be made after considering all prognostic and predictive factors and following discussion with the patient regarding the potential benefits and side effects of treatment.

NICE diagnostic guidance on gene expression profiling and expanded immunohistochemistry tests for guiding adjuvant chemotherapy decisions in early breast cancer management recommends using Oncotype DX (Genomic Health) in specific patient groups for guiding chemotherapy decisions for people with ER-positive, lymph node negative and HER2-negative early breast cancer.

In addition to Oncotype DX, NICE is aware of the following tests that appear to fulfil a similar function to EndoPredict:

- Prosigna (Nanostring Technologies)
- MammaPrint (Agendia)
- BluePrint
- TargetPrint
- NexCourse Breast IHC4 (Geneoptix).

**Costs and use of the technology**

Information on the cost of using the technology has been provided by the manufacturer. The cost of EndoPredict including all extraction reagents and consumables is £1000 per test if run locally, and £1500 per test if carried out at Myriad in Munich. The VERSANT kPCR module costs £40,000, but the manufacturer offers a free reagent agreement if the laboratory agrees to use EndoPredict for 3 to 5 years and returns the PCR module at the end of the contract term (or renews the contract). The Nottingham Predictive Index, the Adjuvant! Online algorithm and the PREDICT tool are free to use.

No other practical difficulties have been identified in using or adopting EndoPredict.

**Likely place in therapy**

EndoPredict would be used as an alternative to current prognostic tools in the planning of therapy for women with early-stage ER-positive/HER2-negative breast cancer with either no or up to 3 positive lymph nodes and who belong to an intermediate, clinically unclear risk group (according to Nottingham Prognostic Index or other decision-making tools or protocols currently used in the NHS). If adopted, EndoPredict could help to identify a subpopulation at low risk of developing long-term metastases (up to 10 years after diagnosis) and avoid unnecessary chemotherapy.

**Specialist commentator comments**

According to 1 specialist commentator, EndoPredict is not a technology that would add value to any point of the current treatment pathway for ER-positive/HER2-negative population, because of its similarity with the Oncotype DX technology – already endorsed by NICE diagnostics guidance. The commentator also notes that even though there are cost savings suggested by the health economic analyses, these would be unlikely to apply to the NHS. Drawing from their professional experience, the commentator explained that there is more conservative use of chemotherapy in the UK in comparison with other countries and hence the introduction of EndoPredict on a large scale is likely to lead to an increase in costs that will probably not be accompanied by the predicted cost-savings. Two specialist commentators stated that 28,000 patients is an overestimation of the
number of people per year for whom EndoPredict would be helpful. The number is a crude estimate and does not take factors such as age into consideration.

One specialist commentator noted that the majority of oncologists in the UK use either Adjuvant! Online or PREDICT, with fewer people using the Nottingham Predictive Index. The same commentator suggested that there would be more value in using EndoPredict together with PREDICT or Adjuvant! Online and explained that often the patients might have a low risk of breast cancer metastasis based on a clinical and pathology evaluation but a high score with such a prognostic test. Another commentator emphasised that, in general, EPclin provides more information than the current Adjuvant! Online estimation.

One specialist commentator noted that EndoPredict is a second generation multi-gene expression test which appears to perform better in predicting late metastasis (that is, more than 5 years after diagnosis) due to the inclusion of oestrogen receptor regulation genes which are absent from the other tests such as Oncotype DX. The same commentator explained that, in their personal experience, EndoPredict has been useful in guiding chemotherapy decision-making, in cases where treatment decisions were debatable on clinical and pathology evidence alone. The commentator concluded that they would recommend EndoPredict over Oncotype DX based on the fact that Oncotype DX does not use the clinical parameters and is less effective in predicting late metastasis.

One specialist commentator stated that using only clinical and pathology parameters does not accurately assess the biology of the tumour and that the importance of a test such as EndoPredict is that it improves clinicians' ability to identify people who should be offered chemotherapy. They noted that the data from the Martin et al. (2014) trial in the clinical evidence section is most relevant to how EndoPredict would be used and that the patients with low-risk scores (25%) may be able to avoid cytotoxic chemotherapy.

One of the specialist commentators also stated that the exclusion of men from using EndoPredict is discriminatory and a result of the limited evidence of the use of the test in men. They added that many new research studies allow the inclusion of men with breast cancer.

One specialist commentator noted that the best way to test the clinical utility of EndoPredict in a prospective way is represented by the OPTIMA trial. The trial, which currently does not investigate EndoPredict, is seeking to test multi-gene arrays in ER-positive/HER2-negative breast cancer populations who would at present receive chemotherapy as routine in the UK. It is a randomised trial comparing standard chemotherapy followed by endocrine therapy with test-directed therapy (low risk score no chemotherapy; high risk score chemotherapy).
Equality considerations

NICE is committed to promoting equality, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others. In producing guidance and advice, NICE aims to comply fully with all legal obligations to:

- promote race and disability equality and equality of opportunity between men and women
- eliminate unlawful discrimination on grounds of race, disability, age, sex, gender reassignment, pregnancy and maternity (including women post-delivery), sexual orientation, and religion or belief (these are protected characteristics under the Equality Act 2010).

EndoPredict is intended only for women with breast cancer; it is not indicated for men. Women of white family origin and those aged over 50 years are at increased risk of breast cancer. Sex, race and age are protected characteristics defined in the Equality Act 2010. People with cancer are protected under the Equality Act from the point of diagnosis.

Evidence review

Clinical and technical evidence

Regulatory bodies

A search of the Medicines and Healthcare Products Regulatory Agency website revealed no manufacturer Field Safety Notices or Medical Device Alerts for this device. No reports of adverse events were identified from a search of the US Food and Drug Administration (FDA) database: Manufacturer and User Device Facility Experience (MAUDE).

Clinical evidence

Of 132 relevant papers identified, 117 were excluded because they were duplicates, case studies, reviews, or news articles. There were 15 appropriate studies that used EndoPredict. This briefing includes 3 analytical validity and 6 clinical validity studies (see evidence selection).

Three studies demonstrated the analytical validity of EndoPredict (Poremba et al. 2014, Denkert et al. 2012, Kronenwett et al. 2012). Poremba et al. (2014) examined the effects of pre-analytical variables such as time to fixation, fixation time, temperature and tissue section storage time on the EndoPredict results in 138 assays. The study showed that the pre-analytical variables tested did not influence the gene expression analysis with the EndoPredict assay. Denkert et al. (2012)
evaluated the performance of EndoPredict in several different molecular pathology laboratories and sought to determine the number of laboratories that have implemented the test successfully. Ten FFPE breast cancer tumour blocks were sent to 7 laboratories for RNA extraction (totalling 70 study samples). All of the laboratories successfully implemented EndoPredict. The test was reproducible and the EP scores correlated positively with the reference values. Kronenwett et al. (2012) examined the analytical performance characteristics, precision, repeatability and reproducibility of EndoPredict and found that the test showed reproducible performance characteristics with minor variation between laboratories.

Bertucci et al. (2014) analysed the predictive value of the EndoPredict gene signature in terms of the pathological complete response to chemotherapy. The study data were selected from 7 retrospective gene expression datasets and included a population of ER-positive/HER2-negative breast cancer patients (n=553) treated with anthracycline-based neo-adjuvant chemotherapy followed by surgery. The authors defined pathological complete response as either grade 1 (the total disappearance of tumour both on macroscopic and microscopic examination) or grade 2 (presence of in-situ carcinoma of the breast, no invasive tumour, and no tumour found in the lymph nodes) in accordance with Chevallier grading (Chevallier et al. 1995, Fumagalli et al. 2012). The authors also evaluated disease free survival. The EP score was associated with a pathological complete response rate of 7% in the low-risk group and 17% in the high-risk group (p<0.001). High-risk tumours (classified by EP score) were associated with poorer survival with 5-year disease free survival equal to 73% (95% confidence interval [CI] 63 to 85) versus 88% (95% CI 81 to 95) in the low-risk group (p=0.015). No association was found with patient age, pre-chemotherapy tumour size, lymph node status, histological type, or progesterone receptor status in univariate analysis. Multivariate analysis, including tumour grade and the EP score as a continuous variable, found that only the EP score remained associated with pathological complete response. The EP score, both as a categorical variable and as a continuous variable, correlated with the tumour response to chemotherapy and tumours with an EP high-risk score were associated with higher response rate to chemotherapy than tumours with a low-risk score.

Dubsky et al. (2013a) assessed the clinical relevance of EPclin in relation to established clinical guidelines. The prospective-retrospective study included 1702 breast cancer patients from the ABCSG-6 and ABCSG-8 randomised controlled trials with a median follow-up time of 63 months. Patients received either tamoxifen for 5 years (n=1029) or tamoxifen for 2 years followed by anastrozole for 3 years (n=673). All patients were retrospectively assigned to risk categories based on EPclin score, the German S3 guidelines, National Comprehensive Cancer Network guidelines, and St Gallen guidelines. The study assessed the 10 year recurrence risk based on EPclin classification in comparison with the 3 clinical guidelines. The hazard ratio for low versus intermediate or high risk patients was highest for EPclin (5.11, 95% CI 3.48 to 7.51) followed by St...
Gallen (2.78, 95% CI 1.50 to 5.14), German S3 (2.20, 95% CI 1.16 to 4.19) and National Comprehensive Cancer Network (2.16, 95% CI 0.80 to 5.85). The EPclin score was the best predictor of risk at 10 year follow-up for low-risk versus intermediate/high risk groups when compared with German S3, National Comprehensive Cancer Network, and St Gallen classification recommendations.

Dubsky et al. (2013b) also performed a prospective-retrospective study on the same population of ER-positive/HER2-negative post-menopausal women from the ABCSG-6 and ABCSG-8 trials described in the Dubsky et al. (2013a) paper. The authors retrospectively classified patients to EP risk categories and sought to identify late recurrence in the study population. Kaplan–Meier analysis of late recurrence demonstrated that the EP low-risk group had a significantly improved clinical outcome in both early (0–5 years; hazard ratio [HR] 2.80, 95% CI 1.81 to 4.34, p<0.001) and late time intervals (greater than 5 years; HR 3.28, 95% CI 1.47 to 7.24, p=0.002). Multivariate analysis showed that EndoPredict is an independent prognostic parameter after adjusting for age, grade, lymph node status, tumour size and Ki67 in both time intervals. Concordance index (c-index) was calculated for different clinical and pathological parameters and improved with the addition of the EP score. The EPclin score had the highest c-index (0.786). The EP score and EPclin were both associated with early and late disease recurrences.

Martin et al. (2014) validated the EP score in ER-positive/HER2-negative patients who were treated with adjuvant chemotherapy followed by hormone therapy. The study also evaluated whether EP scores could predict the efficacy of incorporating weekly paclitaxel into anthracycline-based regimens. Participants (n=555) from the randomised GEICAM 9906 trial were evaluated for distant metastasis-free survival. Patients received either 6 cycles of fluorouracil, epirubicin and cyclophosphamide (FEC) or 4 cycles of FEC followed by 8 weekly courses of paclitaxel (FEC-P), as well as endocrine therapy if they had hormone-receptor positive disease. EP scores classified 25% of patients as low risk. The estimated rates of metastasis-free survival at 10 years were 93% for the EP score-based low-risk group and 70% for the EP score-based high-risk group, with an absolute risk reduction of 23% (HR 4.8, 95% CI 2.5 to 9.6; p<0.0001). Adding weekly paclitaxel treatment did not significantly reduce the risk of relapse in the 555 patients analysed (HR 1.1, 95% CI 0.8 to 1.6; p=0.6067). The EPclin score c-index estimate was highest (0.70) in comparison to all analysed combination strategies. Nodal status and EP score remained significant in the multivariate model, suggesting that the EP score is an independent predictor of MFS.

Muller et al. (2013) investigated the performance of the EP test in clinical practice, and performed a retrospective evaluation of the impact of the test on treatment decisions in breast cancer. The EP test was performed in tumour samples from 167 patients with ER-positive/HER2-negative breast cancer. The test was successfully performed in 99% of all samples. Clinicians were sent a
questionnaire for retrospective evaluation of treatment decisions. In over one third of cases (37.7%) the results of the EP test led to a change in planned therapy. For a quarter of patients (25.4%) the originally planned chemotherapy was omitted based on the results. The study showed that EndoPredict changed clinicians' treatment decisions and in some cases avoided chemotherapy.

Zhao et al. (2014) compared the prognostic power of 9 breast cancer gene signatures (Intrinsic, PAM50, 70-gene, 76-gene, Genomic-Grade-Index, 21-gene-Recurrence-Score, EndoPredict, Wound-Response and Hypoxia) in relation to ER status and follow-up time. A cohort of 912 patients (947 breast tumours) was used to evaluate how well the signatures predict distant metastasis-free survival. Using continuous risk scores to predict distant metastasis-free survival, PAM50 had the highest c-index of 0.658 with 95% CI 0.64 to 0.68, followed by GGI (0.656), WR (0.651), RS (0.648), EP (0.648), 76-gene (0.642), 70-gene (0.612), Intrinsic (0.598) and Hypoxia (0.525). EndoPredict had significant prognostic power in ER-positive/HER2-negative untreated cancers, as well as the complete set (treated and untreated).

Recent and ongoing studies

**ISRCTN69220108** – This UK-based study (the Brighton Trial) will examine the impact of EndoPredict on clinical decision making in low to medium risk ER-positive/HER2-negative early breast cancer, by comparing chemotherapy decisions before and after information from EndoPredict is added. It also aims to explore patient attitudes surrounding risk and satisfaction when it is used. The study began recruiting in July 2015 and is expected to be completed in May 2016.

**ABCSG-34** – This is a prospective, open, randomised, phase-II study of a therapeutic cancer vaccine (L-BLP25, Stimuvax) in women with early stage primary breast cancer due to undergo neoadjuvant chemotherapy. One core from the biopsy will be used to test EndoPredict. The testing of EndoPredict is conducted by a laboratory of ABCSG.research, a tumour bank of ABCSG. Patient recruitment is complete. Follow-up and analysis is ongoing.

**Costs and resource consequences**

One cost-effectiveness analysis was identified for EndoPredict. The study by Blank et al. (2015) used a Markov model to determine the health economic impact and incremental cost-effectiveness of EPclin-based risk stratification alone or in combination with clinical guidelines (German S3, National Comprehensive Cancer Network, and St Gallen) to decide which patients would receive chemotherapy. Costs were calculated from a German third-party payer perspective. The study was modelled based on the ABCSG-6 and ABCSG-8 patient cohort. Per patient costs included medical
resource use and took into account medical interventions related to high- and low-risk patients for disease-free, metastatic, and end-of-life therapy.

The average life-long cost per patient treated according to clinical guidelines (in 2015 prices, converted from Euros reported in the study) ranged from £27,575 (German S3) to £29,364 (National Comprehensive Cancer Network). EPclin-based risk stratification in combination with guidelines reduced the average life-long cost per patient to between £24,904 (St Gallen/EPclin) and £25,089 (National Comprehensive Cancer Network/EPclin). In terms of effectiveness, German S3 yielded the most quality-adjusted life years (QALYs) among guideline-based strategies at 13.169 and EPclin alone was superior to all other strategies (13.173). The German S3 guideline was used as the reference for cost-effectiveness analysis and the combined strategies of EPclin and each guideline as well as EPclin alone were dominant (cost less and are more effective in comparison to the reference); in contrast, the St Gallen and National Comprehensive Cancer Network guidelines alone were dominated (more costly and less effective in comparison to the German S3 reference guideline). Sensitivity analyses showed a variation in rank order of the combined strategies, but their advantage remained. Blank et al. (2015) stated that varying the discount rate, the disutility due to chemotherapy, hazard rates, and the prevalence of some high-risk classifications had a strong impact on costs and effects, but cost effectiveness was essentially unaffected. The results showed that combining EPclin with clinical guidelines can increase QALYs and decrease lifetime costs per patient in comparison to guideline strategies alone.

If EndoPredict were adopted in the NHS, the test would incur additional acquisition costs but no additional investigations would be needed beyond those used in current prognostic algorithms. The additional cost could be offset if it guided a decision to not offer chemotherapy when it would otherwise be offered using current decision-making methods. Cancer Research UK statistics report that about 49,936 of women were diagnosed with breast cancer in the UK in 2011; around two-thirds of breast cancers are ER-positive and 75 to 85% of these breast cancers are HER2-negative (National Institute for Health Research 2015). Therefore, EndoPredict has a maximum potential usage in approximately 28,000 people per year.

According to the manufacturer, EndoPredict is currently available in 6 NHS sites and will be available at additional NHS sites in the future as part of the ISRCTN69220108 trial, also called the Brighton Trial. No other additional facilities or technologies are needed alongside the test.

**Strengths and limitations of the evidence**

Three of the clinical validity EndoPredict studies included in this briefing are based on the prospective-retrospective analysis of archived material from 3 multicentre randomised controlled
trials (GEICAM 9906, ABCSG-6 and ABCSG-8). These trials had long-term follow-up for 10 years and included women with primary breast cancer receiving endocrine therapy.

The gold standard for validating a prognostic score would be a prospective clinical trial, in which the test itself is the intervention under investigation. However, specimens collected in the context of randomised clinical trials and stored for future use are the best source of specimens for retrospective technical and prognostic studies (Simon et al. 2009, McShane 2012).

A limitation to the evidence is that Dubsky et al. (2013a) and Dubsky et al. (2013b) study the same populations: 1702 patients from ABCSG-6 and ABCSG-8 studies. The trials addressed endocrine treatment questions in the absence of adjuvant or neo-adjuvant chemotherapy. This creates selection bias toward clinically low-risk ER-positive post-menopausal breast cancer patients: only 4% of women had high-grade differentiation. Blank et al. (2015) modelled a hypothetical cohort based on the patient characteristics from the same population. The costs were examined from a German third-party payer perspective, thus limiting relevance to the NHS. The analysis lacked comparisons to other genetic tests, and the authors stated this was due to a lack of results of tests in the ABCSG-6 and ABCSG-8 patients.

The Martin et al. (2014) study had a relatively small sample size (555 patients), which may limit the generalisability of the results. In a previous GEICAM study, a sample size of 1250 was calculated as necessary to detect an expected absolute difference in favour of the paclitaxel arm of the study (Martin et al. 2008). The smaller sample size may have limited its ability to obtain significant results regarding the efficacy of taxanes.

Bertucci et al. (2014) and Muller et al. (2013) are retrospective studies and therefore are subject to bias. Muller et al. (2013) sent a questionnaire to clinicians for retrospective evaluation of treatment decisions. This introduces the risk of recall bias, as clinicians' recollections of how they planned to manage individuals' care prior to receiving the test results may have been skewed by the results themselves.

Follow-up time from the Bertucci et al. (2014) study was relatively short (40 months) for an ER-positive/HER2-negative population, where it is not uncommon for metastases to occur 5–10 years after diagnosis. In addition, follow-up information was only available for 299 patients (53% of the original cohort). Therefore, the findings from this study might not be robust enough to draw a clear conclusion about the EPclin score and associated longer term outcomes. While no other study reported the proportion of patients with long-term follow-up, Dubsky et al. (2013a, 2013b) report the 10-year risk reduction, but with a median follow-up time of only 63 months.
Seven of the publications received financial support or disclosed competing interests from the manufacturer, and this introduces the potential for bias in the reporting of outcomes. Two publications (Zhao et al. 2014, Bertucci et al. 2014) have no conflicts of interest.

Relevance to NICE guidance programmes

NICE has issued the following guidance:

- Early and locally advanced breast cancer: diagnosis and treatment (2009) NICE guideline CG80
- Advanced breast cancer (update): diagnosis and treatment (2014) NICE guideline CG81
- Early and locally advanced breast cancer (2015) NICE pathway
- Familial breast cancer (2015) NICE pathway
- Intraoperative tests (RD-100i OSNA system and Metasin test) for detecting sentinel lymph node metastases in breast cancer (2013) NICE diagnostics guidance 8
- Gene expression profiling and expanded immunohistochemistry tests for guiding adjuvant chemotherapy decisions in early breast cancer management: MammaPrint, Oncotype DX, IHC4 and Mammostrat (2013) NICE diagnostics guidance 10
- Breast cancer (2011) NICE quality standard 12

References


Deutsche Krebs gesell schaft e.V. [German Cancer Society] and German Society for Gynecology and Obstetrics (DGGG) (2008) Interdisciplinary S3 Guidelines for the Diagnosis, Treatment and follow-up Care of Breast Cancer


National Institute for Health Research (2015) Palbociclib for advanced or metastatic, oestrogen-receptor positive, HER2-negative, recurrent breast cancer – second line. Horizon Scanning Research & Intelligence Centre HSRIC ID: 9396


Search strategy and evidence selection

Search strategy

For the clinical evidence

Embase 1980 to 2015 Week 29, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present; searched 27 July 2015.

1. EndoPredict.mp.


3. Sividon diagnostics.mp.

4. EP score.mp.

5. EPclin score.mp.

6. 1 or 2 or 3 or 4 or 5

7. Multigene test.mp.

8. Gene expression assay.mp. or gene expression assay/

9. 7 or 8

10. Breast cancer.mp. or breast cancer/

11. Breast neoplasms.mp. or breast cancer neoplasms/
12. 10 or 11

13. 6 and 9

14. 6 and 12

15. 13 or 14

16. Limit 15 to English language

17. Limit 16 to yr= "2005-Current"

For the economic evidence

Embase 1980 to 27 July 2015, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present; searched 27 July 2015.

Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present; searched 27 July 2015.

1. EndoPredict.mp.


3. Sividon diagnostics.mp.

4. EP score.mp.

5. EPclin score.mp.


8. Breast cancer.mp.

10. 1 or 2 or 3 or 4 or 5

11. 6 or 7

12. 8 or 9

13. 10 and 11

14. 10 and 12

15. 13 or 14

16. Limit 15 to English language

17. Limit 16 to yr= "2005-Current"

18. (Cost* or economic*).mp.

19. 17 and 18

Cochrane Database of Systematic Reviews: Issue 7 of 12, July 2015; Cochrane Central Register of Controlled Trials: Issue 6 of 12, June 2015; Database of Abstracts of Reviews of Effect: Issue 2 of 4, April 2015

Health Technology Assessment Database: Issue 2 of 4, April 2015


**Evidence selection**

**For the clinical evidence**

- Total number of publications reviewed: 132
- Total number of publications considered relevant: 15
- Total number of publications selected for inclusion in this briefing: 9
Exclusion criteria: abstracts, literature reviews, reviews, case studies, comment letters, news articles, technical evaluations, overlapping populations examining similar outcomes, studies focusing on a single gene, or retrospective studies with fewer than 50 participants.

For the economic evidence

- Total abstracts: 75
- Duplicates: 8
- Abstracts reviewed: 67
- Full papers reviewed: 8
- Studies for review: 1

Appendix

Contents

Data tables

**Table 1**: Overview of the Poremba et al. (2014) study

**Table 2**: Overview of the Denkert et al. (2012) study

**Table 3**: Overview of the Kronenwett et al. (2012) study

**Table 4**: Overview of the Bertucci et al. (2014) study

**Table 5**: Overview of the Dubsky et al. (2013a) study

**Table 6**: Overview of the Dubsky et al. (2013b) study

**Table 7**: Overview of the Martin et al. (2014) study

**Table 8**: Overview of the Muller et al. (2013) study

**Table 9**: Overview of the Zhao et al. (2014) study
Table 10: Overview of the Blank et al. (2015) study

### Table 1 Overview of the Poremba et al. (2014) study

<table>
<thead>
<tr>
<th>Study component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives/hypotheses</td>
<td>To examine the effects of pre-analytical variations such as time to fixation, fixation time, and tissue section storage time on the results of EndoPredict.</td>
</tr>
<tr>
<td>Study design</td>
<td>The design is not made clear in the study publication. The authors studied pre-analytical variations to test their impact on the results of EndoPredict.</td>
</tr>
<tr>
<td>Setting</td>
<td>Tests were carried out within a laboratory in Germany.</td>
</tr>
<tr>
<td>Inclusion/exclusion criteria</td>
<td>No explicit inclusion/exclusion criteria were outlined in the publication.</td>
</tr>
<tr>
<td>Primary outcomes</td>
<td>Analysis of the analytical performance characteristics of the EP test through assessment of TTF, FT, TCC, and SST.</td>
</tr>
<tr>
<td>Statistical methods</td>
<td>A reference EP score was calculated based on test results of samples with assumed standard tissue handling procedures (10 min at 20°C prior to fixation and 20 hours in formalin). To account for the precision of EndoPredict results, a deviation of ±0.75 EP score units from the reference EP score (3-fold standard deviation of the EP score) was defined as acceptance limits for all other samples of the TTF and FT studies. Pearson correlation coefficients (r) were calculated to compare EP test results of paired tissue samples with different TCC or different storage conditions.</td>
</tr>
<tr>
<td>Patients included</td>
<td>39 patients; positive nodal status in 10 patients (negative in 28 patients, 1 unknown); all patients had ER-positive/HER2-negative breast cancer. EPclin scores were only calculated and compared for 38 tumour samples, since the nodal status of one patient was not available.</td>
</tr>
</tbody>
</table>
Results

TTF: EP scores of all measurements were compared to the mean EP score obtained from tissue sections that had been stored for 10 min at 20°C prior to fixation (reference conditions). EP scores for all but 1 sample were within the reference EP score interval generated by reference conditions. The mean deviation of these samples was 0.37 EP score units (range −0.50 to 0.70). The risk classification by EndoPreict was identical for all specimens with a time to fixation of up to 12 hours, irrespective of storage temperature. Only the specimen with a storage time of 24 hours at 20°C showed a large decrease in EP score.

FT: EP scores with FT of 1 hour to up to 5 days were within the reference interval with a mean deviation of −0.02 EP score units from the reference EP score (range −0.40 to 0.40); FT did not affect EP test results.

TCC: Relative expression levels of 8 genes were compared between paired whole tissue sections and tumour-enriched specimens. Pearson correlation coefficients ranging from 0.90 to 0.98 demonstrate significant correlation of expression levels of the EP genes in tissue sections of different TCC. EPclin scores were also highly correlated (r=0.98) with a mean deviation of −0.10 EPclin score units. Risk classification based on EP score showed an overall agreement of 87%.

SST: Tissue sections of 10 breast tumours were stored for 12 months at +4°C or for 12 months at +20°C. The Pearson correlation coefficient was >0.99 when comparing EP scores of samples processed without prior storage and samples stored at +4°C. Mean deviation between stored and non-stored sections was 0.02 (range −0.50 to 0.30) EP score units for storage at +4°C and 0.04 (range −0.80 to 0.50) EP score units for storage at +20°C. The risk classification by EndoPredict was identical for stored and non-stored sections for either condition.

Conclusions

The study showed that the variables tested did not influence the gene expression analysis with the EndoPredict assay.

Abbreviations: EP, EndoPredict; FT, fixation time; SST, tissue section storage time TCC, tumour cell content; TTF, time to fixation.

<table>
<thead>
<tr>
<th>Study component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2 Overview of the Denkert et al. (2012) study</td>
<td></td>
</tr>
</tbody>
</table>

© NICE 2018. All rights reserved. Subject to Notice of rights (https://www.nice.org.uk/terms-and-conditions#notice-of-rights).
<table>
<thead>
<tr>
<th>Objectives/ hypotheses</th>
<th>To evaluate the performance of the test in different molecular pathology laboratories and to determine the number of laboratories that have implemented the EndoPredict test successfully.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study design</td>
<td>Prospective, multicentre study.</td>
</tr>
<tr>
<td>Setting</td>
<td>7 laboratories in Germany, Austria and Switzerland.</td>
</tr>
<tr>
<td>Inclusion/exclusion criteria</td>
<td>Identically sized sections (10 micrometres) from 10 FFPE breast cancer tumour blocks were sent to 7 laboratories for RNA extraction (70 study samples); selection criteria for these institutions were (1) experience in high-volume diagnostic molecular pathology and breast pathology, (2) located in different countries (Germany, Austria, or Switzerland); (3) different organisational structures (university hospital or large specialised private institution).</td>
</tr>
<tr>
<td>Primary outcomes</td>
<td>Successful implementation of the EndoPredict test, achieved if the absolute difference between EP scores determined in the laboratory and the corresponding reference EP scores is below 1.0 EP score units for at least 9 out of 10 samples. For each block, a reference EP score was generated in the manufacturer's (Sividon Diagnostics') laboratory by calculating the mean of the measurements of 4 further sections of each tumour.</td>
</tr>
<tr>
<td>Statistical methods</td>
<td>Pearson correlation coefficient between the reference EP scores and the EP scores reported by participating laboratories.</td>
</tr>
<tr>
<td>Specimens included</td>
<td>n=10, tumour content ranged from 30–80%, with a mean of 63%. Tumours were grade 1 (20%), grade 2 (60%), and grade 3 (20%).</td>
</tr>
<tr>
<td>Results</td>
<td>All of the laboratories successfully implemented the EndoPredict test. The range of EP scores of the 10 tumours was from 2.7 to 12.1. The EP scores measured by the individual participants showed correlation with the reference values, respectively, as reflected by Pearson correlation coefficients ranging from 0.987 to 0.999. The Pearson correlation coefficient between the reference EP scores and all EP scores reported by the participating laboratories (across all blocks and laboratories) was 0.994.</td>
</tr>
<tr>
<td>Conclusions</td>
<td>The data show that EndoPredict is reproducible and the EP scores showed positive correlation with the reference values.</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; EP, EndoPredict; FFPE, formalin-fixed paraffin embedded.
<table>
<thead>
<tr>
<th>Study component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives/hypotheses</td>
<td>To perform a comprehensive analytical validation of the performance characteristics of the EndoPredict test and its verification in a molecular-pathological routine laboratory.</td>
</tr>
<tr>
<td>Study design</td>
<td>The design is not made clear in the study publication. The authors state that EndoPredict was analytically validated according to the adapted guideline MM17-A of the CLSI.</td>
</tr>
<tr>
<td>Setting</td>
<td>Tests were carried out within a laboratory.</td>
</tr>
<tr>
<td>Inclusion/exclusion</td>
<td>No explicit inclusion/exclusion criteria were outlined in the publication. No single tumour sample could be identified with a high enough expression of all genes to be tested. Therefore, a large pool of control RNA was generated to be repeatedly analysed with consistent results. For that purpose, anonymised residual FFPE tumour specimens from 20 breast cancer tumours were screened by the EndoPredict test for gene expression levels of the respective genes. As a result, a pooled reference RNA preparation (1:1:1 mixture) was generated from total RNA isolated from 1 EP low risk and 2 EP high risk tumours.</td>
</tr>
</tbody>
</table>
| Primary outcomes        | Analysis of the analytical performance characteristics of the EP test – assessment of LOB, LOD, input range and PCR efficiency.  
The precision of the EP test.  
The repeatability and reproducibility of the EP test.  
The inter-laboratory variation of performance. |
| Statistical methods     | For EP scores 95% confidence intervals (CI) were calculated. For comparison of EP test results between 2 different laboratories Pearson correlation coefficient ($R^2$) was calculated and agreement of measurements was analysed by Bland-Altman methods. |
| Patients included       | The sample size and profile were not outlined in the publication.                                                                                                                                              |
Results

The RNA input range was between 0.16 and 18.5 ng/μl and did not change the EP score by more than 1 unit in the majority of the samples. Analysis of precision (variation of day, day time, instrument, operator, reagent lots) resulted in a standard deviation of reproducibility of 0.16 EP score units on a scale from 0 to 15.

Conclusions

The study provides the analytical performance characteristics of the EP test that can be used as a reference for analytical verification of the test in molecular pathological laboratories. The EP test showed reproducible performance characteristics with minor variation between laboratories.

Abbreviations: CLSI, Clinical and Laboratory Standards Institute; EP, EndoPredict; LOB, limit of blank; LOD, limit of detection; PCR, polymerase chain reaction.

<table>
<thead>
<tr>
<th>Study component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives/hypotheses</td>
<td>To evaluate whether a high risk of breast cancer relapse according to EP was associated with response to chemotherapy.</td>
</tr>
<tr>
<td>Study design</td>
<td>Retrospective multicentre study</td>
</tr>
<tr>
<td>Setting</td>
<td>Seven gene expression data sets of breast cancer samples profiled using DNA microarrays. Median follow-up 40 months.</td>
</tr>
<tr>
<td>Inclusion/exclusion criteria</td>
<td>Selection criteria of data sets were the presence of at least 10 profiled samples representing pre-treatment invasive breast carcinoma, treated with anthracycline-based neoadjuvant chemotherapy followed by surgery including lumpectomy or mastectomy and axillary lymph node dissection, and documentation of pathological response. Selection criteria for this study included ER-positive/HER2-negative patients.</td>
</tr>
<tr>
<td>Primary outcomes</td>
<td>Degree of pathological response to neoadjuvant chemotherapy defined as a categorical variable (pCR versus non-pCR). DFS was calculated from the date of diagnosis to the date of first loco-regional or metastatic relapse, or death.</td>
</tr>
</tbody>
</table>
Correlations between tumour groups and histoclinical features were calculated with Fisher's exact test or t-test when appropriate. Survival curves were obtained using the Kaplan–Meier method and compared with the log-rank test. Univariate and multivariate analyses for pCR prediction were done using generalised linear models.

Patients included
n=553, 46% ≤ 50 years old.
93% had clinical tumour size T2–4 and 88% (n=489) had no pCR.

Results
High-risk tumours were associated with poorer survival with 5-year DFS equal to 73% (95% CI 63 to 85) versus 88% (95%CI 81 to 95) in the low-risk group (p = 0.015, log-rank test).

EP classification was associated with a pCR rate lower in the low-risk group (7%; 19 out of 283) than in the high-risk group (17%; 45 out of 270; p<0.001, Fisher's exact test). Seventy percent (45 out of 64) of patients with pCR were assigned to the EP high-risk group, versus 48% of patients without pCR (225 out of 468). The correlation persisted when EP score was analysed as a continuous variable (p<0.001, t-test).

No association was found between pCR and patients' age, pre-chemotherapy tumour size, lymph node status, histological type, and PR status. Multivariate analysis, including the EP score as a continuous variable and tumour grade, showed the EP score remained associated with pCR, whereas grade did not.

Conclusions
EP score, as a categorical variable and as a continuous variable, correlated with the pathological response to chemotherapy. Tumours with EP high-risk score are associated with higher response rate to chemotherapy than tumours with a low-risk score.

Abbreviations: CI, confidence interval; DFS, disease free survival; EP, EndoPredict; pCR, pathological complete response; PR, progesterone receptor.

Table 5 Overview of the Dubsky et al. (2013a) study

<table>
<thead>
<tr>
<th>Study component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives/hypotheses</td>
<td>To assess the relevance of the EPclin test by exploring whether it can be used to stratify patients more accurately than common clinical guidelines.</td>
</tr>
<tr>
<td>Study design</td>
<td>Prospective-retrospective multicentre study.</td>
</tr>
<tr>
<td>------------------------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>Setting</td>
<td>Participants enrolled between 1996 and 2004 in the ABCSG-6 (tamoxifen-only arm) and enrolled between 1990 and 1995 in the ABCSG-8 trial.</td>
</tr>
<tr>
<td>Inclusion/exclusion criteria</td>
<td>This study included postmenopausal women with ER-positive, HER2-negative breast cancer, recruited as part of the ABCSG-6 and ABCSG-8 randomised controlled trials. Patients received either tamoxifen for 5 years (n=1029) or tamoxifen for 2 years followed by anastrozole for 3 years (n=673). The median follow-up time was 63 months for the combined cohort.</td>
</tr>
<tr>
<td>Primary outcomes</td>
<td>EPclin score classification of risk of 10-year recurrence. Risk categories were retrospectively assigned according to the prespecified EPclin cut-off value (EPclin score &lt;3.3 were classified as low risk for distant recurrence, EPclin score ≥3.3 were stratified as high risk) and compared to the risk classification in the German S3, NCCN, and St Gallen guidelines.</td>
</tr>
<tr>
<td>Statistical methods</td>
<td>No statistical methods were listed in the study.</td>
</tr>
<tr>
<td>Patients included</td>
<td>1702 patients, median age 63.8 years (range 41.5–80.7). 68% negative nodal status, 27% 1–3 positive nodes, 5% &gt;4 positive nodes.</td>
</tr>
</tbody>
</table>
### Results

In Kaplan–Meier analysis, all stratifiers delineated a group of patients with extremely low risk: S3 low risk showed an absolute freedom of distant recurrence of 94.7% (90.5%–98.9%) after 10 years of follow-up. NCCN, St Gallen and EPclin results were 94.5% (88.9%–100%), 96.9% (94.9%–98.9%) and 95.3% (93.4%–97.3%), respectively.

Log-rank tests, HRs and ARRs were calculated at 10 years of follow-up for low-risk groups versus intermediate-/high-risk patients.

**German S3:**
- Long-rank test: \( p=0.014 \)
- HR: 2.20 (1.16–4.19)
- ARR: 7.9% (3.0%–12.9%)

**NCCN:**
- Long-rank test: \( p=0.119 \)
- HR: 2.16 (0.80–5.85)
- ARR: 6.9% (0.9%–13%)

**St Gallen:**
- Long-rank test: \( p<0.001 \)
- HR: 2.78 (1.50–5.14)
- ARR: 11.2% (7.7%–14.7%)

**EPclin:**
- Long-rank test: \( p<0.001 \)
- HR: 5.11 (3.48–7.51)
- ARR: 18.7% (13.5%–23.9%)

The majority (82%–94%) of all clinically assigned low-risk patients were also classified as low risk by the EPclin. The majority of intermediate/high risk patients classified by the German S3 guidelines were classified as low risk by the EPclin (58%); similarly, 61% of patients stratified as intermediate/high risk by NCCN 2007 and 58% of those stratified as intermediate/high risk by St Gallen 2011 were reclassified to low risk by EPclin.
All patients stratified as intermediate or high risk by clinical guidelines were stratified by EPclin in a Kaplan–Meier analysis. At 10 years, the distant recurrence rates for patients EPclin-low risk and EPclin-high risk were:

- 5% (2%–7%) and 24% (19%–29%) in the German S3 intermediate-/high-risk group
- 5% (3%–7%) and 23% (18%–28%) in the NCCN high-risk group
- 5% (3%–8%) and 25% (20%–30%) in the St Gallen intermediate-/high-risk group.

Conclusions

EPclin score had the highest ARR at 10 year follow-up for low-risk versus intermediate/high risk groups when compared with German S3, NCCN, and St Gallen. It also had the highest HR. EPclin is able to increase classification accuracy in comparison with the three guidelines studied.

Abbreviations: ABCSG, Austrian Breast and Colorectal Cancer Study Group; ARR, absolute risk reductions; ER, oestrogen receptor; HR, hazard ratio; NCCN, National Comprehensive Cancer Network.

Table 6 Overview of the Dubsky et al. (2013b) study

<table>
<thead>
<tr>
<th>Study component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives/hypotheses</td>
<td>To assess whether the EP score identifies late relapse events in ER-positive, HER2-negative breast cancer patients.</td>
</tr>
<tr>
<td>Study design</td>
<td>Prospective-retrospective multicentre study.</td>
</tr>
<tr>
<td>Setting</td>
<td>Participants enrolled between 1996 and 2004 in the ABCSG-6 trial (tamoxifen-only arm) and enrolled between 1990 and 1995 in the ABCSG-8 trial.</td>
</tr>
<tr>
<td>Inclusion/exclusion criteria</td>
<td>This study included postmenopausal women with ER-positive, HER2-negative breast cancer tumours, recruited as part of the ABCSG-6 and ABCSG-8 randomised controlled trials. Patients received either tamoxifen for 5 years (n=1029) or tamoxifen for 2 years followed by anastrozole for 3 years (n=673).</td>
</tr>
<tr>
<td>Primary outcomes</td>
<td>Early (0–5 years) and late (&gt;5 years) recurrence.</td>
</tr>
</tbody>
</table>
Statistical methods
Metastasis rates were estimated using the Kaplan–Meier method. All reported p-values were results of two-sided tests. P-values <5% were considered statistically significant. The c-index was used to assess the prognostic performance of the EP signature and clinic-pathological parameters.

Patients included
1702 patients, median age 63.8 years (range 41.5–80.7).
68% negative nodal status, 27% 1–3 positive nodes, 5% >4 positive nodes.

Results
49% (n=832) of all patients were classified as low risk according to the EP score. Kaplan–Meier analysis of distant recurrence demonstrated that the EP low-risk group had a significantly improved clinical outcome in the early (0–5 years; HR: 2.80, 95% CI 1.81 to 4.34, p<0.001) and late time interval (>5 years; HR: 3.28, 95% CI 1.47 to 7.24, p=0.002).

Multivariate analysis showed that EP is an independent prognostic parameter after adjustment for age, grade, lymph node status, tumour size and Ki67 in the first and second time interval.

The combination of clinical parameters resulted in a c-index of 0.644 in the combined cohort. The addition of the EP score to the combination of clinicopathological parameters resulted in a c-index of 0.716 (p<0.001). C-index improved prediction of late metastases by adding the prognostic information of the EP score. Adjuvant! Online was also improved by adding the EP test information (p<0.001). EPclin showed the best performance in predicting late relapse events with a c-index of 0.786.

Conclusions
EP and EPclin were significantly associated with early and late disease recurrences.

Abbreviations: ABCSG, Austrian Breast and Colorectal Cancer Study Group; ARR, absolute risk reductions; c-index, concordance index; CI, confidence interval; EP, EndoPredict; ER, oestrogen receptor; HR, hazard ratio; Ki67, proliferation index.

Table 7 Overview of the Martin et al. (2014) study

<table>
<thead>
<tr>
<th>Study component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives/hypotheses</td>
<td>To validate the EP score and whether EP results could predict the efficacy of incorporating weekly paclitaxel into anthracycline-based regimens.</td>
</tr>
</tbody>
</table>
### Study design

Prospective-retrospective multicentre study.

### Setting

Participants enrolled from 1999 to 2002 in the GEICAM 9906 trial.

### Inclusion/exclusion criteria

The patients included in this study participated in the **GEICAM 9906 trial**, a randomised phase III trial which compared two adjuvant chemotherapy regimens after breast cancer surgery in node-positive (stages II–III) patients with ER-positive/HER2-negative tumour samples.

Exclusion: Previous chemotherapy, hormone therapy, and/or radiotherapy for breast cancer; bilateral breast cancer; previous or current malignancies (except for basal skin carcinoma, cervical in-situ carcinoma, or superficial bladder carcinoma).

### Primary outcomes

Distant MFS, defined as the interval between the date of randomisation and the date of distant metastatic recurrence or death due to disease progression as the first event.

### Statistical methods

Deaths due to any causes other than disease progression were censored. Metastasis rates and overall survival were estimated using the Kaplan–Meier method. A log-rank test was used to compare MFS and overall survival between EP risk groups and between treatment arms. Cox proportional hazards models were used to calculate HRs and their 95% CIs for all analysed endpoints.

Unbiased c-index estimates were calculated for the different combinations of the clinic-pathological parameters (age, grade, nodal status, tumour size, treatment arm and ER and PR status and Ki67 index) as well as EP score combined with clinical variables and EPclin score to evaluate their contributions to prognostic classification.

### Patients included

555 patients, 45% age <50 years old.
45% of the women had T1-stage cancer and 64% had nodal status N1.
Results

The GEICAM 9066 study compared 6 courses of FEC (600/90/600) versus 4 FEC courses followed by 8 weekly paclitaxel administrations (100 mg/m²).

The estimated rates of MFS at 10 years were 93% for the EP score-based low-risk group (9 events in 141 patients) and 70% for the EP score-based high-risk group (110 events in 414 patients), with an absolute risk reduction of 23% (HR=4.8, 95% CI=2.5 to 9.6; p<0.0001).

EP score-based risk categorisation was associated with overall survival (HR=3.9, 95% CI=2.0 to 7.5; P0.0001).

EPclin score dichotomised patients into low- and high-risk categories with an absolute risk reduction of 28% (P<0.0001). The low-risk group of EPclin test results (74/555) had no metastatic events and an overall survival rate of 99%.

Adding weekly paclitaxel treatment did not significantly reduce the risk of relapse in the 555 patients (HR=1.1, 95% CI=0.8 to 1.6; p=0.6067). MFS differences between treatment arms also failed to reach statistical significance in both the EP high- and low-risk groups. The interaction between the EP score and treatment arm was also non-significant (p=0.71).

The combination of clinic-pathological parameters resulted in a c-index estimate of 0.65. The addition of the EP score to the combination of the clinic-pathological markers resulted in a significant improvement in the predictive accuracy (P<0.0018) and a c-index estimate of 0.67. The EPclin score c-index estimate was 0.70.

Conclusions

The EPclin score c-index was highest in comparison to all analysed combination strategies. Nodal status and EP score remained significant in the multivariate model, suggesting that EP score is an independent predictor of MFS.

Abbreviations: CI, confidence interval; c-index, concordance index; EP, EndoPredict; ER, oestrogen receptor; FEC, Fluorouracil, Epirubicin and Cyclophosphamide; GEICAM, Grupo Español para la Investigación del Cáncer de Mama [Spanish Group for the Investigation of Breast Cancer]; HR, hazard ratio; Ki67, proliferation index; MFS, metastasis-free survival; PR, progesterone receptor.

Table 8 Overview of the Muller et al. (2013) study

<table>
<thead>
<tr>
<th>Study component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Objectives/hypotheses</strong></td>
<td>To investigate the performance of the EP test in clinical practice and perform a retrospective evaluation of the impact of this new test on treatment decisions in breast cancer.</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Study design</strong></td>
<td>Retrospective, single centre study.</td>
</tr>
<tr>
<td><strong>Setting</strong></td>
<td>EP tests from 167 patients over the course of 1 year (August 2011–July 2012).</td>
</tr>
<tr>
<td><strong>Inclusion/exclusion criteria</strong></td>
<td>Inclusion/exclusion criteria were not detailed in the study publication.</td>
</tr>
<tr>
<td><strong>Primary outcomes</strong></td>
<td>Rate of successfully performed assays, time taken to perform tests, impact of EP on changes in therapy decisions.</td>
</tr>
<tr>
<td><strong>Statistical methods</strong></td>
<td>The correlation between EP score and tumour grade was analysed using the Jonckheere–Terpstra test for trends. The correlation between EP and proliferation activity was analysed using the Wilcoxon–Mann–Whitney test.</td>
</tr>
<tr>
<td><strong>Patients included</strong></td>
<td>167 patients. The median age at time of diagnosis was 54 years (range: 30–78 years), the median age in the subgroup with therapy data was 55 years (range: 30–76 years).</td>
</tr>
<tr>
<td><strong>Results</strong></td>
<td>The test could be successfully performed in 99% of all samples. The EPclin-based estimated median 10-year-risk for metastases with endocrine therapy alone was 11% for the whole cohort. The estimated median risk for the EPclin low group was 7%; for the EPclin high risk group it was 19%. More than 50% of tests were performed in 3 or fewer days. Clinicians were sent a questionnaire for retrospective evaluation of treatment decisions. In over one third (37.7%) the results of the EP assay led to a change of planned therapy. For a quarter of patients (25.4%) the originally planned chemotherapy could be omitted based on the result of the multi-gene assay.</td>
</tr>
<tr>
<td><strong>Conclusions</strong></td>
<td>The authors concluded that results showed that the EP assay could be routinely performed outside centralised molecular pathology laboratories and the results led to a change in treatment decisions.</td>
</tr>
</tbody>
</table>

Abbreviations: EP, EndoPredict assay.
Table 9 Overview of the Zhao et al. (2014) study

<table>
<thead>
<tr>
<th>Study component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives/hypotheses</td>
<td>To assess and compare prognostic power of 9 breast cancer gene signatures (Intrinsic, PAM50, 70-gene, 76-gene, Genomic-Grade-Index, 21-gene-Recurrence-Score, EndoPredict, Wound-Response and Hypoxia) in relation to ER status and follow-up time.</td>
</tr>
<tr>
<td>Study design</td>
<td>Prospective-retrospective multicentre study.</td>
</tr>
<tr>
<td>Setting</td>
<td>The gene expression dataset (n=947) is a collection of 6 published breast cancer microarray datasets. The datasets were retrieved from Gene Expression Omnibus and ArrayExpress. Additional clinical information on these 947 samples was collected, including additional information on ER status, node status, tumour size, and DMFS follow-ups. Follow-up time was divided into 3 intervals: first 5 years, 5–10 years, and beyond 10 years.</td>
</tr>
<tr>
<td>Inclusion/exclusion criteria</td>
<td>Inclusion criteria were not explicitly specified in the study publication. Patients were excluded if they had experienced an event before the start of the interval. It is unclear what the authors defined as an event.</td>
</tr>
<tr>
<td>Primary outcomes</td>
<td>Prognostic power across follow-up time and agreement between signatures. DMFS was used as a clinical endpoint.</td>
</tr>
<tr>
<td>Statistical methods</td>
<td>Continuous risk scores were used to predict DMFS. The c-index was chosen to compare the predictive strength of the signatures. Pearson correlation coefficients were used to assess similarity for risk assessment among gene signatures. Cumulative regression plots were used to investigate the nature of time-dependency in ER-positive tumours. The assumption of time-independent proportional hazard was examined for the ER-positive group and the ER-negative group separately using a univariate Cox model with signature risk scores as covariate.</td>
</tr>
</tbody>
</table>
Patients included: An overall cohort of 912 patients (947 breast tumours) was used to evaluate the signatures for prediction of DMFS. 912 participants had available DMFS status. No data on median age or gender was given.

Results: Using the continuous risk scores to predict DMFS, PAM50 had the highest c-index of 0.658 with 95% CI 0.64 to 0.68, followed by GGI (0.656), WR (0.651), RS (0.648), EP (0.648), 76-gene (0.642), 70-gene (0.612), Intrinsic (0.598) and Hypoxia (0.525).

The correlations between individual signatures were above 0.4, except those involving Hypoxia and between 76-gene and Intrinsic ($\rho=0.23$), indicating reasonably good concordance across the signatures.

Within the first 5 years, all signatures except for Hypoxia had significant positive effects ($p<0.0001$) in the ER-positive group; while in the ER-negative group, Hypoxia ($p<0.0001$), WR ($p=0.021$) and 76-gene ($p=0.023$) were the only classifiers with significant positive effects on DMFS prediction. After 5 years, the signatures had little prognostic power.

Conclusions: Most of the signatures were strong risk predictors for DMFS during the first 5 years of follow-up.

Abbreviations: c-index, concordance index; DMFS, Distant Metastasis Free Survival; ER, oestrogen receptor; GGI, genomic grade index; RS, 21-gene-recurrence-score; WR, Wound-Response.

### Table 10 Overview of the Blank et al. (2015) study

<table>
<thead>
<tr>
<th>Study component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives/ hypotheses</td>
<td>To determine the health economic impact and incremental cost effectiveness of EPclin-based risk stratification in combination with clinical guidelines (German S3, St Gallen, and the 2007 NCCN) to decide on chemotherapy use.</td>
</tr>
<tr>
<td>Study design</td>
<td>Life-long Markov state transition model.</td>
</tr>
<tr>
<td>Setting</td>
<td>A hypothetical cohort of early ER-positive/HER2-negative breast cancer patients with the same characteristics as those in ABCSG6 and ABCSG8 clinical trials. Low- and high-risk patient paths were modelled using transitions from the disease-free state to metastasis or death, and from metastasis to death. Parametric time-to-event modelling was used to define the transitions, and time-dependent hazards were estimated from the ABCSG6/8 10-year distant metastasis-free survival and OS data.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Inclusion/exclusion criteria</td>
<td>Patients were randomised to either tamoxifen for 5 years, or tamoxifen for 2 years followed by anastrozole for 3 years, without chemotherapy.</td>
</tr>
<tr>
<td>Primary outcomes</td>
<td>Cost-effectiveness of EPclin-based risk stratification in combination with clinical guidelines (German S3, St Gallen, and the 2007 NCCN) to decide on chemotherapy use in terms of QALYs and LYG.</td>
</tr>
<tr>
<td>Statistical methods</td>
<td>Summary statistics and decision tree economic model; probabilistic sensitivity analysis (second order Monte Carlo simulation).</td>
</tr>
<tr>
<td>Patients included</td>
<td>1619 patients were used for health economic analysis from the ABCSG6 and ABCSG8 trials.</td>
</tr>
<tr>
<td>Results</td>
<td>The average life-long cost per patient treated according to clinical guidelines (in 2015 prices, converted from 2010 Euros) ranged from £27,575 (German S3) to £29,364 (NCCN). EPclin-based risk stratification in combination with guidelines reduced the average life-long cost per patient to between £24,904 (St Gallen/EPclin) and £25,089 (NCCN/EPclin). QALYs ranged from 13.169 (German S3, reference) to 13.173 (EPclin alone). LYG ranged from 16.968 (St Gallen/EPclin) to 17.018 (NCCN). Using German S3 guideline as the reference strategy, St Gallen and NCCN strategies were dominated (more costly and less clinically effective) and the German S3/EPclin, St Gallen/EPclin, NCCN/EPclin, and EPclin alone were dominant strategies (less costly and more effective).</td>
</tr>
<tr>
<td>Conclusions</td>
<td>The results showed that combining EPclin with clinical guidelines can increase QALYs and decrease lifetime costs per patient in comparison to guideline strategies alone.</td>
</tr>
</tbody>
</table>
About this briefing

Medtech innovation briefings summarise the published evidence and information available for individual medical technologies. The briefings provide information to aid local decision-making by clinicians, managers and procurement professionals.

Medtech innovation briefings aim to present information and critically review the strengths and weaknesses of the relevant evidence, but contain no recommendations and are not formal NICE guidance.

Development of this briefing

This briefing was developed for NICE by King's Technology Evaluation Centre (KiTEC). The interim process & methods statement sets out the process NICE uses to select topics, and how the briefings are developed, quality-assured and approved for publication.

Project team

- King's Technology Evaluation Centre (KiTEC)
- Medical Technologies Evaluation Programme, NICE

Peer reviewers and contributors

- Anastasia Chalkidou, Senior Health Technology Assessor, KiTEC
- Lamprini Kaftantzi, Health Technology Assessor
- Naomi Herz, Health Technology Assessor and Health Economist, KiTEC
- Stephen Keevil, Director, KiTEC
- Cornelius Lewis, Director, KiTEC
- Viktoria McMillan, Centre Manager, KiTEC
Specialist commentators

The following specialist commentators provided comments on a draft of this briefing:

- Helena Earl, Consultant Medical Oncologist, Cambridge University Hospitals NHS Foundation Trust
- Andrew Wardley, Consultant Medical Oncologist, The Christie NHS Foundation Trust
- Clive Wells, Consultant Histopathologist, University College London Hospitals NHS Foundation Trust

Declarations of interest

- Helena Earl and Clive Wells declared no conflicts of interest.
- Andrew Wardley has directorships with Manchester Breast Cancer and Andrew Wardley Limited. He performs fee paid work for companies producing cancer drugs and tests including Genomic Health but not for Sividon Diagnostics or Myriad Diagnostics.

Copyright

© National Institute for Health and Care Excellence, 2015. All rights reserved. NICE copyright material can be downloaded for private research and study, and may be reproduced for educational and not-for-profit purposes. No reproduction by or for commercial organisations, or for commercial purposes, is allowed without the written permission of NICE.

ISBN: 978-1-4731-1519-4