



# Virtual clinical trial to compare cancer detection using combinations of 2D mammography, digital breast tomosynthesis and synthetic 2D imaging

Alistair Mackenzie<sup>1</sup> · Emma L. Thomson<sup>1,2</sup> · Melissa Mitchell<sup>1,2</sup> · Premkumar Elangovan<sup>1</sup> · Chantal van Ongeval<sup>3</sup> · Lesley Cockmartin<sup>4</sup> · Lucy M. Warren<sup>1</sup> · Louise S. Wilkinson<sup>5</sup> · Matthew G. Wallis<sup>6</sup> · Rosalind M. Given-Wilson<sup>7</sup> · David R. Dance<sup>1,2</sup> · Kenneth C. Young<sup>1,2</sup>

Received: 18 March 2021 / Revised: 7 June 2021 / Accepted: 1 July 2021  
© European Society of Radiology 2021

## Abstract

**Objectives** This study was designed to compare the detection of subtle lesions (calcification clusters or masses) when using the combination of digital breast tomosynthesis (DBT) and synthetic mammography (SM) with digital mammography (DM) alone or combined with DBT.

**Methods** A set of 166 cases without cancer was acquired on a DBT mammography system. Realistic subtle calcification clusters and masses in the DM images and DBT planes were digitally inserted into 104 of the acquired cases. Three study arms were created: DM alone, DM with DBT and SM with DBT. Five mammographic readers located the centre of any lesion within the images that should be recalled for further investigation and graded their suspiciousness. A JAFROC figure of merit (FoM) and lesion detection fraction (LDF) were calculated for each study arm. The visibility of the lesions in the DBT images was compared with SM and DM images.

**Results** For calcification clusters, there were no significant differences ( $p > 0.075$ ) in FoM or LDF. For masses, the FoM and LDF were significantly improved in the arms using DBT compared to DM alone ( $p < 0.001$ ). On average, both calcification clusters and masses were more visible on DBT than on DM and SM images.

**Conclusions** This study demonstrated that masses were detected better with DBT than with DM alone and there was no significant difference ( $p = 0.075$ ) in LDF between DM&DBT and SM&DBT for calcifications clusters. Our results support previous studies that it may be acceptable to not acquire digital mammography alongside tomosynthesis for subtle calcification clusters and ill-defined masses.

## Key Points

- The detection of masses was significantly better using DBT than with digital mammography alone.
- The detection of calcification clusters was not significantly different between digital mammography and synthetic 2D images combined with tomosynthesis.
- Our results support previous studies that it may be acceptable to not acquire digital mammography alongside tomosynthesis for subtle calcification clusters and ill-defined masses for the imaging technology used.

**Keywords** Screening · Mammography · Cancer

✉ Alistair Mackenzie  
alistairmackenzie@nhs.net

<sup>1</sup> National Coordinating Centre for the Physics in Mammography, Royal Surrey NHS Foundation Trust, Guildford, UK

<sup>2</sup> Department of Physics, University of Surrey, Guildford, UK

<sup>3</sup> Department of Radiology, UZ Leuven, Herestraat 49, B-3000 Leuven, Belgium

<sup>4</sup> Department of Imaging and Pathology, Division of Medical Physics and Quality Assessment, KU Leuven, Herestraat 49, B-3000 Leuven, Belgium

<sup>5</sup> Oxford Breast Imaging Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

<sup>6</sup> Cambridge Breast Unit, Cambridge University Hospitals NHS Foundation Trust, Cambridge & NIHR Cambridge Biomedical Research Centre, Cambridge, UK

<sup>7</sup> Department of Radiology, St George's Healthcare NHS Trust, London, UK

## Abbreviations

DF	Degrees of freedom
DM	Digital mammography
FoM	Figure of merit
FRF	False recall fraction
JAFROC	Jackknife alternative free-response receiver operating characteristics
LDF	Lesion detection fraction
MGD	Mean glandular dose
SM	Synthetic 2D images

## Introduction

Several clinical trials have investigated if synthetic 2D image (SM) (created from the tomosynthesis images) can be used instead of acquiring 2D digital mammograms (DM) in combination with digital breast tomosynthesis (DBT) [1–7]. This would reduce the radiation dose received by women, but could affect cancer detection rate. The SM images are likely to be technically inferior due to extra geometric blurring during the continuous acquisition of projection images under different angles or due to electronic noise accumulating from multiple low-dose projection images [8]. Clinical trials have not found any statistical difference in cancer detection between using DM or SM combined with DBT [9–11]. However, such clinical trials have potential issues with achieving sufficient numbers of subtle cases and can have many confounding factors, such as different women in the study arms, different imaging systems and image readers, which may result in a different distribution of lesions in the study arms. It may be that only relatively large differences could be found to be statistically significant, even with large numbers of women. Another approach is to use validated virtual clinical trials (VCTs) that include the insertion of simulated lesions into images (either mammograms or images of breast phantoms) which are then interpreted by experienced mammography readers. Such studies reduce the number of variables and can find statistically significant and valuable results over a reduced time-frame and allow easier stratification by lesion type [12].

This study was designed to assess the performance of the combination of DBT and SM on the detection of realistic subtle lesions (calcification clusters and masses) compared to that of DM alone or combined with DBT. The use of subtle lesions improves the sensitivity of the study, which enables the detection of smaller differences than would be possible in all but the largest standard clinical trials.

## Materials and methods

Ethical approval was obtained for this study as part of the OPTIMAM project as well as local ethical committee approval for the retrospective collection of the cases at the test site.

### Image collection

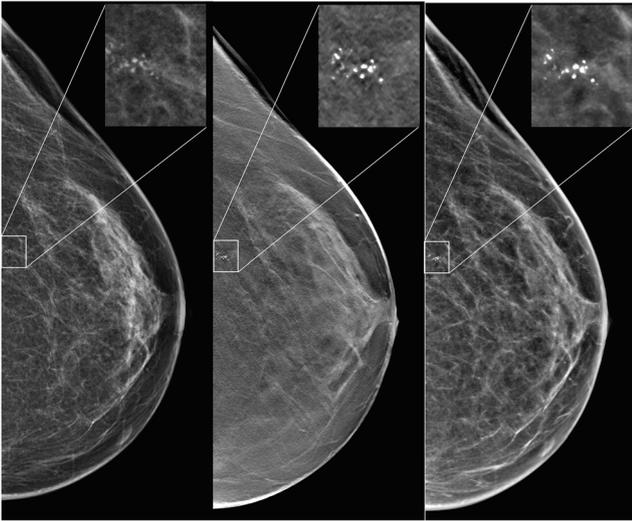
The images used in this study were collected from a Siemens Inspiration mammography system (Siemens Healthcare) at UZ Leuven (UZL), Belgium.

A set of 300 cases recalled for a suspected cancer in one breast with images acquired between June 2017 and October 2017 was identified. The images of the contralateral breast were used for the study. An expert radiologist, with 27 years of mammography reading experience, assessed the images to confirm that there were no features of either malignant or benign lesions. Images with memorable features such as distinctive calcifications or clips were also excluded from the dataset. A dataset of 166 cases, using a single view (either cranio-caudal or medio-lateral oblique), was randomly selected for the study and included breasts of all BI-RADS density categories. From follow-up, 160 of the cases had no new cancer diagnosis in the breast used within 2 years, while 6 cases did not have any follow-up.

### Insertion of calcification clusters and masses in images

Models of calcification clusters were created from images of calcifications segmented from micro-CT images of biopsy samples [13]. Voxel models of simulated round masses with ill-defined borders were created using a fractal growth method known as diffusion limited aggregation, where they were shown to be indistinguishable from real masses [14]. The mass diameter ranged from 4 to 8 mm. The insertion sites within the breasts were randomly chosen and covered the complete breast volume. There were limitations applied such as insertions were more than 5 mm from the skin line and at a height of between a fifth and four-fifths of the compressed breast thickness.

A breast density map was acquired for each image using Volpara software (version 1.5.4, Volpara Health Technologies Ltd.) [15]. The average density at the location of lesion insertion was estimated. This ensured that the lesion was inserted with the correct contrast. The simulated lesions were inserted into the unprocessed DM and DBT projection images using validated methods [16]. Registration of the DM images and the central DBT projection images was undertaken, to insert the lesions in the corresponding locations in the SM and DM images. The DBT series, with 1 mm separation between the planes, and synthetic 2D images were reconstructed from these projection images. Examples of the appearance of a calcification cluster and mass in DM, DBT and SM images are shown in Figs. 1 and



**Fig. 1** Extracted calcification cluster used in this study: DM (left), one DBT plane (middle), synthetic 2D (right)

2. Only images that were considered realistic by an expert radiologist were included in the study.

## Reader study

The study dataset comprised:

- 62 normal cases
- 64 cases with simulated calcification clusters with a total of 80 calcification clusters inserted
- 40 cases, each with one simulated ill-defined mass

A reader study was conducted using the full study dataset based on imaging modality combinations:

- Arm D: DM only
- Arm D&T: DM and DBT displayed simultaneously
- Arm S&T: SM and DBT displayed simultaneously.

**Fig. 2** Extracted ill-defined mass used in this study: DM (left), one DBT plane (middle), synthetic 2D (right)



The 166 cases in the three study arms were then randomly arranged into 18 sets, such that each set contained cases from the three study arms. All images were viewed by five readers experienced in reporting screening mammograms and DBT images acquired on Siemens systems. The readers had between 1 and 21 years' experience in reading breast images and between 1 and 5 years' experience in reading DBT cases and each reading over 3500 screening mammography cases annually. Readers were recruited from St George's Hospital, London and UZL, Leuven. The cases were shown to the readers using the MedXViewer software [17] on a workstation with two 5 megapixel reporting monitors (Barco MDMG (Barco NV)).

The readers viewed the image sets consecutively, but in a different order from the other readers. To reduce memory effects, the MedXViewer software enforced a minimum of 14 days between viewing images of the same case in a different study arm. Readers could use panning and magnification facilities on the MedXViewer software. Initially, the images were scaled to fit the monitor and the reader could scroll through the DBT stack of individual planes.

The task of the readers was to mark the centre of any lesion within the images that should be recalled for further investigation. The readers were not made aware of the design of the study. Readers marked the lesion in the DBT plane with the highest lesion conspicuity. The readers had a choice of two images to mark in study arms D&T and S&T; in those cases, they were asked to mark the lesion location on the image in which the lesion was most clearly seen. Following a mark, the reader was asked:

- Question 1: What type of lesion is this? Calcification cluster; mass; calcification cluster + mass
- Question 2: What is your level of confidence that this lesion should be recalled? 1–100%
- Question 3: How visible is the lesion in the other image? Not visible; less visible; similar visibility

The coordinates of the marks made by the readers and their answers to the questions were automatically saved by the MedXViewer software.

### Data analysis of reader study

The region of interest (ROI) for each lesion was defined as the margin of the inserted lesion. A mark was classed as a lesion localisation when it was inside a ROI and the lesion type matched the response to question 1. If a mark was not within a ROI or the lesion type was misidentified, it was classified as a non-lesion localisation. JAFROC-4.2 software ([www.devchakraborty.com](http://www.devchakraborty.com)) [18] was used to evaluate the marks. The calculation of jackknife alternative free-response receiver operating curves (JAFROC) was undertaken for random readers and random cases; i.e. the readers and cases were thought to be representative of their respective populations. A global *F*-test using the null hypothesis that all the arms are equal was included in the JAFROC testing. The equally weighted JAFROC figure of merit (FoM) was used to estimate the overall effectiveness of each imaging modality. The statistical difference between the FoMs was undertaken using the Dorfman-Berbaum-Metz analysis of variance [19].

The lesion detection fraction (LDF) was computed for calcification clusters and masses separately from the lesion localisation fraction. The false recall fraction (FRF) was computed from the non-localisation fraction of marks made in cases that did not have an inserted lesion.

The results for question 3 were collated for the D&T and S&T arms. The 3-point score of the comparative visibility of the lesions between 2D and DBT was converted to a 5-point scale of 'Not visible in DBT', 'Less visible in DBT', 'Similar', 'More visible in DBT' and 'Only visible in DBT'. The conversion depended on the score and if the 2D (DM or SM) or DBT image was marked.

### Results

The alternative free-response receiver operating characteristic (AFROC) curves for the combined scores of the readers for calcification clusters and masses are shown in Fig. 3a and b, while Fig. 3c and d show their equivalent JAFROC FoM. The global *F*-test of the FoM produced *F*-values equal to 1.97 (degrees of freedom (DF) of 2 and 79) and 30.7 (DF of 2 and 39) for the calcification clusters and masses respectively. The resultant *p* values between the study arms were 0.18 and < 0.0001 for calcification clusters and masses respectively; i.e. the difference was only statistically significant for the masses. Figure 3e and f show the JAFROC FoM for each reader; although the absolute values vary between readers, broadly they produce the same comparison, with the exception of

reader 1 who found arms D and D&T significantly better than S&T ( $p \leq 0.0008$ ).

The LDFs are shown in Fig. 4. The highest LDFs were 43% (D&T) and 59% (S&T) for calcification clusters and masses respectively. Sixty-eight of the 80 clusters were marked by at least one reader and 37 of 40 masses in at least one of the study arms. Only one of 80 calcification clusters and one of 40 masses were marked by all readers in the three study arms. This indicates that the lesions included in the study were very subtle. The percentage differences between the JAFROC FoMs and LDFs are shown in Tables 1 and 2. No significant differences were found for the detection of calcification clusters. The detection of masses in DM images alone was significantly worse than either of the arms which included DBT.

Figure 5 shows the FRF for non-lesion localisation marks of calcification clusters and masses in the normal images. The only significant differences shown in Tables 1 and 2 were that the D arm has a significantly lower FRF for masses than the other two study arms ( $p < 0.0012$ ).

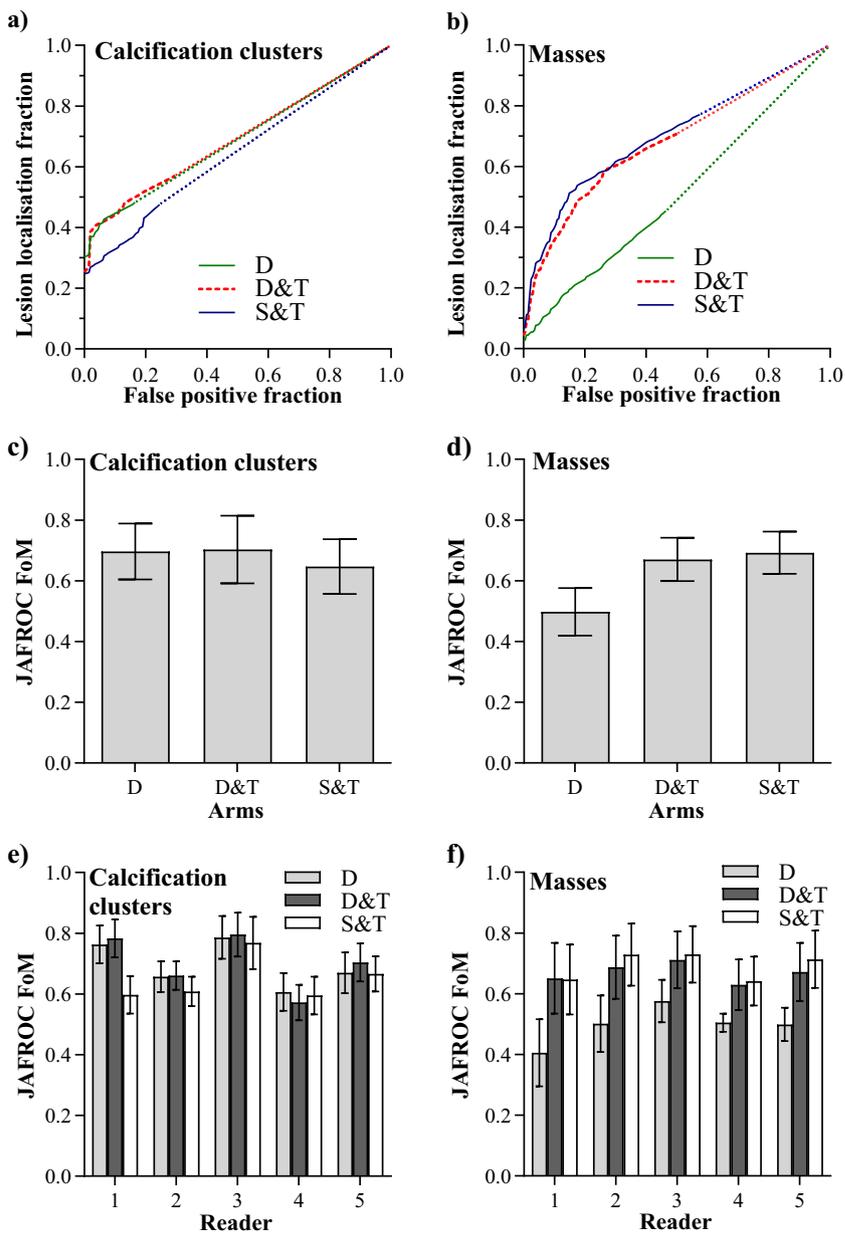
Figure 6 shows the visibility of the lesions in the DM and SM images compared to the DBT images. For calcification clusters, the most common selected option was that the clusters in DM and SM images were similar in visibility to the DBT image. The next most popular option was that the clusters were more easily seen in the DBT. Calcifications clusters were marked as only visible on DM in 12/80 clusters, and marked as only visible on DBT in 6/80 clusters. For masses, there was a clear preference for DBT over DM and SM images, and there were no masses only seen in DM or SM images but 18/40 cases where the mass was only visible in DBT.

### Discussion

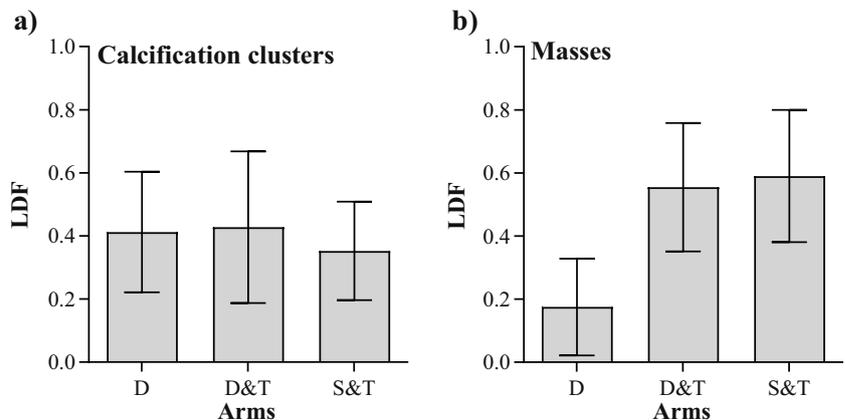
This VCT was designed to compare detection of subtle lesions between different combinations of mammographic imaging techniques. All of the lesions used in this study were visible in at least one of the image types but not all of the lesions were recalled as they were often on the cusp of the recall/not recall decision. By concentrating on subtle cases, this VCT avoids the need for very large numbers of cases to demonstrate the difference in cancer detection rate using combinations of DM, SM and DBT. This study has shown that the detection of masses was significantly improved with DBT over DM alone. In addition, lesion detection with SM combined with DBT was not significantly lower than that with DM and DBT for calcification clusters.

We used realistic small masses that were difficult to detect on DM imaging. In common with other publications [6, 20, 21], there was a clear difference in the detection of masses between DM alone and when DBT was included. The visibility of masses is clearer with DBT due to the reduction in the

**Fig. 3** Reader-averaged AFROC curves for (a) calcification clusters and (b) masses. Solid and dashed lines = measured data, dotted lines = straight line extension to (1,1) defining the area under the curve which is included in the FoM. c, d JAFROC FoMs averaged for all readers for each arm combination. e, f JAFROC FoMs for each reader and arm combination. Error bars indicate 95% confidence intervals



**Fig. 4** Lesion detection fraction (LDF) for lesion localisation marks for (a) calcification clusters and (b) masses. Error bars indicate 95% confidence intervals



**Table 1** Calcification clusters: the percentage difference in JAFROC-weighted FoM, lesion detection fraction (LDF) and false recall fraction (FRF). *p* values are given in parentheses

Difference between arms	FoM	LDF	FRF
D&T – D	0.9% (0.84)	3.6% (0.72)	30% (0.61)
S&T – D	– 7.1% (0.13)	– 15% (0.15)	20% (0.73)
S&T – D&T	– 8.0% (0.09)	– 18% (0.075)	– 7.7% (0.86)

overlap of the background breast parenchyma. The use of DBT greatly increased the detection of masses and in this case the combination of DM imaging with DBT did not improve mass detection compared with DBT with SM.

The use of imaging technology will not only result in the detection of malignant masses but also allow the detection of benign masses that can appear similar to malignant masses. This study was focussed on testing the detection of small ill-defined masses, rather than distinguishing between malignant and benign lesions. A significant increase ( $p \leq 0.0012$ ) in the FRF was found for the two study arms including DBT images compared to digital mammography alone. A clinical study would be required to evaluate the overall impact of DBT on detection of extra cancers and over-investigation.

Four readers found no difference in the detection of calcification clusters when SM images were used instead of DM images alongside DBT, while one reader found that including DM images significantly improved detection ( $p \leq 0.0008$ ). A larger study is needed to demonstrate if the average reduction in detection becomes significant for these subtle and small calcification clusters. Most malignant calcification clusters can be in situ or invasive cancers but more often represent ductal carcinoma in situ (DCIS). The detection of low-grade DCIS has been associated with overtreatment. It has been shown that the use of DBT can increase this over-diagnosis [22].

Consistent with other works, we found that the detection of lesions increased with the use of DBT compared to DM alone due to the extra masses detected, while there was little or no difference in the detection of calcification cluster rates [6, 21, 23, 24]. Houssami [9] reviewed five studies [2–7] comparing DBT&DM and DBT&SM and concluded that in population screening cancer detection rates do not differ between DBT&DM (range, 5.45–8.5/1000 screens) and DBT&SM (range, 5.03–8.8/1000 screens). She found that there was a wide range in recall rates between the studies but only small differences were found between the modalities. Houssami

concluded that it was safe to transition to only acquiring the DBT images without the DM images. However, on close examination of the papers, four of the five studies showed a non-significant decrease in cancer detection rate for SM&DBT vs DM&DBT. She also reported on studies using enhanced data sets [25–27], and three of the studies showed non-significant lower scores in the comparison of SM&DBT against 2D&DBT. Martín et al [10] showed a non-significant decrease of 6.7% in cancer detection when SM replaced DM imaging. This indicates that small differences in lesion detection using SM rather than DM with DBT are difficult to prove with a clinical trial. Our study only used subtle cases and has also shown a non-significant drop in lesion detection for calcification clusters.

Our study included a comparison of the visibility of the calcification clusters between DBT and DM images. There is evidence that the smallest calcifications within a cluster that are visible on DM may not be visible in DBT or SM images [28, 29]. The SM images are constructed from DBT and suspicious regions may be enhanced in the SM image compared with DM image. Calcification clusters typically contain a range of calcification sizes and even if some calcifications are not visible within a cluster using DBT, the larger calcifications may be enhanced and aid detection [30]. Our work showed that the visibility of calcification clusters was considered similar between DBT and DM images, but also that some subtle clusters were more visible in DBT than DM images. This is consistent with Korhonen et al [31]. Many publications [9, 32] support the idea of not acquiring a DM image in conjunction with DBT images, which has the advantage of reducing the radiation dose of the examination. However, there is potentially a loss in detection of sub-types of lesions, e.g. calcification clusters [3]; the results of our study show that this may not be the case for the technology used here.

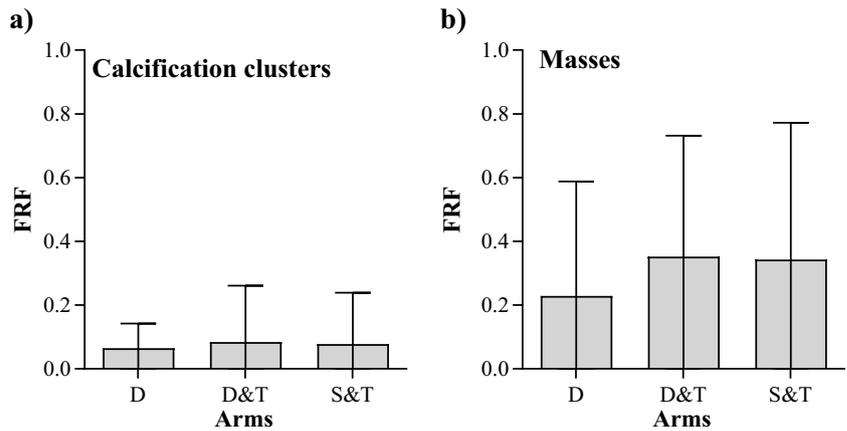
The use of VCTs can aid faster investigations into specific areas of interest and more valid comparison between different

**Table 2** Masses: the percentage difference in JAFROC-weighted FoM, LDF and FRF relative to the second arm in the first column. *p* values are given in parentheses

Difference between arms	FoM	LDF	FRF
D&T – D	35% (< 0.0001*)	210% (< 0.0001*)	54.2% (0.0005*)
S&T – D	39% (< 0.0001*)	240% (< 0.0001*)	50% (0.0012*)
S&T – D&T	3.0% (0.46)	6.3% (0.43)	– 2.7% (0.77)

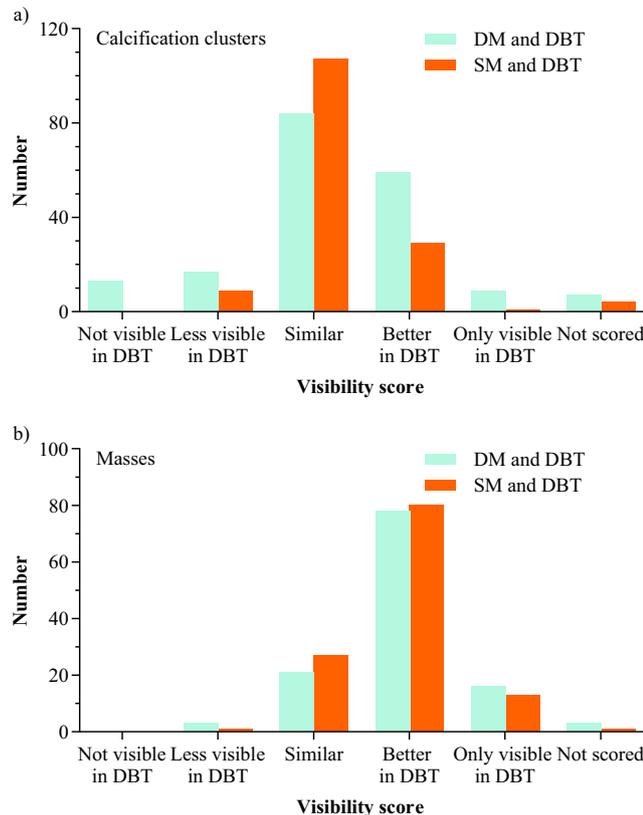
\*Significant difference  $p < 0.05$

**Fig. 5** False recall fraction (FRF) for all non-lesion localisation marks in cases without lesions. **a** Calcification clusters. **b** Masses. Error bars indicate 95% confidence intervals



imaging modalities, and allow deeper investigation into the nuances of image interpretation than conventional clinical trials. However, a VCT should not be considered directly comparable with a clinical trial, rather as adding sophisticated information that aids interpretation of clinical trials. Perception and interpretation are both important initial steps in the assessment of a mammogram/DBT, which may be modified by comparison with priors and the opposite breast. As we have used one breast and one view with no priors, then we only examined this first

stage. This study examines the detection of lesions and did not include benign lesions to examine the discriminatory power of the different study arms. Although the lesions were simulated, the process of simulating and inserting these lesions has been validated [16]. In addition, all of the lesions were checked for realism by an expert radiologist. This study is limited to one imaging system, as the appearance of the lesion may have been different with other systems, due to the scan angle, the DBT reconstruction algorithm [33], 2D post-processing [34, 35] and SM algorithm.



**Fig. 6** Total number of marks of classification of visibility of lesions in 2D images (DM or SM) compared to DBT in arms D&T and S&T. **a** Calcification clusters. **b** Masses

This study has examined the effect of using different types of mammographic images on lesion detection. No significant differences in the detection of calcification clusters were found between the arms of the study, but there was significantly better detection of masses when DBT was included. This work indicates that the acquisition of a DM image alongside a DBT image may not be necessary. Using a study with only subtle lesions, our results support the results of many previous studies that it may be acceptable to not acquire digital mammography alongside tomosynthesis for subtle calcification clusters and ill-defined masses.

**Acknowledgements** We thank UZ Leuven for the use of the images. We thank the observers for reading the images in this study. We thank Volpara Inc. for the use of their software. Ethical approval was obtained for this study as part of the OPTIMAM project as well as local ethical committee approval for the retrospective collection of the cases at the test site.

**Funding** This study has received funding from the Cancer Research UK: OPTIMAM2 project (grant number: C30682/A17321).

**Declarations**

**Guarantor** The scientific guarantor of this publication is Prof. Kenneth C. Young (ken.young@nhs.net).

**Conflict of interest** The authors of this manuscript declare relationships with the following companies:

Chantal Van Ongeval: research and travel agreements with Siemens Healthineers; research agreement with GE Healthcare.

Lesley Cockmartin's lab has research agreements with Siemens Healthineers and GE Healthcare.

Matthew Wallis: This research was supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

**Statistics and biometry** Two of the authors (LMW, AM) have significant statistical expertise for this type of study.

**Informed consent** Written informed consent was waived by the Institutional Review Board.

**Ethical approval** Institutional Review Board approval was obtained.

**Study subjects or cohorts overlap** Five cases were used in a previous publication (<https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10952/109520U/An-observer-study-to-assess-the-detection-of-calcification-clusters/10.1117/12.2506895.full>). In those cases, either the images used were a different view or a different lesion was inserted.

#### Methodology

- not applicable (prospective/retrospective)
- experimental
- multicentre study

## References

- Wallis MG, Moa E, Zanca F, Leifland K, Danielsson K (2012) Two-view and single-view tomosynthesis versus full-field digital mammography: high-resolution X-ray imaging observer study. *Radiology* 262:788–796. <https://doi.org/10.1148/radiol.11103514>
- Aujero MP, Gavenonis SC, Benjamin R, Zhang Z, Holt JS (2017) Clinical performance of synthesized two-dimensional mammography combined with tomosynthesis in a large screening population. *Radiology* 283:70–76. <https://doi.org/10.1148/radiol.2017162674>
- Freer PE, Riegert J, Eisenmenger L et al (2017) Clinical implementation of synthesized mammography with digital breast tomosynthesis in a routine clinical practice. *Breast Cancer Res Treat* 166:501–509. <https://doi.org/10.1007/s10549-017-4431-1>
- Bernardi D, Macaskill P, Pellegrini M et al (2016) 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* 17:1105–1113. [https://doi.org/10.1016/S1470-2045\(16\)30101-2](https://doi.org/10.1016/S1470-2045(16)30101-2)
- Zuckerman SP, Conant EF, Keller BM et al (2016) Implementation of synthesized two-dimensional mammography in a population-based digital breast tomosynthesis screening program. *Radiology* 281:730–736. <https://doi.org/10.1148/radiol.2016160366>
- Skaane P, Sebuodegård S, Bandos AI et al (2018) Performance of breast cancer screening using digital breast tomosynthesis: results from the prospective population-based Oslo Tomosynthesis Screening Trial. *Breast Cancer Res Treat* 169:489–496. <https://doi.org/10.1007/s10549-018-4705-2>
- Skaane P, Bandos AI, Eben EB et al (2014) Two-view digital breast tomosynthesis screening with synthetically reconstructed projection images: comparison with digital breast tomosynthesis with full-field digital mammographic images. *Radiology* 271:655–663. <https://doi.org/10.1148/radiol.13131391>
- Mackenzie A, Marshall NW, Hadjipanteli A, David R Dance, Bosmans H, Young KC (2017) Characterisation of noise and sharpness of images from four digital breast tomosynthesis systems for simulation of images for virtual clinical trials. *Phys Med Biol* 62:2376–2397. <https://doi.org/10.1088/1361-6560/aa5dd9>
- Houssami N (2017) Evidence on synthesized two-dimensional mammography versus digital mammography when using tomosynthesis (three-dimensional mammography) for population breast cancer screening. *Clin Breast Cancer* 18:255–260. <https://doi.org/10.1016/j.clbc.2017.09.012>
- Romero Martín S, Raya Povedano JL, Cara García M, Romero ALS, Garriguet MP, Benito MA (2018) Prospective study aiming to compare 2D mammography and tomosynthesis + synthesized mammography in terms of cancer detection and recall. From double reading of 2D mammography to single reading of tomosynthesis. *Eur Radiol* 28:2484–2491. <https://doi.org/10.1007/s00330-017-5219-8>
- Abdullah P, Alabousi M, Ramadan S et al (2020) Synthetic 2D mammography versus standard 2D digital mammography: a diagnostic test accuracy systematic review and meta-analysis. *AJR Am J Roentgenol*. <https://doi.org/10.2214/AJR.20.24204>
- Bakic PR, Myers KJ, Glick SJ, Maidment ADA (2016) Virtual tools for the evaluation of breast imaging: state-of-the science and future directions. In: Tingberg A, Lång K, Timberg P (eds) *Breast Imaging*. IWDW 2016 Lecture Notes in Computer Science 9699: 518–524. [https://doi.org/10.1007/978-3-319-41546-8\\_65](https://doi.org/10.1007/978-3-319-41546-8_65)
- Shaheen E, Van Ongeval C, Zanca F et al (2011) The simulation of 3D microcalcification clusters in 2D digital mammography and breast tomosynthesis. *Med Phys* 38:6659–6671. <https://doi.org/10.1118/1.3662868>
- Rashidnasab A, Elangovan P, Yip M et al (2013) Simulation and assessment of realistic breast lesions using fractal growth models. *Phys Med Biol* 58:5613–5627. <https://doi.org/10.1088/0031-9155/58/16/5613>
- García E, Diaz O, Martí R et al (2017) Local breast density assessment using reacquired mammographic images. *Eur J Radiol* 93:121–127. <https://doi.org/10.1016/j.ejrad.2017.05.033>
- Elangovan P, Mackenzie A, Warren LM et al (2019) Validation of modelling tools for simulating wide-angle DBT systems. In: Bosmans H, Chen G-H, Gilat Schmidt T (eds) *Proc.SPIE Medical Imaging*. SPIE, pp 109482E-1–109482E10
- Looney PT, Young KC, Halling-Brown MD (2015) MedXViewer: providing a web-enabled workstation environment for collaborative and remote medical imaging viewing, perception studies and reader training. *Radiat Prot Dosimetry* 169:32–37. <https://doi.org/10.1093/rpd/ncv482>
- Chakraborty DP, Berbaum KS (2004) Observer studies involving detection and localization: modelling, analysis and validation. *Med Phys* 31:2313–2330. <https://doi.org/10.1118/1.1769352>
- Dorfman DD, Berbaum KS, Metz CE (1992) Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method. *Invest Radiol* 27:723–731. <https://doi.org/10.1097/00004424-199209000-00015>
- Ciatto S, Houssami N, Bernardi D et al (2013) Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *Lancet Oncol* 14:583–589. [https://doi.org/10.1016/S1470-2045\(13\)70134-7](https://doi.org/10.1016/S1470-2045(13)70134-7)
- Zackrisson S, Lång K, Rosso A et al (2018) One-view breast tomosynthesis versus two-view mammography in the Malmö Breast Tomosynthesis Screening Trial (MBTST): a prospective, population-based, diagnostic accuracy study. *Lancet Oncol* 19:1493–1503. [https://doi.org/10.1016/S1470-2045\(18\)30521-7](https://doi.org/10.1016/S1470-2045(18)30521-7)

22. Evans A, Vinnicombe S (2017) Overdiagnosis in breast imaging. *Breast* 31:270–273. <https://doi.org/10.1016/J.BREAST.2016.10.011>
23. Gilbert FJ, Tucker L, Young KC (2016) Digital breast tomosynthesis (DBT): a review of the evidence for use as a screening tool. *Clin Radiol* 71:141–150. <https://doi.org/10.1016/j.crad.2015.11.008>
24. Giampietro RR, Cabral MVG, Lima SAM, Weber SAT, Dos Santos Nunes-Nogueira V (2020) Accuracy and effectiveness of mammography versus mammography and tomosynthesis for population-based breast cancer screening: a systematic review and meta-analysis. *Sci Rep* 10:7991. <https://doi.org/10.1038/s41598-020-64802-x>
25. Gilbert FJ, Tucker L, Gillan MGC et al (2015) Accuracy of digital breast tomosynthesis for depicting breast cancer subgroups in a UK retrospective reading study (TOMMY Trial). *Radiology* 277:697–706. <https://doi.org/10.1148/radiol.2015142566>
26. Zuley ML, Guo B, Catullo VJ et al (2014) Comparison of two-dimensional synthesized mammograms versus original digital mammograms alone and in combination with tomosynthesis images. *Radiology* 271:664–671. <https://doi.org/10.1148/radiol.13131530>
27. Mariscotti G, Durando M, Houssami N et al (2017) Comparison of synthetic mammography, reconstructed from digital breast tomosynthesis, and digital mammography: evaluation of lesion conspicuity and BI-RADS assessment categories. *Breast Cancer Res Treat* 166:765–773. <https://doi.org/10.1007/s10549-017-4458-3>
28. Mackenzie A, Kaur S, Elangovan P et al (2018) Comparison of synthetic 2D images with planar and tomosynthesis imaging of the breast using a virtual clinical trial. In: Nishikawa RM, Samuelson FW (eds) *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*. SPIE, 10577:0H-1–9. <https://doi.org/10.1117/12.2293070>
29. Ikejimba LC, Glick SJ, Choudhury KR, Samei E, Lo JY (2016) Assessing task performance in FFDM, DBT, and synthetic mammography using uniform and anthropomorphic physical phantoms. *Med Phys* 43:5593–5602. <https://doi.org/10.1118/1.4962475>
30. Rodriguez-Ruiz A, van Engen R, Michielsen K et al (2018) How does wide-angle breast tomosynthesis depict calcifications in comparison to digital mammography? A retrospective observer study. In: Krupinski EA (ed) *14th International Workshop on Breast Imaging (IWBI 2018)*. SPIE, pp 107181T1–107181T11
31. Korhonen KE, Conant EF, Cohen EA, Synnestvedt M, McDonald ES, Weinstein SP (2019) Breast cancer conspicuity on simultaneously acquired digital mammographic images versus digital breast tomosynthesis images. *Radiology* 292:69–76. <https://doi.org/10.1148/radiol.2019182027>
32. Alabousi M, Wadera A, Kashif Al-Ghita M et al (2021) Performance of digital breast tomosynthesis, synthetic mammography, and digital mammography in breast cancer screening: a systematic review and meta-analysis. *J Natl Cancer Inst* 113:680–690. <https://doi.org/10.1093/jnci/djaa205>
33. Wu T, Moore RH, Rafferty EA, Kopans DB (2004) A comparison of reconstruction algorithms for breast tomosynthesis. *Med Phys* 31:2636–2647. <https://doi.org/10.1118/1.1786692>
34. Warren LM, Given-Wilson RM, Wallis MG et al (2014) The effect of image processing on the detection of cancers in digital mammography. *AJR Am J Roentgenol* 203:387–393. <https://doi.org/10.2214/AJR.13.11812>
35. Zanca F, Jacobs J, Van Ongeval C et al (2009) Evaluation of clinical image processing algorithms used in digital mammography. *Med Phys* 36:765–775. <https://doi.org/10.1118/1.3077121>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.