Gene expression profiling and expanded immunohistochemistry tests to guide the use of adjuvant chemotherapy in breast cancer management: a systematic review and cost-effectiveness analysis

S Ward,[1]* A Scope,[1] R Rafia,[1] A Pandor,[1] S Harnan,[1] P Evans[1] and L Wyld[2]

[1]School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK
[2]Royal Hallamshire Hospital, Sheffield, UK

*Corresponding author

Scientific summary

The use of adjuvant chemotherapy in breast cancer management

Health Technology Assessment 2013; Vol. 17: No. 44
DOI: 10.3310/hta17440

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Prognostic tools such as the Nottingham Prognostic Index (NPI) and Adjuvant! Online are currently used in the UK to assist decision-making relating to adjuvant chemotherapy for women with early breast cancer at intermediate or high risk of recurrence following primary surgery. These tools use pathological parameters, for example tumour size, grade and lymph node status in the case of NPI, with the addition of oestrogen receptor (ER) status, age and comorbidity for Adjuvant! Online. Such tools are imperfect and some women with early breast cancer may be over- or undertreated, resulting in unnecessary use of chemotherapy for some women or avoidable deaths in women for whom chemotherapy was withheld.

Gene expression profiling (GEP) and expanded immunohistochemistry (IHC) (or protein expression) tests aim to improve the targeting of chemotherapy by more accurately identifying patients who will gain most benefit from it. These tests either aim to more accurately measure the risk of cancer recurrence by incorporating a wider range of biomarkers than standard clinicopathological algorithms or seek to identify breast cancer subtypes, which provide information on recurrence risk.

Nine tests were included in this assessment, as per the National Institute for Health and Care Excellence (NICE) scope. Six use GEP technology: the Randox Breast Cancer Array (Randox Laboratories, Crumlin, UK), MammaPrint® (Agendia, Amsterdam, the Netherlands), BluePrint™ (Agendia, Amsterdam, the Netherlands), the PAM50 gene expression assay (ARUP Laboratories, Salt Lake City, UT, USA), OncotypeDX™ (Genomic Health Inc., Redwood City, CA, USA) and the Breast Cancer Index℠ (bioTheranostics Inc., San Diego, CA, USA); and three use IHC technology: IHC4 [The National Institute for Health Research (NIHR) Specialist Biomedical Research Centre (BRC) for Cancer is a partnership between The Royal Marsden NHS Foundation Trust and The Institute of Cancer Research (ICR); see http://www.cancerbrc.org/Highlights/Breast_Cancer_highlights/index.shtml], Mammostrat® (Clarient Inc., Aliso Viejo, CA) and NPI plus (NPI+) (University of Nottingham, Nottingham, UK).

Objective

The objective of this study was to evaluate the clinical effectiveness and cost-effectiveness of GEP and expanded IHC tests compared with existing prognostic tools in guiding the use of adjuvant chemotherapy in women with early breast cancer in England and Wales.

Methods

A systematic review of the evidence on the clinical effectiveness of nine GEP and expanded IHC tests to guide the use of chemotherapy in breast cancer management was conducted. For two of the tests (OncotypeDX and MammaPrint) the review updated two existing systematic reviews. Several electronic databases (including MEDLINE, MEDLINE In-Process & Other Non-Indexed Citations, EMBASE and The Cochrane Library) were searched from January 2002 to May 2011 (for the OncotypeDX and MammaPrint tests searches were conducted from January 2009). Outcome measures included analytical validity, clinical validity and clinical utility. The study by Altman (2001) was used to assess the methodological quality of included studies (Altman D. Systematic reviews in health care: systematic reviews of evaluations of prognostic variables. *BMJ* 2001;**323**:224–8).

A systematic review of economic evaluations was also undertaken. In addition, two economic evaluations were submitted by Genomic Health and Clarient for the use of OncotypeDX and Mammostrat in the UK respectively.

A probabilistic model was developed by the External Assessment Group (EAG) using a lifetime horizon. Following a review of the evidence available, only four of the nine tests were included in the economic evaluation. Analysis was undertaken for women with ER-positive (ER+), lymph node-negative (LN–) and human epidermal growth factor receptor type 2-negative (HER2–) early breast cancer from a NHS perspective. These tests were assessed as an addition to existing prognostic tools. A subgroup analysis was conducted in women with a NPI score ≤3.4 and women with a NPI score >3.4. The model used UK-specific data where possible.

In the comparator arm of the economic model, the proportion of patients receiving chemotherapy under current practice was informed by cancer registry data, reflecting the use of current prognostic tools such as NPI and Adjuvant! Online to guide the use of chemotherapy. In the intervention arm the targeting of patients to receive chemotherapy was dependent on the classification of risk by the new test. The natural history of breast cancer was then simulated using a cohort state transition model, taking into account the reduction in the risk of recurrence associated with chemotherapy. Evidence for the benefit of chemotherapy (reduction in the risk of recurrence) by risk group for the new tests was taken directly from the studies identified through the systematic review of the literature, despite the identified limitations of the studies. Patients were able to move between five possible health states – recurrence free, distant recurrence, local recurrence, long-term adverse events and death (from breast cancer, adverse events or other causes). Results were reported in terms of cost per quality-adjusted life-year (QALY).

Results

Nature, description and quality of the available evidence

The literature searches identified 5993 citations, of which 32 full-text papers or abstracts (representing 30 studies) were included in the review. Supplementary information submitted by the manufacturers was also presented. This evidence was summarised but was not subjected to the systematic review process. Additional studies that did not meet the inclusion criteria for the systematic review were used to populate the economic model.

The study populations were generally heterogeneous in the nature of their inclusion criteria although the majority of evidence examined ER+, LN– populations. Most studies included a small number of participants, although a few studies included over 1000 patients. Follow-up was short or not reported for a large number of studies. Only six studies were specific to a UK population (three for OncotypeDX, one for NPI+, one for IHC4 and one for Mammostrat).

Summary of the benefits and risks of gene expression profiling and expanded immunohistochemistry tests

OncotypeDX

Clinical

Previous systematic reviews OncotypeDX was reported to be furthest along the validation pathway. In terms of clinical validity these reviews reported evidence that the OncotypeDX recurrence score was significantly correlated with disease-free-survival and overall survival. One study was reported that reported a significant benefit from the use of chemotherapy in the OncotypeDX high-risk group, although it was highlighted within the review that the study may have been subject to bias.

Current review The current review identified 12 additional studies on the OncotypeDX test. Further larger studies have now reported, which support the prognostic capability of the OncotypeDX test. In particular, one large-scale UK-based study, in postmenopausal women with ER+, LN– early breast cancer, reported that an increase in risk score was significantly associated with an increased risk of distant recurrence. Furthermore, the evidence base has been extended to include the LN+ population, and there are the beginnings of an evidence base for the validity of OncotypeDX in different populations such as Japanese patients. Four studies presented evidence on the impact of OncotypeDX on clinical decision-making, indicating that the use of OncotypeDX leads to changes in decision-making for between 31.5% and 38% of patients. However, only one of these studies was UK based and limitations in relation to study design were identified.

Economic

Two economic studies were identified. Both studies compared the use of OncotypeDX with Adjuvant! Online. These studies were non-UK studies and were not considered generalisable to the UK setting. The economic evaluation submitted by Genomic Health estimated the incremental cost for treatment guided using OncotypeDX to be £6232 per QALY gained compared with current clinical practice in the UK, although a number of limitations with regard to the analysis were highlighted.

A de novo economic model was built by the EAG and estimated the cost per QALY gained to be £29,502 compared with current clinical practice, assuming that the test was offered to all woman with ER+, LN–, HER2– early breast cancer, under our base-case assumptions. This analysis assumed OncotypeDX to be predictive of the benefit of chemotherapy, based on evidence from the Paik *et al.* study, although weaknesses relating to this study are highlighted. (Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, *et al*. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 2006;**24**:3726–34.) A subgroup analysis was performed and showed that the incremental cost-effectiveness ratio (ICER) for OncotypeDX compared with current clinical practice was reduced to £9774 per QALY gained if OncotypeDX was to be offered to women with a (NPI>3.4) only. Compared with current clinical practice, OncotypeDX had a 12.4% (all women) and 91.6% (NPI>3.4) probability of being considered cost-effective when using a threshold of £20,000 per QALY gained respectively, although the quality of the data in the model was considered relatively weak. Key areas of uncertainly relate to assumptions about the benefits of chemotherapy in terms of relative risk reduction by risk group, the risk of recurrence over time and the impact of the new test on decision-making. The ICER increased substantially and was greater than £20,000 per QALY gained for both analyses when assuming the same relative reduction in the risk of recurrence from chemotherapy for all patients irrespective of the OncotypeDX recurrence score classification, that is, assuming that the test is prognostic only.

MammaPrint

Clinical

Previous systematic reviews There is a range of studies evaluating the prognostic ability of MammaPrint in heterogeneous populations; however, the previous reviews indicated that evidence relating to the clinical validity of MammaPrint was not always conclusive or supportive of the prognostic value of the test. In terms of clinical utility, the previous reviews identified one non-UK study which suggested that MammaPrint had an impact on clinical decision-making.

Current review Our review identified seven additional studies on the MammaPrint test. Four studies reported that the MammaPrint score is a strong independent prognostic factor and may provide additional value to standard clinicopathological measures, although the populations in all of these studies were relatively small. Six non-UK studies evaluated the clinical utility of MammaPrint. Five of the studies reported on test reclassification against currently used guidelines and one reported that treatment advice for 40% of patients would change, assuming that all patients classified as high risk and no patients classified as low risk would receive chemotherapy. However, none of the studies provided evidence of actual changes

in treatment decisions following introduction of the test. A study on the benefit of chemotherapy by MammaPrint risk group was identified but omitted from the systematic review because it was based on a pooled analysis of six primary studies.

Economic

An analysis was carried out by the EAG to evaluate the use of MammaPrint in England and Wales but because of the limitations in the evidence available this was considered exploratory only and no base-case ICER was presented.

PAM50

Clinical

The evidence base for PAM50 is still relatively immature. The current review identified two analytical validity studies (reported in abstract form only) comparing the PAM50 test with standard IHC measurements. Four studies evaluated the clinical validity of PAM50; two of these are as yet unpublished. No evidence on clinical utility was identified.

Economic

The EAG did not model treatment guided using PAM50 because of gaps in the evidence base.

Mammostrat

Clinical

The current review identified three studies that provided data to support the use of the Mammostrat test as an independent prognostic tool for women with ER+, tamoxifen-treated breast cancer. Although the evidence base for the Mammostrat test is relatively immature, these studies included a large sample size, appeared to be of reasonable quality and provided data from a UK setting (one study). One study was identified for clinical utility but limitations were identified relating to this study.

Economic

The EAG conducted an exploratory analysis using the same model structure as for the OncotypeDX evaluation and unpublished data from a small sample from a non-UK population; however, because of the limitations in the evidence base, any conclusions drawn from this analysis are subject to significant uncertainty.

IHC4

Clinical

No studies on analytical validity of the test were identified. The current review identified one study on the clinical validity of IHC4, which reports that the IHC4 score is a highly significant predictor of distant recurrence. This study was based on a large sample size and detailed the development of the test in one cohort and the external validation of the test in an independent cohort. The study also reported evidence comparing IHC4 with OncotypeDX. The review did not identify any published evidence on the clinical utility of IHC4 in terms of its impact on treatment decisions or its ability to predict chemotherapy benefit by risk group.

Economic

The EAG evaluated the cost-effectiveness of IHC4 in parallel with that of OncotypeDX as there was direct evidence between the two tests in a UK population from the same data source used to evaluate the cost-effectiveness of OncotypeDX. The IHC4 test was predicted to be dominant compared with current clinical practice in patients with ER+, LN–, HER2– early breast cancer, providing more QALYs at a lower cost. An incremental analysis was conducted comparing OncotypeDX, IHC4 and current clinical practice. When the treatment decision using OncotypeDX was compared with that using IHC4, the ICER

for OncotypeDX increased to £64,111 per QALY gained if the tests were to be offered to all women and £31,125 if the tests were to be offered to women with a NPI >3.4 only. IHC4 was predicted to remain dominant assuming the test to be prognostic only, that is, all women receiving chemotherapy derive the same relative benefit in terms of reduction in distant recurrences. However, because the evidence base for IHC4 is less developed than that for OncotypeDX, additional assumptions were required and the results are subject to greater uncertainty.

Nottingham Prognostic Index plus, Breast Cancer Index, BluePrint and Randox Breast Cancer Array

Clinical

Based on the limited available data identified for these tests, no firm conclusions can be drawn about their analytical validity, clinical validity (prognostic ability) and clinical utility. Further evidence on the prognostic and predictive ability of all of these tests is required.

Economic

No studies were identified in the systematic review of the economic literature. The EAG did not model treatment guided using these tests because of significant gaps in the evidence base.

Discussion

Strengths and limitations of the analyses and uncertainties

Clinical

Two of the tests (OncotypeDX and MammaPrint) have a reasonably large evidence base, although there are some methodological weaknesses relating to this evidence in terms of heterogeneity of patient cohorts and issues arising from the retrospective nature of the evidence, such as the relevance of the evidence to current methods of diagnosis, treatment and standards of care. The evidence base for OncotypeDX is considered to be the most robust. The MammaPrint evidence is typically based on observational data (small cohort studies) rather than randomised data, increasing the risk of selection bias. Both IHC4 and Mammostrat present early evidence of the prognostic ability of the tests based on large UK-based validation cohorts. Further evidence is required on the clinical utility of all of these tests, and on UK-based populations where this is not currently available. The evidence base for the remaining five tests has significant gaps and is considered less robust.

Economic

Four of the nine tests were included in the economic evaluation by the EAG. The model used UK-specific evidence where possible, including the baseline use of chemotherapy, the risk of distant recurrence/recurrence and reclassification with the new test, so that its conclusions would be relevant to the UK setting. Our analysis focused on patients with ER+, LN–, HER2– early breast cancer as use of the tests in this population is supported by the most robust clinical evidence. Women with a NPI ≤3.4 and women with a NPI >3.4 were modelled separately to account for the prognostic value of the current treatment decision based on clinicopathological parameters and to allow a scenario assuming that the test was offered to a subgroup of the population at intermediate risk to be conducted.

However, there are significant limitations with regard to the economic analyses. Results of all of the analyses have to be interpreted with caution and the results cannot be compared directly between tests. Given that no studies following patients from initial diagnosis through to final health outcomes were identified for any of the tests, the economic model needed to combine clinical data from several different sources in order to model how the results from the new tests translate into final outcomes in the form of QALYs. This resulted in significant uncertainties that were not adequately captured with the

probabilistic sensitivity analysis – data used in the model were not always based on UK populations and were not always specifically based on the ER+, LN–, HER2– population of interest. Differences in the age of the study populations and the endocrine and chemotherapy regimens used in the studies compared with those in the model introduced further uncertainty. One key area of uncertainty is whether the tests are prognostic or predictive of the benefit of chemotherapy (i.e. do they allow identification of high-risk patients who would derive a greater relative benefit from chemotherapy). The ICER was very sensitive to this assumption. There were particular concerns relating to the studies used to estimate the benefit associated with chemotherapy for patients categorised by risk group by the new tests, in relation to both the study design and the populations included in these studies. The evidence base on the impact of the new tests on the selection of patients to receive chemotherapy was also lacking or not considered generalisable to the UK population. Univariate sensitivity analyses indicated that the ICER was sensitive to these assumptions.

A greater number of assumptions were required to model IHC4 compared with OncotypeDX because of data limitations for IHC4. There were more significant gaps in the evidence for MammaPrint and Mammostrat, and any conclusions that can be drawn from these exploratory analyses are subject to considerable uncertainty.

Conclusions

The OncotypeDX and MammaPrint tests have a reasonably large evidence base, although there are some methodological weaknesses relating to this evidence in terms of heterogeneity of patient cohorts and the use of retrospective data. The evidence base for OncotypeDX is considered to be the most robust. Two of the tests (IHC4 and Mammostrat) have presented early evidence of the prognostic ability of the tests, based on large UK-based validation cohorts, but further research is required. The clinical utility evidence for GEP and expanded IHC tests is limited by the lack of large prospective studies in UK populations. PAM50, BluePrint, Breast Cancer Index, NPI+ and Randox Breast Cancer Array have only limited clinical evidence to date.

The economic analysis suggests that the use of the new tests may result in small increases in QALYs compared with currently used prognostic tools, but current limitations in the evidence base introduce significant uncertainty in the results. A key area of uncertainty is whether tests are prognostic only or identify high-risk patients who will benefit more relatively from chemotherapy (from reductions in the risk of recurrence) than low-risk patients. The economic analyses suggested that, of the four tests considered, treatment guided using IHC4 has the greatest potential to be cost-effective at a £20,000 threshold, given the low cost of the test; however, further evidence on IHC4 is needed and the exact cost of using the test in the NHS needs to be investigated further. Although the OncotypeDX test has been shown to have the potential to be cost-effective at the £20,000 threshold for patients with a NPI >3.4, further evidence is needed on the impact on decision-making in the UK and to clarify the predictive ability of the test specifically in an ER+, LN–, HER– population receiving current endocrine and chemotherapy regimens.

Implications for service provision

The impact of sending large numbers of samples to central testing facilities for pathology services, in terms of tissue tracking, pathologist and technical staff time, data input on receipt, etc., would need to be explored. Tests requiring the use of fresh tissue require a major change in practice with regard to the handling of tissue, with significant implications for service configuration and costs. The addition of expanded IHC-based tests is likely to fit more easily with current practice in the NHS. Quality assurance issues would need to be addressed, for example for the Ki-67 component of the IHC4 test, before these tests could be considered for use in clinical practice in the NHS.

The main research priorities relate to the reliability and reproducibility of the IHC4 test, along with further evidence of the prognostic ability of IHC4 compared with NPI and Adjuvant! Online. Further evidence on the predictive ability of all of the tests is also required. In addition, evidence to improve the understanding of the impact of these tests (for tests that provide a risk score/category and tests that provide subtype information only) on the management of patients in a UK population is urgently needed.

Study registration

This study is registered as PROSPERO 2011:CRD42011001361, available from www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42011001361.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Publication

Ward S, Scope A, Rafia R, Pandor A, Harnan S, Evans P, *et al*. Gene expression profiling and expanded immunohistochemistry tests to guide the use of adjuvant chemotherapy in breast cancer management: a systematic review and cost-effectiveness analysis. *Health Technol Assess* 2013;**17**(44).

Health Technology Assessment

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Five-year impact factor: 5.804

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index and is assessed for inclusion in the Database of Abstracts of Reviews of Effects.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: www.hta.ac.uk/

This report

The research reported in this issue of the journal was commissioned and funded by the HTA programme on behalf of NICE as project number 10/125/01. The protocol was agreed in May 2011. The assessment report began editorial review in October 2011 and was accepted for publication in March 2012. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).