PET and MRI for the assessment of axillary lymph node metastases in early breast cancer

Positron emission tomography (PET) and magnetic resonance imaging (MRI) for the assessment of axillary lymph node metastases in early breast cancer: systematic review and economic evaluation

KL Cooper,[1]* Y Meng,[1] S Harnan,[1] SE Ward,[1] P Fitzgerald,[1] D Papaioannou,[1] L Wyld,[2] C Ingram,[2] ID Wilkinson[2] and E Lorenz[2]

[1]School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK
[2]Sheffield Teaching Hospitals, University of Sheffield, Sheffield, UK

*Corresponding author

Executive summary

Health Technology Assessment 2011; Vol. 15: No. 4
DOI: 10.3310/hta15040

Health Technology Assessment
NIHR HTA programme
www.hta.ac.uk

Executive summary

Background

Evaluation of axillary lymph node metastases is important for breast cancer staging and treatment planning. Axillary lymph node dissection (ALND) has been considered the gold standard for identifying axillary metastases and controlling spread, but has a high risk of morbidity including long-term lymphoedema. Sentinel lymph node biopsy (SLNB) and 4-node sampling (4-NS) have a lower risk of morbidity. Current guidelines from the National Institute for Health and Clinical Excellence recommend SLNB or 4-NS where ultrasound-guided biopsy is negative. ALND is indicated where ultrasound-guided biopsy, SLNB or 4-NS are positive.

Objectives

The objectives of this assessment were to evaluate the diagnostic accuracy, cost-effectiveness and effect on patient outcomes of positron emission tomography (PET), with or without computed tomography (CT), and magnetic resonance imaging (MRI) in the evaluation of axillary lymph node metastases in patients with newly diagnosed early-stage breast cancer. PET and MRI are assessed firstly as a replacement for SLNB or 4-NS, and secondly as an additional test prior to SLNB or 4-NS.

Methods

A systematic review was undertaken to identify studies reporting sensitivity and specificity of PET or MRI for the assessment of axillary lymph node metastases in early-stage breast cancer. The following databases were searched in April 2009: MEDLINE, MEDLINE In-Process & Other Non-Indexed Citations, EMBASE, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Cochrane Database of Systematic Reviews (CDSR), Cochrane Central Register of Controlled Trials (CENTRAL), Database of Abstracts of Reviews of Effects (DARE), NHS Economic Evaluation Database (NHS EED), HTA database, Science Citation Index and BIOSIS previews. Research registers and conference proceedings were also searched. PET studies with < 20 analysable patients were excluded, while MRI studies of all sizes were included (as there were fewer with a large sample size). Study quality was assessed using the quality assessment of diagnostic accuracy studies checklist. A bivariate random effects approach was used for the meta-analysis of pairs of sensitivity and specificity data.

A decision model was developed to investigate the benefits, harms and cost-effectiveness of PET and MRI, either as a replacement for SLNB or 4-NS or as an additional test prior to SLNB or 4-NS. Both SLNB and 4-NS are currently used in the UK. Comparison of SLNB with 4-NS was not part of the remit of this assessment; therefore, the two baseline strategies (SLNB and 4-NS) were considered separately. It was assumed that, prior to these investigations, all patients underwent clinical examination and axillary ultrasound (and ultrasound-guided biopsy where ultrasound was positive). The number of correct and incorrect diagnoses, costs and impact on quality-adjusted life-years (QALYs) due to cancer recurrence and adverse effects were determined for each strategy. Model results are presented in terms of net health benefit and cost per incremental QALY gained. Probabilistic sensitivity analyses were undertaken.

Results

Summary of clinical results

Diagnostic accuracy of positron emission tomography

Across 26 studies evaluating PET or PET/CT ($n = 2591$ patients), the mean sensitivity was 63% [95% confidence interval (CI) 52% to 74%; range 20%–100%] and mean specificity 94% (95% CI 91% to 96%; range 75%–100%). Of the seven studies evaluating PET/CT ($n = 862$), the mean sensitivity was 56% (95% CI 44% to 67%) and mean specificity 96% (95% CI 90% to 99%). Of the 19 studies evaluating PET only ($n = 1729$), the mean sensitivity was 66% (95% CI 50% to 79%) and mean specificity 93% (95% CI 89% to 96%). PET performed less well for small metastases; the mean sensitivity was 11% (95% CI 5% to 22%) for micrometastases (≤ 2 mm; five studies; $n = 63$), and 57% (95% CI 47% to 66%) for macrometastases (> 2 mm; four studies; $n = 111$). The smallest metastatic nodes detected by PET measured 3 mm, while PET failed to detect some nodes measuring > 15 mm. Studies in which all patients were clinically node-negative showed a trend towards lower sensitivity of PET compared with studies with a mixed population.

Diagnostic accuracy of magnetic resonance imaging

There were nine studies evaluating MRI ($n = 307$ patients), many reporting more than one set of results according to different criteria for positivity. Based on the highest reported sensitivity and specificity per study, the mean sensitivity was 90% (95% CI 78% to 96%; range 65%–100%) and mean specificity 90% (95% CI 75% to 96%; range 54%–100%). Across five studies evaluating ultrasmall super-paramagnetic iron oxide (USPIO)-enhanced MRI ($n = 93$), the mean sensitivity was 98% (95% CI 61% to 100%) and mean specificity 96% (95% CI 72% to 100%). Across the three studies of gadolinium-enhanced MRI ($n = 187$), the mean sensitivity was 88% (95% CI 78% to 94%) and mean specificity 73% (95% CI 63% to 81%). In the single study of in vivo proton magnetic resonance spectroscopy ($n = 27$), the sensitivity was 65% (95% CI 38% to 86%) and specificity 100% (95% CI 69% to 100%). Therefore, USPIO-enhanced MRI showed a trend towards higher sensitivity and specificity than gadolinium-enhanced MRI. No data were presented according to size of metastases for MRI.

Adverse effects and contraindications

No adverse effects were reported for PET. Studies of MRI reported only mild-to-moderate adverse effects including mild rash following USPIO administration, claustrophobia and back pain. Cautions and contraindications exist for both PET (pregnancy) and MRI (allergy to contrast agents, renal or liver dysfunction, pacemakers and other metallic implants), and some patients are unable to undergo MRI due to claustrophobia.

Summary of cost-effectiveness and benefits versus risks

The results of the decision modelling suggest that the most cost-effective strategy is to replace axillary sampling (SLNB or 4-NS) with MRI. This strategy dominates the baseline SLNB and 4-NS strategies, generating higher QALYs and lower costs. SLNB and 4-NS are avoided for all patients, leading to fewer adverse effects, including lymphoedema, which has an assumed lifelong impact on quality of life. However, MRI has lower sensitivity than SLNB and 4-NS [leading to more false-negatives (FNs)] and a lower specificity [leading to more false-positives (FPs)]. Patients with a FN diagnosis will not receive ALND or adjuvant therapy, leading to a higher risk of cancer recurrence. Patients with a FP diagnosis will receive ALND unnecessarily, with the accompanying increased risk of adverse effects. At the population level, the model results suggest that the MRI replacement strategy costs less, and the health benefits gained by the majority of patients outweigh the negative impact on survival and quality of life of a small proportion of patients. This strategy may however be rejected on clinical grounds, owing to the increase in number of FP and FN cases compared with current practice.

If the replacement strategies are rejected on clinical grounds, the most cost-effective strategy is predicted to be the baseline 4-NS strategy (compared with 4-NS as the baseline) or the use of MRI prior to SLNB (compared with SLNB as the baseline). For strategies that place MRI or PET before axillary sampling, patients with true-positive PET or MRI results receive immediate ALND without the need to carry out a separate SLNB or 4-NS procedure (although the requirement for two procedures may also be averted through the use of intraoperative cytology following axillary sampling). The number of FN cases is also reduced owing to the use of two sequential tests. The disadvantage is that, due to the lower specificity of PET and MRI compared with SLNB and 4-NS, there are still more patients with FP results who receive ALND unnecessarily, with the accompanying increased risk of adverse effects. The total QALYs generated by the strategies of adding MRI prior to 4-NS or SLNB are very similar to those in the baseline strategies and these results are not considered to be robust, based on the quality of the data available.

The model results indicate that, as would be expected, patients with FN results have the lowest survival rates because they do not receive ALND or chemotherapy that may reduce the risk of recurrence. Compared with the two baseline strategies, the overall survival rates are lower for the MRI or PET replacement strategies (due to an increase in FNs) and higher for the strategies where MRI or PET are placed before 4-NS and SLNB (due to a decrease in FNs). However, the absolute differences in overall survival rate among tested diagnostic strategies are relatively small, since the absolute number of patients with FN results only accounts for a small proportion of all patients.

Sensitivity analysis suggests that the MRI replacement strategy remains the most cost-effective strategy in the majority of the one-way sensitivity analyses undertaken. The sensitivity analyses indicate that the cost-effectiveness of the PET replacement strategy is significantly affected by the assumption relating to long-term utility decrements for lymphoedema (for which few high-quality data exist) and the probability of relapse for FN patients. If the PET and MRI replacement strategies are excluded, the cost-effectiveness of the strategies of adding MRI prior to 4-NS or SLNB is affected by the assumed MRI sensitivity and specificity, the probability of relapse for FN patients, and the costs of SLNB. The results relating to addition of MRI prior to 4-NS or SLNB are not considered to be robust.

Conclusions

Implications for service provision

The included studies demonstrate a significantly higher sensitivity for MRI than for PET, with USPIO-enhanced MRI providing the highest sensitivity. However, as there were no studies directly comparing PET and MRI, caution should be taken when comparing these estimates. Sensitivity of PET was reduced for smaller metastases. Specificity was similar for PET and MRI. Sensitivity and specificity of PET and MRI varied widely between studies, and MRI studies were relatively small and varied in their methods; therefore, results should be interpreted with caution. All patients currently receive ultrasound prior to other investigations; the sensitivity of PET appears similar to that of ultrasound and so may provide little additional benefit, while the sensitivity of MRI appears slightly higher. Specificity of PET and USPIO-enhanced MRI appear slightly higher than for ultrasound.

PET and MRI have lower sensitivity and specificity than SLNB and 4-NS, but are associated with fewer adverse events. Decision modelling suggests that the most cost-effective strategy may be MRI rather than SLNB or 4-NS, reducing costs and increasing QALYs due to fewer adverse events for the majority of patients. However, this strategy may be clinically unacceptable due

to higher numbers of FN cases (leading to higher risk of recurrence) and FP cases (leading to unnecessary ALND). If this strategy is rejected on clinical grounds, the model suggests that the most cost-effective strategy may be the baseline 4-NS strategy (compared with 4-NS as the baseline) or the use of MRI prior to SLNB (compared with SLNB as the baseline). However, the results relating to addition of MRI prior to 4-NS or SLNB are not considered to be robust based on the quality of the input parameters available, and further work is required to provide more reliable estimates.

Suggested research priorities

If the use of MRI is deemed clinically acceptable (either to replace SLNB or 4-NS or as an additional test), then further large, well-conducted studies of MRI, particularly using USPIO, would be useful to obtain more robust data on sensitivity and specificity, adverse effects and the optimum criteria for defining a node as metastatic. In addition, further data on the long-term impacts of lymphoedema on cost and patient utility would be valuable, as well as studies of the comparative effectiveness and cost-effectiveness of SLNB and 4-NS. More robust UK cost data is needed for 4-NS and SLNB, as well as the cost of MRI and PET techniques.

Funding

This study was funded by the Health Technology Assessment programme of the National Institute of Health Research.

Publication

Cooper KL, Meng Y, Harnan S, Ward SE, Fitzgerald P, Papaioannou D, *et al*. Positron emission tomography (PET) and magnetic resonance imaging (MRI) for the assessment of axillary lymph node metastases in early breast cancer: systematic review and economic evaluation. *Health Technol Assess* 2011;**15**(4).

NIHR Health Technology Assessment programme

The Health Technology Assessment (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 research findings from the HTA programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA programme is needs led in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, from the public and consumer groups and from professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA programme then commissions the research by competitive tender.

Second, the HTA programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Third, through its Technology Assessment Report (TAR) call-off contract, the HTA programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer reviewed by a number of independent expert referees before publication in the widely read journal series *Health Technology Assessment*.

Criteria for inclusion in the HTA journal series

Reports are published in the HTA journal series 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 referees 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.

The research reported in this issue of the journal was commissioned and funded by the HTA programme on behalf of NICE as project number 08/35/01. The protocol was agreed in January 2009. The assessment report began editorial review in November 2009 and was accepted for publication in June 2010. 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 referees 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.

The views expressed in this publication are those of the authors and not necessarily those of the HTA programme or the Department of Health.

ISSN 1366-5278

Published by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk), on behalf of NETSCC, HTA.
Printed on acid-free paper in the UK by Henry Ling Ltd, The Dorset Press, Dorchester.