Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer: a systematic review and economic analysis

Sue Harnan,[1]* Paul Tappenden,[1] Katy Cooper,[1] John Stevens,[1] Alice Bessey,[1] Rachid Rafia,[1] Sue Ward,[1] Ruth Wong,[1] Robert C Stein[2,3] and Janet Brown[4]

[1]Health Economics and Decision Science, School of Health and Related Research, University of Sheffield, Sheffield, UK
[2]University College London Hospitals Biomedical Research Centre, London, UK
[3]Research Department of Oncology, University College London, London, UK
[4]Department of Oncology and Metabolism, University of Sheffield, Sheffield, UK

*Corresponding author s.harnan@sheffield.ac.uk

Declared competing interests of authors: Robert C Stein is the chief investigator of the OPTIMA (Optimal Personalised Treatment of breast cancer using Multi-parameter Analysis) trial and has published results from the OPTIMA Prelim study.

Published June 2019
DOI: 10.3310/hta23300

Scientific summary

Adjuvant chemotherapy decisions in early-stage breast cancer

Health Technology Assessment 2019; Vol. 23: No. 30
DOI: 10.3310/hta23300

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

Scientific summary

Background

Breast cancer is the most commonly diagnosed cancer in women in England and Wales. In 2014, 55,222 new cases of breast cancer were diagnosed. Treatment usually involves surgery to remove the primary tumour and any involved lymph nodes; this may be followed by radiation therapy, endocrine therapy and/or chemotherapy with or without trastuzumab (Herceptin®, Roche Products Ltd) depending on tumour and patient variables. A proportion of patients also receive neoadjuvant therapy prior to surgery. Although chemotherapy can reduce the likelihood of cancer recurrence and death for women with breast cancer, it may have considerable adverse effects. Improved information on a patient's risk of recurrence (i.e. prognostic risk) and/or likely response to chemotherapy (i.e. predictive benefit) may help target chemotherapy to those patients who will benefit the most. Avoiding chemotherapy in patients at low risk of recurrence, who would therefore obtain limited absolute benefit, avoids the unpleasant side effects of chemotherapy and reduces expenditure on both the chemotherapy itself and the treatment of these adverse effects. Tumour profiling tests aim to improve the use of chemotherapy in breast cancer by improving the categorisation of patients in accordance with risk and the identification of those patients who will gain most benefit from chemotherapy.

Objectives

The overall aim of the assessment was to address the question 'Do tumour profiling tests used for guiding adjuvant chemotherapy decisions in patients with early-stage breast cancer represent a clinically effective and cost-effective use of NHS resources?'. This includes an update of the systematic review and cost-effectiveness analysis that informed National Institute for Health and Care Excellence (NICE) Diagnostics Guidance (DG) 10.

The objectives of the assessment were to:

- conduct a systematic review of the published evidence on the effectiveness and cost-effectiveness of five tumour profiling tests with or without clinicopathological factors [EndoPredict® (Myriad Genetics Ltd, London, UK), onco*type* DX® (Genomic Health, Inc., Redwood City, CA, USA), MammaPrint® (Agendia, Inc., Amsterdam, the Netherlands), immunohistochemistry 4 (IHC4) and Prosigna® (NanoString Technologies, Inc., Seattle, WA, USA)] to guide decisions about adjuvant chemotherapy
- develop a health economic model to assess the cost-effectiveness associated with the use of tumour profiling tests compared with current prognostic tools to guide the use of adjuvant chemotherapy in early-stage breast cancer from the perspective of the NHS and Personal Social Services (PSS).

Methods

This report was commissioned by the National Institute for Health Research Health Technology Assessment (HTA) programme as project number 16/30/03. A registered protocol of the systematic review (CRD42017059561) is available on the PROSPERO website: www.crd.york.ac.uk/prospero/display_record.asp?ID=CRD42017059561 (accessed 10 May 2018).

Clinical evidence review methods

A systematic review was undertaken, which included results from a search of nine databases in February 2017 plus other sources including a review published in 2013. The review included studies assessing clinical effectiveness of the five tumour profiling tests to guide decisions about adjuvant chemotherapy in

people with early-stage breast cancer, with a focus on those with hormone receptor-positive (HR+), human epidermal growth factor receptor 2-negative (HER2–) stage I or II cancer with zero to three positive lymph nodes. Outcomes included prognostic performance (whether or not recurrence and survival outcomes differ between test risk groups), prediction of chemotherapy benefit (whether or not effect of chemotherapy differs between test risk groups), clinical utility (the impact of prospective use of the test on recurrence and survival) and decision impact (changes in chemotherapy recommendations pre/post test).

Cost-effectiveness methods

The External Assessment Group (EAG) undertook a review of existing economic analyses published since NICE DG10. The EAG also reviewed and critically appraised economic analyses of onco*type* DX, MammaPrint and EndoPredict, which were provided during the course of the appraisal.

In addition, the EAG developed a de novo health economic model to assess the cost-effectiveness of onco*type* DX, MammaPrint, Prosigna, EndoPredict Clinical (EPClin) and IHC4 plus clinical factors (IHC4+C), each compared with current practice. The health economic analysis was undertaken from the perspective of the NHS and PSS and was largely based on the model developed to inform NICE DG10. The EAG model adopts a hybrid decision tree/Markov structure. The model parameters were informed by a number of sources, including a bespoke analysis of the Translational substudy of the Arimidex, Tamoxifen, Alone or in Combination (TransATAC) trial, the Microarray In Node-negative Disease may Avoid ChemoTherapy (MINDACT) trial, a bespoke analysis of the National Cancer Registration and Analysis Service data set, a bespoke survey disseminated by the UK Breast Cancer Group (UKBCG), the NHS England Access Scheme Database, standard costing sources and other literature.

Results

Clinical evidence results

The review included 153 studies across all five tests and across all outcomes listed in the NICE scope. Four of these were data sets provided by a company as commercial-in-confidence or academic-in-confidence data and could not be presented in this report.

Among studies of lymph node-negative (LN0) patients receiving endocrine monotherapy, percentages of patients categorised as high risk ranged from 9% to 33% across all five tests. In lymph node-positive (LN+) patients, three tests [Prosigna/ROR-PT (risk of recurrence based on Prediction Analysis of Microarray 50 subtype information plus proliferation score plus tumour size), EPClin and IHC4+C] categorised far more (38% to 76%) LN+ patients than LN0 patients as high risk among studies of endocrine monotherapy, whereas onco*type* DX categorised a similar number as high risk in the LN0 and LN+ groups. However, onco*type* DX categorised more patients as low risk in LN+ than other tests (57% in onco*type* DX vs. 4% to 28% in other tests), but with worse 10-year distant recurrence/relapse-free survival/distant recurrence/relapse-free interval outcomes (82% in onco*type* DX vs. 95% to 100% in other tests).

In terms of prognostic performance, all tests had statistically significant prognostic power in unadjusted analyses in LN0 and LN+ populations. However, recurrence score–pathology–clinical (RSPC) was only validated in LN0 patients in a cohort that had been used in part to derive onco*type* DX, and unadjusted analyses using clinical cut-off points were not reported in the validation sets for IHC4 or IHC4+C. All tests provided additional prognostic information in addition to the most commonly used clinicopathological factors and in addition to clinical treatment score and Nottingham Prognostic Index (NPI) in LN0, although data were not reported by risk group for onco*type* DX. Results were more varied in LN+ patients.

There was some evidence of differential chemotherapy benefit between risk groups for onco*type* DX, as shown by significant interaction tests between risk group and chemotherapy treatment in unadjusted analyses. Adjusted interaction tests were significant/borderline significant in the National Surgical Adjuvant Breast and Bowel Project (NSABP) B20 study (LN0 patients, significant in HER2– patients), whereas in the

Southwest Oncology Group (SWOG)-8814 (LN+ patients), they were significant when adjusted for some clinicopathological variables individually, but not when adjusting for ER determined by Allred status. However, part of the NSABP B20 cohort was the derivation cohort for onco*type* DX and this may bias results in favour of observing an interaction. Onco*type* DX RSPC (onco*type* DX plus age, tumour size and grade) was prognostic but not statistically significantly predictive for chemotherapy benefit and was not tested in an entirely independent validation cohort.

Evidence relating to the ability of MammaPrint to predict benefit from chemotherapy was extremely limited. Although the effect of chemotherapy was significant in high-risk groups and not significant in low-risk groups, interaction tests between risk groups and chemotherapy treatment were not significant, suggesting that there was no statistically significant difference in the effect of chemotherapy between risk groups.

The MINDACT randomised controlled trial (RCT) for MammaPrint was, at the time of writing, the only RCT to have reported (in full) the use of a test in clinical practice compared with clinical practice only. It reported that, for patients with a high Modified Adjuvant! Online (mAOL) score (clinical risk) and a low MammaPrint risk, chemotherapy gave a non-significant absolute benefit of 1.5% in 5-year distant metastasis-free survival (DMFS) ($p = 0.267$). This met the primary objective in that the lower bound of the 95% confidence interval for 5-year DMFS in the no-chemotherapy group was $\geq 92\%$. This finding was interpreted by the authors as implying that patients who had a high clinical risk but a low MammaPrint risk could potentially avoid chemotherapy. In patients who were low-mAOL with a high MammaPrint risk, chemotherapy gave an absolute benefit of 0.8%. This could be interpreted to mean that MammaPrint would not be a useful test in mAOL low-risk patients, as it would not alter treatment decisions.

For onco*type* DX and MammaPrint, evidence from observational, non-comparative studies assessing the impact of the test used prospectively in clinical practice suggested that recurrence/survival outcomes in low-risk groups were acceptable even with low rates of chemotherapy. There was no similar evidence relating to the other tests.

Decision impact studies reported that the percentage of patients with any change in chemotherapy recommendation or decision pre/post test ranged from 27% to 49% across UK studies (these included onco*type* DX, EndoPredict and IHC4+C) and from 5% to 70% across European studies (these included all tests except IHC4). The net change in the percentage of patients with a chemotherapy recommendation or decision pre/post test ranged from an increase of 1% to a decrease of 23% among UK studies and from a change of 0% to a decrease of 64% across European studies.

Concordance between tests was not fully reviewed as it was not within the scope of the assessment, but one UK study [Optimal Personalised Treatment of early breast cancer usIng Multi-parameter Analysis preliminary (OPTIMA Prelim)] that compared onco*type* DX, MammaPrint, Prosigna and IHC4 concluded that although tests assigned similar proportions of patients to low/intermediate- and high-risk categories, test results for an individual patient could differ markedly depending on which test was used.

Data relating to anxiety and health-related quality of life (HRQoL) were limited as most studies did not include a comparator, instead adopting a pre–post test design. Anxiety generally reduced post test, but it is unclear if this would happen equally after a treatment decision made in accordance with clinical factors. HRQoL improved in some analyses.

Microarray studies support conclusions from studies using the commercial versions of the assays in suggesting that onco*type* DX, MammaPrint and EndoPredict can discriminate between high- and low-risk patients regardless of lymph node status (there were no relevant microarray studies for EndoPredict or IHC4).

Cost-effectiveness results

The EAG's base-case model suggests the following results.

Oncotype DX

Within the subgroup of LN0 patients with a NPI of ≤ 3.4, the incremental cost-effectiveness ratio (ICER) for oncotype DX versus current practice is expected to be £122,725 per quality-adjusted life-year (QALY) gained (£34,245 per QALY gained assuming a predictive benefit). Within the subgroup of LN0 patients with a NPI of > 3.4 and the one to three positive lymph nodes (LN1–3) subgroup, oncotype DX is expected to be dominated by current practice (conversely, oncotype DX dominates current practice if a predictive benefit is assumed). The results generated using the EAG model are primarily driven by the modelled reduction in the use of adjuvant chemotherapy using the oncotype DX test. When based on the same evidence sources, the Genomic Health, Inc. (Redwood City, CA, USA) model produces broadly similar results.

IHC4 plus clinical factors

Within the subgroup of LN0 patients with a NPI of ≤ 3.4, the ICER for IHC4+C versus current practice is expected to be £2654 per QALY gained. Within the subgroup of LN0 patients with a NPI of > 3.4 and the LN1–3 subgroup, IHC4+C is expected to dominate current practice.

Prosigna

Within the subgroup of LN0 patients with a NPI of ≤ 3.4, the ICER for Prosigna versus current practice is expected to be £91,028 per QALY gained. Within the subgroup of LN0 patients with a NPI of > 3.4 and the LN1–3 subgroup, the ICERs for Prosigna versus current practice are £26,058 and £28,731 per QALY gained, respectively.

EndoPredict Clinical

Within the subgroup of LN0 patients with a NPI of ≤ 3.4, the ICER for EPClin versus current practice is expected to be £147,419 per QALY gained. Within the subgroup of LN0 patients with a NPI of > 3.4, the ICER for EPClin versus current practice is expected to be £46,788 per QALY gained. Within the LN1–3 subgroup, the ICER for EPClin versus current practice is expected to be £21,458 per QALY gained.

MammaPrint

Within the overall MINDACT population, the ICER for MammaPrint versus current practice is expected to be £131,482 per QALY gained. Within the mAOL high-risk subgroup, MammaPrint is expected to be dominated by current practice. Within the mAOL low-risk subgroup, the ICER for MammaPrint versus current practice is expected to be £414,202 per QALY gained.

Discussion

Strengths and limitations in the clinical evidence base

The evidence base was large but included only one RCT of a test being used in clinical practice compared with usual clinical practice that had reported results in full (MINDACT, for MammaPrint). A number of reanalyses of RCTs, which are generally considered to be high-quality sources of data, were also included in the reviews of prognosis and prediction of chemotherapy benefit. However, nearly all studies excluded patients who did not have enough tissue sample, meaning that patients with small tumours are, in theory, likely to be under-represented.

Many studies were funded by industry and this should be borne in mind when interpreting the evidence base.

Many studies were observational in nature, and these are subject to confounding, whereby exclusion of patients who received chemotherapy is likely to introduce bias as these patients are likely to be systematically different in terms of known (and potentially unknown) prognostic variables. Equally, studies that included patients who received chemotherapy may underestimate prognostic effect.

There were some key gaps in the literature for IHC4+C and RSPC. Notably, IHC4+C and RSPC have only been validated in one cohort each, and this was not entirely independent in the case of RSPC. There are known problems with conducting the analyses required for IHC4, and it is unclear whether or not the absolute IHC4 values obtained would be similar across centres.

Much of the evidence base relates to unadjusted analyses, which do not address the crucial question of whether or not a test has additional value over clinicopathological factors. When adjusted analyses were conducted, the clinicopathological variables included were not always consistent and it is unclear if all important factors were included in all analyses.

There were relatively limited data relating to the ability of onco*type* DX and MammaPrint to predict benefit from chemotherapy, and some of the analyses conducted were also subject to criticisms about adjustment for relevant confounders and use of the derivation cohort. This means that there remains uncertainty about whether or not the tests are associated with a predictive benefit from chemotherapy.

The evidence base relating to the impact of tests on treatment decisions (decision impact studies) was limited in that use of chemotherapy differs across countries and there were no UK studies for two of the tests (MammaPrint and Prosigna) and only one UK study for another two of the tests (EndoPredict and IHC4+C).

Strengths and limitations relating to the health economic analysis

The EAG model has a number of strengths:

- For all tests, risk classification and DMFS probabilities are derived from the same source (TransATAC or MINDACT).
- Within the LN0 intermediate-risk subgroup (NPI of > 3.4, analysis of three-level tests), the probability of receiving chemotherapy with and without the test is based on the NHS England Access Scheme Database; this is likely to best reflect how the three-level tumour profiling tests would be used in clinical practice in England.
- The model structure is consistent with that of other published models of tumour profiling tests: when similar data inputs are used, the EAG model produces similar results to the previous EAG model and the Genomic Health model.
- Extensive deterministic sensitivity analyses have been conducted to explore the impact of uncertainty on the model results.

The model is also subject to several limitations, most of which stem from uncertainties in the evidence base. The main limitations and uncertainties relating to the cost-effectiveness analysis are:

- With the exception of onco*type* DX in the subgroup of LN0 patients with a NPI of > 3.4 (clinical intermediate risk), the evidence surrounding the pre- and post-test chemotherapy probabilities is subject to considerable uncertainty. This has the propensity to influence the conclusions regarding the cost-effectiveness of all tests.
- There is uncertainty regarding whether or not onco*type* DX and MammaPrint are predictive of chemotherapy benefit; the inclusion of such effects are likely to strongly influence economic conclusions drawn from the analysis.
- The analysis of MammaPrint is based on a different data source to that used in the other four tests.
- The TransATAC study, which was used to estimate test risk classification and DMFS probabilities, was the derivation study for IHC4. For this reason, there is potential for the overestimation of prognostic performance for this test.

Implications for service provision

The per-test costs for Prosigna provided by NanoString Technologies (used in the EAG economic analyses) are based on an efficient level of throughput. This may not hold if centres do not undertake the anticipated number of tests (e.g. in smaller centres or if multiple tumour profiling tests are available). Furthermore, as NanoString Technologies does not offer a centralised testing service, local testing services will need to be established.

The IHC4 test is not currently commercially available. Standardisation of IHC4 and quality assurance programmes are required before this test may be used routinely within the NHS.

Suggested research priorities

- There is uncertainty regarding whether or not onco*type* DX and MammaPrint are predictive of chemotherapy benefit. Further studies that adjust for all relevant clinicopathological factors in validation cohorts are required.
- There is limited evidence demonstrating long-term impacts resulting from the use of the five tumour profiling tests. Future studies assessing the comparative long-term impact of the tests compared with risk prediction tools commonly used in clinical practice would be valuable.
- There is uncertainty regarding the cost-effectiveness of all five tests included in the NICE scope. It is noteworthy that under the assumption of no predictive chemotherapy benefit the inclusion of additional data collected through the NHS England Access Scheme Database has a significant impact on the conclusions previously drawn from the onco*type* DX analysis within NICE DG10 (moving from an ICER of £22,572 per QALY gained to a situation in which onco*type* DX is dominated in the subgroup of LN0 patients with a NPI of > 3.4). Additional UK-based data collection relating to pre- and post-test chemotherapy use for EPClin, IHC4+C, Prosigna and MammaPrint may be important in reducing existing uncertainty surrounding the cost-effectiveness of these tests.

Study registration

This study is registered as PROSPERO CRD42017059561.

Funding

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

Health Technology Assessment

NICE TAR and DAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.513

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

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

Editorial contact: journals.library@nihr.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: http://www.nets.nihr.ac.uk/programmes/hta

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 16/30/03. The protocol was agreed in March 2017. The assessment report began editorial review in November 2017 and was accepted for publication in August 2018. 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 and Social Care. 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 and Social Care.

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