



A randomised trial of screening with digital breast tomosynthesis plus conventional digital 2D mammography versus 2D mammography alone in younger higher risk women

DOI:

[10.1016/j.ejrad.2017.06.018](https://doi.org/10.1016/j.ejrad.2017.06.018)

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Maxwell, A., Michell, M., Lim, Y., Astley, S., Wilson, M., Hurley, E., Evans, D., Howell, A., Iqbal, A., Kotre, J., Duffy, S., & Morris, J. (2017). A randomised trial of screening with digital breast tomosynthesis plus conventional digital 2D mammography versus 2D mammography alone in younger higher risk women. *European Journal of Radiology*. <https://doi.org/10.1016/j.ejrad.2017.06.018>

Published in:

European Journal of Radiology

Citing this paper

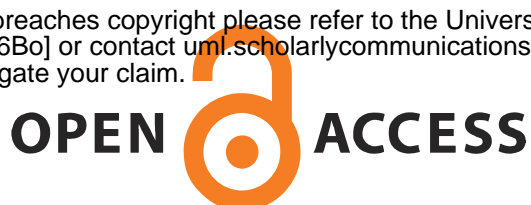
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact um.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



Accepted Manuscript

Title: A randomised trial of screening with digital breast tomosynthesis plus conventional digital 2D mammography versus 2D mammography alone in younger higher risk women

Authors: Anthony J. Maxwell, Michael Michell, Yit Y. Lim, Susan M. Astley, Mary Wilson, Emma Hurley, D. Gareth Evans, Anthony Howell, Asif Iqbal, John Kotre, Stephen Duffy, Julie Morris



PII: S0720-048X(17)30273-5
DOI: <http://dx.doi.org/doi:10.1016/j.ejrad.2017.06.018>
Reference: EURR 7878

To appear in: *European Journal of Radiology*

Received date: 8-6-2017
Revised date: 22-6-2017
Accepted date: 26-6-2017

Please cite this article as: Maxwell Anthony J, Michell Michael, Lim Yit Y, Astley Susan M, Wilson Mary, Hurley Emma, Evans D Gareth, Howell Anthony, Iqbal Asif, Kotre John, Duffy Stephen, Morris Julie. A randomised trial of screening with digital breast tomosynthesis plus conventional digital 2D mammography versus 2D mammography alone in younger higher risk women. *European Journal of Radiology* <http://dx.doi.org/10.1016/j.ejrad.2017.06.018>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A randomised trial of screening with digital breast tomosynthesis plus conventional digital 2D mammography versus 2D mammography alone in younger higher risk women

Anthony J Maxwell^{a,b*} anthony.maxwell@manchester.ac.uk, Michael Michell^c michael.michell@nhs.net, Yit Y Lim^a yit.lim@uhsm.nhs.uk, Susan M Astley^{b,a} sue.astley@manchester.ac.uk, Mary Wilson^a mary.wilson@uhsm.nhs.uk, Emma Hurley^a, emma.hurley@uhsm.nhs.uk, D Gareth Evans^{a,d}, gareth.evans@cmft.nhs.uk, Anthony Howell^{a,e} tony.howell@ics.manchester.ac.uk, Asif Iqbal^c asif.iqbal@nhs.net, John Kotre^f kotre@hotmail.co.uk, Stephen Duffy^g s.w.duffy@qmul.ac.uk, Julie Morris^{h,i} julie.morris@manchester.ac.uk

^a Nightingale Centre, University Hospital of South Manchester, Manchester, M23 9LT, UK

^b Division of Informatics Imaging & Data Sciences, School of Health Sciences, University of Manchester, Manchester, M13 9PT, UK

^c Breast Radiology Department, King's College Hospital, London, SE5 9RS, UK

^d Genomic Medicine, Manchester Academic Health Sciences Centre, University of Manchester and Central Manchester Foundation Trust, Manchester M13 9WL, UK

^e Manchester Cancer Research Centre, Division of Cancer Sciences, School of Medical Sciences, University of Manchester, Manchester, M20 4QL, UK

^f Christie Medical Physics and Engineering, The Christie NHS Foundation Trust, Manchester, M20 4BX, UK

^g Wolfson Institute of Preventive Medicine, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London, EC1M 6BQ, UK

^h Medical Statistics, University Hospital of South Manchester, Manchester, M23 9LT, UK

ⁱ Centre for Biostatistics, School of Health Sciences, University of Manchester, Manchester, M13 9PL, UK

*Corresponding author

Highlights

- 2D mammography plus DBT did not reduce false positive rate at incident screening
- DBT initially increased reader uncertainty
- Effect of DBT on false positive rate at prevalent screening remains uncertain

ABSTRACT

Introduction: Digital breast tomosynthesis has been shown to increase invasive cancer detection rates at screening compared to full field digital (2D) mammography alone, and some studies have reported a reduction in the screening recall rate. No prospective randomised studies of DBT have previously been published. This study compares recall rates with 2D mammography with and without concurrent DBT in women in their forties with a family history of breast cancer undergoing incident screening.

Materials and methods: Asymptomatic women aged 40 to 49 who had previously undergone mammography for an increased risk of breast cancer were recruited in two screening centres. Participants were randomised to screening with 2D mammography only at the first study screen followed a year later by screening with 2D plus DBT, or vice versa. Recall rates were compared using an intention to treat analysis. Reading performance was analysed for the larger centre.

Results: 1227 women were recruited. 1221 first screens (604 2D, 617 2D+DBT) and 1124 second screens (558 2D+DBT, 566 2D) were analysed. Eleven women had screen-detected cancers: 5 after 2D, 6 after 2D+DBT. The false positive recall rates were 2.4% for 2D and 2.2% for 2D+DBT ($p=0.89$). There was a significantly greater reduction between rounds in the number of women with abnormal reads who were not recalled after consensus / arbitration with 2D+DBT than 2D ($p=0.023$).

Conclusion: The addition of DBT to 2D mammography in incident screening did not lead to a significant reduction in recall rate. DBT may increase reader uncertainty until DBT screening experience is acquired.

Keywords: Breast cancer; Digital breast tomosynthesis (DBT); Digital mammography; Screening; Recall rate; Family history

INTRODUCTION

In recent years digital breast tomosynthesis (DBT) has become established as a useful adjunct to, and in some centres a replacement for, conventional full-field digital mammography (FFDM). Imaging the breast with DBT as a stack of contiguous thin (usually 1 mm thick) slices allows separation of overlying opacities, thus making cancers more obvious and resolving composite densities. Several large trials together with in-service observational studies have demonstrated a significantly higher sensitivity of combined FFDM and DBT compared to FFDM alone together with the potential to reduce the false positive rate [1].

False positive mammography is a significant concern, as it results in one of the tangible harms of mammographic screening. Many women who are recalled for further investigation experience high levels of anxiety in the short term [2] [3], in addition to the inconvenience and expense of attending a clinic appointment which ultimately brings no health benefit to the woman. False positive recalls have been reported to occur particularly in younger women and those with risk factors [4], although the UK Age Trial found that the false positive recall rates among women undergoing annual screening in their forties were little different to those of older women [5]. Younger women in general have denser breasts than older women, and this might therefore be expected to result in more frequent overlap of glandular tissue to produce composite densities which can mimic cancers. False positive rates in the UK National Health Service Breast Screening Programme are more common in women undergoing a first (prevalent) screen than those undergoing repeat (incident screening) (7.9% v. 3.0% [6]), as the lack of a prior mammogram for comparison can increase uncertainty about benign glandular densities.

The study presented here was designed to determine whether the addition of DBT to FFDM for women aged 40-49 attending the family history screening services in two large UK breast

centres results in a reduction in the number of false positive screens. These younger women are often anxious because of their increased breast cancer risk [7], and measures which potentially reduce the need for recall in this group are worthy of investigation.

MATERIALS AND METHODS

Trial design

The primary study was a prospective parallel group study of asymptomatic women undergoing two rounds of incident screening a year apart, conducted in two large UK breast screening centres. The study utilised a crossover design with 1:1 block randomisation to either conventional 2-dimensional full-field digital mammography alone (hereafter referred to as 2D mammography) or 2D mammography plus digital breast tomosynthesis (referred to as 2D+DBT), stratified by centre. The crossover at the second screen was to give equitable access to DBT and to reduce disparities in the radiation dose which participants received within the study.

Women undergoing prevalent screening were also recruited into a non-randomised cohort observational arm and underwent a single combined 2D+DBT examination. Recruitment of these women was terminated in the second year of the trial due to fertility.

The trial was approved by the UK National Research Ethics Service (North West - Greater Manchester West committee). Written informed consent was obtained from the participants prior to mammography.

Participants

Asymptomatic women aged 40 to 49 years previously assessed in a family history clinic (FHC) as being at moderate or high risk of breast cancer as defined by NICE (the National Institute for Health and Care Excellence) [8] ($\geq 3\%$ 10-year risk between the ages of 40 and 50 and/or a lifetime risk from age 20 of $\geq 17\%$) were eligible for inclusion. We have previously demonstrated an expected survival benefit of screening in this group [9].

Exclusion criteria were the inability to give informed consent, the presence of breast implants, pregnancy and a personal history of breast cancer.

The process for identifying and recruiting eligible women differed between the two centres:

Centre A: The FHC and mammography facility are co-located, and all newly referred women and the majority of previously screened women had an FHC consultation and clinical breast examination immediately prior to mammography. Some previously screened women were scheduled only for mammography due to clinic capacity limitations. Eligible women were identified from the family history clinic database and from new referrals. They were sent an invitation letter and participant information leaflet with the clinic or mammogram appointment.

Centre B: Women were referred for mammography from a local genetics centre and several outside FHCs and entered onto the National Breast Screening Computer System (NBSS). Eligible women were identified from NBSS and sent an invitation letter and participant information leaflet with the mammogram appointment.

In both centres a daily list of women requiring mammography was generated in advance and potential recruits were assigned a code number. This number was used as the study number for those women who consented to participate. Written informed consent was taken by a radiographer prior to mammography.

Recruitment in Centre A was from April 2014 to August 2015 and in Centre B from September 2014 to November 2015. The differential start dates at the two centres was due to the different study set-up durations. Follow-up (second study round) mammography continued until October 2016 and November 2016 respectively.

Randomisation

Randomisation was performed after written informed consent using a commercial web-based randomisation service (www.sealedenvelope.com). The participant's study number

was the sole identifier. Randomisation was with random permuted blocks within the strata with stratification by study site. A mandatory checklist of exclusion criteria was completed before each randomisation could take place. The outcome of the randomisation was displayed on the web page and was manually recorded on the clinic list. It was also automatically sent by email to the randomising radiographer.

Women requiring a second mammogram within the study (one year after randomisation and the first study mammogram) were identified in advance on the daily mammography lists together with their assigned examination.

Each participant was registered on a bespoke web-based database using her study number only for identification. The date, participant's age at entry, trial centre, screen type (prevalent or incident), whether the woman was also undergoing annual screening MRI and the assigned first examination for randomised incident women were entered.

Interventions

Mammography was performed on Hologic Selenia Dimensions machines. Mediolateral oblique (MLO) and craniocaudal (CC) views of each breast were performed. Women underwent either 2D mammography alone or combined 2D+DBT under the same compression (so-called 'combo' mode). The imaging examinations were conducted by experienced specialist mammography radiographers who were fully trained in accordance with UK National Health Service Breast Screening Programme standards and who had additional specific training on the mammographic equipment and techniques used in the study.

Women undergoing prevalent screening underwent a single combined 2D+DBT examination. Those undergoing incident screening were randomised to undergo either 2D mammography alone or 2D+DBT as the first examination. Incident women undergoing a

second mammographic examination within the study a year later had the alternate examination (i.e. 2D only if they had previously had 2D+DBT, and vice versa). Once a mammographic examination was performed, examination details (including date, examination performed and number of DBT slices in each projection if applicable) were entered against the participant's entry on the database.

Images were viewed on either Sectra PACS workstations (Sectra AB, Linköping, Sweden) or Hologic SecurView workstations (Hologic Inc., Marlborough, MA, USA), both with 5MP dual monitors. All examinations were independently double read by two radiologists or a radiologist and a breast physician. All readers were of consultant or equivalent status with experience of mammography reading in the NHSBSP for a minimum of 2 years, reading a minimum of 5000 mammograms per annum and participating annually in the PERFORMS self-assessment test [10]. All readers underwent prior training in DBT image interpretation, either through participation in the TOMMY trial [11] or through attendance at an NHSBSP-approved DBT course.

The intention was for all the mammographic interpretations to be entered in real time by each reader in the on-line database, so that for women who had undergone a combined 2D+DBT examination, the 2D interpretation was recorded before the DBT examination was viewed. However, technical problems at Centre B precluded live data entry on a consistent basis, and therefore only the final recall / no recall decision is available for each participant from this centre. In Centre A, the location and description of any perceived abnormalities was recorded on the database together with the imaging score using the British Society of Breast Radiology (BSBR - previously the Royal College of Radiologists Breast Group) five-point classification system [12]. An overall normal or abnormal opinion of the combined examination was then recorded by each reader. Where there was disagreement in the final opinion, a consensus discussion took place between the readers, with arbitration by a third reader if necessary.

Women whose mammograms were reported as abnormal were recalled for further investigation in accordance with each centre's standard procedure. The findings at assessment were recorded, together with whether needle biopsy was performed. Histopathological details of cancers diagnosed were recorded.

Known interval cancers (those diagnosed in screened women between scheduled screening examinations) occurring within a year of a mammographic examination within the study were recorded.

Outcomes

The primary outcome measure for participants undergoing incident screening at trial entry was the recall rate for further assessment with 2D mammography and 2D+DBT respectively. For participants undergoing prevalent screening at trial entry, the planned primary outcome measure was the recall rate for further assessment.

Planned secondary outcome measures included cancer detection rates and individual reader recall rates. For prevalent screens it was intended to examine the number of women whose 2D mammograms were regarded as abnormal (recorded prior to viewing the DBT images) relative to the overall decision based on viewing the 2D+DBT images together.

Sample size

These were based on historic data from Centre A.

Incident screening: A 7% recall rate was anticipated for incident screening with FFDM alone, reducing to 4.9% with the addition of DBT. For 90% power to detect a difference of this magnitude as significant using McNemar or conditional logistic regression estimates (5% level, two-sided testing) would require 1305 women and 2610 screens.

Prevalent screening: In a two-year study, it was anticipated that 528 women having a prevalent screen would be recruited. An 11.5% recall rate was anticipated for prevalent screening with FFDM alone, reducing to 8% with the addition of DBT. These numbers were insufficient for a comparative trial with adequate power, hence a single arm study of combined 2D+DBT was performed. A study size of 528 would give approximately 80% power for a one-sample test of these recall rates with an expected confidence interval of less than 2.3% in either direction. In addition, the sequential reading of first the 2D images and then the 2Ds together with the DBTs would allow a direct comparison between the two techniques, although without the robustness of a randomised study.

Statistical methods

A generalised estimating equation regression model with an exchangeable correlation structure was used to compare recall and biopsy rates between 2D only and 2D+DBT examination strategies. This analysis accounted appropriately for the crossover design of the study, namely, each woman was examined using either 2D only or 2D+DBT at the first of two incident screening rounds and by the alternative examination strategy at the second incident screening round. Differences between sites (Centres A and B) and the effect of the order of the examination strategy were also assessed. Analysis was by intention to treat (i.e. by randomised examination strategy allocation).

RESULTS

Incident screening

A total of 1227 women were recruited and randomised, 608 to 2D mammography only at first screen and 619 to 2D+DBT. See Figure 1 for the CONSORT diagram. The mean age at recruitment was 44 years (range 40-49 years).

Technical failures occurred with the mammograms of four women who were excluded from further participation. Two women were randomised to 2D+DBT but incorrectly underwent 2D mammography only as the first study examination (both had 2D+DBT at the second screen). The results are therefore presented on an intention to treat basis. Two further women underwent 2D mammography at the first screen on another manufacturer's mammography equipment although both underwent a correctly performed 2D+DBT examination at the second screen. Their first screens were therefore excluded from analysis.

A second screen in the study was performed on 1124 women after a mean interval of 372 days (range 184 to 642) from the first. The reasons for women not undergoing a second screen were: cancer diagnosis at the first screen (5); diagnosis of an interval cancer between screens (1); cancelled appointment (5); did not attend appointment (21); declined further participation in the study (5); discharged from the family history clinic, care transferred or moved away (28); mammography performed elsewhere between planned study screens (12); incorrect examination performed (4); underwent risk-reducing surgery (1); underwent implant surgery (1); pregnant (3); died (1); not identified as a participant (12).

Cancer detection rates and false positive recall and biopsy rates

Table 1 shows the number of women screened, the number of cancers detected and the number of false positive recalls in each arm.

Eleven women were diagnosed with breast cancer as a result of a screening mammography examination performed in the study, six after 2D+DBT and five after 2D only. Two women were diagnosed with interval cancers 37 weeks and 3 weeks after a 2D+DBT study examination respectively.

There was no difference between the groups for either the overall recall rates or the false positive recall rates (2D 2.8% and 2.4%, 2D+DBT 2.7% and 2.2%; Table 2). There was a significant difference in false positive recall rates between sites ($p=0.003$) but no evidence of an order effect ($p=0.15$).

Fourteen women in Centre A underwent biopsy following FP recall: 7 after 2D only and 7 after 2D+DBT. The corresponding numbers for Centre B are 1 after 2D and 2 after 2D+DBT. There was no significant difference in biopsy rates between the screening modalities ($p=0.82$) or between the sites ($p=0.32$) and no evidence of an order effect ($p=0.21$).

Discordant examinations

For Centre A, data were available for the results of both readers. A discordant examination was defined as one or both readers recording an examination as abnormal but the decision was made not to recall the woman for further assessment following consensus discussion and/or arbitration by a third reader. The number of discordant examinations is shown in Table 3.

There was no significant difference between the discordant examination rates for 2D and 2D+DBT for either the first or the second round. The decrease in discordant examination rates between the first and second round was greater for 2D+DBT than for 2D only (interaction round vs examination strategy; $p=0.023$).

Prevalent screening

A total of 109 women were recruited, 50 in Centre A and 59 in Centre B. The mean age at recruitment was 43 years (range 40 – 49 years). One woman in Centre A was subsequently excluded due to a missing image. Table 4 shows the number of women recalled for further assessment. None of these women were diagnosed with breast cancer. No further analysis has been performed due to the low numbers.

DISCUSSION

There is now a substantial body of published literature which demonstrates that digital breast tomosynthesis has a higher sensitivity for cancer detection than conventional full field digital mammography alone [13] [14] [15] [16] [17] [18]. It appears to be particularly useful in women with moderately dense (BI-RADS C) breasts [19] in whom the sensitivity of 2D mammography is reduced by masking by normal breast tissue, although a higher sensitivity for cancer detection by DBT has been demonstrated even in women with fatty breasts [13]. This has led to the adoption of DBT for primary screening in many centres, particularly in the USA and some European countries. It seems unlikely, however, that the high invasive cancer detection rates seen with DBT in some studies (40% higher than 2D in the Oslo study [13]) will be sustained at subsequent screens as this would exceed the overall number of cancers (screen-detected plus interval) diagnosed in the screened population, even allowing for some overdiagnosis. This prevalent screen effect of supplemental imaging technologies is seldom discussed in the literature, and the long-term detection rates with repeated DBT screening remain to be seen.

To our knowledge, the study presented here is the first prospective randomised controlled trial comparing recall rates with 2D mammography alone and 2D plus DBT. It demonstrates that the addition of DBT does not result in a significant reduction in recall rate at incident screening, at least in specialist UK breast centres with low recall rates with 2D mammography. The participants were women in their forties whose breasts are relatively dense compared to those of older women [20] [21] [22], and although they are not typical in terms of their cancer risk it is reasonable to conclude that a similar study in older women in the UK population screening age group (50 to 70 years) would yield a similar result. Given the uncertainty surrounding the cancer detection rates of DBT at repeated screening and the substantially increased reading times relative to 2D mammography alone, this study does not provide support for its routine use in incident screening at present.

Our findings are at variance with a number of published studies, some of which showed higher [15] [17] and some lower [14] [16] [23] [24] [25] [26] [27] false positive rates with DBT compared to 2D mammography alone. In the Oslo study the false positive rate before arbitration was lower for 2D+DBT but the actual number of false positive recalls was higher [13]. Many of the studies that have shown a reduced recall rate with DBT had a higher recall rate with 2D mammography than ours [16] [23] [24] [26] [27] [28] and thus had a larger proportion of women who could potentially benefit. There are a number of factors which determine recall rate, including the prevailing culture within the institution and the training, experience and attitude of the individual readers. Depending on how it is used, double reading of screening examinations may increase or decrease false positive rates [29] [30]. Published studies of screening DBT have utilised various reading methods, including double reading by two radiologists [24], single reading with computed-aided detection (CAD) [16] [26] and, in a multicentre study of over 450,000 women, presumably either single reading or a mixture of methods [27]. Our system of consensus / arbitration (whereby women are not automatically recalled on the basis of the most suspicious opinion) may have contributed to keeping down the recall rate in both study arms.

We have demonstrated that increased reader uncertainty was caused by DBT in the first study screening round, although this reversed in the second round. This was despite the fact that all readers were trained in DBT and almost all had experience of using it routinely in screening assessment prior to the start of the study. It may be related to the need to develop confidence in dismissing asymmetric densities on their DBT appearances alone, as two studies have shown that resolution with DBT of asymmetric densities seen on 2D mammography were the biggest driver of the reduction in false positive recalls [16] [26] and interobserver agreement has been shown in one study to be higher with 2D+DBT than 2D mammography alone [28]. This 'learning curve' effect did not have a significant impact upon the false positive recall rate in our study due to the routine use of consensus discussion of

cases where the readers disagreed, together with arbitration by a third reader when necessary. However, it may have been a factor in those studies that showed an increased recall rate with DBT, and indeed the Malmo study (which did show a higher false positive recall rate with DBT) also found evidence of a screening DBT learning curve, with a progressive decline in the DBT false positive rate as the study progressed [31].

Although our study is relatively small compared to some, its strength lies in its randomised design. Most published studies have been cohort observational studies using an incremental reading strategy, whereby the 2D mammogram was read first followed by the DBT. In at least one [14], the reduction in false positive recalls was a theoretical one only, and the women were recalled anyway on the basis of an abnormal 2D mammogram. However, perhaps the most significant factor explaining the difference between the effect of DBT on recall rate in this study and most others is the failure of many published studies to differentiate between women undergoing a first (prevalent) screen and those undergoing a subsequent (incident) screen. These are fundamentally different types of screening, as the availability of previous mammograms for women undergoing incident screening results in a recall rate with conventional digital mammography in the UK National Health Service Breast Screening Programme that is less than half of that for women undergoing prevalent screening [6]. It is unfortunate that insufficient numbers of women were recruited to the prevalent screening arm in our study, although it is notable that even in the larger STORM-2 trial with 9672 participants [17] 93% of screens were incident, precluding separate analysis of prevalent screens. Women undergoing prevalent screening may be those most likely to benefit from the use of DBT.

Although most participants underwent two screening examinations within the study, only one of these included DBT. It is not known, therefore, whether the availability of a prior DBT screening examination (rather than just a prior 2D mammogram) would have reduced the false positive recalls in women undergoing 2D+DBT. To the best of our knowledge, this

limitation also applies to all of the other prospective screening DBT studies published to date.

The PROSPECTS trial (Prospective Randomised Trial of Digital Breast Tomosynthesis (DBT) Plus Standard 2D Digital Mammography (2DDM) or Synthetic 2D Digital Mammography (S 2D) Compared to Standard 2D Digital Mammography in Breast Cancer Screening) will commence shortly in the UK. This is a multicentre study which will randomise 100,000 women aged 50 to 70 undergoing population screening to either 2D mammography alone or 2D plus DBT (with synthetic 2D) over two screening rounds. The study will measure the cost effectiveness of screening with 2D+DBT compared to 2D mammography alone and will provide further data to address the key issue of the effect of DBT on recall rates in normal risk women undergoing prevalent and incident screening.

In conclusion, this randomised controlled trial performed in two specialist breast centres demonstrates that the addition of DBT to 2D mammography has no significant effect on the false positive recall rate in women in their forties with an increased risk of breast cancer undergoing incident screening when used in conjunction with double reading with robust consensus/arbitration. It does suggest that a learning curve exists, with some increased reader uncertainty resulting from DBT initially, but reversing with increasing experience. The potential remains for DBT to reduce the recall rate at prevalent screening, but this question has yet to be answered.

CONFLICT OF INTEREST STATEMENT

We confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

ACKNOWLEDGEMENTS

The following individuals contributed significantly to the conduct of this study:

Image reading - Ursula Beetles, Sara Bundred, Megan Bydder, Susan Dyson, David Evans, Soujanya Gadde, Jane Golliger, Anil Jain, Satchithananda Keshthra, Juliet Morel, Caroline Parkin, Clare Peacock, Vivien Phillips, Rumana Rahim, Rema Wasan.

Study administration – Jayne Beesley, Jill Fox.

The study was approved by the UK National Research Ethics Service (study reference 14/NW/0053) and was registered with the UK Clinical Research Network (UKCRN ID 16312) and the ISRCTN registry (ISRCTN37806452).

The study was funded by the Genesis Breast Cancer Prevention charity (now known as Prevent Breast Cancer) – grant number GA13-007. The funder had no involvement in the study design, the collection, analysis and interpretation of data, the writing of the report or in the decision to submit the article for publication.

The study was sponsored by the University Hospital of South Manchester NHS Foundation Trust.

The authors thank Dr Philip Tresadern for database and web interface development and management and the radiographic and family history clinic staff at both sites.

REFERENCES

- [1] F.J. Gilbert, L. Tucker, K.C. Young,¹; Digital breast tomosynthesis (DBT): a review of the evidence for use as a screening tool, *Clin Radiol*. 71 (2016) 141–150. doi:10.1016/j.crad.2015.11.008.
- [2] J. Brett, C. Bankhead, B. Henderson, E. Watson, J. Austoker,¹; The psychological impact of mammographic screening. A systematic review, *Psycho-Oncology*. 14 (2005) 917–938. doi:10.1002/pon.904.
- [3] A. Bolejko, P. Hagell, C. Wann-Hansson, S. Zackrisson,¹; Prevalence, Long-term Development, and Predictors of Psychosocial Consequences of False-Positive Mammography among Women Attending Population-Based Screening, *Cancer Epidemiol Biomarkers Prev*. 24 (2015) 1388–1397. doi:10.1158/1055-9965.EPI-15-0060.
- [4] H.D. Nelson, E.S. O'Meara, K. Kerlikowske, S. Balch, D. Miglioretti,¹; Factors Associated With Rates of False-Positive and False-Negative Results From Digital Mammography Screening: An Analysis of Registry Data, *Ann Intern Med*. 164 (2016) 226–235. doi:10.7326/M15-0971.
- [5] L.E. Johns, S.M. Moss,¹; False-positive results in the Randomized Controlled Trial of Mammographic Screening from Age 40 (“Age” trial), *Cancer Epidemiol Biomarkers Prev*. 19 (2010) 2758–2764. doi:10.1158/1055-9965.EPI-10-0623.
- [6] ¹; Health and Social Care Information Centre, Breast Screening Programme, England, Statistics for 2014-15, (2016) 1–60.
- [7] G. Rees, A. Fry, A. Cull, S. Sutton,¹; Illness perceptions and distress in women at increased risk of breast cancer, *Psychology & Health*. 19 (2004) 749–765. doi:10.1007/BF02895155.
- [8] A. McIntosh, C. Shaw, G. Evans, N. Turnbull, N. Bahar, M. Barclay, D. Easton, J. Emery, J. Gray, J. Halpin, P. Hopwood, J. McKay, C. Sheppard, M. Sibbering, W. Watson, A. Wailoo,¹; A. Hutchinson (2004 updated 2006 and 2013) Clinical Guidelines and Evidence Review for The Classification and Care of Women at Risk of Familial Breast Cancer, London: National Collaborating Centre for Primary Care/University of Sheffield. NICE guideline CG164. <https://www.nice.org.uk/Guidance/CG164>.
- [9] A. Maurice, D.G.R. Evans, A. Shenton, L. Ashcroft, A. Baidam, L. Barr, et al.,¹; Screening younger women with a family history of breast cancer – does early detection improve outcome? *Eur J Cancer*. 42 (2006) 1385–1390. doi:10.1016/j.ejca.2006.01.055.
- [10] H.J. Scott, A.G. Gale,¹; Breast screening: PERFORMS identifies key mammographic training needs, *Br J Radiol*. 79 Spec No 2 (2006) S127–33. doi:10.1259/bjr/25049149.
- [11] F.J. Gilbert, L. Tucker, M.G.C. Gillan, P. Willsher, J. Cooke, K.A. Duncan, et al.,¹; 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, *Health Technol Assess*. 19 (2015) i–xxv– 1–136. doi:10.3310/hta19040.
- [12] A.J. Maxwell, N.T. Ridley, G. Rubin, M.G. Wallis, F.J. Gilbert, M.J. Michell, et al.,¹; The Royal College of Radiologists Breast Group breast imaging classification, *Clin Radiol*. 64 (2009) 624–627. doi:10.1016/j.crad.2009.01.010.

- [13] P. Skaane, A.I. Bandos, R. Gullien, E.B. Eben, U. Ekseth, U. Haakenaasen, et al.,¹; Comparison of Digital Mammography Alone and Digital Mammography Plus Tomosynthesis in a Population-based Screening Program, *Radiology*. 267 (2013) 47–56. doi:10.1148/radiol.12121373.
- [14] S. Ciatto, N. Houssami, D. Bernardi, F. Caumo, M. Pellegrini, S. Brunelli, et al.,¹; Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study, *Lancet Oncol*. 14 (2013) 583–589. doi:10.1016/S1470-2045(13)70134-7.
- [15] K. Lång, I. Andersson, A. Rosso, A. Tingberg, P. Timberg, S. Zackrisson,¹; Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: results from the Malmö Breast Tomosynthesis Screening Trial, a population-based study, *Eur Radiol*. 26 (2016) 184–190. doi:10.1007/s00330-015-3803-3.
- [16] M.A. Durand, B.M. Haas, X. Yao, J.L. Geisel, M. Raghu, R.J. Hooley, et al.,¹; Early clinical experience with digital breast tomosynthesis for screening mammography, *Radiology*. 274 (2015) 85–92. doi:10.1148/radiol.14131319.
- [17] D. Bernardi, P. Macaskill, M. Pellegrini, M. Valentini, C. Fantò, L. Ostillio, et al.,¹; Breast cancer screening with tomosynthesis (3D mammography) with acquired or synthetic 2D mammography compared with 2D mammography alone (STORM-2): a population-based prospective study, *Lancet Oncol*. (2016). doi:10.1016/S1470-2045(16)30101-2.
- [18] A.S. Tagliafico, M. Calabrese, G. Mariscotti, M. Durando, S. Tosto, F. Monetti, et al.,¹; Adjunct Screening With Tomosynthesis or Ultrasound in Women With Mammography-Negative Dense Breasts: Interim Report of a Prospective Comparative Trial, *J Clin Oncol*. 34 (2016) 1882–1888. doi:10.1200/JCO.2015.63.4147.
- [19] E.A. Rafferty, M.A. Durand, E.F. Conant, D.S. Copit, S.M. Friedewald, D.M. Plecha, et al.,¹; Breast Cancer Screening Using Tomosynthesis and Digital Mammography in Dense and Nondense Breasts, *JAMA*. 315 (2016) 1784–1786. doi:10.1001/jama.2016.1708.
- [20] C.M. Checka, J.E. Chun, F.R. Schnabel, J. Lee, H. Toth,¹; The Relationship of Mammographic Density and Age: Implications for Breast Cancer Screening, *AJR Am J Roentgenol*. 198 (2012) W292–W295. doi:10.2214/AJR.10.6049.
- [21] P. Skippage, L. Wilkinson, S. Allen, N. Roche, M. Dowsett, R. a'Hern,¹; Correlation of Age and HRT Use with Breast Density as Assessed by Quantra™, *Breast J*. 19 (2012) 79–86. doi:10.1111/tbj.12046.
- [22] P.F. Johnson, J.C. Sergeant, M. Bydder, A. Maxwell, S. Whiteside, P. Stavrinou, et al.,¹; PB. 25: Relationship between volumetric breast density, age and hormonal factors, *Breast Cancer Res*. 15 (2013) Suppl 1 P25.
- [23] B.M. Haas, V. Kalra, J. Geisel, M. Raghu, M. Durand, L.E. Philpotts,¹; Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening, *Radiology*. 269 (2013) 694–700. doi:10.1148/radiol.13130307.
- [24] S.V. Destounis, A. Arieno, R. Morgan,¹; Initial Experience with Combination Digital Breast Tomosynthesis Plus Full Field Digital Mammography or Full Field Digital Mammography Alone in the Screening Environment, *J Clin Imaging Sci*. 4 (2014) 9. doi:10.4103/2156-7514.127838.

- [25] D. Bernardi, F. Caumo, P. Macaskill, S. Ciatto, M. Pellegrini, S. Brunelli, et al.,;1; Effect of integrating 3D-mammography (digital breast tomosynthesis) with 2D-mammography on radiologists' true-positive and false-positive detection in a population breast screening trial, *Eur J Cancer*. 50 (2014) 1232–1238. doi:10.1016/j.ejca.2014.02.004.
- [26] A.P. Lourenco, M. Barry-Brooks, G.L. Baird, A. Tuttle, M.B. Mainiero,;1; Changes in Recall Type and Patient Treatment Following Implementation of Screening Digital Breast Tomosynthesis, *Radiology*. 274 (2015) 337–342. doi:10.1148/radiol.14140317.
- [27] S.M. Friedewald, E.A. Rafferty, S.L. Rose, M.A. Durand, D.M. Plecha, J.S. Greenberg, et al.,;1; Breast Cancer Screening Using Tomosynthesis in Combination With Digital Mammography, *JAMA*. 311 (2014) 2499–2507. doi:10.1001/jama.2014.6095.
- [28] L.A. Carbonaro, G. Di Leo, P. Clauser, R.M. Trimboli, N. Verardi, M.P. Fedeli, et al.,;1; Impact on the recall rate of digital breast tomosynthesis as an adjunct to digital mammography in the screening setting. A double reading experience and review of the literature, *Eur J Radiol*. 85 (2016) 808–814. doi:10.1016/j.ejrad.2016.01.004.
- [29] E.D.C. Anderson, B.B. Muir, J.S. Walsh, A.E. Kirkpatrick,;1; The efficacy of double reading mammograms in breast screening, *Clin Radiol*. 49 (1994) 248–251.
- [30] S. Ciatto, D. Ambrogetti, R. Bonardi, S. Catarzi, G. Risso, M. Rosselli Del Turco, et al.,;1; Second reading of screening mammograms increases cancer detection and recall rates. Results in the Florence screening programme, *J Med Screen*. 12 (2005) 103–106. doi:10.1258/0969141053908285.
- [31] K. Lång, M. Nergården, I. Andersson, A. Rosso, S. Zackrisson,;1; False positives in breast cancer screening with one-view breast tomosynthesis: An analysis of findings leading to recall, work-up and biopsy rates in the Malmö Breast Tomosynthesis Screening Trial, *Eur Radiol*. 26 (2016) 3899–3907. doi:10.1002/ijc.29593.

Figure Legend

Figure 1. CONSORT flow diagram

TABLES

Table 1. Number of women in each centre who underwent incident screening with each modality (by intention to treat) and number who underwent false positive (FP) recall in each of the two screening rounds within the study together with numbers of cancers detected.

	Round	Modality					
		2D only			2D+DBT		
		Screened	FP recall	Cancer	Screened	FP recall	Cancer
Centre A	1	428	19	0	438	14	4
	2	401	7	3	401	9	1
	Both	829	26	3	839	23	5
Centre B	1	176	1	1	179	0	0
	2	165	1	1	157	3	1
	Both	341	2	2	336	3	1
Both centres	Both	1170	28	5	1175	26	6

Table 2. Percentage of women undergoing false positive recall in each of the centres and overall, including 95% confidence intervals.

Percentage (95% CI)			
	2D only	2D + DBT	p value
Centre A	3.1% (2.2%, 4.6%)	2.7% (1.8%, 4.1%)	0.63
Centre B	0.6% (0.2%, 2.3%)	0.9% (0.3%, 2.7%)	0.64
Overall	2.4% (1.7%, 3.4%)	2.2% (1.9%, 3.8%)	0.89

Table 3. Number of examinations reported abnormal by one or both readers but the woman was not recalled for assessment following consensus/arbitration.

	Number of discordant examinations (Percentage; 95% CI)		p value
	2D only	2D+DBT	
First study	14/428	22/438	p=0.20
screen	(3.3%; 2.0% - 5.4%)	(5.0%; 3.3% - 7.5%)	
Second study	12/401	8/401	p=0.36
screen	(3.0%; 1.7% - 5.2%)	(2.0%; 1.0% - 3.9%)	

Table 4. Number of women in each centre who underwent prevalent screening and number recalled for further assessment.

	Number screened	Number recalled	95% CI
Centre A	49	4 (8.2%)	3.2% - 19.2%
Centre B	59	3 (5.1%)	1.7% - 13.9%
Overall	108	7 (6.5%)	3.2% - 12.8%