The TOMMY trial: a comparison of TOMosynthesis with digital MammographY in the UK NHS Breast Screening Programme – a multicentre retrospective reading study comparing the diagnostic performance of digital breast tomosynthesis and digital mammography with digital mammography alone

Fiona J Gilbert,[1*] Lorraine Tucker,[1] Maureen GC Gillan,[2] Paula Willsher,[1] Julie Cooke,[3] Karen A Duncan,[4] Michael J Michell,[5] Hilary M Dobson,[6] Yit Yoong Lim,[7] Hema Purushothaman,[8] Celia Strudley,[9] Susan M Astley,[10] Oliver Morrish,[11] Kenneth C Young[9] and Stephen W Duffy[12]

[1]Department of Radiology, University of Cambridge, Cambridge, UK
[2]Aberdeen Biomedical Imaging Centre, University of Aberdeen, Aberdeen, UK
[3]Jarvis Breast Centre, Guildford, UK
[4]North East Scotland Breast Screening Centre, Aberdeen, UK
[5]Breast Radiology, King's College Hospital, London, UK
[6]West of Scotland Breast Screening Service, Glasgow, UK
[7]The Nightingale Centre, University Hospital South Manchester, Manchester, UK
[8]Department of Radiology, St Bartholomew's Hospital, London, UK
[9]National Co-ordinating Centre for Physics of Mammography, Royal Surrey County Hospital, Guildford, UK
[10]Department of Imaging Science and Biomedical Engineering, University of Manchester, Manchester, UK
[11]East Anglian Regional Radiation Protection Service, Cambridge University Hospitals, Cambridge, UK
[12]Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK

*Corresponding author

Declared competing interests of authors: Dr Michell reports personal fees, non-financial support and grants from Hologic, the supplier of the mammographic and tomographic equipment, outside the submitted work. Dr Astley reports grants from NIHR during the conduct of the study; non-financial support from Matakina; and workshop training sessions and non-financial support from Hologic outside the submitted work. No others declared.

Published January 2015
DOI: 10.3310/hta19040

Scientific summary

A comparison of TOMosynthesis with digital MammographY

Health Technology Assessment 2015; Vol. 19: No. 4
DOI: 10.3310/hta19040

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

Scientific summary

Background

Although breast screening with mammography is recognised as the most effective method of detecting early-stage breast cancer and reducing breast cancer mortality, up to 30% of cancers are not detected by standard screening and this percentage is higher in dense breasts and in women under 50 years of age. A major limitation of breast screening mammography is that overlapping fibroglandular tissue can decrease the visibility of abnormalities or mimic abnormalities. As a result, some cancers are missed and there are unnecessary recalls, assessments and psychological stress.

Women with increased breast density have an increased risk of developing breast cancer, have lower screening programme sensitivity, and tend to have larger screen-detected and interval cancers. This is of concern for the NHS Breast Screening Programme (NHSBSP) as it extends the programme screening to include younger women and to offer screening mammography to women with a family history (FH) of breast cancer.

Digital breast tomosynthesis (DBT) is a newly developed three-dimensional (3D) imaging technique that has the potential to improve the accuracy of mammography by reducing interference from breast tissue overlap. This facilitates differentiation between malignant and non-malignant features and could decrease the number of false-positive recalls, associated health-care costs and patient anxiety.

The optimal role for DBT in the diagnosis and assessment of breast cancer is still uncertain. Current evidence indicates that DBT should be used as an adjunct to two-dimensional (2D) mammography. However, the additional radiation exposure required to obtain a standard 2D mammogram could be avoided if a synthetic 2D mammogram could be reconstructed from the images acquired from DBT imaging.

Objectives

- To compare the diagnostic accuracy of using DBT in conjunction with 2D or synthetic 2D, against standard 2D mammography in a retrospective reading study.
- To determine if the use of DBT in conjunction with 2D or synthetic 2D improves the accuracy of detection of (1) small or subtle breast cancers, (2) cancers in women with dense breasts, (3) cancers presenting as soft-tissue masses and (4) cancers presenting as microcalcifications.
- To assess the performance of two automated breast density software programs against observer-based visually assessed breast density.
- To assess the association of breast density with cancer incidence.

Study population

The study was conducted in six UK NHSBSP centres. Women (aged 47–73 years) recalled for further assessment after routine breast screening and women (aged 40–49 years) with moderate/high of risk of developing breast cancer who were attending annual mammography screening were recruited into the trial after giving written informed consent.

Intervention

All participants underwent standard two-view [mediolateral oblique (MLO) and craniocaudal (CC)] 2D mammography of both breasts and two-view (MLO and CC) DBT imaging acquired in a single examination under the same degree of breast compression. Image-processing software generated a synthetic 2D mammogram from the acquired 3D images.

Retrospective reading study

In an independent retrospective reading study, readers reviewed (1) 2D, (2) 2D + DBT or (3) synthetic 2D + DBT images for a case without access to original screening mammograms or prior examinations, and readers were blinded to the outcome status of each case.

Sensitivities and specificities were calculated for each of the three reading arms, for all cases combined, and then for subgroups by visually assessed breast density and dominant radiological feature. In addition, sensitivity to cancers was calculated for subgroups by lesion size [invasive and ductal carcinoma in situ (DCIS) separately] and histological grade. Sensitivity and specificity were compared between reading arms using McNemar methods, and receiver operating characteristic (ROC) analysis was performed to compare the diagnostic accuracy of the three reading arms (2D alone vs. 2D + DBT vs. synthetic 2D + DBT) by calculating the area under the curve and *p*-values.

Results

A total of 8869 participants were recruited over a 21-month period. This data set comprised 7684 cases recruited sequentially from assessment clinics and 1185 cases from women aged 40–49 years at moderate/high risk of developing breast cancer (subsequently referred to as FH cases). The latter group provided a cohort with higher breast density for subanalysis of the impact of breast density on the diagnostic accuracy of DBT.

After exclusions, there were 7060 subjects for analysis comprising 6020 (1158 cancers) assessment cases and 1040 (two cancers) FH cases. Reading data were available for 6928 (98%) cases by 2D only, 6960 (99%) by 2D + DBT and 6654 (94%) by synthetic 2D + DBT. The analysis included cases read in only two arms of the study to avoid introducing bias and was repeated using only cases which were read in all three arms. This produced identical results.

For all subjects combined, sensitivity was 87% [95% confidence interval (CI) 85% to 89%] for 2D only, 89% (95% CI 87% to 91%) for 2D + DBT and 88% (95% CI 86% to 90%) for synthetic 2D + DBT. The difference in sensitivity between 2D and 2D + DBT was of borderline significance ($p = 0.07$) and for synthetic 2D + DBT there was no significant difference ($p = 0.6$). Sensitivity was significantly higher for 2D + DBT than for 2D alone for invasive tumours of size 11–20 mm, for women with breast density of 50% or more ($p = 0.03$), and for age range 50–59 years ($p = 0.01$). A similar increase in sensitivity ($p = 0.01$) was seen for 2D + DBT in grade 2 invasive tumours (but not grade 1 or grade 3), and for lesions in which the dominant radiological feature was a mass ($p = 0.04$). For synthetic 2D + DBT, there was significantly ($p = 0.006$) higher sensitivity than for 2D alone in invasive cancers of size 11–20 mm. Cancers missed by 2D alone tended to be of size 11–20 mm, or to have a mass as the major radiological sign, compared with the other two reading modalities; cancers missed by 2D + DBT were less likely than the other two reading modalities to be of grade 2 or to have density less than 50%. The borderline improvement in cancer detection in the DBT reading arms compared with 2D alone differs from recent prospective screening studies, but this may be as a result of differences in case selection and study design.

Specificity was 58% (95% CI 56% to 60%) for 2D, 69% (95% CI 67% to 71%) for 2D + DBT and 71% (95% CI 69% to 73%) for synthetic 2D + DBT. Specificity was significantly higher for 2D + DBT and for DBT + synthetic 2D than for 2D ($p < 0.001$ in both cases). With ROC analysis, a significant increase in diagnostic accuracy ($p < 0.001$) for 2D + DBT and synthetic 2D + DBT compared with 2D alone was observed, most likely as a result of the marked improvement in specificity. The increases in specificity for the DBT reading arms were observed in all subgroups of breast density and dominant radiological feature and across all age groups ($p < 0.001$ in all cases). In all three reading arms, specificity tended to be lower for microcalcifications and higher for distortion/asymmetry, and synthetic 2D + DBT was inferior to both 2D and 2D + DBT in the detection of microcalcifications and DCIS of size 11–20 mm.

The breast density substudy evaluating the performance of two automated breast density software programs against observer-based visually assessed breast density indicated a high degree of variation in density as scored by readers, whereas the two commercially available software packages appeared to provide a more reliable assessment. The results confirmed the strong relationship between volumetric density and increased risk of breast cancer that has already been reported but highlighted the need for further research in this area to develop an accurate and reproducible method of breast density measurement for the assessment of breast cancer risk.

Conclusions

This reading study showed that the performance of 2D with DBT was better than 2D alone in terms of specificity, with a marginal improvement in sensitivity, and that synthetic 2D was comparable to conventional 2D when used with DBT. Case selection bias in the study design limits extrapolation of the results to a screening population. However, the observation that integrated 2D + DBT imaging was equally effective across all age groups and breast densities (and in particular for women aged 50–59 years and for breast density $\geq 50\%$) could be advantageous to the planned age extension of the NHSBSP and to screening mammography of the younger cohort of FH women. In addition, the potential to reduce false-positive recalls could also benefit both screening programmes and the screening population. However, further research is required to evaluate the practicalities and costs of implementing 2D + DBT in a screening setting and to undertake further comparison with 2D and synthetic 2D for different lesion types and breast densities. Prognostic modelling on existing data sets could be used to predict the impact on outcomes and mortality. Validated measures of breast density are required to assess personalised breast cancer risk and screening recommendations. In addition, the performance of DBT systems from different manufacturers needs to be evaluated.

Study registration

This trial is registered as ISRCTN73467396.

Funding

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

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 5.116

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: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 09/22/182. The contractual start date was in December 2010. The draft report began editorial review in February 2014 and was accepted for publication in August 2014. 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.

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