

## 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, Lorraine Tucker, Maureen GC Gillan, Paula Willsher, Julie Cooke, Karen A Duncan, Michael J Michell, Hilary M Dobson, Yit Yoong Lim, Hema Purushothaman, Celia Strudley, Susan M Astley, Oliver Morrish, Kenneth C Young and Stephen W Duffy*



**National Institute for  
Health Research**



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

<sup>1</sup>Department of Radiology, University of Cambridge, Cambridge, UK

<sup>2</sup>Aberdeen Biomedical Imaging Centre, University of Aberdeen, Aberdeen, UK

<sup>3</sup>Jarvis Breast Centre, Guildford, UK

<sup>4</sup>North East Scotland Breast Screening Centre, Aberdeen, UK

<sup>5</sup>Breast Radiology, King's College Hospital, London, UK

<sup>6</sup>West of Scotland Breast Screening Service, Glasgow, UK

<sup>7</sup>The Nightingale Centre, University Hospital South Manchester, Manchester, UK

<sup>8</sup>Department of Radiology, St Bartholomew's Hospital, London, UK

<sup>9</sup>National Co-ordinating Centre for Physics of Mammography, Royal Surrey County Hospital, Guildford, UK

<sup>10</sup>Department of Imaging Science and Biomedical Engineering, University of Manchester, Manchester, UK

<sup>11</sup>East Anglian Regional Radiation Protection Service, Cambridge University Hospitals, Cambridge, UK

<sup>12</sup>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

This report should be referenced as follows:

Gilbert FJ, Tucker L, Gillan MGC, Willsher P, Cooke J, Duncan KA, *et al*. 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* 2015;**19**(4).

*Health Technology Assessment* is indexed and abstracted in *Index Medicus/MEDLINE*, *Excerpta Medica/EMBASE*, *Science Citation Index Expanded (SciSearch®)* and *Current Contents®/Clinical Medicine*.



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/](http://www.publicationethics.org/)).

Editorial contact: [nihredit@southampton.ac.uk](mailto:nihredit@southampton.ac.uk)

The full HTA archive is freely available to view online at [www.journalslibrary.nihr.ac.uk/hta](http://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](http://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.

**© Queen's Printer and Controller of HMSO 2015. This work was produced by Gilbert *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.**

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

## Editor-in-Chief of *Health Technology Assessment* and NIHR Journals Library

**Professor Tom Walley** Director, NIHR Evaluation, Trials and Studies and Director of the HTA Programme, UK

### NIHR Journals Library Editors

**Professor Ken Stein** Chair of HTA Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

**Professor Andree Le May** Chair of NIHR Journals Library Editorial Group (EME, HS&DR, PGfAR, PHR journals)

**Dr Martin Ashton-Key** Consultant in Public Health Medicine/Consultant Advisor, NETSCC, UK

**Professor Matthias Beck** Chair in Public Sector Management and Subject Leader (Management Group), Queen's University Management School, Queen's University Belfast, UK

**Professor Aileen Clarke** Professor of Public Health and Health Services Research, Warwick Medical School, University of Warwick, UK

**Dr Tessa Crilly** Director, Crystal Blue Consulting Ltd, UK

**Dr Peter Davidson** Director of NETSCC, HTA, UK

**Ms Tara Lamont** Scientific Advisor, NETSCC, UK

**Professor Elaine McColl** Director, Newcastle Clinical Trials Unit, Institute of Health and Society, Newcastle University, UK

**Professor William McGuire** Professor of Child Health, Hull York Medical School, University of York, UK

**Professor Geoffrey Meads** Professor of Health Sciences Research, Faculty of Education, University of Winchester, UK

**Professor John Powell** Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK

**Professor James Raftery** Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

**Dr Rob Riemsma** Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

**Professor Helen Roberts** Professor of Child Health Research, UCL Institute of Child Health, UK

**Professor Helen Snooks** Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Please visit the website for a list of members of the NIHR Journals Library Board:  
[www.journalslibrary.nihr.ac.uk/about/editors](http://www.journalslibrary.nihr.ac.uk/about/editors)

**Editorial contact:** [nihredit@southampton.ac.uk](mailto:nihredit@southampton.ac.uk)

# Abstract

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

<sup>1</sup>Department of Radiology, University of Cambridge, Cambridge, UK

<sup>2</sup>Aberdeen Biomedical Imaging Centre, University of Aberdeen, Aberdeen, UK

<sup>3</sup>Jarvis Breast Centre, Guildford, UK

<sup>4</sup>North East Scotland Breast Screening Centre, Aberdeen, UK

<sup>5</sup>Breast Radiology, King's College Hospital, London, UK

<sup>6</sup>West of Scotland Breast Screening Service, Glasgow, UK

<sup>7</sup>The Nightingale Centre, University Hospital South Manchester, Manchester, UK

<sup>8</sup>Department of Radiology, St Bartholomew's Hospital, London, UK

<sup>9</sup>National Co-ordinating Centre for Physics of Mammography, Royal Surrey County Hospital, Guildford, UK

<sup>10</sup>Department of Imaging Science and Biomedical Engineering, University of Manchester, Manchester, UK

<sup>11</sup>East Anglian Regional Radiation Protection Service, Cambridge University Hospitals, Cambridge, UK

<sup>12</sup>Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK

\*Corresponding author [fjg28@medschl.cam.ac.uk](mailto:fjg28@medschl.cam.ac.uk)

**Background:** Digital breast tomosynthesis (DBT) is a three-dimensional mammography technique with the potential to improve accuracy by improving differentiation between malignant and non-malignant lesions.

**Objectives:** The objectives of the study were to compare the diagnostic accuracy of DBT in conjunction with two-dimensional (2D) mammography or synthetic 2D mammography, against standard 2D mammography and to determine if DBT improves the accuracy of detection of different types of lesions.

**Study population:** 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 attending annual mammography screening were recruited after giving written informed consent.

**Intervention:** All participants underwent a two-view 2D mammography of both breasts and two-view DBT imaging. Image-processing software generated a synthetic 2D mammogram from the DBT data sets.

**Retrospective reading study:** In an independent blinded retrospective study, readers reviewed (1) 2D or (2) 2D + DBT or (3) synthetic 2D + DBT images for each case without access to original screening mammograms or prior examinations. Sensitivities and specificities were calculated for each reading arm and by subgroup analyses.

**Results:** Data were available for 7060 subjects comprising 6020 (1158 cancers) assessment cases and 1040 (two cancers) family history screening cases. Overall 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 ( $p = 0.6$ ). 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 in both DBT reading arms for all subgroups of age, density and dominant radiological feature ( $p < 0.001$  all cases). In all reading arms, specificity tended to be lower for microcalcifications and higher for distortion/asymmetry. Comparing 2D + DBT to 2D alone, sensitivity was significantly higher: 93% versus 86% ( $p < 0.001$ ) for invasive tumours of size 11–20 mm. Similarly, for breast density 50% or more, sensitivities were 93% versus 86% ( $p = 0.03$ ); for grade 2 invasive tumours, sensitivities were 91% versus 87% ( $p = 0.01$ ); where the dominant radiological feature was a mass, sensitivities were 92% and 89% ( $p = 0.04$ ). For synthetic 2D + DBT, there was significantly ( $p = 0.006$ ) higher sensitivity than 2D alone in invasive cancers of size 11–20 mm, with a sensitivity of 91%.

**Conclusions:** The specificity of DBT and 2D was better than 2D alone but there was only marginal improvement in sensitivity. The performance of synthetic 2D appeared to be comparable to standard 2D. If these results were observed with screening cases, DBT and 2D mammography could benefit to the screening programme by reducing the number of women recalled unnecessarily, especially if a synthetic 2D mammogram were used to minimise radiation exposure. Further research is required into the feasibility of implementing DBT in a screening setting, prognostic modelling on outcomes and mortality, and comparison of 2D and synthetic 2D for different lesion types.

**Study registration:** Current Controlled Trials ISRCTN73467396.

**Funding:** This project was funded by the NIHR Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 19, No. 4. See the HTA programme website for further project information.

# Contents

<b>List of tables</b>	<b>xiii</b>
<b>List of figures</b>	<b>xv</b>
<b>Glossary</b>	<b>xvii</b>
<b>List of abbreviations</b>	<b>xix</b>
<b>Plain English summary</b>	<b>xxi</b>
<b>Scientific summary</b>	<b>xxiii</b>
<b>Chapter 1 Introduction</b>	<b>1</b>
Background	1
Digital breast tomosynthesis	1
Development and process	1
Diagnostic accuracy	2
Cancer detection and recall rate	2
<i>Oslo Tomosynthesis Screening Trial</i>	2
<i>Screening with Tomosynthesis OR standard Mammography trial</i>	3
<i>Malmö Breast Tomosynthesis Screening Trial</i>	3
Reader performance	3
Rationale for study	4
Research objectives	4
<i>Primary objectives</i>	4
<i>Secondary objectives</i>	4
<b>Chapter 2 Methodology</b>	<b>5</b>
Study design	5
Study settings and population	6
Recruitment	6
<i>Inclusion criteria</i>	7
<i>Exclusion criteria</i>	7
Intervention	7
Image acquisition	7
Radiation dose	7
Reader experience	8
Image management	10
<i>Anonymisation of images</i>	10
<i>Image storage at trial office</i>	10
<i>Digital breast tomosynthesis quality control</i>	10
<i>Prospective data collection</i>	10
<i>Review of images from assessment clinic cases</i>	10
<i>Review of images from high-/moderate-risk screening mammography cases</i>	11
<i>Assessment of breast density</i>	11
<i>Histopathology data</i>	11
<i>Prospective data checks</i>	11

Retrospective reading study	11
<i>Study data set</i>	11
<i>Reading arms</i>	12
<i>Synthetic two-dimensional images</i>	12
<i>Randomisation process and distribution of images to readers</i>	12
Retrospective data collection	12
<i>Image review</i>	12
<i>Collation and cleaning of data</i>	13
<i>Retrospective data checks</i>	13
Statistics and analyses	13
<i>Sample size</i>	13
<i>Sensitivity</i>	13
<i>Specificity</i>	14
<i>Analysis</i>	15
<b>Chapter 3 Reader study</b>	<b>17</b>
Introduction	17
Aims	17
Method	17
Results	19
Discussion	27
<b>Chapter 4 Breast density assessment</b>	<b>29</b>
Introduction	29
Measurement of breast density	29
Materials and methods	30
Results	31
<i>Total breast volume</i>	32
<i>Fibroglandular tissue volume</i>	33
<i>Volumetric breast density</i>	34
<i>Breast density and cancer risk</i>	40
Discussion	42
<b>Chapter 5 Results</b>	<b>45</b>
Recruitment	45
Data sets	45
Analysis	47
Receiver operating characteristic analysis	54
<b>Chapter 6 Discussion</b>	<b>59</b>
<b>Chapter 7 Conclusions and future research</b>	<b>63</b>
Future research	64
<b>Acknowledgements</b>	<b>65</b>
<b>References</b>	<b>69</b>



<b>Appendix 1</b> Participant information sheet for women recalled for assessment	<b>81</b>
<b>Appendix 2</b> Participant information sheet for moderate-/high-risk women as a result of family history	<b>85</b>
<b>Appendix 3</b> Consent form	<b>89</b>
<b>Appendix 4</b> Image management report	<b>91</b>
<b>Appendix 5</b> Flow diagram indicating management of image files	<b>95</b>
<b>Appendix 6</b> Quality control report on imaging equipment	<b>97</b>
<b>Appendix 7</b> Prospective data collection form: assessment	<b>107</b>
<b>Appendix 8</b> Prospective data collection form: family history	<b>109</b>
<b>Appendix 9</b> Prospective data collection form: pathology	<b>111</b>
<b>Appendix 10</b> Randomisation process for retrospective reading study	<b>113</b>
<b>Appendix 11</b> Retrospective study data collection form: two-dimensional	<b>115</b>
<b>Appendix 12</b> Retrospective study data collection form: two-dimensional and three-dimensional	<b>117</b>
<b>Appendix 13</b> Invitation letter for assessment women	<b>119</b>
<b>Appendix 14</b> Invitation letter for moderate- or high-risk women as a result of family history	<b>121</b>
<b>Appendix 15</b> Letter to general practitioner advising of trial participation	<b>123</b>
<b>Appendix 16</b> National Co-ordinating Centre for the Physics of Mammography tomosynthesis commissioning protocol for the TOMMY trial	<b>125</b>
<b>Appendix 17</b> Radiographer quality control for the TOMMY trial	<b>129</b>



# List of tables

<b>TABLE 1</b> Radiation dose for study	<b>8</b>
<b>TABLE 2</b> Summary of reader experience	<b>9</b>
<b>TABLE 3</b> Potential differential sensitivity of 2D and 2D + DBT; anticipated percentage distribution of cancers identified by each imaging modality	<b>14</b>
<b>TABLE 4</b> Potential differential specificity; anticipated percentage of non-cancers ruled out by each imaging modality	<b>14</b>
<b>TABLE 5</b> Cancers diagnosed at assessment tabulated by detection method	<b>15</b>
<b>TABLE 6</b> Test set by case type	<b>17</b>
<b>TABLE 7</b> Recall rates based on a suspicion score of 3 or more	<b>18</b>
<b>TABLE 8</b> Suspicion classification (based on Maxwell <i>et al.</i> )	<b>18</b>
<b>TABLE 9</b> Proportion of cancer cases identified as suspicious or malignant, and the proportion of normal/benign cases scored normal, benign or probably benign	<b>19</b>
<b>TABLE 10</b> Change in performance between the two reads ranked according to the number of DBT reads undertaken between the two time points	<b>21</b>
<b>TABLE 11</b> Change in performance between the two reads ranked according to the number of digital mammograms read in the year prior to the first DBT read	<b>22</b>
<b>TABLE 12</b> Performance at the first read ranked by experience of reading FFDM in the previous 12 months	<b>23</b>
<b>TABLE 13</b> Performance at the first read ranked by years of experience reading mammograms	<b>24</b>
<b>TABLE 14</b> Recall rates (suspicion score of 3 or more) by site	<b>24</b>
<b>TABLE 15</b> Cancer cases identified as suspicious or malignant (score 4 or 5) by site	<b>25</b>
<b>TABLE 16</b> Summary results for the total cohort of cases. Unless indicated, numbers are median (minimum–maximum) values	<b>31</b>
<b>TABLE 17</b> Breast composition measures using Volpara	<b>40</b>
<b>TABLE 18</b> Breast composition measures using Quantra	<b>42</b>
<b>TABLE 19</b> Odds ratios by quartile of Quantra percentage density	<b>42</b>
<b>TABLE 20</b> Recruitment by centre	<b>45</b>
<b>TABLE 21</b> Recruitment summary	<b>46</b>

<b>TABLE 22</b> Participant characteristics	<b>48</b>
<b>TABLE 23</b> Cancer detection rate in three arms of reading study	<b>51</b>
<b>TABLE 24</b> Detection status of cancers all three imaging modalities	<b>53</b>
<b>TABLE 25</b> Characteristics of invasive cancers missed by each imaging modality and of those found by all three	<b>53</b>
<b>TABLE 26</b> Mean glandular dose to the standard breast and contrast-to-noise ratio averaged over the seven systems and coefficient of variation, for conventional and tomosynthesis images	<b>98</b>
<b>TABLE 27</b> Mean glandular dose measurements for all seven systems during the trial	<b>99</b>
<b>TABLE 28</b> Contrast-to-noise ratio measurements for all seven systems during the trial	<b>100</b>
<b>TABLE 29</b> Mean modulation transfer function measurements for the seven systems	<b>101</b>
<b>TABLE 30</b> Average threshold gold thickness and coefficient of variation for 2D and tomosynthesis CDMAM images	<b>102</b>
<b>TABLE 31</b> Summary of patient dose data gathered from each centre	<b>105</b>
<b>TABLE 32</b> Mean glandular dose to 53-mm standard breast and average MGD for oblique views of 50- to 60-mm breasts	<b>105</b>
<b>TABLE 33</b> Summary of patient dose measurements at all sites, all views, all breast thicknesses	<b>106</b>
<b>TABLE 34</b> Summary of remedial levels for radiographer QC tests	<b>136</b>

# List of figures

<b>FIGURE 1</b> Case collection	5
<b>FIGURE 2</b> Reading study	6
<b>FIGURE 3</b> Proportion of cancer cases recalled, plotted against the proportion of normal/benign cases not recalled at time point 1 (◆) and time point 2 (■)	20
<b>FIGURE 4</b> Proportion of cancer cases identified as suspicious or malignant, and the proportion of normal/benign cases scored normal, benign or probably benign at time point 1 (◆) and time point 2 (■)	20
<b>FIGURE 5</b> Performance at first read vs. FFDM load in previous 12 months	25
<b>FIGURE 6</b> Performance at the first read ranked by number of years of experience reading mammograms	25
<b>FIGURE 7</b> Recall rates by site at first (T1) and second (T2) reads	26
<b>FIGURE 8</b> Cancer cases identified as suspicious/malignant by site	26
<b>FIGURE 9</b> Histograms showing the distribution of breast volume across the study population as measured by (a) Quantra and (b) Volpara	32
<b>FIGURE 10</b> Scatterplot comparing Quantra and Volpara scores for total breast volume	32
<b>FIGURE 11</b> Histograms showing the distribution of fibroglandular tissue volume across the study population as measured by (a) Quantra and (b) Volpara	33
<b>FIGURE 12</b> Scatterplot comparing Quantra and Volpara scores for fibroglandular tissue volume	34
<b>FIGURE 13</b> Histograms showing the distribution of volumetric breast density across the study population as measured by (a) Quantra and (b) Volpara	35
<b>FIGURE 14</b> Scatterplot comparing Quantra and Volpara scores for volumetric breast density	35
<b>FIGURE 15</b> Histograms showing the distribution of area-based breast density across the study population from (a) the observers and (b) the Quantra software	36
<b>FIGURE 16</b> Scatterplot comparing the observer and Quantra scores for the area-based breast density	36
<b>FIGURE 17</b> Scatterplots comparing the observer area-based breast density scores with the volumetric measurements from (a) Quantra and (b) Volpara	37

<b>FIGURE 18</b> Graphs showing the decrease in breast density with age for the study group for the four available density measures	<b>38</b>
<b>FIGURE 19</b> Odds ratios by quartile of Volpara absolute dense volume	<b>41</b>
<b>FIGURE 20</b> Odds ratios by quartile of Quantra absolute dense volume	<b>41</b>
<b>FIGURE 21</b> Cumulative recruitment	<b>46</b>
<b>FIGURE 22</b> Study flow chart	<b>47</b>
<b>FIGURE 23</b> Receiver operating characteristic curves for the three reading arms	<b>54</b>
<b>FIGURE 24</b> Receiver operating characteristic curves for the three reading arms in those cases with visually assessed density less than 50%	<b>55</b>
<b>FIGURE 25</b> Receiver operating characteristic curves for all three reading arms in those with visually assessed density of 50% or more	<b>55</b>
<b>FIGURE 26</b> Receiver operating characteristic curves for the three reading arms in those cases where the dominant radiological sign was a mass	<b>56</b>
<b>FIGURE 27</b> Receiver operating characteristic curves for the three reading arms in those cases with microcalcifications as the dominant radiological feature	<b>56</b>
<b>FIGURE 28</b> Receiver operating characteristic curves for all three reading arms in those with distortion or ASD as the dominant radiological feature	<b>57</b>
<b>FIGURE 29</b> Flow diagram indicating the management of image files	<b>95</b>
<b>FIGURE 30</b> Mean glandular dose to the standard breast, for (a) conventional exposures and (b) tomosynthesis exposures	<b>98</b>
<b>FIGURE 31</b> Contrast-to-noise ratio for (a) conventional exposures and (b) DBT exposures	<b>99</b>
<b>FIGURE 32</b> Modulation transfer function measurements for the seven systems	<b>101</b>
<b>FIGURE 33</b> Contrast detail curves from each of the systems for (a) conventional and (b) tomosynthesis images	<b>102</b>
<b>FIGURE 34</b> Variation in 2D mammography and tomosynthesis dose and SNR from analysis of weekly QC images of 45-mm PMMA	<b>103</b>
<b>FIGURE 35</b> Patient dose data from all sites (all views)	<b>106</b>

# Glossary

**Aetna** A US health-care company providing a wide range of health-care products and services including diagnostic tests and medical treatment.

**BRCA gene status** An inherited genetic mutation of either the *BRCA1* or the *BRCA2* gene affecting the production of tumour suppressor proteins. Damage to deoxyribonucleic acid can result in genetic alterations and significantly increases the lifetime risk of developing breast cancer.

**Breast Imaging Reporting and Data System scale** A widely accepted risk assessment and quality assurance tool in mammography, particularly in the USA. A scale of numerical codes (0–6) is used to indicate a film reader's level of suspicion of malignancy.

**DCMTK** (Toolkit from dicom.offis.de) A software application widely used in hospitals and as a tool for research projects to convert and store Digital Imaging and Communication in Medicine image files and software servers.

**Diagnostic Reference Levels** The Ionising Radiation (Medical Exposure) Regulations 2000 requiring all UK hospitals to adhere to safe diagnostic radiation standards and levels.

**Estimated mean glandular dose** An estimation of the radiation dose received in an exposure based on the assumption that the breast comprises equal proportions of adipose and fibroglandular tissue.

**Hologic SecurView DX Workstation** Visual display of digital images on monitors optimised to read two-dimensional and three-dimensional images.

**Hologic Selenia Dimensions System** Chosen manufacturer's model of mammography imaging system used in this trial to acquire two-dimensional and digital breast tomosynthesis images.

**Interval cancer** A cancer detected and diagnosed in the interval between a routine screening mammogram and the subsequent routine screen. The cancer may have been present at initial screening (false negative) or as a result of rapid growth in the interval period between screening and detection.

**International Workshop on Digital Mammography** International workshop allowing discussion of innovative mammography, tomosynthesis, computed tomography, magnetic resonance imaging and computer-aided detection imaging and image-processing technologies.

**Lymph node status** Measure of the extent of axillary lymph node involvement in cases where breast cancer spreads through the lymphatic system, outside the breast.

**MACRO** The electronic data capture and management system widely employed in clinical research to store and manage data collected on forms.

**Mammographic density** This refers to the proportion of fibroglandular tissue in the breast and low-density fatty breast tissue.

**Microcalcification clusters** Tiny calcium deposits detected on a mammogram and visualised primarily as fine white flecks. The size, shape and pattern of these clusters are significant in determining potential for malignancy.

**National Co-ordinating Centre for the Physics of Mammography** National body that monitors image quality and ensures imaging performance meets minimum NHS Breast Screening Programme standards for image quality.

**PERFORMS** (PERsonal perFORmance in Mammographic Screening) System designed to assess the image interpretation skills and provide ongoing education to film readers working in the national breast screening programme. Annual self-assessment of a caseload of screening mammograms is undertaken and feedback provided on diagnostic accuracy.

**Sensitivity of mammography** The percentage of breast cancers that are actually detected by mammography in a given population of women, which are true cancers.

**Specificity of mammography** The percentage of negative results for breast cancer in a given population of women who do not have the disease.

**Synthetic two-dimensional image** Simulated two-dimensional image created from a summation of individual digital breast tomosynthesis image slices.

**Volumetric breast density** Computer-calculated percentage score of dense fibroglandular tissue volume to the overall breast volume.



## List of abbreviations

2D	two-dimensional	MLO	mediolateral oblique
3D	three-dimensional	MRI	magnetic resonance imaging
AEC	automatic exposure control	MTF	modulation transfer function
ASD	asymmetrical density	NCCPM	National Co-ordinating Centre for the Physics of Mammography
AUC	area under the curve	NHSBSP	NHS Breast Screening Programme
BI-RADS	Breast Imaging Reporting and Data System	OR	odds ratio
CC	craniocaudal	PACS	picture archiving and communications system
CDR	Central Data Repository	PI	principal investigator
CI	confidence interval	PMMA	polymethyl methacrylate
CNR	contrast-to-noise ratio	QA	quality assurance
DBT	digital breast tomosynthesis	QC	quality control
DCIS	ductal carcinoma in situ	RCT	randomised controlled trial
DICOM	Digital Imaging and Communication in Medicine	ROC	receiver operating characteristic
FDA	Food and Drug Administration	ROI	region of interest
FFDM	full-field digital mammography	SNR	signal-to-noise ratio
FH	family history	STORM	Screening with Tomosynthesis OR standard Mammography
HICS	Hologic Image Collection System	TOMMY	TOMosynthesis with digital MammographY in the UK NHS Breast Screening Programme
HVL	half-value layers	VAS	visual analogue scale
IT	information technology		
MGD	mean glandular dose		



## Plain English summary

**B**reast screening is recognised as the best way to detect early-stage breast cancer and reduce the number of deaths from this disease. In a standard breast screening radiograph (mammogram), overlapping breast tissue may hide some cancers or make normal tissue appear abnormal.

Digital breast tomosynthesis (DBT) takes multiple low-dose radiographs of the breast that are processed by a computer to reconstruct a DBT image. This allows abnormalities in the breast be seen more clearly and could make it easier to see small cancers and decrease the number of 'false alarms'. DBT images are usually read with a standard two-dimensional (2D) mammogram. This double radiation exposure could be avoided if a 2D mammogram could be created from DBT images.

The study compared the accuracy of reading (1) a 2D mammogram with (2) 2D mammogram with a DBT or (3) synthetic 2D mammogram with a DBT to identify breast cancer. Data from 7061 cases were analysed: 6021 cases from women (47–73 years) recalled after routine screening for further tests and 1040 cases from women (40–49 years) with a family history of breast cancer attending annual breast screening.

The results of the study indicated that the use of a combination of 2D + DBT or synthetic 2D + DBT produces a small increase in the number of cancer cases detected and could help reduce the number of women who are recalled for unnecessary tests, that is false alarms, which could reduce health-care costs and patient anxiety.



# 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.





# Chapter 1 Introduction

## Background

Breast screening with mammography is recognised as the most effective method of detecting early-stage breast cancer and reducing breast cancer mortality. A recent meta-analysis of 11 randomised trials concluded that there was a 20% relative risk reduction in breast cancer mortality in women invited to screening.<sup>1</sup> However, one of the primary limitations of standard two-dimensional (2D) mammography is that overlapping dense fibroglandular tissue within the breast can decrease the visibility of malignant abnormalities or simulate the appearance of an abnormality. This reduces the sensitivity of screening and increases the number of false-positive recalls.<sup>2</sup> It has been shown that 15–30% of cancers are not detected by standard screening,<sup>3</sup> and this percentage is higher in women aged under 50 years<sup>4</sup> and in women with dense breasts.<sup>5–7</sup> Women with dense breasts have reduced screening programme sensitivity<sup>8–11</sup> and tend to have larger screen-detected and interval cancers.<sup>8,12,13</sup> These issues are of concern for the UK NHS Breast Screening Programme (NHSBSP) as it extends the screening age range to include younger, pre- or perimenopausal women who are known to have a higher proportion of dense breast tissue<sup>14,15</sup> and is also potentially problematic for women at moderate or high risk of developing familial breast cancer who attend annual mammography.<sup>14</sup>

## Digital breast tomosynthesis

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 overlapping shadows from breast tissue that degrade the image quality in standard 2D projection imaging. This should improve the visibility of cancers and facilitate the differentiation between malignant and non-malignant features. The expectation is that small cancers, which may be obscured by normal fibroglandular tissue in standard 2D projection imaging, could be more readily detected using DBT, particularly in women with radiologically dense breasts. In addition, by facilitating the analysis of superimposed breast structures, DBT may enable the reader to identify features such as asymmetrical density (ASD) on 2D imaging as normal composite shadows and thereby decrease the number of false-positive recalls<sup>16–19</sup> and the associated health-care costs<sup>20</sup> and patient anxiety.<sup>21,22</sup>

## Development and process

The fundamentals of tomographic imaging were established in the 1930s, but clinical applications of tomosynthesis in mammography did not evolve until several decades later, following the development of flat-panel digital display detectors, rapid computer processing and advances in reconstruction and post-processing algorithms.<sup>23</sup> In DBT, a sequence of projection images is obtained by moving the position of the X-ray tube and making exposures at regular intervals/angles. The angular range used varies from one manufacturer to another. In the Hologic Selenia Dimensions System (Hologic Inc., Bedford, MA, USA) used in this study, 11 exposures are taken over an angular range of 15 degrees. The exposure used for each projection image is relatively small so that the overall mean glandular dose (MGD) for DBT is comparable with that of conventional 2D imaging. The projection images acquired by the detector are processed by reconstruction algorithms to produce a pseudo-3D tomographic image of the breast in which each reconstructed image (or slice) shows the tissues sharply for that plane and blurs out details in higher and lower planes. Typically, this reconstruction is done to produce images with a 1-mm slice thickness although increased slice thickness can be used. Image quality of DBT is highly dependent on system geometry and the choice of optimal image acquisition, reconstruction and display parameters.<sup>24–27</sup>

Some manufacturers employ a larger angular range, which would theoretically improve the depth resolution between planes, but at the expense of in-plane resolution.<sup>25,26</sup> A viewing workstation provides tools that enable the reader to scroll vertically through the tomographic images as well as to compare them with the corresponding 2D images.

## Diagnostic accuracy

The superiority of DBT over standard 2D projection imaging in terms of lesion visibility and margin detection was first demonstrated in experimental studies using phantoms and mastectomy specimens.<sup>28–30</sup> Improved lesion visibility, size and classification compared with standard film or digital mammography have been reported<sup>31–38</sup> and the possibility that DBT could reduce the need for additional mammographic views at least for non-calcified lesions<sup>39–42</sup> has been suggested. There are mixed reports of the sensitivity of DBT for the detection of microcalcifications<sup>32,43–51</sup> that may be partly as a result of the different techniques used for image reconstruction and the need to combine image slices into thicker slabs for optimal visualisation of microcalcification clusters. These observations suggested that DBT was unlikely to be used as a stand-alone imaging modality if a 2D mammogram was required for optimal microcalcification assessment.<sup>25,52,53</sup>

## Cancer detection and recall rate

Higher sensitivity of DBT, either alone or in combination with 2D mammography, has been reported<sup>16,19,45,48,54,55</sup> although no change<sup>56</sup> or reductions in sensitivity of between 4% and 9% were also noted in some comparison studies of one-view and two-view DBT versus 2D or DBT added to 2D.<sup>33,44,57</sup> Several studies have shown increased specificity of DBT compared with 2D<sup>16,33,45</sup> and in combination with 2D mammography.<sup>16,20,43,46,58</sup> In a multireader multicentre trial using receiver operating characteristic (ROC) analysis, Rafferty *et al.*<sup>17</sup> reported increased diagnostic accuracy of 2D + DBT compared with 2D alone, particularly in the detection of invasive cancers, and a reduction in the false-positive recall rate and a large retrospective evaluation of screening mammography in 13,158 women<sup>59</sup> reported a significantly lower recall rate with 2D + DBT compared with 2D alone, especially for women aged < 50 years and those with dense breasts. However, the latter study was underpowered to demonstrate any significant increase in cancer detection rate. Rose *et al.*<sup>60</sup> conducted an observational study to assess changes in screening performance measures after the introduction of DBT. Six radiologists interpreted 13,586 screening mammography studies without DBT and 9499 studies with the addition of DBT. Routine use of 2D + DBT resulted in a significant reduction in recall rate and, although changes in other performance measures, for example cancer detection rate (in particular, earlier detection of invasive cancers), biopsy rate and positive predictive value, did not reach statistical significance, this was attributed to study design limitations.

In general, published studies have demonstrated the potential for DBT to decrease recall rates and possibly increase cancer detection rates.<sup>61,62</sup> Some of the conflicting results and uncertainties from early studies may be attributed to differences in study methodology, case composition and the use of prototype DBT systems from different manufacturers with different configurations.<sup>63–65</sup> The Houssami and Skaane review<sup>64</sup> did conclude that the available data did support investment of new large-scale randomised population screening trials.

Several population-based screening trials are now in progress, although none is randomised:

### Oslo Tomosynthesis Screening Trial

The Oslo Tomosynthesis Screening Trial, the first large-scale trial to implement DBT, compares various screen-reading protocols. Interim analysis of data from 12,631 participants reported a 27% increase in cancer detection rate across all breast densities, a significant increase (40%) in the detection of invasive cancers and a 15% decrease in false-positive recall rate with 2D + DBT mammography compared with 2D alone.<sup>19</sup>

### Screening with Tomosynthesis OR standard Mammography trial

The results of the Italian STORM (Screening with Tomosynthesis OR standard Mammography) trial<sup>66</sup> were consistent with interim data from the Oslo trial. In a population-based screening programme, a significant increase in cancer detection was observed across all age groups and breast densities comparing sequential 2D screen reading and combined 2D + DBT. In addition, the study showed that 2D + DBT had the potential to reduce the false-positive recall rate by 17%.

### Malmö Breast Tomosynthesis Screening Trial

The Malmö Breast Tomosynthesis Screening Trial (<https://clinicaltrials.gov/ct2/show/NCT01091545>) is conducting a paired analysis of the sensitivity and specificity of DBT compared with 2D in a population-based screening programme in Sweden. Preliminary results indicated a 15% increase in sensitivity with DBT compared with 2D mammography but with a slight (3%) increase in recall rates.

These screening studies provide the best evidence available to date that the combined use of DBT and 2D may increase the number of cancers detected, with a large impact on decreasing the number of unnecessary recalls.

## Reader performance

The need for substantial reader training with DBT has been acknowledged.<sup>16,67,68</sup> Multireader studies have demonstrated improved performance, using DBT in combination with 2D mammography compared with 2D mammography alone, by radiologists with a range of experience, as measured by reduction in recall rate and ROC analysis.<sup>17,58,69</sup> However, Wallis *et al.*,<sup>46</sup> using a photon-counting DBT system, reported that the addition of DBT improved the performance only of inexperienced readers. Reading and reporting times are considerably longer (almost twofold) using DBT,<sup>16,46,68,70</sup> with a reading time of 91 seconds for 2D + DBT compared with 45 seconds for 2D alone in a screen-reading setting<sup>19</sup> because of the number of images to be reviewed. This has resource implications if DBT were to be introduced into screening practice, although the impact on workflow could be mitigated if recall rates were shown to be decreased and unnecessary diagnostic assessments avoided.<sup>71</sup>

A number of publications have summarised and reviewed both the technical and the clinical aspects of DBT and have speculated on its potential application in breast imaging.<sup>23,25–27,61,63,64,72,73</sup> The AETNA policy report,<sup>74</sup> a combined recommendation from the American College of Radiologists, the American Cancer Society and the American College of Obstetrics and Gynaecologists, considers that breast tomosynthesis imaging is experimental and investigational because of insufficient evidence of its effectiveness. An Australian health technology review acknowledged that DBT systems were likely to replace 2D in screening when existing 2D equipment is due for replacement<sup>75</sup> and an overview of the evidence and issues to be considered in relation DBT and breast cancer screening raised concerns regarding additional radiation exposure and the learning curve for radiologists to accurately interpret DBT results.<sup>65</sup>

The use of DBT in combination with 2D requires an approximate doubling of radiation exposure. Hologic have developed image-processing software [C-View™ (Hologic Inc., MA, USA, 2011)] to simulate a conventional 2D mammogram by generating a synthetic 2D image from each set of tomosynthesis slices,<sup>70,76</sup> thus eliminating the need for the acquisition of additional 2D exposures. One published study<sup>70</sup> has reported lower sensitivity using synthetic 2D + DBT than using 2D + DBT but the authors acknowledged that greater diagnostic accuracy may be achieved with newer versions of the software. Synthetic 2D is currently being evaluated within the Oslo trial.<sup>77,78</sup>

## Rationale for study

The need for robust studies using clinically relevant methodology to further inform the optimal role, if any, for DBT in the diagnosis and assessment of breast cancer has been highlighted.<sup>63–65,79</sup> Of primary importance is further research to address whether DBT should be used as an adjunct to standard 2D or as a stand-alone technology.

To assist in the choice of DBT system to be used in the study, a working party was set up at the start of the project in 2010 by some of the grant applicants to review the status of DBT system technology. When the working party report was published,<sup>80</sup> seven companies were involved in the development of commercial DBT systems, and the group concluded that a multivendor evaluation of DBT systems in the study would be premature at this stage, as some of vendors' DBT technology was still being developed. The Hologic DBT system was relatively well established, was technically stable and was the only commercially available system with US Food and Drug Administration (FDA) approval at the start of the trial. The Hologic DBT system allows both 2D and DBT images to be acquired in a single examination with the same breast position and compression (combination mode), enabling direct comparison of the performance of the two imaging modalities. This reduces any bias that could occur as a result of different positioning of the breast when taking the second image. The rapid acquisition time (< 10 seconds) minimises patient discomfort and patient movement, which could result in blurring of the acquired images, and a fast image-processing time (3–4 seconds) minimises the impact on clinic workflow. In addition, the system coregisters the 2D and DBT images, thus facilitating image interpretation and lesion localisation.

## Research objectives

The purpose of the study was to test the diagnostic accuracy of DBT as an adjunct to standard 2D mammography as a primary screening tool.

### Primary objectives

- To compare the diagnostic accuracy of using DBT with 2D, and DBT with synthetic 2D, with standard 2D in women aged 47–73 years.
- To determine if the addition of DBT to 2D or synthetic 2D improves the accuracy of detection of small or subtle breast cancers.
- To determine if diagnostic accuracy improves in women with dense breasts by using DBT with 2D.

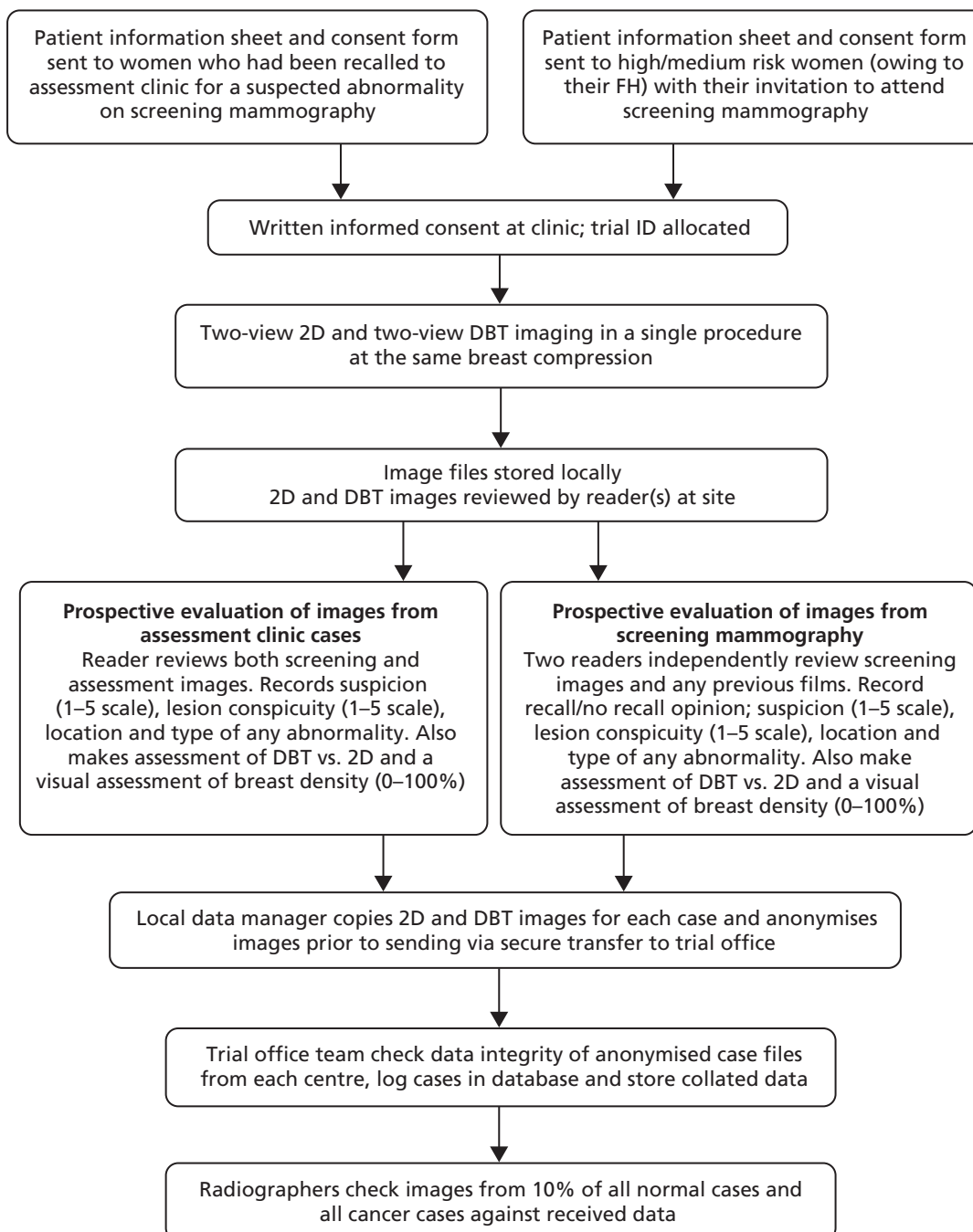
### Secondary objectives

- To compare the diagnostic accuracy of DBT with 2D, and DBT with synthetic 2D, against standard 2D in:
  - women aged 40–49 years with a moderate or high risk of breast cancer as a result of family history (FH) and who attend annual screening mammography
  - cancers presenting as soft-tissue masses
  - 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 relationship of breast density with cancer incidence.

## Chapter 2 Methodology

### Study design

This study is a multicentre retrospective reading study comparing the diagnostic performance of 2D + DBT versus 2D and synthetic 2D + DBT versus 2D. The overall study design is shown in *Figures 1* and 2.



**FIGURE 1** Case collection.

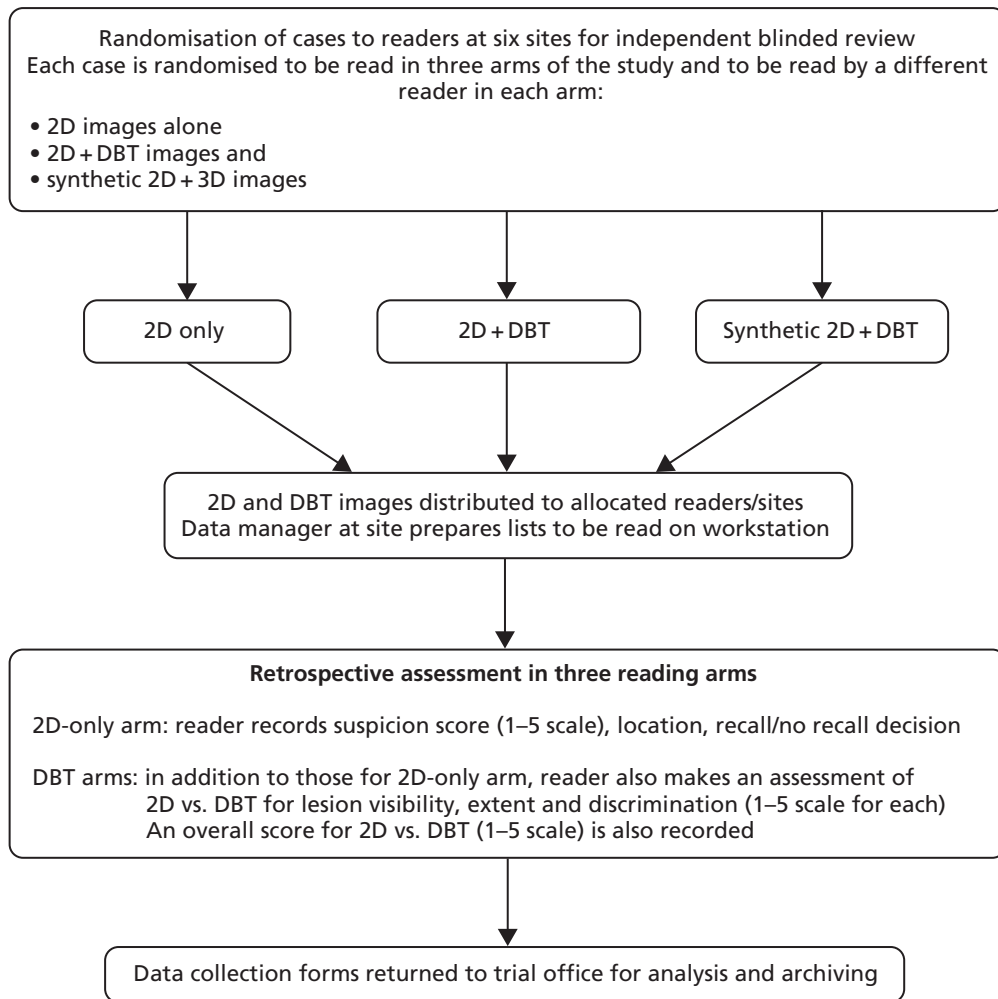


FIGURE 2 Reading study.

## Study settings and population

Participants were recruited from six NHSBSP centres in the UK and from the symptomatic breast service in Aberdeen. They comprised women aged 47–73 years recalled to an assessment clinic for a mammographic abnormality detected at routine breast screening and also women below 50 years of age with a FH of breast cancer who attended annual mammography screening.<sup>81</sup> The purpose of targeting this group of women was to compile a data set with a relatively high proportion of cancer cases (estimated to be approximately 18%). It was expected there would be around 50% of cases with overlapping tissues on standard 2D mammography that simulated suspicious features but were actually normal breast tissue.

## Recruitment

A participant information leaflet (see *Appendices 1* and *2*) outlining the potential benefits and risks of the study was sent to women who would be suitable for inclusion in the study, and written informed consent (see *Appendix 3*) was obtained on attendance at the assessment clinic or screening mammography appointment.

### Inclusion criteria

- Women aged 47–73 years attending routine breast screening (either film or digital mammography) and recalled for further assessment.
- Women aged 40–49 years with FH of breast cancer and invited to attend annual breast screening mammography.
- Women who have had a previous diagnosis of breast cancer attending screening.

### Exclusion criteria

- Any woman unable to give informed consent, including anyone unable to understand the nature and purpose of the study.
- Any woman with breast implants.
- Any woman who was pregnant.

### Intervention

All participants underwent standard two-view [mediolateral oblique (MLO) and craniocaudal (CC)] DBT imaging and two-view (MLO and CC) 2D mammography of both breasts. For participants recruited in assessment clinics, imaging was conducted prior to any additional investigations deemed necessary in the assessment clinic. Following DBT and 2D imaging, women resumed the normal pathway through the assessment or screening clinic. The DBT images were available for management of the women during the clinic or shortly afterwards. Any subsequent management followed standard assessment clinic or screening centre procedures.

### Image acquisition

Both the DBT and the standard 2D imaging were performed as a single procedure at the same breast compression on a Hologic Selenia Dimensions Digital Mammography Unit (Hologic Inc., Bedford, MA, USA). This important feature of the system eliminated any bias that could have occurred as a result of different positioning of the breast when taking the second image. Differences in compression could significantly influence the detectability of cancers as a result of the resulting differences in the superimposition of tissues. The rapid acquisition sequence also minimises patient discomfort and patient movement, which could result in blurring of the acquired image. Radiographers were experienced specialist mammography radiographers, fully trained in accordance with NHSBSP standards, who had received additional specific training on the DBT equipment used in the study.

### Radiation dose

The radiation dose for DBT was additional to normal procedures and some of the dose from the 2D imaging may have been additional if local protocol was to take fewer initial images at assessment (some of this dose may have been offset by not having to acquire supplementary, coned or magnification images).

The additional lifetime risk of inducing a breast cancer as a result of a single two-view mammography examination is estimated to be approximately 1 in 20,000 between the ages of 50 and 70 years.<sup>82,83</sup> For this trial, the total MGD was estimated at 7 mGy (*Table 1*), giving rise to an estimated 1 in 10,000 risk of cancer induction (assuming an induction rate of 14 per million per mGy). In practice, some of this dose would have been received during normal assessment procedures (estimated at 1.5 to 3.0 mGy depending on local practice); therefore the additional dose ranged from 4.0 to 5.5 mGy. The total dose for the trial falls just within the diagnostic reference level for standard two-view 2D mammography.



**TABLE 1** Radiation dose for study

Procedure	Estimated MGD for typical breast (50 to 60 mm thick)	Diagnostic reference level <sup>a</sup>
Two-view 2D	3 mGy	7 mGy
Two-view DBT	4 mGy	Not available
Total study dose	7 mGy	Not applicable

<sup>a</sup> Diagnostic reference level is 3.5 mGy for one oblique view; this has been doubled for two views.

Some of the trial participants were women aged 40–49 years with FH of breast cancer who were attending annual surveillance mammography.<sup>81</sup> The radiation risk implications of cancer screening in this cohort was reviewed, with benefits expected to substantially exceed risks down to at least the age of 40.<sup>82</sup> In the trial, the total dose including DBT was approximately 7 mGy. Overall, the additional radiation dose involved was very low and within the range currently accepted for routine screening.

### Reader experience

Readers from each participating centre were a mixture of radiologists, advanced practitioner radiographers and breast clinicians, representative of current reading practice in the NHSBSP. All readers had a proven track record of film reading in the NHSBSP, including:

- mammographic film reading for a minimum of 2 years
- reading a minimum of 5000 mammograms per annum
- annual participation in PERsonal PerFORMance in Mammographic Screening (PERFORMS) self-assessment test
- attendance at assessment clinics and multidisciplinary team meetings.

Readers who are indicated in *Table 2* with a superscript 'a' contributed only to prospective reading and did not participate in the retrospective reading study.

The number of readers available varied from week to week owing to other commitments, annual leave, etc.; the average number per week was 16.

Prior to the start of the trial, reader training consisted of 2 days of applications training from the DBT system manufacturer (Hologic Inc.) and attendance at a 1-day DBT reading course presented by staff from Breast Radiology at King's College Hospital, London. Readers were also asked to read a test set of 80 cases. These same cases were read again at the end of the trial in order to evaluate improvement in reader performance over time. An account of this evaluation is given later in this report (see *Chapter 3*). Over the first 12 months of the study recruitment period readers also gained experience of DBT by reviewing 2D and DBT images acquired at their own site. All readers used Hologic SecurView DW workstations (Hologic Inc., Bedford, MA), optimised to read both 2D and DBT images.



TABLE 2 Summary of reader experience

Site name	Reader code	Reader type <sup>a</sup>	Number of years reading mammograms	Number of mammograms per year	Number of years reading digital mammograms
Aberdeen	A3	1	11	6000	5
Aberdeen	A4	1	4	6000	4
Aberdeen	A1 <sup>a</sup>	1	6	7000	2
Aberdeen	Ax <sup>a</sup>	1	15	6000	4
Aberdeen	A2	1	17	6000	4
Barts	B5	1	3	10,000	3
Barts	B3 <sup>a</sup>	1	15	8000	12
Barts	B4	1	4	10,000	4
Barts	B1	1	4	8000	4
Barts	B2 <sup>a</sup>	2	8	5147	8
Glasgow	G5	1	18	12,000	2
Glasgow	G3	2	10	10,000	1
Glasgow	G1	1	25	5000	5
Glasgow	G4	2	7	13,000	1
Glasgow	G2	1	10	6000	3
Guildford	J4	1	3.5	6000	3
Guildford	J2	1	10	7000	2
Guildford	J1	1	20	8000	3
Guildford	J5	3	7	10,000	2
Guildford	J3	1	18	10,000	1
Manchester	M5	1	17	9000	7
Manchester	M1	1	5	7000	5
Manchester	M3	1	22	7000	2
Manchester	M2	1	24	10,000	7
Manchester	M6	3	6	7000	6
King's College	K1	1	25	13,994	2
King's College	K3 <sup>a</sup>	1	8	6549	2
King's College	K4	1	6	8430	2
King's College	K6	1	23	8061	2
King's College	K7	1	3	8525	2
King's College	K8	2	10	8000	2

<sup>a</sup> Reader type 1 = consultant radiologist; 2 = advanced practitioner radiographer; and 3 = breast physician.

## Image management

Images acquired on the Hologic Selenia Dimensions system were exported to a number of output devices:

- SecurView DX workstation (Hologic Inc, Bedford, MA, USA): Processed 2D and DBT image files were exported to the workstation for review by readers.
- Picture archiving and communications system (PACS): Processed image files that formed part of a patient's screening episode were exported to PACS. Individual centres determined whether or not to store the processed DBT image files on their NHS PACS system. Archiving was initiated following processing of all images collected for the study.
- Hologic SecurXchange (Hologic Inc, Bedford, MA, USA)/Hologic Image Collection System (HICS) (Hologic Inc, Bedford, MA, USA): An output set containing both the raw and the processed DBT and 2D image files was exported to the SecurXchange archiver unit and HICS for anonymising and storage prior to image transfer to the trial office.

### Anonymisation of images

Copies of the 2D and DBT image files for each participant were anonymised prior to transfer to the trial office. A program in the HICS removed patient-identifiable data from the Digital Imaging and Communication in Medicine (DICOM) files. Two output files were generated for each case. The first contained identifiable patient data and a site-specific R2 number and provided the master link between anonymised and identifiable data. The second file was generated by the HICS after the anonymisation process and contained only a list of dates and R2 numbers. Images transferred to the trial office via the Hologic SecurXchange unit were identified by R2 number and R2 ID only.

### Image storage at trial office

Images from each site were collated and stored in the Central Data Repository (CDR) on the University of Cambridge network server in readiness for copying and distribution to sites for the retrospective reading study. A detailed report on image management can be found in *Appendix 4* together with a diagram representing flow of data (see *Appendix 5, Figure 29*).

### Digital breast tomosynthesis quality control

All DBT systems used in the study were tested prior to the start of the trial by the National Co-ordinating Centre for the Physics of Mammography (NCCPM) to ensure that the 2D imaging performance met the minimum standards required by the NHSBSP and to establish baseline DBT and 2D performance.<sup>84</sup> Each system was also tested by physicists on installation, prior to clinical use, and every 6 months for the duration of the trial. Physicists from NCCPM worked with the local physics service at each participating centre to establish quality control (QC) procedures. Full details of QC testing are reported in *Appendix 6*.

In addition, standardised routine QC tests on workstations were undertaken by radiographic staff on a weekly basis in accordance with manufacturer's guidelines.

### Prospective data collection

The purpose of this was to create a ground truth database as a reference for the reading study by collecting detailed information on each case recruited. The database collated data from prospective reading and histopathology reports.

### Review of images from assessment clinic cases

The 2D and DBT images were reviewed in the assessment clinic by one reader and used to inform subsequent patient management. Data were collected prospectively for each case using proforma data collection sheets (see *Appendix 7*). Where a case had multiple lesions, a separate sheet was used for each.

For each case, the position of any lesion seen in the breast was marked with an X on a grid consisting of nine squares. Suspicion was scored on a standard five-point scale.<sup>85</sup> Lesion type (0–5 scale) and lesion conspicuity (0–3 scale) were also recorded. Readers were also asked to give their opinion on the overall performance of 2D versus DBT on a –2 to +2 preference scale and comment on any additional information obtained from the DBT images. They also recorded an assessment of breast density on a 10-cm visual analogue scale (VAS). The outcome of the assessment was recorded.

### **Review of images from high-/moderate-risk screening mammography cases**

The DBT and 2D images were reviewed by two readers as independently and data recorded on proforma data collection sheets (see *Appendix 8*).

Each reader reviewed the DBT and 2D images and, for each case, the position of any lesion seen in the breast was marked with an X on a grid consisting of nine squares. Suspicion was scored on a standard five-point scale.<sup>85</sup> Lesion type (0–5 scale) and lesion conspicuity (0–3 scale) were also recorded. Readers were also asked to give their opinion on the overall performance of DBT versus 2D on a –2 to +2 preference scale and comment on any additional information obtained from the DBT images. They also recorded an assessment of breast density on a 10-cm VAS. An overall decision of recall/no recall was recorded.

### **Assessment of breast density**

Qualitative or quantitative measurement of breast density from 2D mammograms is known to be highly subjective and variable.<sup>86</sup> DBT has the potential to enable direct measurement of volumetric radiological density.<sup>87</sup> Each reader recorded a rating of breast density on a 10-cm VAS from the DBT and 2D images for each case. These were converted into percentages and compared with the automated breast density calculated using Quantra™ Version 2.0 (Hologic Inc., Bedford, MA, USA) and Volpara™ Version 1.4.2 (Matakina Technology Limited, Wellington, New Zealand) software packages. The methods and findings from these comparisons are described later in this report (see *Chapter 4*).

### **Histopathology data**

Histopathology from core biopsy or surgical excision was used as the gold standard to confirm the presence of a cancer. The outcome of assessment procedures, for example core biopsy and/or surgical excision, was collected at each site (see *Appendix 9*) and collated with the reading data for each case to generate a ground truth database.

### **Prospective data checks**

Detailed logic checks were undertaken of all prospective assessment data, particularly in cases where there was a discrepancy between the initial assessment data and the final histology data, to ensure that all cancer cases were identified. The majority of cases that had initially been scored as suspicious of malignancy were subsequently scored as benign based on the DBT images or review of prior images from other centres, or after ultrasound examination had been performed.

In addition, the FH prospective data were thoroughly checked for data accuracy and the principal investigator (PI) at each site was asked to confirm that all cancer pathology data had been sent to the trial office.

## **Retrospective reading study**

### **Study data set**

Images and prospective data collection information, clinical outcome and histopathology data were collated into a ground truth database from which cases could be identified to distribute to centres for the retrospective reading study.

### Reading arms

It has not yet been determined whether or not DBT could be used as a stand-alone imaging modality for breast screening or whether or not it should be used in addition to 2D. Since there is some uncertainty over the visibility of microcalcification clusters,<sup>17,43,44,48,49,51</sup> it has been suggested that a standard 2D mammogram may be required along with two-view DBT for optimal microcalcification assessment.<sup>16,25,47,53,88</sup> However, concerns have been raised regarding the additional radiation dose this would involve.

Software has become available that creates a synthetic 2D image from a single DBT scan, simulating a conventional 2D image. The combination of DBT with 2D requires approximately doubling the radiation dose to the breast being imaged. If it can be demonstrated that synthetic 2D images are satisfactory and comparable to 2D (in combination with DBT), double exposure could potentially be eliminated. Therefore, in this study the diagnostic performance of three imaging regimens were compared:

- (a) two-view 2D
- (b) two-view 2D + two-view DBT
- (c) two-view synthetic 2D + two-view DBT.

### Synthetic two-dimensional images

Two separate batches of DBT images were exported from the repository at the University of Cambridge onto portable encrypted hard drives. These were sent to Hologic Inc. for production of synthetic 2D images using conversion software. On return, all images were uploaded into the repository. For logistical reasons, images from some cases were not sent for conversion.

### Randomisation process and distribution of images to readers

Cases were randomly selected from the study data set to be distributed between readers and centres to minimise bias. Only cases with a complete set of images, that is left and right MLO and left and right CC, were included in the retrospective study. The only exceptions to this were women who had undergone mastectomy. Each reading set comprised approximately 40 cases per reader per week, and consisted of a mix of normal, benign and cancer cases.

Cases were randomised using a program called R, version 2.15.1 (R Foundation for Statistical Computing, Vienna, Austria) (see *Appendix 10*). This program was managed by Cambridge Clinical Trials Unit. The tool required a list of reader IDs, capacity to be allocated to individual readers and a list of anonymised case IDs, selected from the study data set, as an input file. It allocated cases with a logic that the recruiting site and the sites in which each of the three arms were read were always different. A text file was generated by the program listing cases for each reader.

## Retrospective data collection

### Image review

Readers reviewed (1) 2D images or (2) 2D + DBT or (3) synthetic 2D + DBT images for any one case and did not review any cases from their own centre. Cases were read on a workstation without access to the original screening mammograms or prior examinations. Readers were blinded to the outcome status of each case and read cases independently of all other readers.

For the 2D-only arm, the location of any suspicious abnormality in the breast was recorded on a nine-square grid on the data collection proforma (see *Appendix 11*). Suspicion was scored on a five-point scale<sup>85</sup> and a decision to recall or not was recorded.

For both DBT arms, slice numbers where the lesion was best demonstrated were also recorded. Lesion visibility, lesion extent, discrimination and an overall opinion of DBT versus 2D were all scored on a preference scale from -2 to +2. Readers also recorded a decision to recall or not recall based on (1) 2D images alone, (2) DBT images alone and (3) 2D + DBT images combined for each case (see *Appendix 12*).

### **Collation and cleaning of data**

The retrospective study data were collated into a database designed using MACRO v4.2.2.3810 (InferMed, London, UK). Data collected on the retrospective proforma sheets were input by the trial team at Cambridge. The trial data manager checked for missing and erroneous data using query management in MACRO. Further checks were carried out by research radiographers. Data queries were relayed back to originating sites and the response recorded. An initial database lock was performed to check the completeness of the data and the download was sent to the statistician for analysis. All queries from the statistician were resolved before the final data analysis was done.

### **Retrospective data checks**

All cases for which a reader recorded a high level of suspicion that was not reflected in the recall data (i.e. the case was marked as 'No Recall') were returned to the reader to confirm accuracy of data. Similarly, cases for which suspicion was low but which were recorded as 'Recall' were returned to reader for clarification.

## **Statistics and analyses**

### **Sample size**

The power calculations assumed that, for any given cancer case, at least one of the reading arms gave the 'correct' answer (malignant or not). This was generally conservative. Assuming that some cancer cases will have been wrongly classified by all three reading arms, this would tend to reduce the number of discordant observations but would increase the absolute difference within the discordant observations. The latter tends to outweigh the former in terms of power. For the main study, we wished to compare 2D mammography with 2D + DBT and synthetic 2D + DBT to detect as statistically significant any improvement of sensitivity or specificity conferred by either of the DBT combinations. In particular, we wanted to determine if the addition of DBT to 2D or synthetic 2D improves the accuracy of detection of small or subtle cancers and in women with dense breasts.

We also wanted to examine whether or not the addition of DBT could prove to be particularly useful for a number of subgroups:

- women aged 40–49 years with moderate or high risk of familial breast cancer
- cancers presenting as soft-tissue masses
- cancers presenting as microcalcifications.

The sample size calculation was powered to allow statistically significant differences to be evaluated for subgroup analyses.

### **Sensitivity**

The smallest expected subgroup of cancers is likely to be around 15% of the total tumour population. In any given subgroup, we postulated a sensitivity for 2D mammography of 85% and for 2D + DBT of 95%. Assuming that both detect a cancer in 80% of cases, that is discordance between the two imaging modalities of 20%, we would expect the percentages seen in *Table 3* to be observed.

**TABLE 3** Potential differential sensitivity of 2D and 2D + DBT; anticipated percentage distribution of cancers identified by each imaging modality

Detected by 2D + DBT	Detected by 2D mammography		Total (%)
	No (%)	Yes (%)	
No	0	5	5
Yes	15	80	95
Total	15	85	100

With a 5% significance level and two-sided testing, to have 90% power to detect the above difference (5% missed by DBT and 15% missed by 2D) as significant requires at least 38 cancers with discordant findings.<sup>89</sup> Thus, 190 cancers (38/0.2) were needed in the subgroup. As stated above, the smallest subgroup was likely to be approximately 15% of the total; therefore, a total of 1267 cancer cases was required. Approximately 18% of cases recalled for assessment are ultimately found to have breast cancer. This implies a total study size of 7000 assessment cases. A study population of this size would have at least 90% power for any subgroup that is at least 15% of the total study and 80% for any subgroup that is at least 11%. We expected that the difference between 2D and 2D + DBT would be larger than that between 2D and DBT alone, and, therefore, these comparisons would also be sufficiently powered.

### Specificity

It might be reasonable to anticipate that the specificity of 2D would be 93% and that the addition of DBT might improve this to 97%. Assuming 90% agreement between the two imaging modalities, negative assessment outcomes would be as shown in *Table 4*.

For 90% power to detect this as significant, we required 62 discordant negative cases in any given subgroup, that is 620 negative cases in total in any given subgroup. Since the subgroups of interest were all expected to be at least 15% of the total study size, we expected 1050 (15% of 7000) subjects in each subgroup, of whom 861 (82%) would be negative. Thus, there will be > 90% power for the postulated difference in specificity. Again, larger differences between 2D and 2D + DBT would be expected, so these are also sufficiently powered.

**TABLE 4** Potential differential specificity; anticipated percentage of non-cancers ruled out by each imaging modality

Ruled out by 2D + DBT	Ruled out by 2D mammography		Total (%)
	No (%)	Yes (%)	
No	0	3	3
Yes	7	90	97
Total	7	93	100

## Analysis

Sensitivities and specificities were calculated for each of the three reading arms, firstly for all cases combined, and then for subgroups by breast density and dominant radiological feature. In addition, sensitivity to cancers was calculated for subgroups by size [invasive and ductal carcinoma in situ (DCIS) separately] and histological grade. In view of the matched nature of the data (i.e. that each imaging modality is applied to the same cases), analysis of binary outcomes (e.g. presence or absence of a specific feature) was by McNemar methods.<sup>90</sup> This implied that, for a comparison of two imaging modalities, only cases with non-missing data for each modality were included. Typical data for such analysis can be tabulated as in *Table 5*.

The formal comparison of sensitivity of the two imaging modalities depends only on the discordant observations *b*, cancers seen only on 2D, and *c*, cancers seen only on DBT. If both are equally sensitive, *b* and *c* will be approximately equal, that is one modality misses as many cases as the other, although not necessarily the same individual cases. If DBT has superior sensitivity, *c* will tend to be larger than *b*. The McNemar inference depends on the difference between these two discordant totals, *b* and *c*.

A similar comparison of discordant totals among the subjects with a non-cancer outcome of the complete assessment episode was made to assess the significance of the difference in specificities.

The outcomes from the three arms of the reading study were compared with the gold standard of the final histopathological verification of the presence of benign or malignant disease. If a woman was returned to routine screening this was deemed a normal case.

In addition to calculation of sensitivities and specificities for histologically diagnosed cancer, we also estimated and compared ROC curves based on the five-point suspicion score of the radiologists for each imaging modality. As a single measure of accuracy, we used the area under the ROC curve (AUC). Significance testing and calculation of 95% confidence intervals (CIs) on the AUCs were performed using the method of De Long *et al.*<sup>91</sup> Statistical analysis was performed using Stata version 10.0 (Stata Corporation, College Station, TX, USA).

**TABLE 5** Cancers diagnosed at assessment tabulated by detection method

Detected by 2D + DBT	Detected by 2D mammography		Total
	No	Yes	
No	<i>a</i>	<i>b</i>	<i>a + b</i>
Yes	<i>c</i>	<i>d</i>	<i>c + d</i>
Total	<i>a + c</i>	<i>b + d</i>	<i>a + b + c + d</i>

*a*, cancers that were not seen in the reading study by either DBT or 2D; *b*, cancers seen only on 2D, not on DBT; *c*, cancers seen only on DBT; *d*, cancers seen on both 2D and DBT.





## Chapter 3 Reader study

### Introduction

Digital breast tomosynthesis is a relatively new imaging modality, and few readers have extensive experience of it in a screening setting. This study aimed to establish whether or not increased experience of reading DBT images, as measured at the start and end of the TOMMY trial (a comparison of TOMosynthesis with digital MammographY in the UK NHS Breast Screening Programme), altered performance in terms of recall rate and cancer detection. This knowledge is important when designing training regimes for new readers.

### Aims

The specific aims of the reader study were:

1. to investigate variability in performance of readers involved in the TOMMY trial
2. to determine if there was a change in individual performance over the course of the trial
3. to determine if performance is related to prior experience with DBT or digital mammography
4. to determine if change in performance over the course of the trial is related to the number of DBT images read during the trial
5. to determine whether or not change in performance during the course of the trial is related to previous experience.

### Method

In total, 80 DBT cases were identified from those obtained prior to the TOMMY trial at King's College Hospital in London. Of these, 39 cases contained either no abnormality or benign abnormalities, while 41 cases contained a cancer. Ground truth was provided, with cases classified as normal, benign, malignant in situ or malignant invasive (*Table 6*).

Twenty-eight readers from TOMMY trial sites assessed the set of 80 DBT cases, completing a proforma for every lesion identified. Of these, 22 of the readers assessed the image set on two occasions, before the start of the trial and at the end of it. Results are presented for those readers who completed both reads (*Table 7*). The order in which the cases were assessed was randomised for each read.

For these results we have focused on the suspicion score allocated to each case (*Table 8*).

**TABLE 6** Test set by case type

Case type	Number of cases
Normal	7
Benign	32
Malignant (in situ)	15
Malignant (invasive)	26

**TABLE 7** Recall rates based on a suspicion score of 3 or more

Reader study ID	T1		T2		Δ (T2 – T1)	
	Proportion of cancer cases recalled	Proportion of normal/benign cases not recalled	Proportion of cancer cases recalled	Proportion of normal/benign cases not recalled	Proportion of cancer cases recalled	Proportion of normal/benign cases not recalled
R2	0.98	0.64	0.85	0.62	-0.12	-0.03
R3	1.00	0.18	0.95	0.44	-0.05	0.26
R4	0.80	0.56	0.83	0.63	0.02	0.07
R6	0.95	0.54	0.90	0.46	-0.05	-0.08
R7	0.93	0.33	0.93	0.38	0.00	0.05
R10	0.93	0.67	0.93	0.62	0.00	-0.05
R13	0.95	0.26	0.95	0.26	0.00	0.00
R14	0.95	0.54	1.00	0.35	0.05	-0.19
R15	0.95	0.21	0.98	0.54	0.02	0.33
R16	0.93	0.51	0.93	0.49	0.00	-0.03
R17	0.93	0.51	0.95	0.46	0.02	-0.05
R18	1.00	0.08	1.00	0.31	0.00	0.23
R19	1.00	0.18	0.95	0.45	-0.05	0.27
R20	1.00	0.18	1.00	0.46	0.00	0.28
R21	1.00	0.21	0.95	0.38	-0.05	0.18
R22	1.00	0.18	1.00	0.21	0.00	0.03
R23	0.98	0.28	1.00	0.51	0.02	0.23
R24	0.90	0.56	0.93	0.49	0.02	-0.08
R25	0.93	0.54	0.95	0.53	0.02	-0.01
R26	0.98	0.15	0.95	0.25	-0.02	0.10
R27	1.00	0.44	1.00	0.16	0.00	-0.28
R28	0.95	0.44	0.98	0.26	0.02	-0.18
All readers	0.96	0.37	0.95	0.42	0.01	0.01

T1, time point 1; T2, time point 2.

**TABLE 8** Suspicion classification (based on Maxwell *et al.*<sup>85</sup>)

Classification	Suspicion
1	Normal
2	Benign
3	Probably benign (recall)
4	Suspicion (recall)
5	Malignant (recall)

## Results

The results are shown in *Table 7*. A recall corresponded to a suspicion score of 3 or more. The proportion of cancer cases correctly recalled ranged from 0.80 to 1.00 at the first time point and from 0.83 to 1.00 at the second time point. The proportion of normal/benign cases not recalled ranged from 0.08 to 0.67 at the first time point and from 0.16 to 0.63 at the second time point. However, the number of cases that were actually normal was very low (7 out of 80 cases) so we would expect a significant recall rate for normal/benign cases. There was no significant difference between the recall rates before and after the trial ( $p = 0.501$  cancer,  $p = 0.198$  normal/benign).

*Table 9* shows the results with a higher threshold on the suspicion score. The proportion of cancer cases correctly identified as suspicious or malignant ranged from 0.61 to 0.88 at the first time point and from 0.61 to 0.95 at second time point. The proportion of normal/benign cases scored normal, benign or

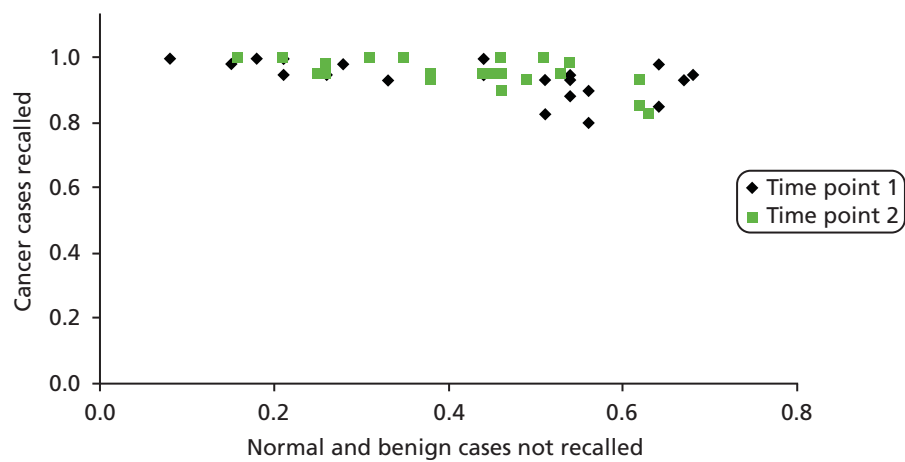
**TABLE 9** Proportion of cancer cases identified as suspicious or malignant, and the proportion of normal/benign cases scored normal, benign or probably benign

Reader study ID	T1		T2		$\Delta$ (T2 – T1)	
	Proportion of cancer cases scored 4 or 5	Proportion of normal/benign cases scored 1 to 3	Proportion of cancer cases scored 4 or 5	Proportion of normal/benign cases scored 1 to 3	Proportion of cancer cases scored 4 or 5	Proportion of normal/benign cases scored 1 to 3
R2	0.71	0.90	0.61	0.87	-0.10	-0.03
R3	0.80	0.82	0.66	0.90	-0.15	0.08
R4	0.63	0.97	0.66	0.95	0.02	-0.03
R6	0.66	0.95	0.68	0.92	0.02	-0.03
R7	0.76	0.87	0.76	0.82	0.00	-0.05
R10	0.71	0.85	0.76	0.87	0.05	0.03
R13	0.73	0.82	0.80	0.62	0.07	-0.21
R14	0.61	0.92	0.68	0.97	0.07	0.05
R15	0.88	0.64	0.88	0.74	0.00	0.10
R16	0.80	0.89	0.76	0.79	-0.04	-0.10
R17	0.71	0.85	0.76	0.82	0.05	-0.03
R18	0.76	0.87	0.95	0.67	0.19	-0.21
R19	0.68	0.95	0.68	0.79	0.00	-0.15
R20	0.83	0.74	0.80	0.82	-0.02	0.08
R21	0.83	0.82	0.80	0.87	-0.02	0.05
R22	0.85	0.74	0.88	0.77	0.02	0.03
R23	0.83	0.82	0.78	0.90	-0.05	0.08
R24	0.71	0.85	0.66	0.92	-0.05	0.08
R25	0.71	0.85	0.76	0.77	0.05	-0.08
R26	0.88	0.62	0.90	0.69	0.02	0.08
R27	0.83	0.77	0.93	0.63	0.10	-0.14
R28	0.71	0.87	0.71	0.85	0.00	-0.03
All readers	0.75	0.84	0.77	0.82	0.03	-0.03

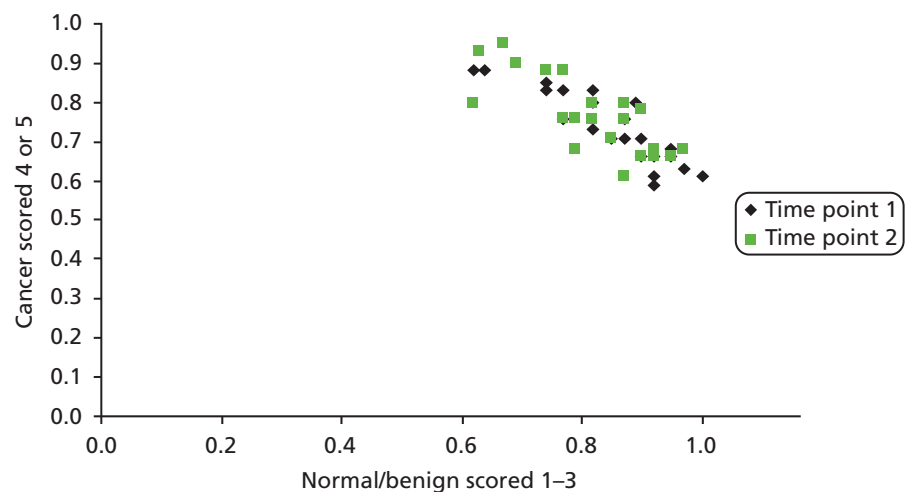
T1, time point 1; T2, time point 2.

probably benign ranged from 0.62 to 1.00 at the first time point and from 0.62 to 0.97 at the second time point. Once again there was no significant difference between time points ( $p = 0.470$  and  $p = 0.339$ , respectively). Data from *Tables 7* and *9* are plotted in *Figures 3* and *4*, respectively.

*Tables 10–12* show the results ranked by reader experience. *Table 10* shows the change in performance between reads ranked according to the volume of DBT undertaken within the trial between the two time points. *Table 11* shows the change in performance ranked according to full-field digital mammography (FFDM) experience (number of mammograms read) in the 12 months prior to the trial and *Table 12* shows the baseline performance ranked by FFDM experience based on the number of mammograms read during the year. *Table 13* shows the results ranked by number of years reading mammograms and *Tables 14* and *15* show the analysis by site. *Figure 4* shows performance at the initial read by FFDM load in the previous 12 months and *Figure 5* by number of years reading mammograms prior to the study. *Figures 6, 7* and *8* illustrate performance by site at the two time points.



**FIGURE 3** Proportion of cancer cases recalled, plotted against the proportion of normal/benign cases not recalled at time point 1 (◆) and time point 2 (■).



**FIGURE 4** Proportion of cancer cases identified as suspicious or malignant, and the proportion of normal/benign cases scored normal, benign or probably benign at time point 1 (◆) and time point 2 (■).

**TABLE 10** Change in performance between the two reads ranked according to the number of DBT reads undertaken between the two time points. Experience of digital mammography film reading in the 12 months prior to the first read is also presented

Reader study ID	Number of digital DBT reads	Number of DBT reads between the two time points	Proportion of cancer cases recalled	Proportion of cancer cases scored 4 or 5	Proportion of normal/benign cases not recalled	Proportion of normal/benign cases scored 1, 2 or 3
			$\Delta(T2 - T1)$	$\Delta(T2 - T1)$	$\Delta(T2 - T1)$	$\Delta(T2 - T1)$
R7	500	115	0.00	0.00	-0.05	0.05
R6	1200	140	-0.05	0.02	-0.03	-0.08
R23	8500	300	0.02	-0.05	0.08	0.23
R10	4063	357	0.00	0.05	0.03	-0.05
R19	5000	452	-0.05	0.00	-0.15	0.28
R21	5643	475	-0.05	-0.02	0.05	0.18
R14	1800	489	0.05	0.07	0.05	-0.21
R13	1200	499	0.00	0.07	-0.21	0.00
R15	75	506	0.02	0.00	0.10	0.33
R26	-	518	-0.02	0.02	0.08	0.10
R16	1000	519	0.00	-0.04	-0.10	-0.03
R2	3000	520	-0.12	-0.10	-0.03	-0.03
R17	800	544	0.02	0.05	-0.03	-0.05
R4	1000	593	0.02	0.02	-0.03	0.07
R24	6700	616	0.02	-0.05	0.08	-0.08
R25	10,000	647	0.02	0.05	-0.08	-0.01
R22	10,000	658	0.00	0.02	0.03	0.03
R3	-	673	-0.05	-0.15	0.08	0.26
R28	6500	674	0.02	0.00	-0.03	-0.18
R27	8000	675	0.00	0.10	-0.14	-0.28
R18	8000	697	0.00	0.19	-0.21	0.23
R20	9000	714	0.00	-0.02	0.08	0.28

**TABLE 11** Change in performance between the two reads ranked according to the number of digital mammograms read in the year prior to the first DBT read. Experience of DBT reading between the two time points is also presented

Reader study ID	Number of digital DBT reads	Number of DBT reads between the two time points	Proportion of cancer cases recalled	Proportion of cancer cases scored 4 or 5	Proportion of normal/benign cases not recalled	Proportion of normal/benign cases scored 1, 2 or 3
			$\Delta(T2 - T1)$	$\Delta(T2 - T1)$	$\Delta(T2 - T1)$	$\Delta(T2 - T1)$
R3	–	673	–0.05	–0.15	0.08	0.26
R26	–	518	–0.02	0.02	0.08	0.10
R15	75	506	0.02	0.00	0.10	0.33
R7	500	115	0.00	0.00	–0.05	0.05
R17	800	544	0.02	0.05	–0.03	–0.05
R4	1000	593	0.02	0.02	–0.03	0.07
R16	1000	519	0.00	–0.04	–0.10	–0.03
R6	1200	140	–0.05	0.02	–0.03	–0.08
R13	1200	499	0.00	0.07	–0.21	0.00
R14	1800	489	0.05	0.07	0.05	–0.21
R2	3000	520	–0.12	–0.10	–0.03	–0.03
R10	4063	357	0.00	0.05	0.03	–0.05
R19	5000	452	–0.05	0.00	–0.15	0.28
R21	5643	475	–0.05	–0.02	0.05	0.18
R28	6500	674	0.02	0.00	–0.03	–0.18
R24	6700	616	0.02	–0.05	0.08	–0.08
R18	8000	697	0.00	0.19	–0.21	0.23
R27	8000	675	0.00	0.10	–0.14	–0.28
R23	8500	300	0.02	–0.05	0.08	0.23
R20	9000	714	0.00	–0.02	0.08	0.28
R22	10,000	658	0.00	0.02	0.03	0.03
R25	10,000	647	0.02	0.05	–0.08	–0.01

**TABLE 12** Performance at the first read ranked by experience of reading FFDM in the previous 12 months

Reader study ID	Number of digital DBT reads	Proportion of cancer cases recalled	Proportion of cancer cases scored 4 or 5	Proportion of normal/benign cases not recalled	Proportion of normal/benign cases scored 1, 2 or 3
R3	–	1.00	0.80	0.18	0.82
R26	–	0.98	0.88	0.15	0.62
R15	75	0.95	0.88	0.21	0.64
R7	500	0.93	0.76	0.33	0.87
R17	800	0.93	0.71	0.51	0.85
R4	1000	0.80	0.63	0.56	0.97
R16	1000	0.93	0.80	0.51	0.89
R6	1200	0.95	0.66	0.54	0.95
R13	1200	0.95	0.73	0.26	0.82
R14	1800	0.95	0.61	0.54	0.92
R2	3000	0.98	0.71	0.64	0.90
R10	4063	0.93	0.71	0.67	0.85
R19	5000	1.00	0.68	0.18	0.95
R21	5643	1.00	0.83	0.21	0.82
R28	6500	0.95	0.71	0.44	0.87
R24	6700	0.90	0.71	0.56	0.85
R18	8000	1.00	0.76	0.08	0.87
R27	8000	1.00	0.83	0.44	0.77
R23	8500	0.98	0.83	0.28	0.82
R20	9000	1.00	0.83	0.18	0.74
R22	10,000	1.00	0.85	0.18	0.74
R25	10,000	0.93	0.71	0.54	0.85

**TABLE 13** Performance at the first read ranked by years of experience reading mammograms

Reader study ID	Years experience reading mammograms	Proportion of cancer cases		Proportion of normal/benign cases not recalled	Proportion of normal/benign cases scored 1, 2 or 3
		Recalled	Scored 4 or 5		
R13	25	0.95	0.73	0.26	0.82
R23	24	0.98	0.83	0.28	0.82
R25	24	0.93	0.71	0.54	0.85
R18	20	1.00	0.76	0.08	0.87
R17	18	0.93	0.71	0.51	0.85
R20	18	1.00	0.83	0.18	0.74
R2	17	0.98	0.71	0.64	0.90
R27	17	1.00	0.83	0.44	0.77
R6	15	0.95	0.66	0.54	0.95
R15	10	0.95	0.88	0.21	0.64
R19	10	1.00	0.68	0.18	0.95
R14	10	0.95	0.61	0.54	0.92
R16	7	0.93	0.80	0.51	0.89
R22	7	1.00	0.85	0.18	0.74
R28	6	0.95	0.71	0.44	0.87
R24	5	0.90	0.71	0.56	0.85
R4	4	0.80	0.63	0.56	0.97
R10	4	0.93	0.71	0.67	0.85
R21	4	1.00	0.83	0.21	0.82
R7	1.5	0.93	0.76	0.33	0.87

**TABLE 14** Recall rates (suspicion score of 3 or more) by site

Site	n	T1		T2		Change (T2 – T1)	
		Proportion of cancer cases recalled	Proportion of normal/benign cases not recalled	Proportion of cancer cases recalled	Proportion of normal/benign cases not recalled	Proportion of cancer cases recalled	Proportion of normal/benign cases not recalled
Aberdeen	5	0.93	0.45	0.89	0.51	-0.04	0.06
Glasgow	5	0.94	0.41	0.96	0.42	0.02	0.01
Jarvis	6	1.00	0.19	0.98	0.39	-0.01	0.20
Manchester	5	0.95	0.43	0.96	0.34	0.01	-0.09

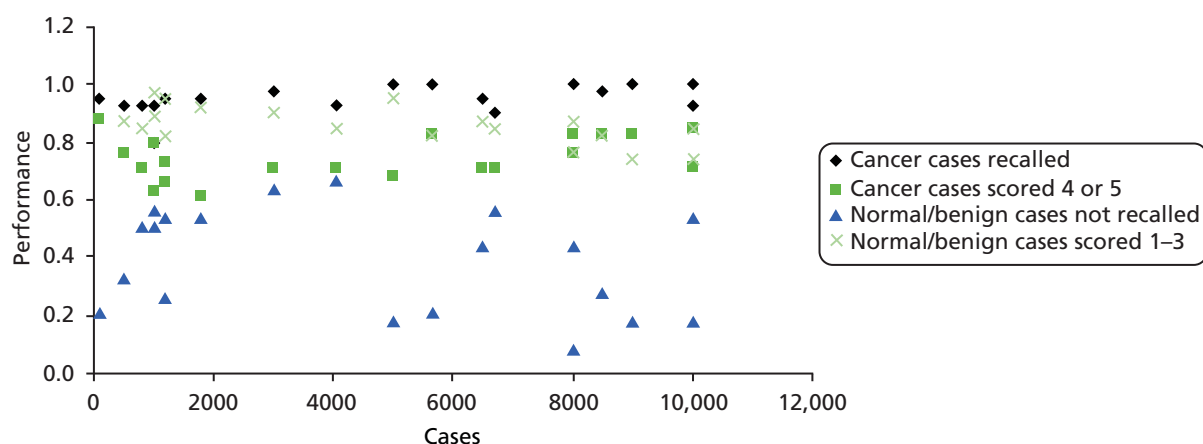
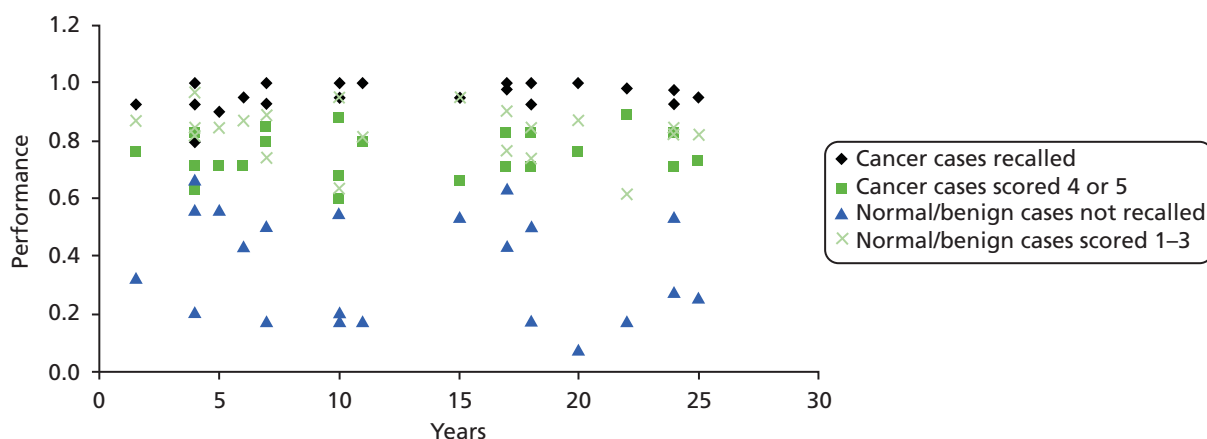
T1, time point 1; T2, time point 2.

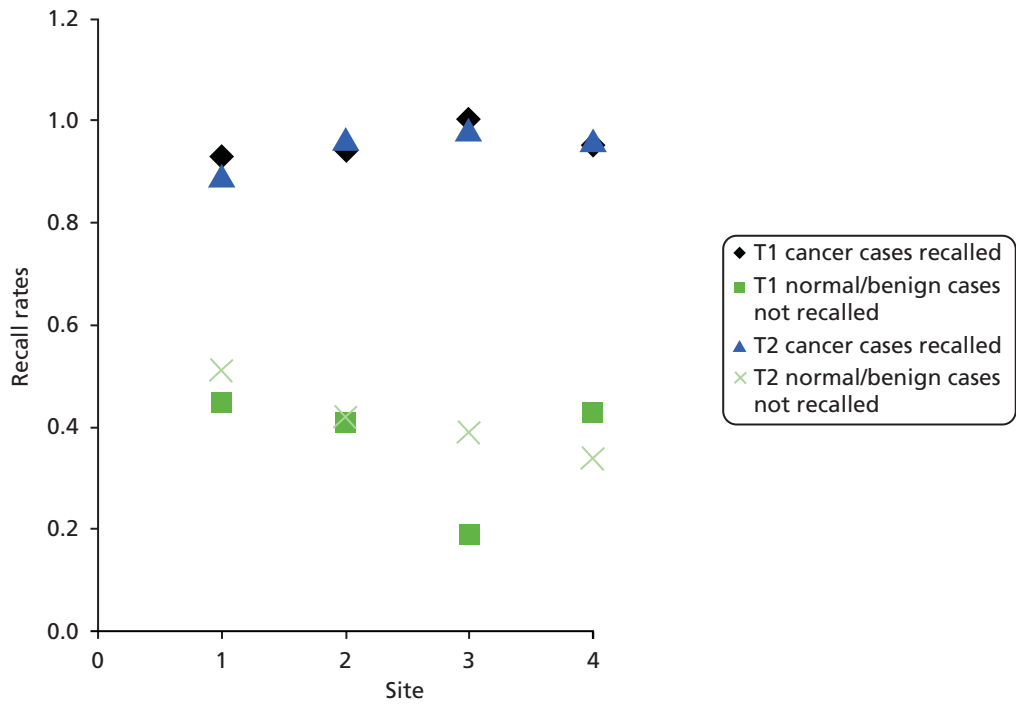


**TABLE 15** Cancer cases identified as suspicious or malignant (score 4 or 5) by site

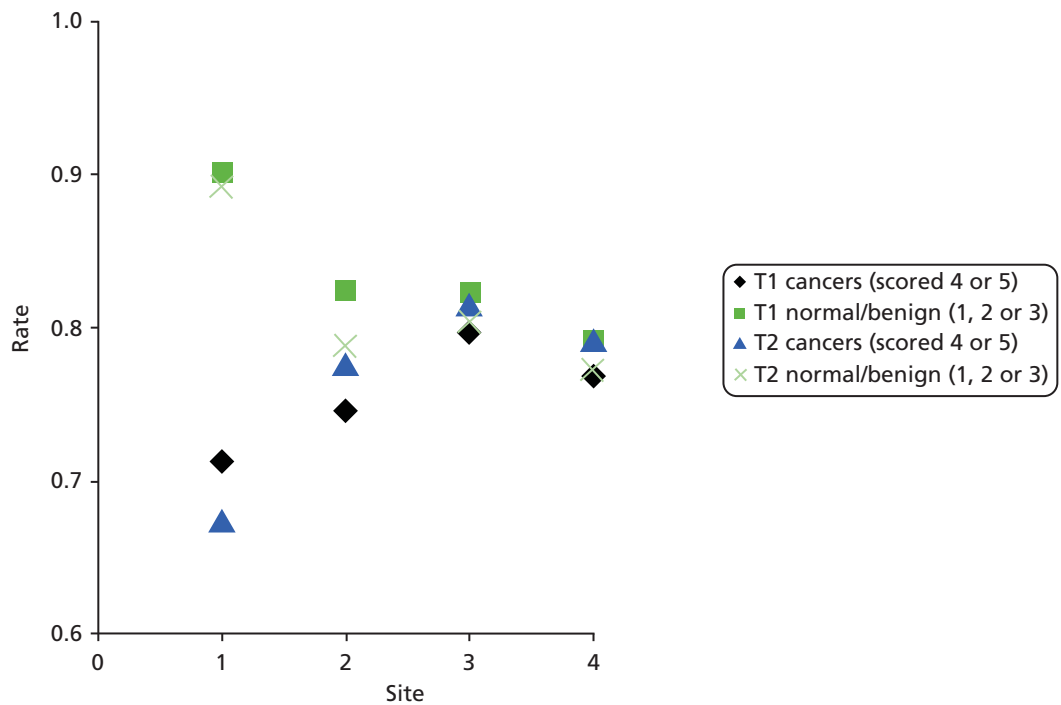
Site	n	T1		T2		Change (T2 – T1)	
		Proportion of cancer cases scored 4 or 5	Proportion of normal/benign cases scored 1, 2 or 3	Proportion of cancer cases scored 4 or 5	Proportion of normal/benign cases scored 1, 2 or 3	Proportion of cancer cases scored 4 or 5	Proportion of normal/benign cases scored 1, 2 or 3
Aberdeen	5	0.71	0.90	0.67	0.89	-0.04	-0.01
Glasgow	5	0.75	0.82	0.78	0.79	0.03	-0.04
Jarvis	6	0.80	0.82	0.82	0.80	0.02	-0.02
Manchester	5	0.77	0.79	0.79	0.77	0.02	-0.02

T1, time point 1; T2, time point 2.

**FIGURE 5** Performance at first read vs. FFDM load in previous 12 months.**FIGURE 6** Performance at the first read ranked by number of years of experience reading mammograms.



**FIGURE 7** Recall rates by site at first (T1) and second (T2) reads. Site 1, Aberdeen; site 2, Glasgow; site 3, Jarvis; and site 4, Manchester. T1, time point 1; T2, time point 2.



**FIGURE 8** Cancer cases identified as suspicious/malignant by site. Site 1, Aberdeen; site 2, Glasgow; site 3, Jarvis; and site 4, Manchester. T1, time point 1; T2, time point 2.

## Discussion

As the data set was rich in difficult cases and benign abnormalities as well as cancers, it was appropriate to look at the proportion of cancers scored 4 or 5 as well as the proportion of cancer cases recalled (scored 3 or more). In particular, one would expect many of the non-cancer cases to be scored as 3 (probably benign), since they contained benign abnormalities.

Even though there was little change with time (*Tables 8 and 9; Figures 2 and 3*), there are differences between readers, particularly when looking at cancer cases scored 4 and 5 (suspicious or malignant). In *Figure 3*, it is apparent that readers choose an operating point that balances the cancer detection rate against the proportion of normal cases scored as normal to probably benign.

Previous experience did not predict performance, in terms of either recall rates or detection of cancer. The greatest variability between readers was seen in the proportion of normal cases not recalled.

When assessing whether or not DBT should be introduced into the screening programme, it is important to consider extra reader time and the impact of reader fatigue. Timing reading with DBT in 2012, when readers were relatively inexperienced, the median time for interpretation of 2D images was 17.0 seconds with an interquartile range of 12.3–23.6 seconds, while for DBT it was 66.0 seconds, with an interquartile range of 51.1–80.5 seconds. The difference was statistically significant ( $p < 0.001$ ). Reading times were significantly longer in FH clinics ( $p < 0.01$ ). Although it took approximately four times longer to interpret DBT than 2D images, the cases were more complex than would be expected for routine screening and had higher mammographic density.<sup>92</sup>

There is some difference when comparing reads at time point 1 and time point 2 for one of the sites (Jarvis). In this instance the proportion of normal/benign lesions that were not recalled increased by 20% over the course of the trial. In Manchester there was a 9% decrease in the same parameter. The proportion of cancers identified as suspicious or malignant (4 or 5) was similar to that of cases identified as normal/benign (1, 2 or 3) at most sites, except for Aberdeen, where the difference was 19% at the first read and 22% at the second read.

The lack of change over time indicates that a larger set of training images is probably unnecessary, although the intersite variations indicate that individuals operate according to local practice even under test conditions.



# Chapter 4 Breast density assessment

## Introduction

Although the risk of developing breast cancer is dependent on the cumulative impact of a wide range of risk factors,<sup>93</sup> increasing mammographic density has been shown to be one of the strongest independent and modifiable predictors of breast cancer risk.<sup>4,94–100</sup>

Women with high breast density have been reported to have a four to six times increased risk of developing breast cancer compared with those with low breast density,<sup>95,101–105</sup> and high breast density has also been linked to an increased risk of cancers not detected at screening,<sup>4,8,96,106,107</sup> larger tumour size<sup>108–110</sup> and positive lymph nodes.<sup>109,111–113</sup> The underlying cause of these links are thought to be numerous, and early studies hypothesised that a significant reason for an increase in breast cancer incidence with higher density breasts was as a result of a 'masking bias' that made mammographic screening less sensitive to cancer detection.<sup>96,101</sup> Later studies, however, have shown that there is increased risk for at least 7–10 years following a screening examination, indicating that 'masking bias' is only one of the mechanisms linking breast density to an increased cancer risk.<sup>94,101,114</sup> In addition, the increased radiation dose required in dense breasts and cumulative lifetime exposure from screening indicates that there may be a less favourable benefit to harm ratio associated with screening of women with dense breasts, particularly in younger women.<sup>93</sup>

These issues are of particular concern for the NHSBSP, as it is now extending its programme<sup>14</sup> to include younger, pre- or perimenopausal women, who are known to have a higher proportion of dense breast tissue,<sup>5,115</sup> and it is also potentially problematic for women aged 40–49 years at moderate or high risk of developing familial breast cancer attending for annual mammography.<sup>81</sup>

It has been suggested that population mammographic screening might be more effective if screening strategies were tailored according to mammographic breast density<sup>79</sup> with more frequent screening or use of magnetic resonance imaging (MRI) or adjunctive ultrasound to improve detection in women with dense breasts.<sup>99,116–119</sup> In this context, a campaign to include breast density in mammography reporting is currently being debated in the USA.<sup>120</sup> Legislation has been passed in several states mandating that breast density be reported using the Breast Imaging Reporting and Data System (BI-RADS) scale, with women with > 50% breast density offered supplemental screening.<sup>121</sup>

## Measurement of breast density

The radiographic appearance of the breast on a mammogram reflects variations in the relative amounts of fat, connective tissue and epithelial tissue and their different X-ray attenuation characteristics, with breast density expressed as a percentage of the mammogram occupied by radiologically dense fibroglandular tissue.<sup>101</sup> Both qualitative and quantitative methods currently used to assess breast density by mammography have limits, since they are based on the projected area rather than the volume of breast tissue, are time-consuming, and are subject to inter- and intrareader variability.<sup>97,122</sup> For breast density assessment to be incorporated into a population-based screening programme such as the NHSBSP, efficient automated methods validated against screening programme outcomes such as sensitivity, interval cancer rates and tumour size at diagnosis are required.

Traditionally, assessment of mammographic density is performed by the radiologist evaluating the mammogram. He/she makes a judgement on the information presented in all the mammographic views in order to present a single score for each examination. Consistency of this measure requires an experienced

observer to be able to correctly assess the relative proportions of glandular and fatty tissue while accounting for variations in breast shape, radiographic texture and the presence of cancer (leading to a localised increase in density). The radiologist is also able to take account of the variation in the radiographic acquisition of the mammogram. These density scores can then be presented as a percentage on a continuous scale or within discrete ranges such as the composition categories used in BI-RADS.<sup>94</sup>

Although moderate agreement has been shown between observers in assigning percentage scores of breast density, studies suggest that training and experience are essential in ensuring that those scores are accurate and reproducible.<sup>123,124</sup> In addition, each manufacturer applies its own image processing, which is often designed to minimise the effect of density in the mammogram, further increasing the difficulty of density assessment (Dr Ralph Highnam, Matakina International Limited, 2013, personal communication). Methods have therefore been investigated to standardise density estimates by use of automated methods. The introduction of FFDM technologies, initially developed for digitised analogue mammograms, has provided an opportunity to implement breast density measurement algorithms.<sup>123–126</sup>

At their simplest, these algorithms work by applying thresholds to the pixel values within the digital image to identify the area of the image that contains the breast and then to determine the proportion of that breast that is dense. For example, the pixel values with the highest signal can be seen to be the areas of the image where no breast tissue has attenuated the primary X-ray beam. The areas of lowest signal, on the other hand, represent areas where the X-rays have passed through a section of tissue that is relatively most attenuating.<sup>127</sup>

Later developments have led to software that estimates the volume of dense fibroglandular tissue rather than just the area projected onto the mammogram. By using the image pixel data in combination with information about its acquisition found in the DICOM data file, sophisticated algorithms are able to provide measurements of the relevant tissue volumes. For example, data regarding the breast thickness and the X-ray exposure's tube potential, current, time, target and filter – in combination with knowledge of the radiation attenuation properties of different tissues – can enable a derivation of the breast composition represented by each pixel.<sup>97,128</sup> Improvements in the measurement are then made through advances in this derivation, for example providing better calibration of the image data and breast thickness estimation.<sup>129</sup>

Two software tools were used in this study. Hologic's Quantra has been validated and found to have good agreement with measurements of breast density from MRI data and reader assessments.<sup>121,128</sup> Matakina's Volpara™ software has similarly found to be in good agreement with MRI and observer measurements of breast density.<sup>121,130</sup>

The aim of this substudy within the trial was to evaluate and utilise two of the commercially available software packages for the measurement of volumetric breast density and compare these with observer-based scores of area density. This data would then be correlated with the results of the retrospective reading study designed to determine if the addition of DBT improved the detection of cancers in women at higher risk of developing breast cancer due to increased breast density. A secondary aim of this study was to assess whether or not there is a relationship between breast density and breast cancer incidence.

## Materials and methods

In the prospective data collection phase of the trial, readers were asked to assess breast density on a 10-cm VAS.<sup>129</sup> The markings on this scale were then converted into a score ranging from 0 to 100%.

The two software packages used to assess the volumetric density of each mammogram were Quantra and Volpara. Each program's output consisted of a number of results, including values for the absolute volume of fibroglandular tissue and overall breast volume as well as the volumetric breast density on a per image basis. Scores from each image of a full examination were then combined to derive a score for each case.

Each software vendor provides an ability to combine these automatically. However, we chose to adapt the underlying logic provided by the Quantra system (Dr Ashwini Kshirsagar, Hologic Inc., 2013, personal communication) to minimise the effect of lesion presence on the result. Their generally applied logic takes the maximum total breast and fibroglandular tissue volumes calculated from the CC and MLO images. The values for left and right breasts are then added and the overall density determined. We believe this to be in line with the behaviour of observers and with findings from prevalence studies that risk may be associated with the density derived on the contralateral mammogram in the absence of prior imaging.<sup>101</sup>

In order to obtain the overall score for each examination, the absolute values of total breast volume and fibroglandular tissue for each of the four views (left and right, CC and MLO) were examined. For cases where no cancer was assessed as being present, the largest results for each breast (from either the CC or the MLO view) were determined and the average volumes of the two breasts were calculated. For cases where cancer was confirmed, results were used from the contralateral breast. If no contralateral data were available, results from the affected breast were used. Volumetric density was calculated from the ratio of the fibroglandular tissue volume to the overall breast volume. The same logic was applied to the scores given by the Quantra system for area breast density, which should nominally be comparable with the observer scores.

With these density scores the aim of this study was to compare the observer's score with the automated techniques (Quantra and Volpara), to establish any age-related correlation with density and to establish if cancer incidence is associated with breast density. Pathology reports were used to confirm cancer cases.

## Results

A total of 8867 women's standard 2D mammograms acquired for the trial were available for analysis. The software was unable to produce scores for every image analysed. Reasons for algorithm failure were varied but included magnification views where there was no background detected; cases in which information given in the DICOM image header was inconsistent with that normally expected in mammography, for example non-female gender selection; and invalid filter types or improbable thicknesses (e.g. 0 cm or > 30 cm).

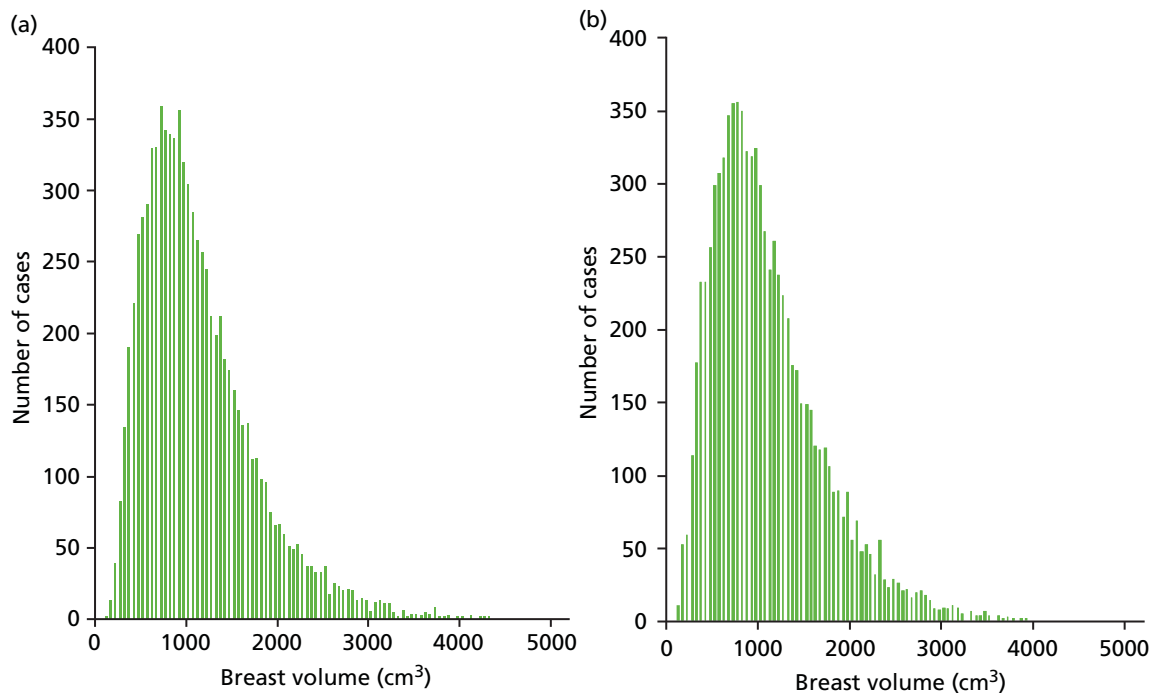
The summary results for the study cohort are shown in *Table 16*. In total, 8391 cases were given an overall density score by the observers on the VAS, 8512 cases were calculated from scores from Quantra analysis of 33,966 images and 8532 cases were calculated from scores from Volpara analysis of 34,755 images.

**TABLE 16** Summary results for the total cohort of cases. Unless indicated, numbers are median (minimum–maximum) values

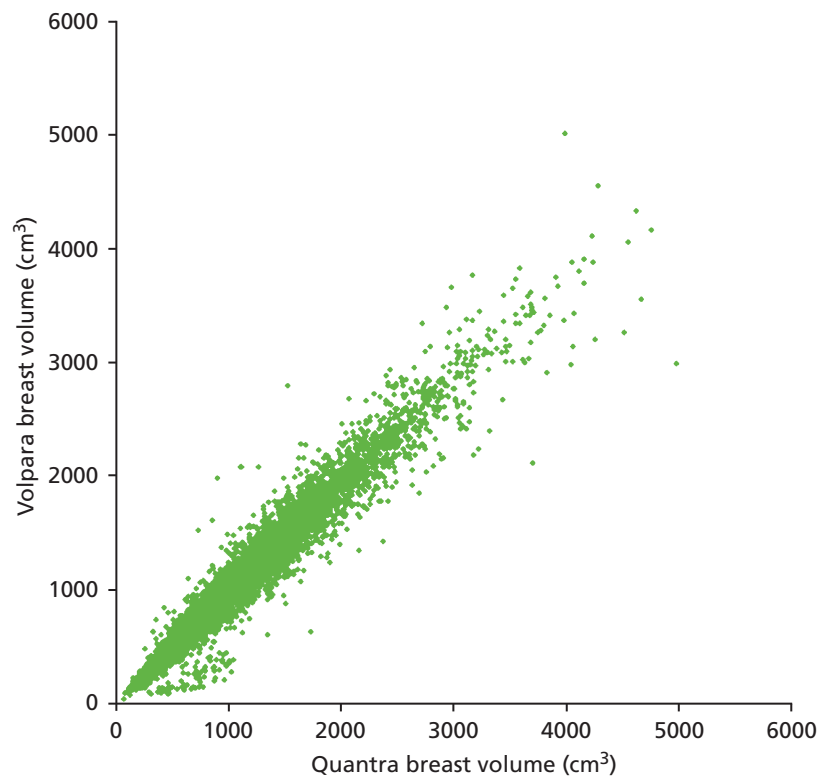
Measurements	Observers	Quantra	Volpara
Cases ( <i>n</i> )	8391	8512	8532
Breast volume (cm <sup>3</sup> )	–	953.5 (73.0–4986.5)	921.4 (33.4–5009.3)
Fibroglandular tissue volume (cm <sup>3</sup> )	–	93.0 (4.0–1024.0)	71.6 (6.8–628.5)
Area breast density (%)	36.8 (0.0–100.0)	14.8 (0.0–76.5)	–
Volumetric breast density (%)	–	9.5 (1.4–56.2)	7.7 (2.5–54.2)

### Total breast volume

Figures 9 and 10 illustrate the distribution and comparison of total breast volume as measured by the Quantra and Volpara software throughout the study population. The two systems show a very good linear correlation with a coefficient of determination of 0.95. The mean difference between the values calculated for each case is 5.04% ( $\pm 0.32\%$ , two standard errors), suggesting good agreement between the two systems.



**FIGURE 9** Histograms showing the distribution of breast volume across the study population as measured by (a) Quantra and (b) Volpara.

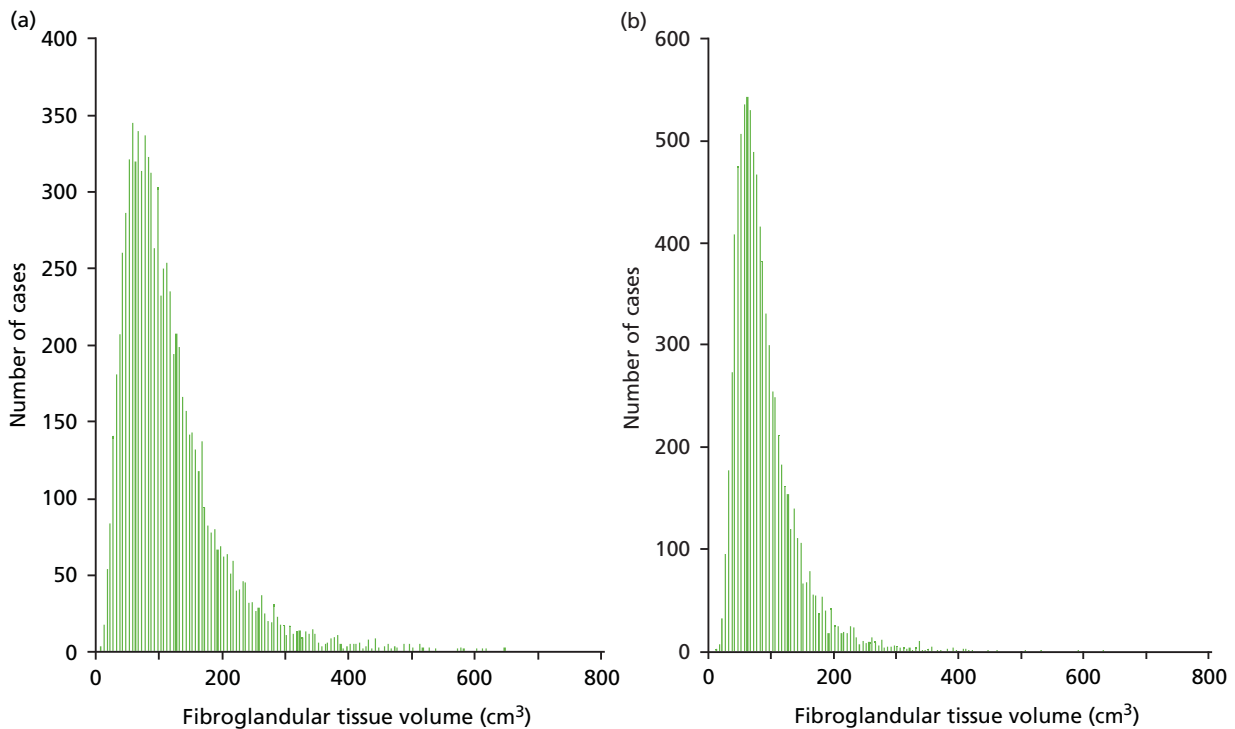


**FIGURE 10** Scatterplot comparing Quantra and Volpara scores for total breast volume.

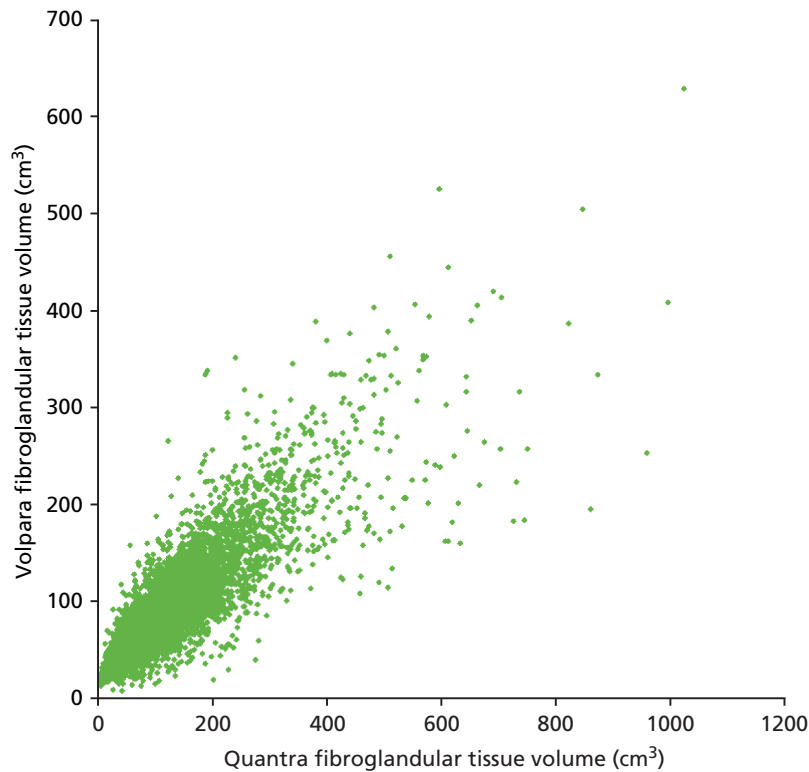


### Fibroglandular tissue volume

Figures 11 and 12 show the results for the fibroglandular tissue volume within the breast. The coefficient of determination for a linear correlation is 0.74 and the mean difference between the values calculated for each case is 21.19% ( $\pm 0.72\%$ ), with Quantra giving the larger of the two values. Our understanding is that this is because of differences in the way that each system handles the density associated with the skin within the assessment of fibroglandular volume. The differences between the two systems appear to get larger as the amount of fibroglandular tissue increases. This is particularly noticeable above 400 cm<sup>3</sup>. In addition, the range of results for fibroglandular volume is greater for the Quantra system, with most values lying between 0 and 400 cm<sup>3</sup> as opposed to Volpara's results lying between 0 and 300 cm<sup>3</sup>.



**FIGURE 11** Histograms showing the distribution of fibroglandular tissue volume across the study population as measured by (a) Quantra and (b) Volpara. The x-axes of each graph are designed to match for clarity but seven cases in the Quantra data set had volumes greater than 800 cm<sup>3</sup>.



**FIGURE 12** Scatterplot comparing Quantra and Volpara scores for fibroglandular tissue volume.

### **Volumetric breast density**

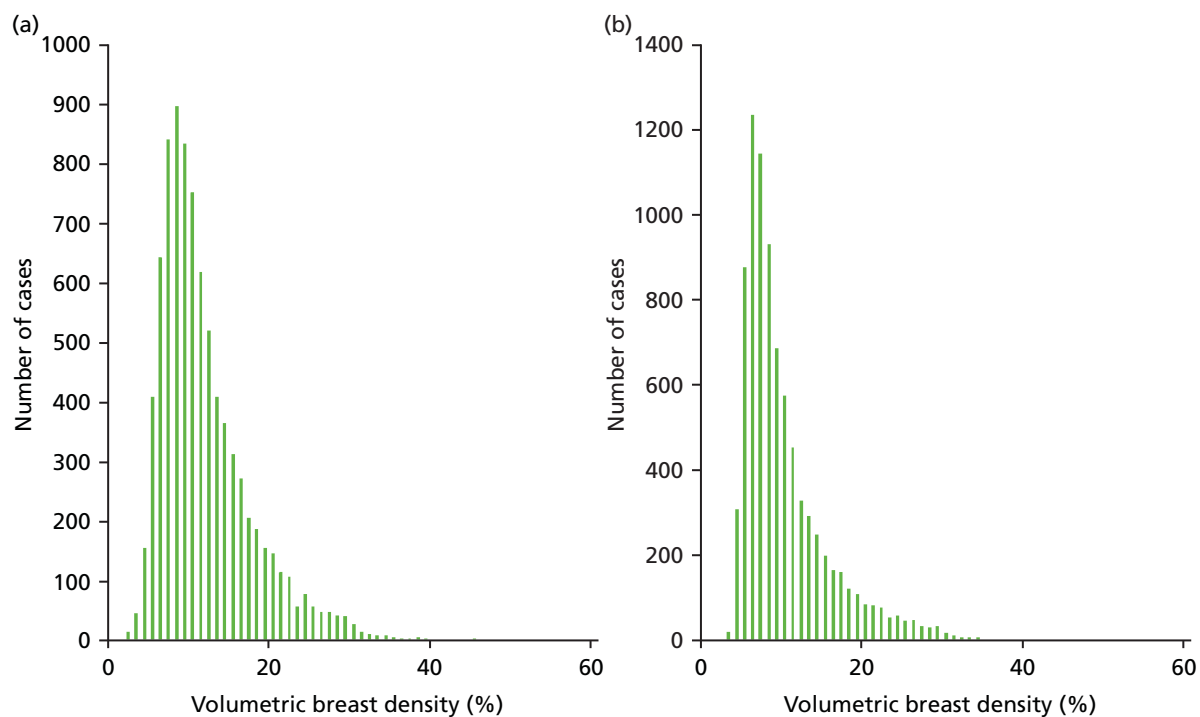
Figures 13 and 14 show the results for the volumetric breast density. The coefficient of determination for a linear correlation is 0.65 and the mean difference between the values calculated for each case is 16.32% ( $\pm 0.69\%$ ).

Figures 15 and 16 show the results for breast density estimated from the projected mammogram, the area-based breast density scored by the observers, and the Quantra software. The Volpara software did not give an area-based density result. There is poor correlation between the two measurements, with a coefficient of determination for a linear correlation of 0.31.

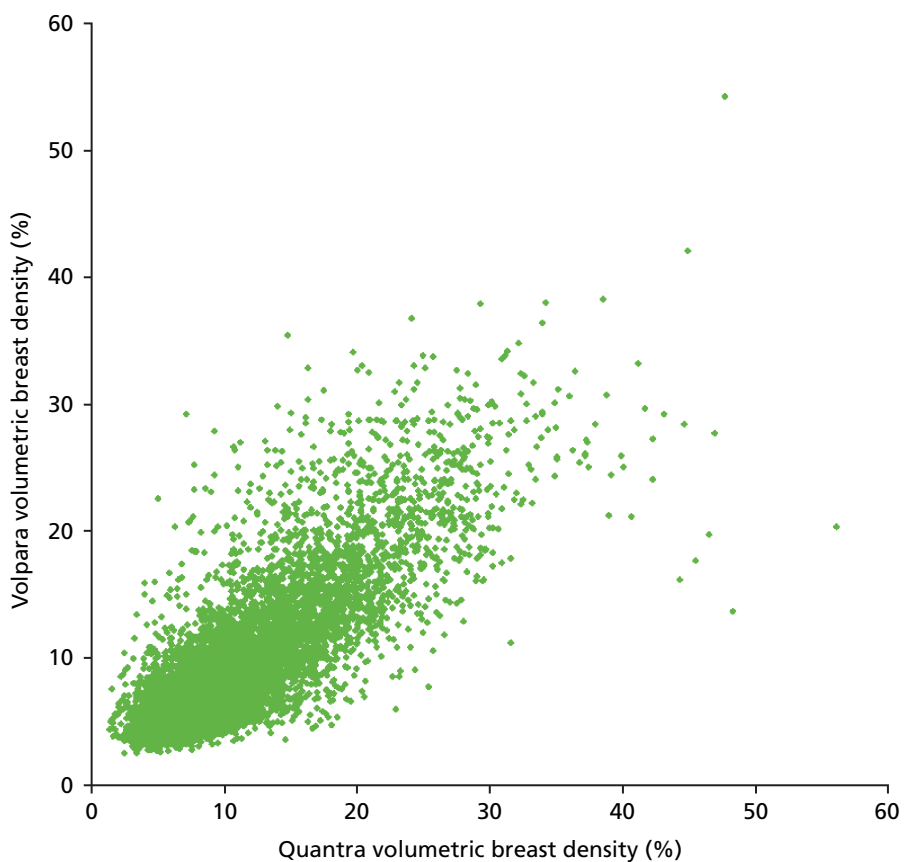
Figure 17 compares the area-based breast density scored by the observers, and the volumetric density measurement from each program. The coefficient of determination for an exponential correlation is 0.33 and 0.38 for the Quantra and Volpara systems, respectively. The large values given at each 5% mark on the observers' histograms in Figures 14 and 16 are a consequence of the way that the VAS scale was processed in two centres, with one rounding results to within the nearest 5% and the other the nearest 10%.

For FH cases, density was assessed by two observers from the same centre, giving us 638 cases with two scores for comparison. In 70% of these cases the score agreed to within 10%; however, 8% of cases disagreed by more than 20%.

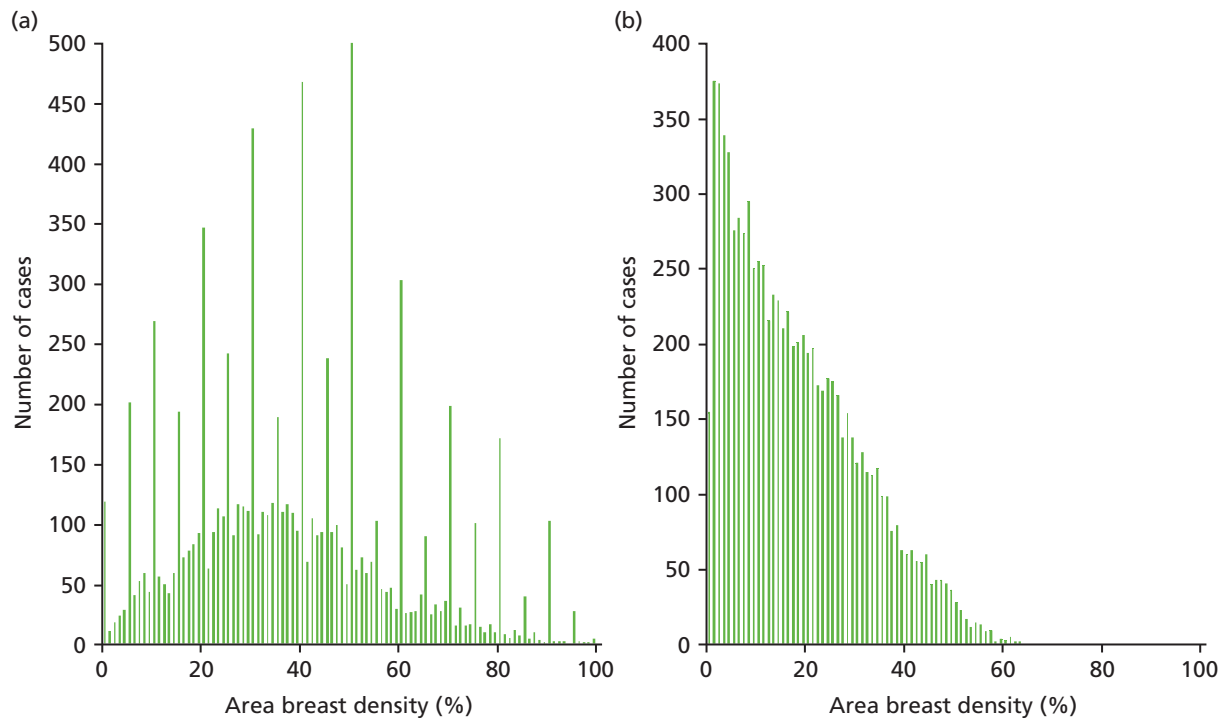
Figure 18 shows the variation with decade of age in the four breast densities measures examined in this work.



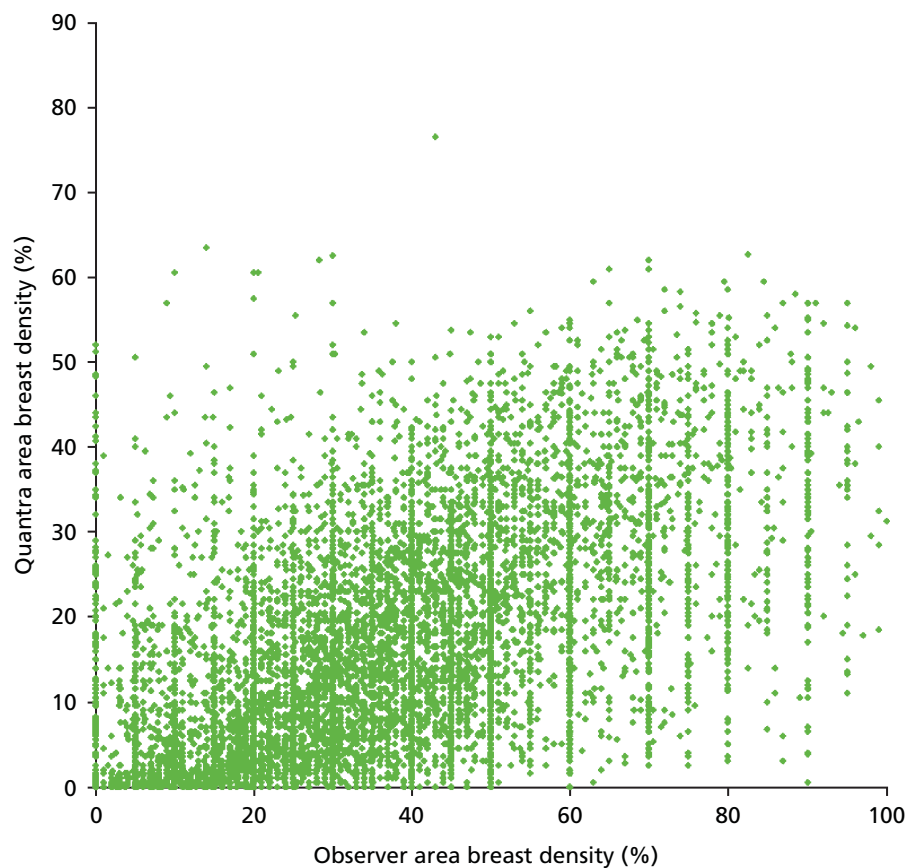
**FIGURE 13** Histograms showing the distribution of volumetric breast density across the study population as measured by (a) Quantra and (b) Volpara.



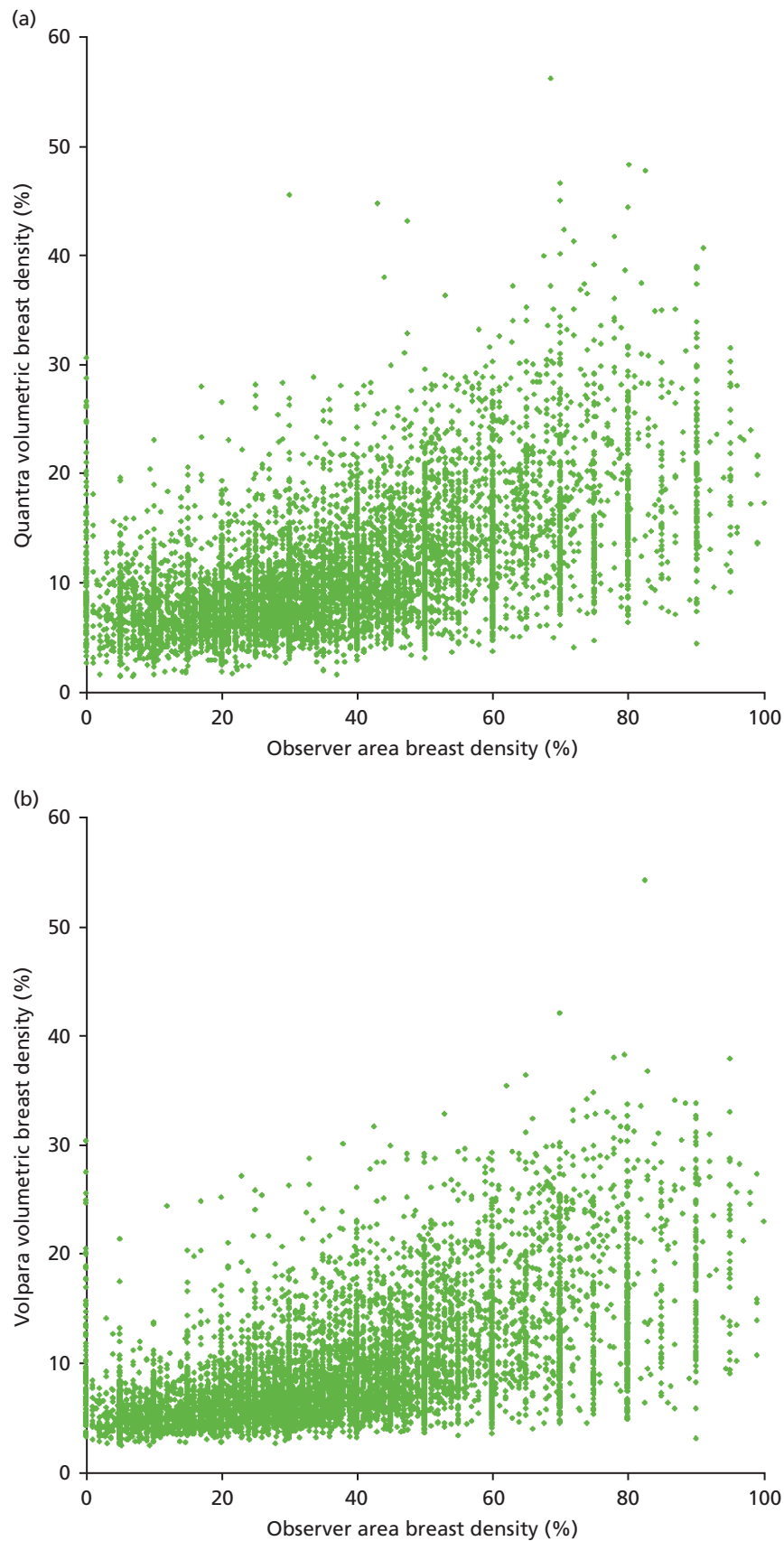
**FIGURE 14** Scatterplot comparing Quantra and Volpara scores for volumetric breast density.



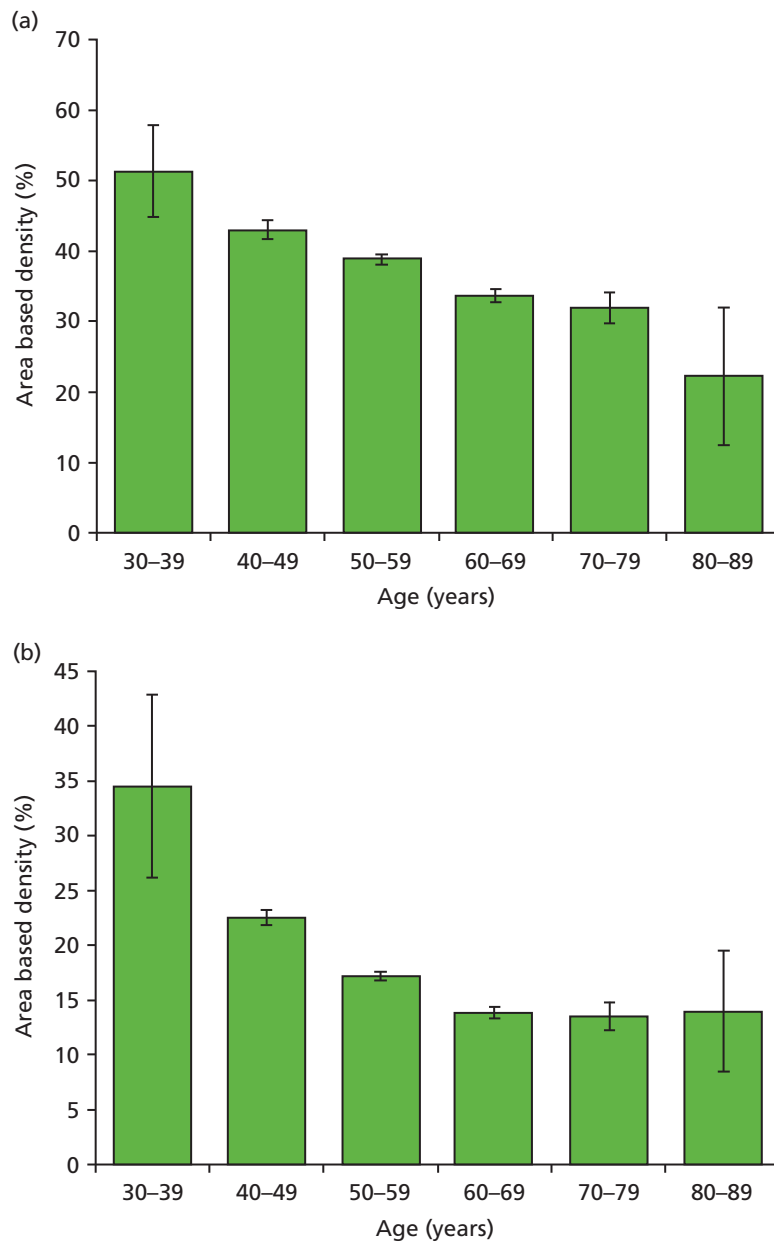
**FIGURE 15** Histograms showing the distribution of area-based breast density across the study population from (a) the observers and (b) the Quantra software.



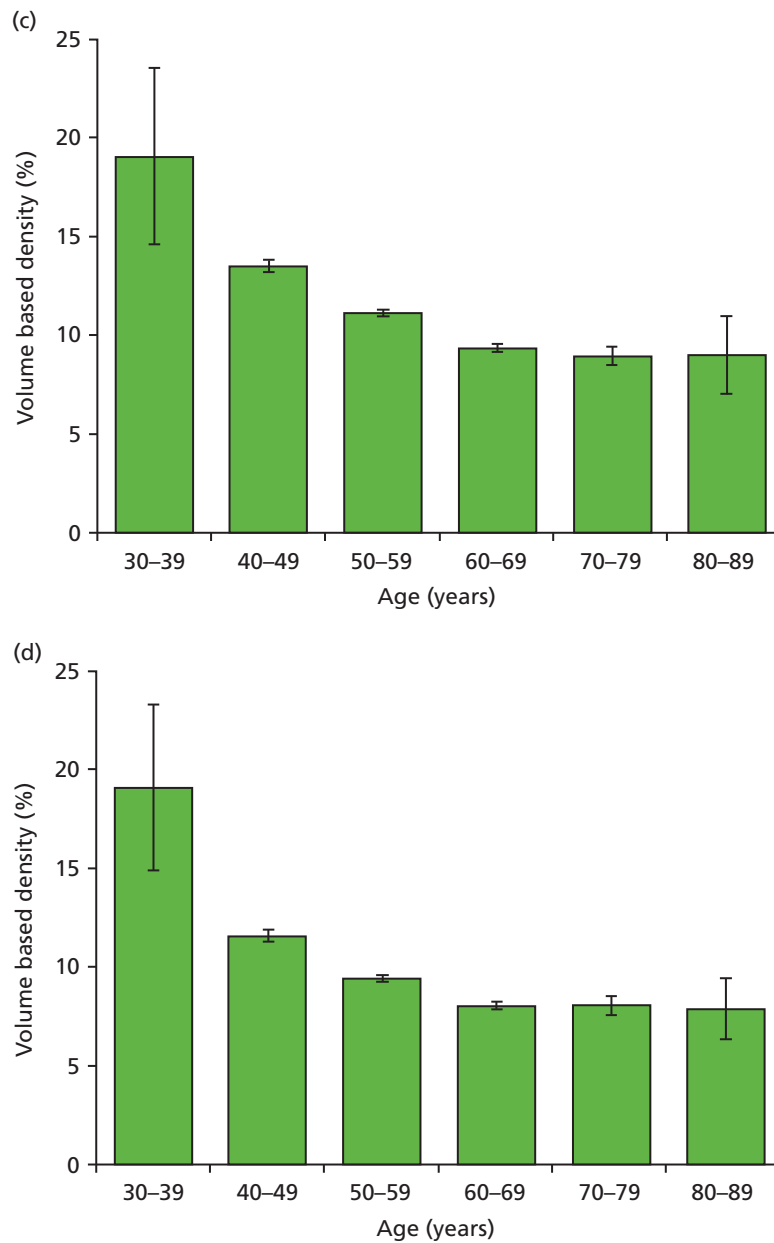
**FIGURE 16** Scatterplot comparing the observer and Quantra scores for the area-based breast density.



**FIGURE 17** Scatterplots comparing the observer area-based breast density scores with the volumetric measurements from (a) Quantra and (b) Volpara.



**FIGURE 18** Graphs showing the decrease in breast density with age for the study group for the four available density measures. Error bars are two standard errors of the mean. (a) Area-based density in the observers; (b) area-based density in the Quantra; (c) volume-based density in Quantra; and (d) volume-based density in Volpara. (*continued*)



**FIGURE 18** Graphs showing the decrease in breast density with age for the study group for the four available density measures. Error bars are two standard errors of the mean. (a) Area-based density in the observers; (b) area-based density in the Quantra; (c) volume-based density in Quantra; and (d) volume-based density in Volpara.

### Breast density and cancer risk

Volpara breast composition data were available for 7019 of the 7060 cases (1157 of the 1160 cancers and 5862 of the 5900 non-cancers). *Table 17* shows mean and standard deviation of breast composition measures using Volpara, by age, cancer status and cohort (assessment or FH screened). The absolute dense volume was generally higher in cancers than in non-cancers and declined with age in all groups. The percentage density showed the same tendencies, although less markedly. Adjusting for age, there was a significant effect of Volpara absolute dense volume on risk of cancer, with a 3% increase in odds of cancer per additional 10 cm<sup>3</sup> of dense tissue [odds ratio (OR) = 1.03, 95% CI 1.01 to 1.05,  $p < 0.001$ ]. The effect of dense volume on risk did not vary significantly by age ( $p = 0.2$ ). *Figure 19* shows the ORs and 95% CIs by quartile of absolute dense volume. There was no significant effect of percentage density on risk ( $p = 0.7$ ).

Corresponding information for Quantra breast composition data was available for 7005 cases (1156 cancers and 5849 non-cancers). *Table 18* shows the means and standard deviations of the Quantra measurements by age, cancer status and cohort. Again, dense volume was generally higher in cancers and declined with age. Adjusting for age, there was a significant effect of Quantra absolute dense volume on cancer risk, with a 2% increase in risk per additional 10 cm<sup>3</sup> of dense tissue (OR = 1.02, 95% CI 1.01 to 1.03,  $p < 0.001$ ). The effect did not vary significantly by age ( $p = 0.1$ ). *Figure 20* shows the ORs by quartile of Quantra density. There was also a significant effect of Quantra percentage density on risk, although this was less significant than that of absolute dense volume, with a 9% increase in risk per 5% increase in per cent density (OR = 1.09, 95% CI 1.02 to 1.16,  $p = 0.01$ ). *Table 19* shows the ORs by quartile of Quantra percentage density.

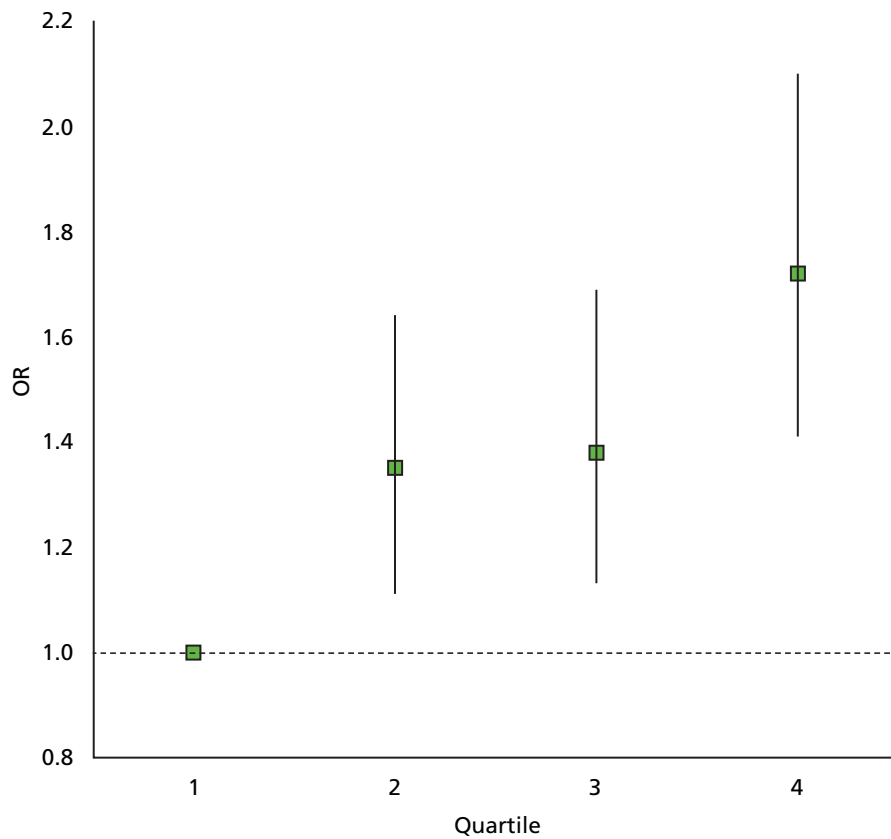
Of the three measures which were significantly associated with breast cancer risk, the highest standardised OR (OR per standard deviation of the measure) was observed for Volpara absolute density, 1.1718, with Quantra absolute density being very close to this, 1.1661. Quantra per cent density had a considerably lower standardised OR, 1.1076.

**TABLE 17** Breast composition measures using Volpara

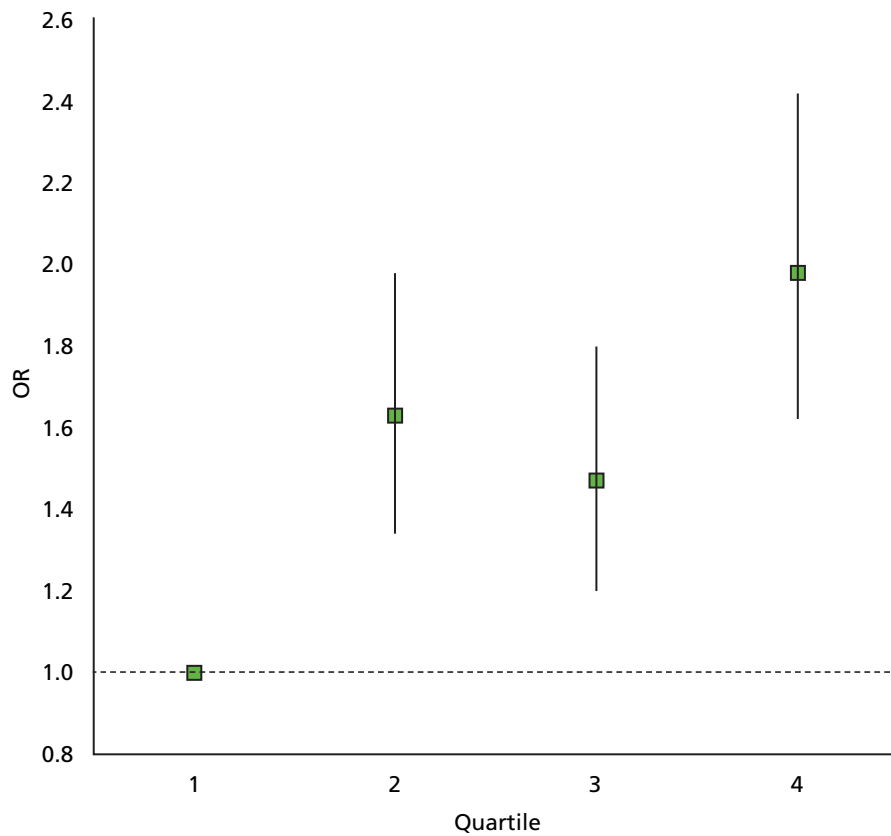
Age (years)	Breast composition measure	Mean (SD) for population			
		Cancers	Assessment non-cancers	FH non-cancers	All non-cancers
< 50	Breast volume (cm <sup>3</sup> )	1063 (663)	1027 (673)	1041 (668)	1037 (669)
	Dense volume (cm <sup>3</sup> )	111 (53)	101 (67)	101 (61)	101 (63)
	% density	13 (7)	12 (6)	12 (7)	12 (7)
	No. of subjects	29	313	942	1255
50–59	Breast volume (cm <sup>3</sup> )	1153 (670)	1034 (614)	983 (597)	1033 (613)
	Dense volume (cm <sup>3</sup> )	94 (54)	84 (50)	84 (51)	84 (50)
	% density	10 (6)	9 (5)	10 (5)	10 (5)
	No. of subjects	460	3092	43	3135
≥ 60	Breast volume (cm <sup>3</sup> )	1092 (562)	1010 (557)	582 (75)	1009 (556)
	Dense volume (cm <sup>3</sup> )	77 (44)	73 (43)	53 (20)	73 (43)
	% density	8 (4)	8 (4)	9 (2)	8 (4)
	No. of subjects	660	1402	3	1405
All ages (including age missing)	Breast volume (cm <sup>3</sup> )	1116 (613)	1028 (601)	1040 (662)	1030 (613)
	Dense volume (cm <sup>3</sup> )	85 (49)	82 (49)	100 (61)	85 (52)
	% density	9 (5)	9 (5)	12 (7)	10 (6)
	No. of subjects	1157	4830	1032	5862

SD, standard deviation.





**FIGURE 19** Odds ratios by quartile of Volpara absolute dense volume. Quartile 1, < 51 cm<sup>3</sup>; quartile 2, 51–71.99 cm<sup>3</sup>; quartile 3, 72–102.99 cm<sup>3</sup>; and quartile 4, > 103 cm<sup>3</sup>. Vertical lines indicate 95% CIs.



**FIGURE 20** Odds ratios by quartile of Quantra absolute dense volume. Quartile 1, < 59 cm<sup>3</sup>; quartile 2, 59–92.99 cm<sup>3</sup>; quartile 3, 93–142.99 cm<sup>3</sup>; and quartile 4, > 143 cm<sup>3</sup>. Vertical lines indicate 95% CIs.

**TABLE 18** Breast composition measures using Quantra

Age (years)	Breast composition measure	Mean (SD) for population			
		Cancers	Assessment non-cancers	FH non-cancers	All non-cancers
< 50	Breast volume (cm <sup>3</sup> )	1118 (723)	1075 (696)	1088 (681)	1085 (685)
	Dense volume (cm <sup>3</sup> )	143 (99)	142 (128)	137 (99)	138 (107)
	% density	14 (7)	14 (7)	14 (7)	14 (7)
	No. of subjects	29	313	938	1251
50–59	Breast volume (cm <sup>3</sup> )	1195 (670)	1079 (636)	1044 (602)	1078 (636)
	Dense volume (cm <sup>3</sup> )	131 (93)	114 (99)	111 (82)	114 (88)
	% density	12 (6)	11 (6)	13 (9)	11 (6)
	No. of subjects	459	3088	43	3131
≥ 60	Breast volume (cm <sup>3</sup> )	1126 (582)	1054 (570)	612 (121)	1052 (570)
	Dense volume (cm <sup>3</sup> )	104 (71)	97 (76)	57 (41)	97 (76)
	% density	9 (5)	9 (5)	9 (5)	9 (5)
	No. of subjects	660	1402	3	1400
All ages (including age missing)	Breast volume (cm <sup>3</sup> )	1156 (642)	1073 (622)	1088 (678)	1075 (632)
	Dense volume (cm <sup>3</sup> )	116 (83)	111 (88)	135 (97)	115 (90)
	% density	11 (5)	11 (6)	14 (7)	11 (6)
	No. of subjects	1156	4821	1028	5849

SD, standard deviation.

**TABLE 19** Odds ratios by quartile of Quantra percentage density

% density	OR	95% CI
< 7	1.00	–
7–9.99	1.20	1.00 to 1.43
10–13.99	1.08	0.88 to 1.33
≥ 14	1.40	1.14 to 1.72

## Discussion

The striking result from this analysis is the lack of correlation between observer scores of breast density and an automated analysis. Observer measurement of breast density has been shown in other studies to suffer from interobserver variability<sup>123,124,131,132</sup> and a recent paper has attempted to provide a correction for differences between observers.<sup>133</sup>

From our results it is clear that observers are able to discriminate different densities in subjects with low automated breast densities. However, when examining the histograms, shown in *Figure 14*, of the area-based measurements from both human observers and software analysis, there is clearly a difference in the distribution of scores. This may in part be a result of the observers applying a semi-volumetric approach to the assessment rather than a purely area-based one (Dr Ralph Highnam, Matakina International Limited, 2013, personal communication); however, when we compare the distribution with those found in the literature for visual and threshold methods,<sup>131,134</sup> it is the software

that produces the more comparable distribution. Results for the FH cases where two observers have scored the same images show that in 70% of cases there was good agreement. In 8% of cases there was, however, a disagreement of more than 20%, which is a result that has been noted elsewhere.<sup>132,135</sup> It should be noted that, although the observers were experienced at reading mammograms, they do not routinely estimate breast density in the NHSBSP. However, in the UK in symptomatic practice, diagnostic mammograms are categorised to a three-point scale (fatty, mixed and dense).

One possible technical reason for the difference in observer-based density scores and automated scores could be the processing of the displayed image. The software analyses the raw digital data. The observers make their estimation of density on processed images optimised for display on the workstation and have the ability to alter the window width and level of the greyscale applied to the image's pixels. This adjustment can significantly alter the image presentation and, therefore, potentially affect the density estimate. In this study, readers were advised not to alter the window levels. Comparing the two automated techniques, it can be seen that there is very good agreement when evaluating the overall breast volume. This is presumably a result of the relative ease with which each algorithm identifies the breast against the background, while each applies its own proprietary corrections to account for compression paddle height and tilt.<sup>121,128,130</sup>

There was a less agreement in the assessment of the fibroglandular volumes. There was relatively good agreement at lower volumes; however, it became poorer as those volumes increased, particularly above 400 cm<sup>3</sup>. This then led to differences in the resulting volumetric density measurements. The reasons for this discrepancy are unclear but are most likely a result of the different methods used by each software tool. Each algorithm allocates differing ratios of fibroglandular, adipose and skin tissues to each pixel based on their relative X-ray attenuation with reference to pixels in the image defined as pure adipose or fatty tissue.<sup>121</sup> As density increases, it becomes harder to identify these reference areas and each manufacturer's solution to this problem is likely to result in differences in the final volumetric density results.<sup>121,128</sup>

Some images produced quite different results as indicated by the outliers present in the fibroglandular volume and volumetric density scores shown in *Figures 11* and *13*. Further review of the images contributing to these outliers suggest that, on rare occasions, 'non-standard' image presentation can result in an incorrect estimation of the relevant tissue volumes while not triggering an error message in the results.

For example, in one left MLO view, the presence of a pacemaker resulted in significantly different results for fibroglandular tissue volume, with Volpara software scoring 425 cm<sup>3</sup> and Quantra software scoring 303 cm<sup>3</sup>, leading to volumetric densities of 19.4% and 13%, respectively, for that image. A mammogram of a visibly extremely dense breast with unusual texture resulted in each system producing very different scores for all of the volume measures, including overall breast volume. The cause of this is unknown but again may be a result of the software requiring a particular reference point where the tissue composition is assumed, for example pixels that are entirely composed of fat. Such reference points may have been difficult to locate in such a dense image. Finally, significant differences in scores were found in an image where tissue was missing from the edge of the mammogram.

Analysis of the relationship between cancer incidence and volumetric density has shown that there is a significant association of increased risk with density. This relationship was seen to be stronger with absolute measurements of the fibroglandular tissue volume than the percentage volume from both the Quantra and Volpara software. The ORs calculated here are lower than those found in the literature;<sup>136</sup> however, this population may be biased and not reflective of the general screening population. In addition, only screen-detected cancers were included here. Interval cancers collected in the future may show this study to have underestimated the risk.

If volumetric density is to be used to estimate breast cancer risk, it is important that the measurements be reliable. It is also important to understand that there are technical differences in the way in which each software package determines the fibroglandular tissue volume, and therefore the density. In addition, any long-term monitoring of breast density may require careful consideration if switching measurement methodology. The effect of the individual machine acquisition parameters and image processing together with the use of different software algorithms is worthy of further investigation to determine the overall effect on density scores in order that longitudinal studies can be undertaken.

The literature indicates that breast density may be an independent risk factor for breast cancer but is also correlated with age and body mass index and may be modified by hormone replacement therapy, FH and ethnicity.<sup>93,101</sup> *Figure 17* illustrates the variation with age, and any attempt to use density to assess risk must take such confounding factors into consideration.

## Chapter 5 Results

### Recruitment

A total of 8869 participants were recruited from the six participating centres over a 21-month period (Tables 20 and 21 and Figure 21). This data set comprised 7684 assessment cases and 1185 cases from women aged 40–49 years at moderate or high risk of developing breast cancer owing to their FH (subsequently referred to as FH cases). The latter group was included to provide a group of cases with higher breast density for subanalysis of the impact of breast density on the diagnostic accuracy of DBT. At the start of recruitment, some centres were still in the process of setting up annual screening (National Institute for Health and Care Excellence Familial Breast Cancer Guidelines) for FH women;<sup>81</sup> therefore, recruitment in this subgroup was at a slower rate initially. However, an interim analysis of data indicated that the trial was approaching its proposed recruitment target of 7000 cases, but with only 12% cancer cases, as opposed to the 18% predicted. The trial management group decided in September 2012 to cease recruitment of FH cases and continue with recruitment of a further 2000 assessment cases to provide sufficient cancer cases to give 90% power.

### Data sets

After exclusions had been applied ( $n = 207$ ), we were left with a prospective data set of 8662 cases (Figure 22). Not all of these could be included in the retrospective reading study, since we were restricted to reading only around 7000 owing to budget and time constraints. In total, 1412 cases were not allocated to the reading study. These included those cases with which there were problems in image transmission resulting in an incomplete set of images. Issues around transmission of images are described in Appendix 1. These appear to have been random and not attributable to any particular type, for example those with larger files. Problems were more likely to be a result of timing of image transfer, speed of connection and various system firewalls. This resulted in a reading study data set of 7250 cases (Figure 22).

There were a few cases ( $n = 190$ ) for which we collected data for only one arm of the study. This was either because the case was not allocated to be read in any other arm (e.g. logistics, not being converted to 2D synthetic, image management issues) or because the data were not received back from sites (lost in transit, logistics). Discounting these resulted in a final data set for analysis of 7060 cases (assessment cases:  $n = 6020$ , FH cases:  $n = 1040$ ; Figure 22).

TABLE 20 Recruitment by centre

Centre	Assessment	FH	All
Aberdeen	641	318	959
Glasgow	2097	0	2097
Manchester	1593	486	2079
King's College	942	115	1057
Barts	1004	18	1022
Guildford	1407	248	1655
Total	7684	1185	8869

TABLE 21 Recruitment summary

Month	Assessment	FH	All	Cumulative
July 2011	77	0	77	77
August 2011	252	17	269	346
September 2011	374	18	392	738
October 2011	339	23	362	1100
November 2011	369	76	445	1545
December 2011	309	75	384	1929
January 2012	521	81	602	2531
February 2012	618	141	759	3290
March 2012	455	96	551	3841
April 2012	409	86	495	4336
May 2012	466	136	602	4938
June 2012	283	83	366	5304
July 2012	375	89	464	5768
August 2012	468	134	602	6370
September 2012	419	113	532	6902
October 2012	422	17	439	7341
November 2012	454	0	454	7795
December 2012	277	0	277	8072
January 2013	413	0	413	8485
February 2013	309	0	309	8794
March 2013	75	0	75	8869
Total	7684	1185	8869	

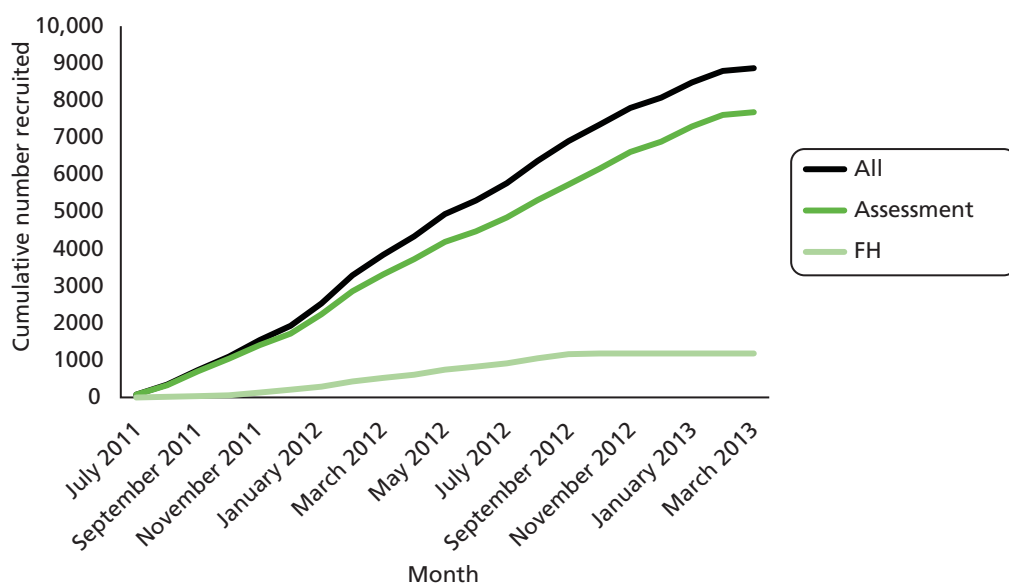
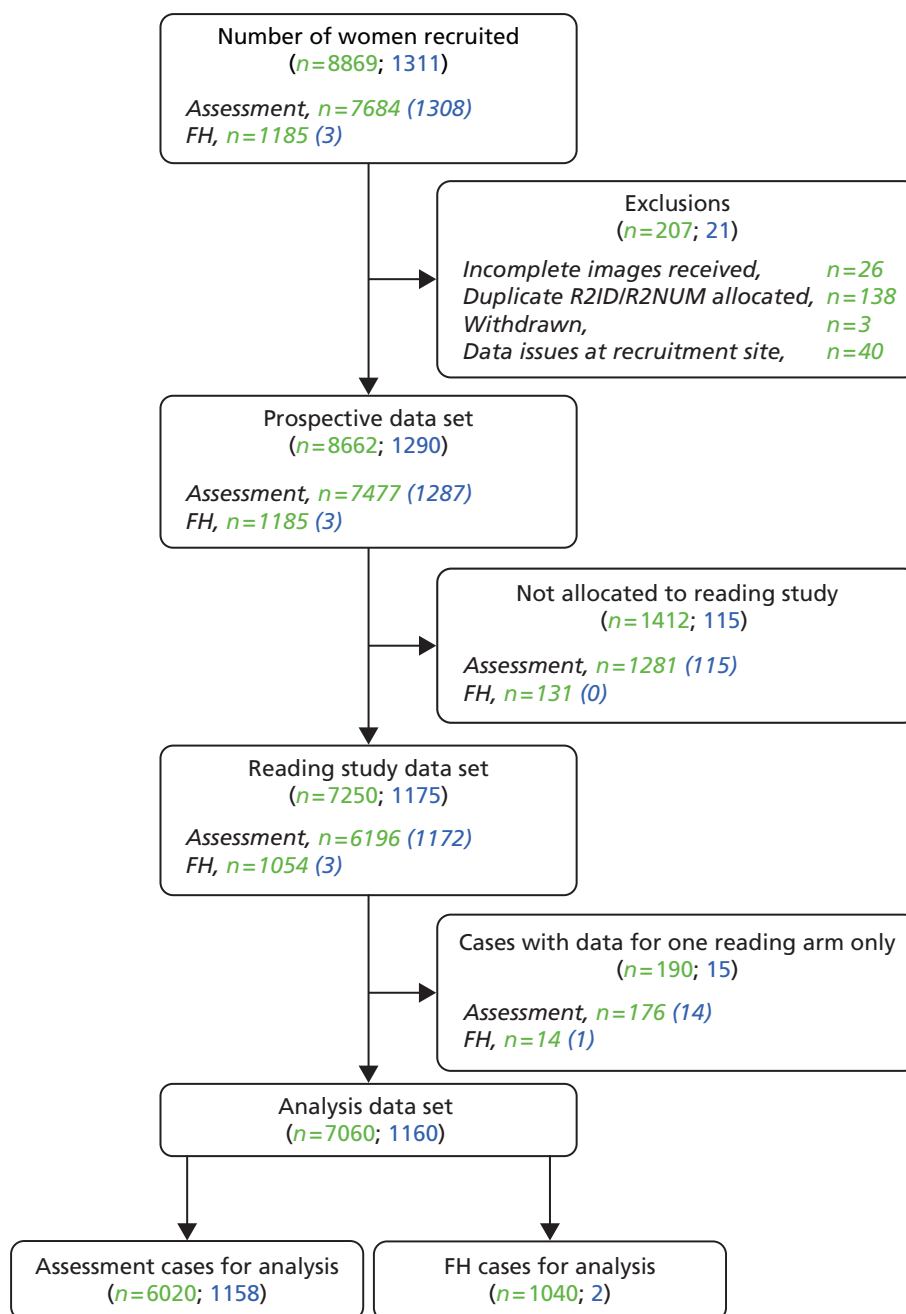


FIGURE 21 Cumulative recruitment.



**FIGURE 22** Study flow chart. Green text represents the total number of women; blue text presents the number of women with cancer.

## Analysis

After exclusions, there were a total of 7060 subjects for analysis. The data set comprised 6020 (1158 cancers) assessment cases and 1040 (two cancers) FH surveillance cases. Reading data were available for 6927 (98%) cases by 2D only, 6959 (99%) by 2D + DBT and 6653 (94%) by 2D synthetic + DBT. *Table 22* shows the characteristics of the study population.

The analysis included cases which were read in only two arms of the study to avoid introducing bias. It was run again using only cases which were read in all three arms and produced identical results.

TABLE 22 Participant characteristics

Characteristics	Assessment cases		FH cases	
	Number randomised (%) <sup>a</sup>	Number of cancers (%) <sup>a</sup>	Number randomised (%) <sup>a</sup>	Number of cancers (%) <sup>a</sup>
<b>Age range (years)</b>				
< 40	3 (< 1)	1 (< 1)	11 (1)	0
40–49	340 (6)	27 (2)	938 (94)	1 (50)
50–59	3568 (59)	462 (40)	44 (4)	1 (50)
60–69	1714 (29)	519 (45)	3 (< 1)	0
≥ 70	364 (6)	141 (12)	0	0
Unknown	31	9	44	0
<b>Breast density (%)</b>				
0–24	1636 (27)	378 (33)	233 (23)	0
25–49	2556 (43)	439 (38)	418 (42)	1 (50)
50–74	1376 (23)	271 (24)	271 (27)	1 (50)
75–100	396 (7)	63 (5)	83 (8)	0
Unknown	56	8	35	0
<b>Cancer type</b>				
Invasive ductal (± DCIS)	–	788 (68)	–	1 (50)
Invasive lobular (± DCIS)	–	109 (9)	–	–
Invasive other (± DCIS)	–	59 (5)	–	–
DCIS	–	203 (18)	–	1 (50)
<b>Cancer size (mm)</b>				
<i>Invasive cancers</i>				
1–5	–	73 (8)	–	0
6–10	–	243 (26)	–	0
11–20	–	434 (46)	–	1 (100)
21–50	–	183 (19)	–	0
> 50	–	10 (1)	–	0
Unknown	–	13	–	0
<b>DCIS</b>				
1–5	–	30 (15)	–	0
6–10	–	30 (15)	–	0
11–20	–	47 (24)	–	0
21–50	–	78 (39)	–	1 (100)
> 50	–	15 (7)	–	0
Unknown	–	3	–	0



TABLE 22 Participant characteristics (continued)

Characteristics	Assessment cases		FH cases	
	Number randomised (%) <sup>a</sup>	Number of cancers (%) <sup>a</sup>	Number randomised (%) <sup>a</sup>	Number of cancers (%) <sup>a</sup>
<b>Cancer grade</b>				
<i>Invasive cancers</i>				
1	–	242 (26)	–	0
2	–	504 (54)	–	2 (100)
3	–	180 (20)	–	0
Unknown	–	30	–	0
<i>DCIS</i>				
Low	–	10 (8)	–	0
Intermediate	–	31 (22)	–	1 (100)
High	–	97 (70)	–	0
Unknown	–	65	–	0
<b>Lymph node status (invasive cancers only)</b>				
Normal	–	514 (58)	–	0
< 4 nodes positive	–	292 (33)	–	1 (100)
4 or more nodes positive	–	77 (9)	–	0
Unknown	–	73	–	0
<b>Dominant radiological feature</b>				
Circumscribed mass	1814 (30)	145 (13)	84 (8)	0
Spiculated mass	712 (12)	508 (44)	3 (< 1)	0
Microcalcification	1006 (17)	282 (24)	40 (4)	1 (50)
Distortion	514 (8)	109 (9)	10 (1)	1 (50)
ASD	1837 (31)	107 (9)	26 (3)	0
None	137 (2)	7 (1)	877 (84)	0
<b>Recruitment site</b>				
Aberdeen	602 (10)	83 (7)	288 (28)	0
Glasgow	1362 (23)	259 (23)	0	0
Manchester	1280 (21)	237 (20)	386 (37)	1 (50)
Guildford	1183(20)	249 (22)	239 (23)	1 (50)
King's College	760 (12)	165 (14)	116 (11)	0
Barts	833 (14)	165 (14)	11 (1)	0

<sup>a</sup> Percentages have been reported for known data.

In addition to the matched data available for 1137 cancer cases which had both 2D and 2D + DBT imaging, unmatched data were available for 18 cases for 2D imaging only and for another five cases for 2D + DBT imaging only. Pooling the matched and the unmatched data and analysing with a modification of the Mantel–Haenszel method<sup>137</sup> resulted in an OR for cancer detection (2D + DBT vs. 2D alone) of 1.363 (95% CI 1.002 to 1.852,  $p = 0.048$ ), indicating significantly better sensitivity for 2D + DBT than 2D alone.

Table 23 shows sensitivity and specificity for specific subgroups and for all subjects combined. For all subjects combined, sensitivity was 87% (95% 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 ( $p = 0.07$ ). There was no significant difference in sensitivity between 2D and synthetic 2D + DBT ( $p = 0.6$ ). 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 synthetic 2D + DBT than for 2D ( $p < 0.001$  in both cases).

The increased specificity with 2D + DBT and synthetic 2D + DBT was observed in all subgroups of density and dominant radiological feature and across all age groups ( $p < 0.001$  in all cases). Specificity tended to be lower by all three modalities for microcalcifications and higher by all three modalities for distortion/ASD density, but the significant improvement in specificity for both 3D modalities was consistently observed in these categories.

Sensitivity was significantly higher ( $p = 0.01$ ) for 2D + DBT than 2D alone for age range 50–59 years. Sensitivity was significantly higher ( $p < 0.001$ ) for 2D + DBT than for 2D alone for invasive tumours of size 11–20 mm, with a sensitivity of 86% (95% CI 82% to 90%) for 2D and 93% (95% CI 90% to 96%) for 2D + DBT.

For the purposes of analysis stratified by density, we classified density as below 50% or 50% and above. We had planned to use 70% density as the cut-off, but there were only 105 cancer cases with density of 70% or more.

Significantly higher sensitivity ( $p = 0.03$ ) was also observed for 2D + DBT for those with density 50% or more, with a sensitivity of 86% (95% CI 82% to 90%) for 2D compared with 93% (95% CI 90% to 96%) for 2D + DBT. A similarly increased sensitivity ( $p = 0.01$ ) was seen for 2D + DBT in grade 2 invasive tumours (but not grade 1 or grade 3), with sensitivity of 87% (95% CI 84% to 90%) for 2D and 91% (95% CI 88% to 94%) for 2D + DBT. A significant increase in sensitivity was observed for 2D + DBT ( $p = 0.04$ ) where the dominant radiological feature was a mass, with 89% (95% CI 86% to 92%) sensitivity for 2D and 92% (95% CI 89% to 95%) for 2D + DBT.

For synthetic 2D + DBT, there was significantly ( $p = 0.006$ ) higher sensitivity than for 2D alone in invasive cancers of size 11–20 mm, with a sensitivity of 92% (95% CI 89% to 95%). No other significant differences in sensitivity were noted.

Table 24 shows the distribution of detection of cancers for all combinations of modalities in those 1112 cancers for which we have reading data for all three reading arms. In total, 1079 (97%) cancers were detected by at least one reading arm; 840 (75%) cancers were detected by all three reading arms. A total of 142 (13%) cancers were missed by 2D, 118 (11%) by 2D + DBT and 136 (12%) by synthetic 2D + DBT. Of these, 33 (3%) cancers were missed in all three reading arms. DCIS cases showed similar distribution to all cancers combined.

Table 25 shows characteristics of these missed cancers for each imaging modality, plus those found by all three modalities. Two major differences among the modalities were that 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 modalities; and cancers missed by 2D + DBT were less likely to be of grade 2 or to have density less than 50% than the other two modalities.

TABLE 23 Cancer detection rate in three arms of reading study

Variables	2D only		2D + DBT		Synthetic 2D + DBT	
	Sensitivity (%) (cancers recalled/total cancers)	Specificity (%) (non-cancers not recalled/total non-cancers)	Sensitivity (%) (cancers recalled/total cancers)	Specificity (%) (non-cancers not recalled/total non-cancers)	Sensitivity (%) (cancers recalled/total cancers)	Specificity (%) (non-cancers not recalled/total non-cancers)
<b>Age (years)</b>						
< 50	86 (25/29)	68 (847/1241)	83 (24/29)	83 (1036/12153)	89 (25/28)	83 (953/1151)
50–59	87 (399/461)	54 (1646/3064)	91 (414/453)	67 (2070/3096)	88 (402/455)	69 (2034/2967)
≥ 60	88 (582/665)	57 (834/1467)	88 (583/660)	67 (980/1468)	88 (571/652)	66 (917/1400)
<b>Invasive size (mm)</b>						
1–10	85 (268/315)	–	85 (262/309)	–	84 (260/308)	–
11–20	86 (371/431)	–	93 (399/429)	–	91 (387/423)	–
> 20	93 (179/193)	–	91 (171/188)	–	90 (171/190)	–
<b>DCIS size (mm)</b>						
1–10	80 (48/60)	–	83 (49/59)	–	85 (51/60)	–
11–20	89 (41/46)	–	83 (39/47)	–	76 (35/46)	–
> 20	91 (86/94)	–	93 (87/94)	–	88 (82/93)	–
<b>Grade (invasive only)</b>						
1	86 (206/240)	–	89 (210/236)	–	88 (205/232)	–
2	87 (435/502)	–	91 (452/495)	–	88 (439/449)	–
3	90 (162/180)	–	88 (157/178)	–	90 (159/176)	–
<b>Breast density (%)</b>						
< 50	88 (715/814)	58 (2275/3930)	89 (717/806)	71 (2800/3969)	88 (706/799)	70 (2629/3747)
≥ 50	86 (286/334)	57 (1002/1759)	93 (299/329)	70 (1232/1767)	87 (286/329)	72 (1220/1694)

continued

TABLE 23 Cancer detection rate in three arms of reading study (continued)

Variables	2D only		2D + DBT		Synthetic 2D + DBT	
	Sensitivity (%) (cancers recalled/total cancers)	Specificity (%) (non-cancers not recalled/total non-cancers)	Sensitivity (%) (cancers recalled/total cancers)	Specificity (%) (non-cancers not recalled/total non-cancers)	Sensitivity (%) (cancers recalled/total cancers)	Specificity (%) (non-cancers not recalled/total non-cancers)
<b>Node status (invasives only)</b>						
Negative	88 (451/513)	–	90 (455/506)	–	89 (448/503)	–
1–3 positive	86 (248/289)	–	88 (252/287)	–	87 (250/285)	–
> 3 positive	86 (66/77)	–	93 (70/75)	–	88 (66/75)	–
<b>Dominant radiological feature</b>						
Soft-tissue mass	89 (580/650)	51 (979/1919)	92 (594/643)	67 (1287/1928)	91 (580/636)	66 (1226/1862)
Microcalcification	88 (249/282)	31 (230/745)	88 (246/279)	39 (293/750)	85 (237/278)	44 (318/723)
Distortion/ASD	71 (171/216)	64 (1363/2125)	82 (175/213)	75 (1614/2145)	82 (176/214)	76 (1540/2027)
<b>All subjects combined</b>						
All subjects	87 (1006/1155)	58 (3307/5772)	89 (1021/1142)	69 (4086/5900)	88 (998/1135)	71 (3904/5518)

**TABLE 24** Detection status of cancers all three imaging modalities

Detected by			Number of cancers (%)	Number of DCIS cases (%)
2D	2D+ DBT	Synthetic 2D + DBT		
No	No	No	33 (3)	6 (3)
No	No	Yes	16 (1)	4 (2)
No	Yes	No	18 (2)	2 (1)
No	Yes	Yes	75 (7)	14 (7)
Yes	No	No	24 (2)	8 (4)
Yes	No	Yes	45 (4)	8 (4)
Yes	Yes	No	61 (6)	15 (8)
Yes	Yes	Yes	840 (75)	143 (71)
Total			1112	200

**TABLE 25** Characteristics of invasive cancers missed by each imaging modality and of those found by all three

Factor	Category	Number of cancers missed by 2D only (%)	Number of cancers missed by 2D + DBT (%)	Number of cancers missed by synthetic 2D + DBT (%)	Number of cancers found by all three modalities (%)
Size (mm)	1–10	47 (39)	47 (50)	48 (47)	207 (30)
	11–20	60 (50)	30 (32)	34 (34)	326 (47)
	> 20	14 (11)	17 (18)	19 (19)	155 (23)
Node status	Negative	62 (54)	51 (56)	55 (56)	388 (60)
	1–3 positive	41 (36)	35 (38)	35 (35)	201 (31)
	> 3 positive	11 (10)	5 (6)	9 (9)	55 (9)
Grade	1	34 (29)	26 (29)	27 (27)	164 (24)
	2	67 (56)	43 (48)	57 (56)	373 (55)
	3	18 (15)	21 (23)	17 (17)	140 (21)
Breast density	< 50%	80 (66)	69 (73)	70 (66)	488 (70)
	≥ 50%	42 (34)	26 (27)	36 (34)	195 (30)
Dominant radiological feature	Mass	68 (55)	45 (47)	52 (49)	479 (69)
	Microcalcification	12 (10)	12 (13)	16 (15)	88 (12)
	Distortion/ASD	42 (34)	37 (39)	37 (35)	126 (18)
	None	1 (1)	1 (1)	1 (1)	4 (1)

## Receiver operating characteristic analysis

Receiver operating characteristic 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 AUC and *p*-value. Areas under the curve were compared using the method of DeLong *et al.*<sup>91</sup>

Figure 23 shows the ROC curve for each reading arm on the same graph.

For 2D alone, the AUC was 0.84 (95% CI 0.83 to 0.86), for 2D + DBT the AUC was 0.89 (95% CI 0.87 to 0.90) and for synthetic 2D + DBT the AUC was 0.87 (95% CI 0.86 to 0.89). Both 2D + DBT and synthetic 2D + DBT had significantly greater AUCs than 2D alone ( $p < 0.001$  in both cases).

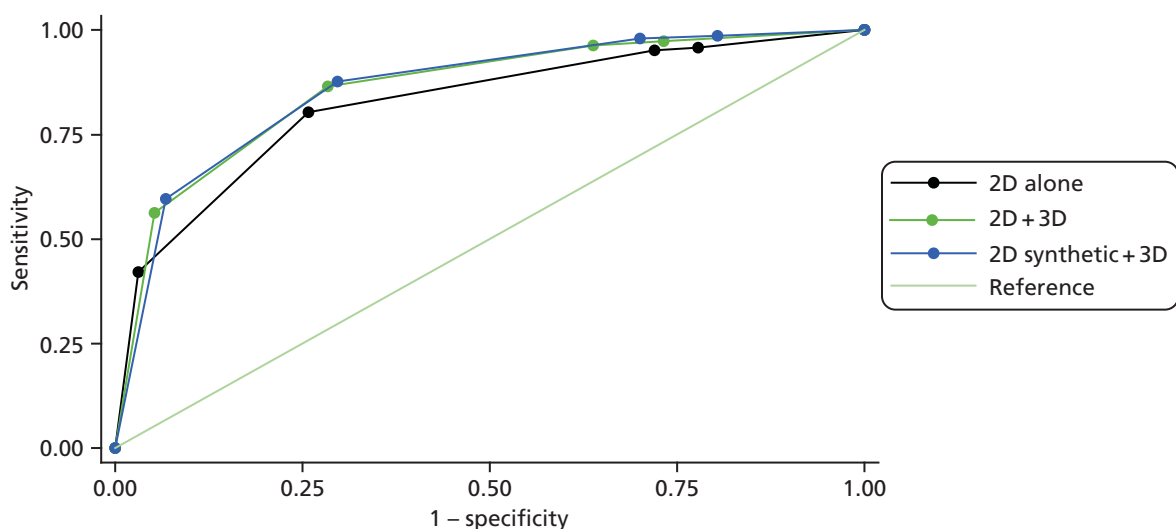
Although, in general, specificity was considerably higher for the two DBT arms at very low levels of sensitivity, 2D alone would have a slightly higher specificity than 2D + DBT or synthetic 2D + DBT. However, thereafter, the ROC curves for the two DBT arms would always be higher than for 2D alone.

ROC analysis was also carried out on specific subgroups:

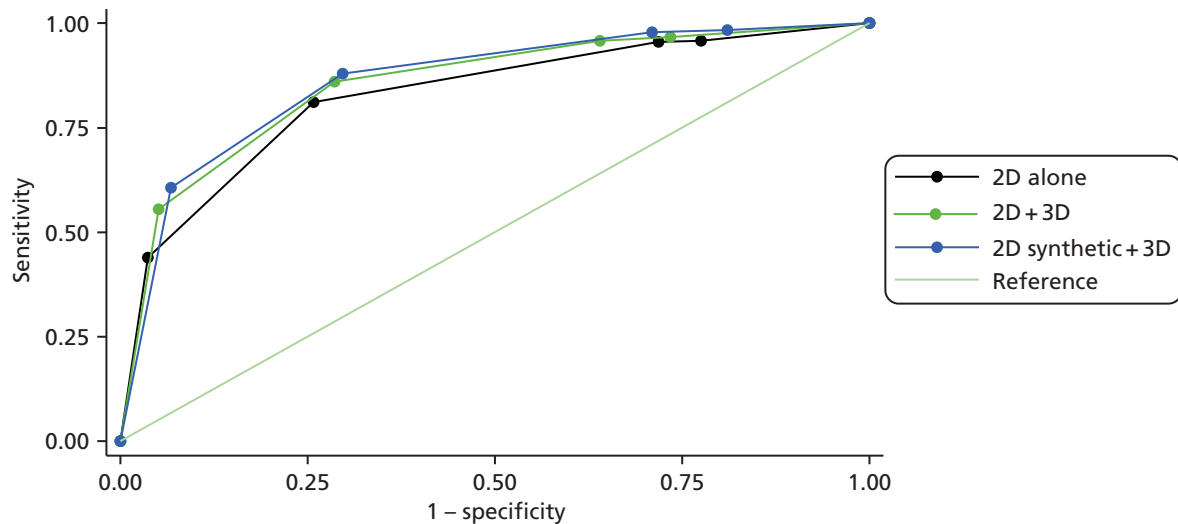
- i. cases with visually assessed density of less than 50%
- ii. cases with visually assessed density of 50% or more
- iii. cases where the dominant radiological sign was a soft-tissue mass
- iv. cases where the dominant radiological feature was microcalcification
- v. cases where the dominant radiological feature was asymmetry or distortion.

Figure 24 shows the ROC curves for the three reading arms in those cases with visually assessed breast density of less than 50%. The AUCs were 0.85 (95% CI 0.83 to 0.87) for 2D alone, 0.89 (95% CI 0.87 to 0.90) for 2D + DBT and 0.88 (95% CI 0.86 to 0.90) for synthetic 2D + DBT. The AUCs for 2D + DBT and synthetic 2D + DBT were significantly greater than that for 2D alone ( $p < 0.001$  in both cases).

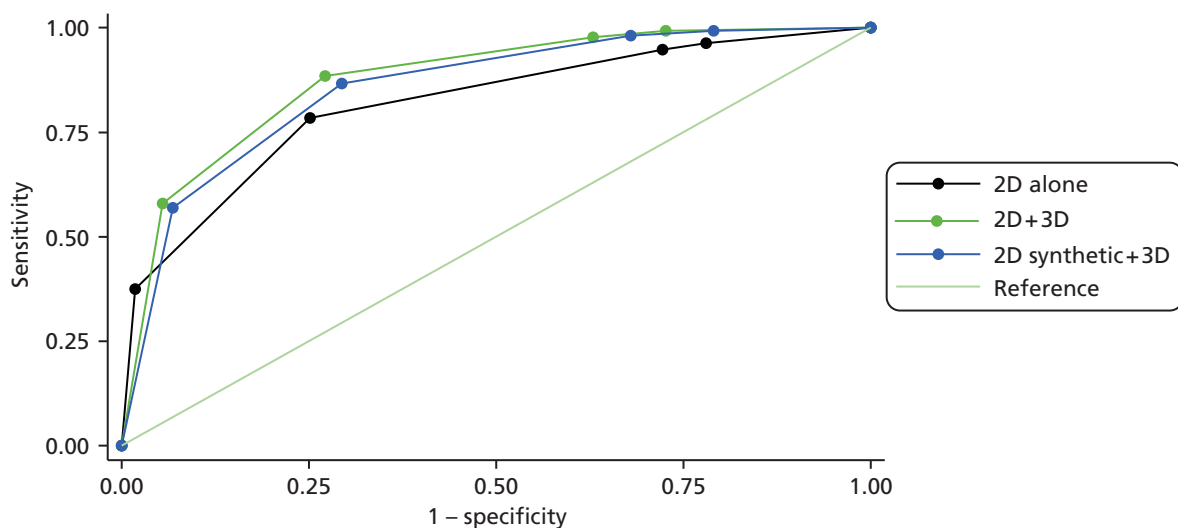
Figure 25 shows the corresponding curves in those with visually assessed density of 50% or more. The AUCs were 0.83 (95% CI 0.80 to 0.86) for 2D alone, 0.89 (95% CI 0.87 to 0.92) for 2D + DBT and 0.87 (95% CI 0.84 to 0.89) for synthetic 2D + DBT. Both 2D + DBT ( $p < 0.001$ ) and synthetic 2D + DBT ( $p = 0.005$ ) had significantly greater AUCs than 2D alone.



**FIGURE 23** Receiver operating characteristic curves for the three reading arms.



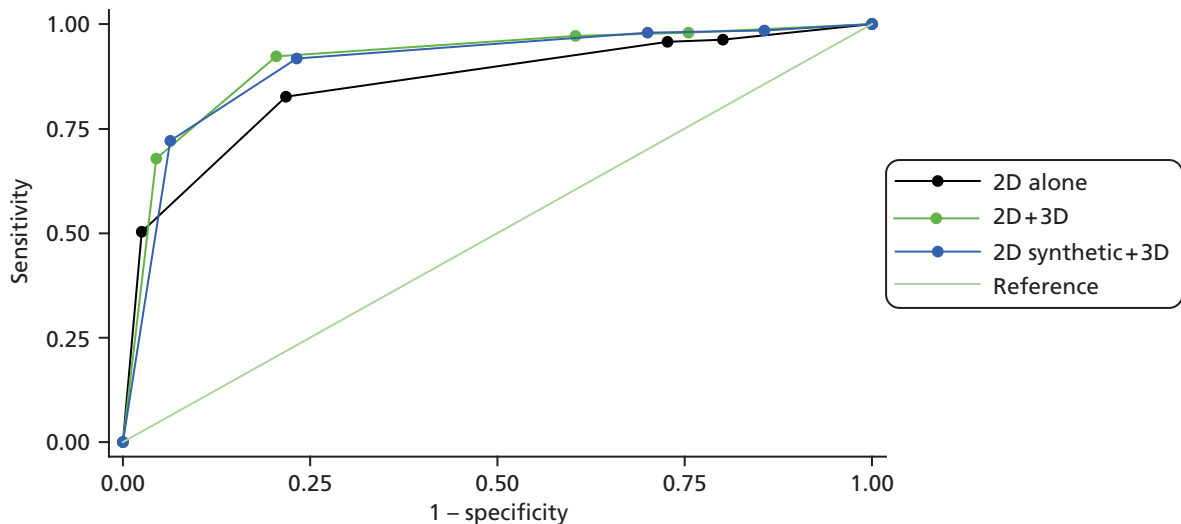
**FIGURE 24** Receiver operating characteristic curves for the three reading arms in those cases with visually assessed density less than 50%.



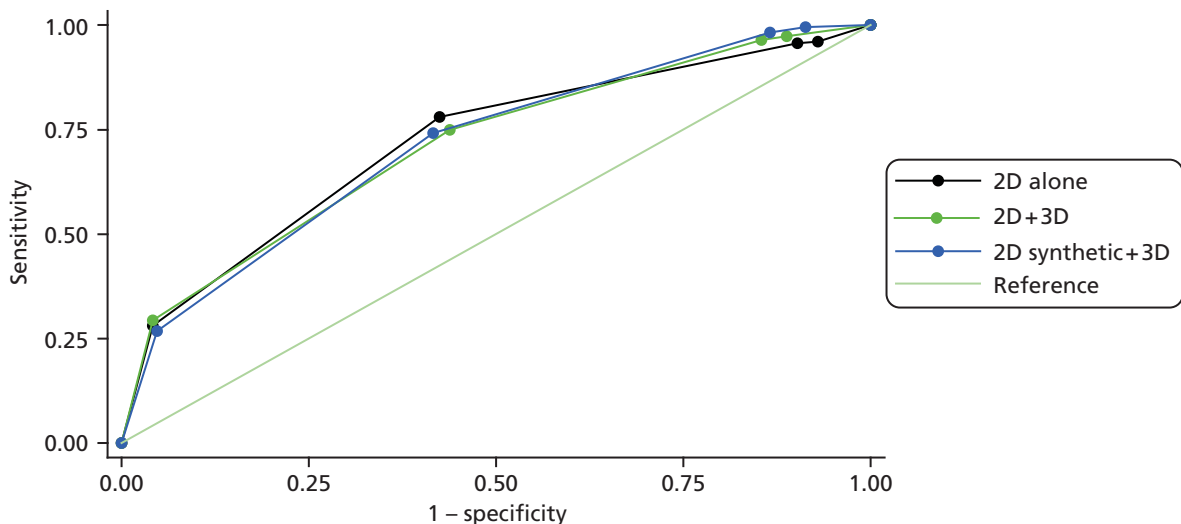
**FIGURE 25** Receiver operating characteristic curves for all three reading arms in those with visually assessed density of 50% or more.

Figure 26 shows the ROC curves in those subjects in whom the dominant radiological feature was a mass. The AUCs were 0.86 (95% CI 0.84 to 0.88) for 2D alone, 0.92 (95% CI 0.90 to 0.94) for 2D + DBT and 0.91 (95% CI 0.89 to 0.93) for synthetic 2D + DBT. The AUCs for 2D + DBT and for synthetic 2D + DBT were significantly greater than for 2D alone ( $p < 0.001$  in both cases).

The corresponding curves for those with microcalcifications as the dominant sign are shown in Figure 27. The AUCs were 0.73 (95% CI 0.69 to 0.77) for 2D alone, 0.73 (95% CI 0.69 to 0.77) for 2D + DBT and 0.72 (95% CI 0.68 to 0.76) for synthetic 2D + DBT. The AUCs did not differ significantly from 2D alone for either 2D + DBT ( $p = 0.9$ ) or synthetic 2D + DBT ( $p = 0.8$ ).



**FIGURE 26** Receiver operating characteristic curves for the three reading arms in those cases where the dominant radiological sign was a mass.

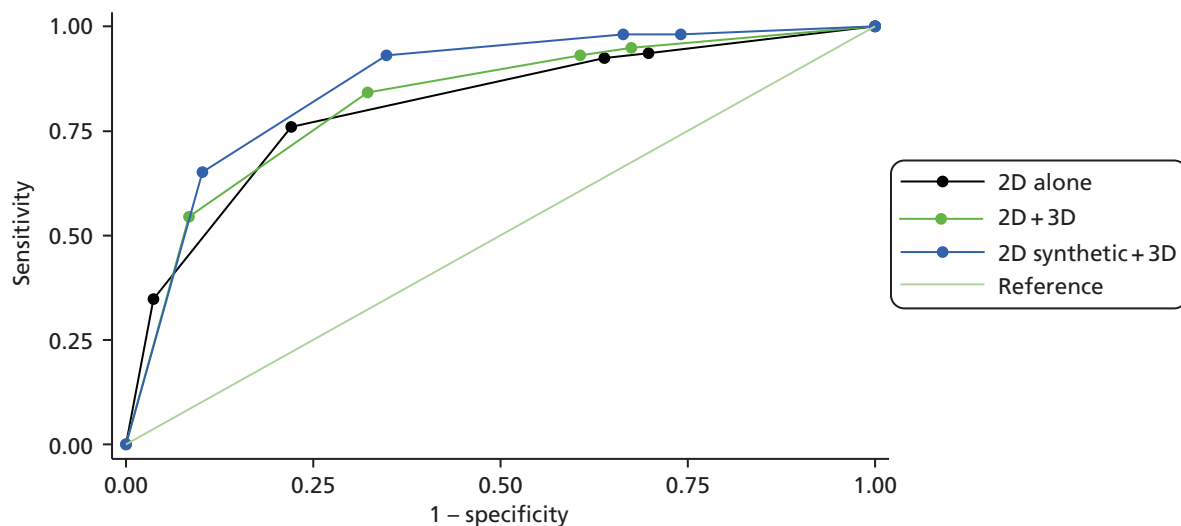


**FIGURE 27** Receiver operating characteristic curves for the three reading arms in those cases with microcalcifications as the dominant radiological feature.

Figure 28 shows the ROC curves for those with distortion or asymmetry as the dominant radiological feature. The AUCs were 0.82 (95% CI 0.78 to 0.86) for 2D alone, 0.86 (95% CI 0.82 to 0.89) for 2D + DBT and 0.87 (95% CI 0.83 to 0.90) for synthetic 2D + DBT. Significantly greater AUCs than for 2D alone were noted for both 2D + DBT ( $p = 0.03$ ) and 2D synthetic + DBT ( $p = 0.002$ ).

These differences in AUC are mostly driven by differences in specificity, including the low AUCs for all three reading arms in those with calcifications as the dominant radiological feature. However, there is clearly a shift to greater confidence in malignancy among cancers with the addition of DBT. With 2D alone, 39% of cancers were scored as 5 (malignant), whereas in both 2D + DBT and synthetic 2D + DBT 53% of cancers were scored as 5.





**FIGURE 28** Receiver operating characteristic curves for all three reading arms in those with distortion or ASD as the dominant radiological feature.



## Chapter 6 Discussion

In this retrospective reading study, our primary paired data analysis demonstrated only a modest improvement in cancer detection rate but a clear improvement in specificity when DBT is used in conjunction with 2D images or synthetic images compared with 2D alone. However, evidence of a significant increase in sensitivity for 2D + DBT compared with 2D alone was noted in a secondary analysis that pooled together the matched and unmatched data for cases where there were incomplete data for one of the reading arms of the study.

Non-significant increases in cancer detection ranging from 4.3% to 49.0% have been reported in several screening studies after the addition of DBT;<sup>59,60,138,139</sup> in these cases lack of statistical significance was largely attributable to sample size limitations. Our results differ from data from two prospective screening trials comparing 2D + DBT and 2D alone. Significant increases in cancer detection rate, of 27% and 34%, were reported in the Oslo<sup>19</sup> and STORM<sup>66</sup> trials, respectively. In addition, a large US retrospective analysis comparing performance measures before and after the introduction of combined 2D + DBT screening has recently reported a 29% increase in cancer detection rate.<sup>140</sup> Our modest improvement in sensitivity reflects the selection bias in our method of case collection. Our data set comprised 85% of cases from women recalled for abnormalities seen on the 2D screening, with a high proportion of cancer cases, and this approach was considered to be the most appropriate at this stage in the evolution of DBT technology when few clinical studies had been published. Thus, we may have underestimated the contribution of DBT, since these cases had already been detected by 2D mammography. We have not included cases where DBT demonstrated cancers from a screening cohort.

By using assessment cases for this study we effectively used 2D imaging to find all the cases in 100% of the screened women. When we used 2D + DBT in the 5% of the women recalled from screening, the readers were, in principle, able to find all the cancers identified as suspicious lesions in 2D for the whole population. As for cancers only detectable by DBT, they could find these only in the 5% recalled. Therefore, we have not used DBT in the other 95% that were not recalled. Thus, there may be 20 times as many cancers detectable only by DBT in the whole screened population. In this study the sensitivity improvement using 2D + DBT was only 2%. In actual screening we can expect up to 20 times more of this type of cancer, leading to up to a 40% increase in cancer detection for 2D + DBT compared with 2D. This calculation shows that, when one takes account of the method of case selection, the sensitivity improvement found here is consistent with those published for screening,<sup>19,66,140</sup> where increases of about 25–40% have been reported. Almost all previous non-screening studies are also affected by such case selection issues.

Another possible explanation for the lack of improvement in sensitivity is that the addition of DBT may have had minimal impact on the performance of this group of experienced, high-volume film readers.

The 19% significant improvement in specificity that we observed with 2D + DBT compared with 2D alone would be consistent with published studies that highlight the potential of screening with 2D + DBT to reduce false-positive recall rates. Nevertheless, the inherent selection bias in our study data set could result in an overestimate of the true impact of DBT on specificity. Reductions in recall rate ranging from 15% to 42%<sup>59,60,138–140</sup> have been reported using 2D + DBT in screening compared with 2D alone. The magnitude of the decrease is likely to vary depending on the recall/arbitration policy in practice, background recall rate and reader experience. Our study, similar to the Oslo<sup>19</sup> and STORM<sup>66</sup> trials, was conducted with experienced readers from screening programmes with a generally low recall rate. Although a greater impact of DBT on recall rate may be expected in the USA, where recall rates are often higher than in the UK or Scandinavia, the largest published study to date<sup>140</sup> reported a more modest 15% reduction in recall rate after the introduction of DBT. Since our study was not a screening trial, it is difficult to predict the degree of improvement in sensitivity or specificity that could be expected in a screening environment.

In comparisons between studies, other methodological differences in study design that could confound the interpretation of results should also be noted. Our study, similar to the Oslo<sup>19</sup> and STORM<sup>66</sup> trials, used paired comparisons of 2D and 2D + DBT data from the same cases whereas other studies<sup>59,60,138-140</sup> compared data from different cohorts of women before and after the introduction of 2D + DBT screening. Differences in reading protocol could also impact on the interpretation of results. For example, our study randomised the 2D and the 2D + DBT images for reading by different readers. In the Oslo trial,<sup>19</sup> cases were randomised for reading by 2D or 2D + DBT and adjusted for reader-specific performance, and, in the STORM trial,<sup>66</sup> cases were sequentially read, first with 2D and then 2D + DBT, by two readers. The multicentre Friedewald study<sup>140</sup> probably provides the most pragmatic evidence to date of the impact of combined 2D + DBT in screening practice, but extrapolation of its generalisability to the UK and other European screening programmes, where independent double reading is standard, is limited.

The inclusion of FH cases was intended to provide a cohort of younger women to examine whether or not the addition of DBT to this group would be advantageous. Theoretically, this group could have given some insight into the use of DBT in a screening setting, although it is recognised that this cohort was not representative of the current UK NHSBSP screening population. However, we were unable to perform a separate formal analysis for this cohort using the initial reading information. We accrued 1040 FH cases, including two cancer cases. We were concerned that the reading protocol during case collection would be difficult to apply rigorously. Hence, readers may not have provided an unbiased evaluation of 2D and then 2D + DBT in the initial prospective data collection aspect of the study to permit evaluation of the impact of DBT on reader performance.

The improvement in specificity for 2D + DBT compared with 2D alone was observed irrespective of mammographic density, age, tumour size and dominant radiological feature and in both invasive cancers and DCIS, as noted previously,<sup>16,59,60</sup> and indicates the potential of integrated DBT to improve specificity by allowing readers to avoid recalling false negative cases. In agreement with published studies,<sup>17,45,141</sup> ROC analysis demonstrated a significant increase in diagnostic accuracy for 2D + DBT compared with 2D alone and was most likely attributable to the marked improvement in specificity.

Planned subgroup analyses indicated that there was better sensitivity of 2D + DBT than 2D alone for the detection of grade 2 invasive cancers, for 11- to 20-mm invasive cancers, and for lesions where the dominant radiological feature was categorised as distortion or ASD. One explanation might be that small grade 2 invasive cancers are slightly less likely than grade 1 cancers to show as masses. The addition of DBT to 2D for cancers > 20 mm appeared to have little impact, but ideally a screening programme would detect invasive lesions smaller than 20 mm, so this supports addition of DBT to a 2D examination. It has been shown in other studies that the use of DBT is more advantageous for lesions with mass, rather than microcalcifications, as the main radiological feature.<sup>17,43,44</sup>

One concern is that the addition of DBT will increase the detection of benign lesions such as radial scars and complex sclerosing lesions,<sup>142</sup> resulting in an increase in false-positives and an increased negative biopsy rate. However, these drawbacks should be outweighed by the increase in cancer detection rate. Another concern is that DBT will detect a higher percentage of grade 1 and 2 cancers than 2D mammography. Tomosynthesis facilitates detection of stellate cancers of grades 1 and 2, as these tend to have greater desmoplastic reaction resulting in more spiculate lesions than grade 3 cancers. This appears to be the case from one study<sup>60</sup> reporting the grade of cancers detected using the two techniques, 2D alone and 2D + DBT, but is not supported by results from another.<sup>19</sup> These are the only two studies we found that have published these data. It would be useful to have a breakdown of cancer grade from other large studies such as Friedewald *et al.*<sup>140</sup> and STORM<sup>66</sup> for comparison before any conclusions can be made.

One notable observation from recent studies<sup>19,66,140</sup> was the preferential increase in the detection of invasive cancers with detection of DCIS unchanged. Since overdiagnosis/overtreatment of low-grade DCIS is frequently quoted as one of the 'harms' facing breast screening programmes, confirmation of these results in other trials with DBT would provide additional evidence in support of DBT as a screening tool.

We did observe a significant improvement in sensitivity with 2D + DBT in women aged 50–59 years and for women with breast density  $\geq 50\%$ . As young women are more likely to have high breast density and the improvement of sensitivity for older women is a strong indicator for younger women, these observations are of relevance when considering an extension of the UK screening programme to women younger than 47 years.<sup>5,14,81,115</sup> Increased breast density has been shown to be one of the strongest independent predictors of breast cancer risk and accurate reproducible methods for the measurement of breast density are being developed and validated.<sup>93</sup> The results of our breast density substudy indicated that there was a high degree of variation in the observed density scored by the image readers, whereas the two commercially available software packages appeared to provide a more reliable assessment of breast density when their output is compared with that found elsewhere in the literature.<sup>131,134</sup> Although the software packages had good agreement in the assessment of overall breast volume, they clearly implemented different algorithms for the measurement of fibroglandular breast volume and at high breast densities the correlation was less robust. We confirmed that there is a strong relationship between volumetric density and increased risk of breast cancer. However, this is probably an underestimate as only cancer cases detected at screening were included. Further analysis of interval and subsequent round cancers would provide a truer estimate of breast cancer risk associated with density. The controversy in the USA over mandatory reporting of breast density<sup>120</sup> highlights the need for further research in this area, as noted by Assi *et al.*<sup>93</sup> At present, there is no reference standard for the measurement of breast density but development of an accurate and reproducible method of breast density measurement will be required if it is to be used to assess breast cancer risk and identify groups of women most likely to benefit from shorter or longer screening intervals.

In May 2013, the FDA approved new software for the creation of a synthetic 2D image (C-View) from DBT images based on data from a non-inferiority trial comparing the performance of synthetic 2D + DBT with DBT alone. Our study design allowed comparison of the diagnostic performance of synthetic 2D + DBT with both 2D and 2D + DBT imaging. This comparison is of clinical relevance, since current evidence favours the use of DBT as an adjunct to 2D mammography rather than a stand-alone imaging modality.<sup>64</sup> There was no significant difference in sensitivity between synthetic 2D + DBT and 2D alone, but specificity was significantly higher. In comparison with 2D + DBT, there was no significant difference in sensitivity or specificity for synthetic 2D + DBT. Subgroup analysis suggested that synthetic 2D + DBT, similarly to 2D + DBT, showed better sensitivity for the detection of 11- to 20-mm invasive cancers than 2D alone. However, synthetic 2D + DBT was not as good as 2D or 2D + DBT in the detection of microcalcifications and DCIS of size 11–20 mm (where the dominant radiological feature is likely to be microcalcifications). Only a few studies have been published to date on the performance of synthetic 2D. Although Gur *et al.*<sup>70</sup> reported lower sensitivity but comparable specificity for synthetic 2D + DBT compared with standard 2D + DBT, two recent studies using the same version of the C-View software as our study, Skaane *et al.*<sup>78</sup> and Zuley *et al.*,<sup>143</sup> demonstrated comparable performance. Inferior performance of synthetic 2D images from a prototype GE Healthcare (subsidiary of General Electric, Little Chalford, Buckinghamshire, UK) DBT system was reported by Locatelli *et al.*,<sup>144</sup> limiting its use for research purposes at present. Hologic have produced a computer-aided detection algorithm for identification of microcalcifications which may improve reader performance (Dr Susan M Astley, Department of Imaging Science and Biomedical Engineering, University of Manchester, 2014, personal communication) but to date there are no published data. In addition, coregistration of synthetic 2D and DBT images on the workstation could also improve image interpretation.<sup>142</sup>

An advantage of our reading study was that there were sufficient numbers of cancers to test sensitivity and specificity in each of the subcategories. The distribution of cancer cases in terms of invasion, size, grade and lymph node status was similar to that found in the UK screening programme. Our study involved 26 high-volume readers with variable years of experience and included radiologists, breast physicians and advanced practitioner radiographers. This reflects the pragmatic current practice in UK screening centres. In addition, three readers, each from different sites, were randomly assigned to read the same case independently in one of the three reading arms, reducing the risk of individual reader bias.

The major limitation of the study was that the study data set was primarily cases that had a mammographic abnormality identified by 2D imaging. Thus, the ratio of cancer cases to normal cases was higher than would be encountered in population screening where there is a ratio of 1 : 150–200; therefore, the interpretation of results may not be applicable to those obtained in a screening environment. In addition, the retrospective reading design meant that the readers knew that their reading decision had no clinical implications and there may have been less concern with the consequence of missing a cancer.

At the start of the trial we invited readers to undertake a DBT-only test set reading to ascertain variability in reader performance and whether or not this changed over the duration of the trial following exposure to reading with DBT. The reader test set study suggested that increased experience of reading with DBT had little effect on reading performance but further analysis of individual reader performance from the retrospective reading study will be undertaken. The reader study highlighted various aspects of reader performance that require further research in terms of inter-reader and intersite variations in operating point. The impact of longer reading times with DBT<sup>16,19,40,46,68</sup> needs to be considered if DBT is to be incorporated into the workflow of the screening programme.

## Chapter 7 Conclusions and future research

Although our study was in agreement with recent published studies in demonstrating that 2D + DBT performed better than 2D alone in terms of specificity, we were unable to demonstrate the same significant increase in sensitivity for 2D + DBT reported in these studies in our retrospective reading study. Our study design clearly limits estimation of the impact of DBT on diagnostic performance and extrapolation of our results to a screening population. Our method of case selection of index test cases meant that the relative benefit of DBT on sensitivity has been underestimated, since these cases have already been detected by 2D screening. Similarly, this data set is likely to have overestimated the impact of DBT on specificity.

In our analysis, combined 2D + DBT imaging was shown to be more effective than 2D alone across all age groups and breast densities, particularly for women aged 50–59 years and for breast density  $\geq 50\%$ . If this effect was translated to population screening practice it would be advantageous to the planned age extension of the NHSBSP for younger women who have denser breast tissue and to screening mammography of the younger cohort of FH women. The potential of combined 2D + DBT imaging to reduce the burden of false-positive recalls and associated diagnostic assessments would also benefit screening programmes. Although this study was not designed to assess DBT in the diagnostic and assessment setting, improved visualisation of benign and malignant features using DBT could reduce the need for additional ultrasound, MRI or supplemental mammographic imaging.<sup>39–42</sup>

The screening population would also benefit if some of the anxiety and stress associated with breast screening from unnecessary recalls and additional testing could be reduced.

Longer-term outcomes, as measured by interval cancer rates, will be of importance, since subsequent-round screening with DBT would be unlikely to show a similar magnitude of improvement in sensitivity. However, it could be expected that earlier detection of small invasive cancers would translate into improved clinical outcomes.

Combined 2D + DBT imaging has been approved for screening in the USA and recently published studies indicate that there is convincing evidence of a clear incremental benefit in invasive cancer detection with the use of DBT. An American College of Radiology statement (22 July 2014) noted:

*While there is strong evidence that tomosynthesis will have an important role in breast imaging, further studies are needed to assess tomosynthesis' relationship to long-term clinical outcomes, including reduced mortality. It will also be important to learn which subgroups of women might benefit most from these exams.*<sup>145</sup>

While we recognise that additional information on the performance of DBT is required, the accepted 'gold standard' of a randomised controlled trial (RCT) comparing 2D and 2D + DBT would be time-consuming and expensive, and could delay implementation of a key imaging technology in screening by 5–7 years.

## Future research

- Since no formal cost-effectiveness/cost–benefit evaluations have been published, we suggest that the feasibility and practicalities of implementing DBT into the workflow of a UK screening setting should be evaluated. This should include a cost-effectiveness evaluation or modelling. This should also include evaluation of the additional costs (e.g. upgrading equipment), providing information technology (IT) support for image archiving, connectivity to PACS and IT systems, increased reading time and potential benefits (e.g. increased cancer detection, particularly small invasive cancers, reductions in false-positive recalls, additional diagnostic testing, and reduction in interval cancers).
- For combined imaging with 2D + DBT to be implemented in screening, the use of synthetic 2D to minimise radiation exposure would be advantageous. The overall non-inferiority of synthetic 2D + DBT to 2D alone, shown in our study and in the publications of Skaane *et al.*<sup>78</sup> and Zuley *et al.*,<sup>143</sup> would justify use of this imaging combination in a RCT in a screening setting. However, before synthetic 2D + DBT could be recommended for screening, further comparative work with synthetic 2D and 2D alone should be undertaken; for example quantifying the effect on sensitivity and specificity for lesions with different radiological appearances, of different pathological types (e.g. DCIS, invasive, different grades of tumour) and in different types of breast (e.g. > 50% density).
- It is important to note that the majority of published studies have utilised the Hologic DBT system. Apart from some reading studies, relatively little literature exists for other commercially available DBT systems. Therefore, more information is required from screening studies with other DBT systems to verify the improvements in sensitivity and specificity shown by the Hologic system. This could be partly achieved by modelling using simulated lesions and observer studies such as OPTIMAM (Optimisation of breast cancer detection using digital radiograph technology).<sup>146</sup>
- Comparison should be made with 2D on the size, grade, and type of cancers detected by DBT, for interval cancers and for cancers detected in the subsequent screening round, where we would not expect the same degree of improvement in the detection of invasive cancers.
- We would recommend prognostic modelling on existing data sets of screen-detected and interval cancers to predict the outcome of cases and projected impact on breast cancer mortality. This could be achieved using, for example, the Nottingham Prognostic Index.<sup>147</sup>
- Mammographic density is one of the major risk factors for breast cancer and has the potential to be incorporated into strategies for personalised risk prediction and screening guidelines. For example, younger women with dense breasts, for whom standard 2D mammography is less effective, could benefit more from screening with DBT, and recommendations to customise breast screening frequency for women with different breast cancer risk profiles could also be made. These require development of a validated method of assessing breast density.
- Further research is required into the ability of DBT to discriminate between early aggressive cancers and slow-growing DCIS to address any concerns about DBT contributing to overdiagnosis.



# Acknowledgements

## Contribution of authors

**Fiona J Gilbert** (Professor of Radiology, University of Cambridge): chief investigator for the TOMMY trial, principal in conception and design of study, interpretation of data and writing final report.

**Lorraine Tucker** (Research Radiographer, University of Cambridge): co-ordinator of the TOMMY trial, responsible for day-to-day management of the trial, conducted QC image review and drafted and made revisions to final report.

**Maureen GC Gillan** (Research Fellow, University of Aberdeen): co-ordinator of the TOMMY trial when it was based in Aberdeen. Instrumental in drafting of trial proposal and trial protocol and assisted in drafting and review of final report.

**Paula Willsher** (Research Radiographer, University of Cambridge): conducted image review and QC checks of prospective data, conducted logic checks on final data and assisted in revision and review of final report.

**Julie Cooke** (Consultant Radiologist, Clinical Director, Jarvis Breast Centre, Guildford): PI at trial site, reader in retrospective reading study and reviewed and contributed to final report.

**Karen A Duncan** (Consultant Radiologist, North East Scotland Breast Screening Centre, Aberdeen): PI at trial site, reader in retrospective reading study and reviewed and contributed to final report.

**Michael J Michell** (Consultant Radiologist, Clinical Director Breast Radiology Department, King's College Hospital, London): PI at trial site. Provided expertise from previous trial for study design, conducted tomosynthesis training for readers and provided images for test set. Participated in retrospective reading study and reviewed and contributed to final report.

**Hilary M Dobson** (Consultant Radiologist, Clinical Director West of Scotland Breast Screening Service, Glasgow): PI at trial site, reader in retrospective reading study and reviewed final report.

**Yit Yoong Lim** (Consultant Radiologist, The Nightingale Centre, University Hospital South Manchester): PI at trial site, reader in retrospective reading study and reviewed final report.

**Hema Purushothaman** (Consultant Radiologist, St Bartholomew's Hospital, London): PI at trial site, reader in retrospective reading study and reviewed final report.

**Celia Strudley** (Principal Physicist, NCCPM, Royal Surrey County Hospital, Guildford): developed and conducted QC tests on tomosynthesis equipment used in the TOMMY trial, produced 6-monthly and final imaging equipment QC reports and reviewed final report.

**Susan M Astley** (Reader in Imaging Science, University of Manchester): involved in trial design. Supervised administration of reader study and analysis of test set results, produced test set report and reviewed final report.

**Oliver Morrish** (Lead Clinical Scientist, East Anglian Regional Radiation Protection Service, Cambridge University Hospitals NHS Foundation Trust): analysed breast density data and produced breast density report.

**Kenneth C Young** (Professor, Head of NCCPM and Director of Research, Royal Surrey County Hospital, Guildford): developed QC tests on tomosynthesis equipment used in TOMMY trial. Reviewed and contributed to final report.

**Stephen W Duffy** (Professor of Cancer Screening, Wolfson Institute of Preventive Medicine, Queen Mary University of London): advised on study design and data collection, conducted statistical analysis for the trial and cancer risk analysis for breast density report. Contributed to interpretation of results and review of final report.

## On behalf of the TOMMY trial teams

### Cambridge

Richard Black (Clinical Scientist): management and distribution of images for trial. Produced report on image management.

Sridevi Nagarajan (Senior Data Manager): collation of prospective and retrospective data, production of trial database and analysis of data.

Emily Dixon (Data Manager): collection, collation and input of data, assisted with co-ordination of trial and management of randomisation of images for retrospective study.

Dave Fyvie (Research Assistant): assistance in collection and input of trial data.

Louise Harlow (Trial Co-ordinator): original team member.

### Aberdeen

Herman Klaasen, Tanja Gagliardi, Gerald Lip (Consultant Radiologists): readers in retrospective reading study.

Jeanette Davidson (Data Manager): collection and collation of trial data and transfer and distribution of images.

George Cameron (Computer Physicist): original team member who set up mechanism for electronic data collection and transfer.

### Glasgow

Arachna Seth, Janet Litherland (Consultant Radiologists), Mabel Morrow, Linda McClure (Clinical Specialists): readers in retrospective reading study.

Ann Mumby (Superintendent Radiographer): continued data manager duties of Julie Murphy, collection and collation of trial data and transfer and distribution of images.

Julie Murphy (Data Manager): original data manager.

### **Guildford**

Caroline Kissin, Caroline Taylor, Kathy Stoner, Phillipa Skippage (Consultant Radiologists), Victoria Cooke (Associate Specialist): readers in retrospective reading study.

Lindsay Mungutroy (Data Manager): advised on electronic data collection and transfer and designed spreadsheet for data collection. Responsible for collection and collation of trial data and transfer and distribution of images.

### **King's College Hospital, London**

Juliet Morel, Jane Goligher, Rumana Rahim (Consultant Radiologists), Susan Dyson (Advanced Practitioner Radiographer): readers in retrospective reading study.

Asif Iqbal (Data Manager): collection and collation of trial data, transfer and distribution of images.

Vivien Phillips (Head of Breast Radiography): support throughout the trial.

### **Manchester**

Mary Wilson (Consultant Radiologist, Director Greater Manchester Breast Screening Programme): original PI for site, contributed to trial set-up, reader in retrospective reader study.

Emma Hurley, Ursula Beetles (Consultant Radiologists), Sara Bundred (Breast Physician): readers in retrospective reading study.

Jin Zhou (Data Manager): collection and collation of trial data, transfer and distribution of images and contributed to conduct, analysis and writing up of reader study.

Elaine Harkness (Postdoctoral Research Associate): contributed to conduct, analysis and writing up of reader study.

Catriona Tate (Data Manager): original data manager, contributed to conduct and analysis of reader study.

### **St Bartholomew's Hospital, London**

Tamara Suaris (Consultant Radiologist): temporary PI during maternity leave, reader in retrospective reading study.

Ilyena Froud (Consultant Radiologist): reader in retrospective reading study.

Francis McNally (Data Manager): collection and collation of trial data and transfer and distribution of images.

## Other acknowledgements

Ralph Highnam (chief executive officer, Matakina International Limited): for technical advice and support regarding operation of Volpara breast density assessment tool.

Ashwini Kshirsagar (Senior Principal Scientist and Manager, Hologic Ltd): for technical support and advice regarding operation of Quantra breast density assessment software.

Miriam Harris (Queen Mary University of London Trials Advisory Group): lay representative who participated in collaborators' meetings.

Professor Janet Dunn (Clinical Trials Unit, University of Warwick), Professor Andrew Evans (Chairperson of Breast Imaging, University of Dundee), Professor Julietta Patnick (Director NHS Cancer Screening Programmes): independent members of the Trial Steering Committee.

## References

1. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012;**380**:1778–86. [http://dx.doi.org/10.1016/S0140-6736\(12\)61611-0](http://dx.doi.org/10.1016/S0140-6736(12)61611-0)
2. Michell MJ. Breast screening review: a radiologist's perspective. *Br J Radiol* 2012;**85**:845–7. <http://dx.doi.org/10.1259/bjr/21332901>
3. Duncan KA, Needham G, Gilbert FJ, Deans HE. Incident round cancers: what lessons can we learn? *Clin Radiol* 1998;**53**:29–32. [http://dx.doi.org/10.1016/S0009-9260\(98\)80030-5](http://dx.doi.org/10.1016/S0009-9260(98)80030-5)
4. Carney PA, Miglioretti DL, Yankaskas BC, Kerlikowske K, Rosenberg R, Rutter CM, *et al.* Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Ann Int Med* 2003;**138**:168–75. <http://dx.doi.org/10.7326/0003-4819-138-3-200302040-00008>
5. Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. *Radiology* 2002;**205**:1165–75.
6. Rosenberg RD, Hunt WC, Williamson MR, Gilliland FD, Wiest PW, Kelsey CA, *et al.* Effects of age, breast density, ethnicity, and estrogen replacement therapy on screening mammographic sensitivity and cancer stage at diagnosis: review of 183,134 screening mammograms in Albuquerque, New Mexico. *Radiology* 1998;**209**:511–18. <http://dx.doi.org/10.1148/radiology.209.2.9807581>
7. Day N, Warren R. Mammographic screening and mammographic patterns. *Breast Cancer Res* 2000;**2**:247–51. <http://dx.doi.org/10.1186/bcr64>
8. Mandelson MT, Oestreicher N, Porter PL, White D, Finder CA, Taplin SH, *et al.* Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Canc Inst* 2000;**92**:1081–7. <http://dx.doi.org/10.1093/jnci/92.13.1081>
9. Chiu SY, Duffy S, Yen AM, Tabar L, Smith RA, Chen HH. Effect of baseline breast density on breast cancer incidence, stage, mortality, and screening parameters: 25-year follow-up of a Swedish mammographic screening. *Cancer Epidemiol Biomarkers Prev* 2010;**19**:1219–28. <http://dx.doi.org/10.1158/1055-9965.EPI-09-1028>
10. van Gils CH, Otten JD, Verbeek AL, Hendriks JH, Holland R. Effect of mammographic breast density on breast cancer screening performance: a study in Nijmegen, the Netherlands. *J Epidemiol Community Health* 1998;**52**:267–71. <http://dx.doi.org/10.1136/jech.52.4.267>
11. Buist DS, Porter PL, Lehman C, Taplin SH, White E. Factors contributing to mammography failure in women aged 40–49 years. *J Natl Canc Inst* 2004;**96**:1432–40. <http://dx.doi.org/10.1093/jnci/djh269>
12. Sala E, Warren R, Duffy S, Welch A, Luben R, Day N. High risk mammographic parenchymal patterns and diet: a case–control study. *Br J Cancer* 2000;**83**:121–6.
13. Nickson C, Kavanagh AM. Tumour size at detection according to different measures of mammographic breast density. *J Med Screen* 2009;**16**:140–6. <http://dx.doi.org/10.1258/jms.2009.009054>
14. Department of Health (DH). *Cancer Reform Strategy*. London: DH; 2007.

15. Moser K, Sellars S, Wheaton M, Cooke J, Duncan A, Maxwell A, *et al.* Extending the age range for breast screening in England: pilot study to assess the feasibility and acceptability of randomization. *J Med Screen* 2011;**18**:96–102. <http://dx.doi.org/10.1258/jms.2011.011065>
16. Gur D, Abrams GS, Chough DM, Ganott MA, Hakim CM, Perrin RL, *et al.* Digital breast tomosynthesis: observer performance study. *AJR Am J Roentgenol* 2009;**193**:586–91. <http://dx.doi.org/10.2214/AJR.08.2031>
17. Rafferty EA, Park JM, Philpotts LE, Poplack SP, Sumkin JH, Halpern EF, *et al.* Assessing radiologist performance using combined digital mammography and breast tomosynthesis compared with digital mammography alone: results of a multicenter, multireader trial. *Radiology* 2013;**266**:104–13. <http://dx.doi.org/10.1148/radiol.12120674>
18. Caumo F, Bernardi D, Ciatto S, Macaskill P, Pellegrini M, Brunelli S, *et al.* Incremental effect from integrating 3D-mammography (tomosynthesis) with 2D-mammography: increased breast cancer detection evident for screening centres in a population-based trial. *Breast* 2014;**23**:76–80. <http://dx.doi.org/10.1016/j.breast.2013.11.006>
19. Skaane P, Bandos AI, Gullien R, Eben EB, Ekseth U, Haakenaasen U, *et al.* Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology* 2013;**267**:47–56. <http://dx.doi.org/10.1148/radiol.12121373>
20. Bernardi D, Ciatto S, Pellegrini M, Tuttobene P, Fanto C, Valentini M, *et al.* Prospective study of breast tomosynthesis as a triage to assessment in screening. *Breast Cancer Res* 2012;**133**:267–71. <http://dx.doi.org/10.1007/s10549-012-1959-y>
21. Brett J, Bankhead C, Henderson B, Watson E, Austoker J. The psychological impact of mammographic screening: a systematic review. *Psycho-oncology* 2005;**14**:917–38. <http://dx.doi.org/10.1002/pon.904>
22. Bond M, Pavey T, Welch K, Cooper C, Garside R, Dean S, *et al.* Psychological consequences of false-positive screening mammograms in the UK. *Evid Based Med* 2013;**18**:54–61. <http://dx.doi.org/10.1136/eb-2012-100608>
23. Dobbins JT 3rd, Godfrey DJ. Digital X-ray tomosynthesis: current state of the art and clinical potential. *Phys Med Biol* 2003;**48**:R65–106. <http://dx.doi.org/10.1088/0031-9155/48/19/R01>
24. Smith A. Full-field breast tomosynthesis. *Radiol Manage* 2005;**27**:25–31.
25. Dobbins JT III. Tomosynthesis imaging: at a translational crossroads. *Med Phys* 2009;**36**:1956–67. <http://dx.doi.org/10.1118/1.3120285>
26. Sechopoulos I. A review of breast tomosynthesis. Part I: the image acquisition process. *Med Phys* 2013;**40**:014301. <http://dx.doi.org/10.1118/1.4770279>
27. Sechopoulos I. A review of breast tomosynthesis. Part II: image reconstruction, processing and analysis, and advanced applications. *Med Phys* 2013;**40**:014302. <http://dx.doi.org/10.1118/1.4770281>
28. Niklason LT, Christian BT, Niklason LE, Kopans DB, Castleberry DE, Opsahl-Ong BH, *et al.* Digital tomosynthesis in breast imaging. *Radiology* 1997;**205**:399–406. <http://dx.doi.org/10.1148/radiology.205.2.9356620>
29. Suryanarayanan S, Karellas A, Vedantham S, Glick SJ, D'Orsi CJ, Baker SP, *et al.* Comparison of tomosynthesis methods used with digital mammography. *Acad Radiol* 2000;**7**:1085–97. [http://dx.doi.org/10.1016/S1076-6332\(00\)80061-6](http://dx.doi.org/10.1016/S1076-6332(00)80061-6)
30. Wu T, Stewart A, Stanton M, McCauley T, Phillips W, Kopans DB, *et al.* Tomographic mammography using a limited number of low-dose cone-beam projection images. *Med Phys* 2003;**30**:365–80. <http://dx.doi.org/10.1118/1.1543934>

31. Skaane P, Gullien R, Bjorndal H, Eben EB, Ekseth U, Haakenaasen U, *et al.* Digital breast tomosynthesis (DBT): initial experience in a clinical setting. *Acta Radiol* 2012;**53**:524–9. <http://dx.doi.org/10.1258/ar.2012.120062>
32. Yang TL, Liang HL, Chou CP, Huang JS, Pan HB. The adjunctive digital breast tomosynthesis in diagnosis of breast cancer. *BioMed Res Int* 2013;**2013**:597253. <http://dx.doi.org/10.1155/2013/597253>
33. Svane G, Azavedo E, Lindman K, Urech M, Nilsson J, Weber N, *et al.* Clinical experience of photon counting breast tomosynthesis: comparison with traditional mammography. *Acta Radiol* 2011;**52**:134–42. <http://dx.doi.org/10.1258/ar.2010.100262>
34. Timberg P, Bath M, Andersson I, Mattsson S, Tingberg A, Ruschin M. In-plane visibility of lesions using breast tomosynthesis and digital mammography. *Med Phys* 2010;**37**:5618–26. <http://dx.doi.org/10.1118/1.3488899>
35. Fornvik D, Zackrisson S, Ljungberg O, Svahn T, Timberg P, Tingberg A, *et al.* Breast tomosynthesis: accuracy of tumor measurement compared with digital mammography and ultrasonography. *Acta Radiol* 2010;**51**:240–7. <http://dx.doi.org/10.3109/02841850903524447>
36. Seo N, Kim HH, Shin HJ, Cha JH, Kim H, Moon JH, *et al.* Digital breast tomosynthesis versus full-field digital mammography: comparison of the accuracy of lesion measurement and characterization using specimens [published online ahead of print 4 September 2013]. *Acta Radiol* 2013. <http://dx.doi.org/10.1177/0284185113503636>
37. Mun HS, Kim HH, Shin HJ, Cha JH, Ruppel PL, Oh HY, *et al.* Assessment of extent of breast cancer: comparison between digital breast tomosynthesis and full-field digital mammography. *Clin Radiol* 2013;**68**:1254–9. <http://dx.doi.org/10.1016/j.crad.2013.07.006>
38. Luparia A, Mariscotti G, Durando M, Ciatto S, Bosco D, Campanino PP, *et al.* Accuracy of tumour size assessment in the preoperative staging of breast cancer: comparison of digital mammography, tomosynthesis, ultrasound and MRI. *Radiol Med* 2013;**118**:1119–36. <http://dx.doi.org/10.1007/s11547-013-0941-z>
39. Hakim CM, Chough DM, Ganott MA, Sumkin JH, Zuley ML, Gur D. Digital breast tomosynthesis in the diagnostic environment: a subjective side-by-side review. *AJR Am J Roentgenol* 2010;**195**:W172–6. <http://dx.doi.org/10.2214/AJR.09.3244>
40. Zuley ML, Bandos AI, Ganott MA, Sumkin JH, Kelly AE, Catullo VJ, *et al.* Digital breast tomosynthesis versus supplemental diagnostic mammographic views for evaluation of noncalcified breast lesions. *Radiology* 2013;**266**:89–95. <http://dx.doi.org/10.1148/radiol.12120552>
41. Tagliafico A, Astengo D, Cavagnetto F, Rosasco R, Rescinito G, Calabrese M. One-to-one comparison between digital spot compression view and digital breast tomosynthesis. *Eur Radiol* 2012;**22**:539–44. <http://dx.doi.org/10.1007/s00330-011-2305-1>
42. Noroozian M, Hadjiiski L, Rahnama-Moghadam S, Klein KA, Jeffries DO, Pinsky RW, *et al.* Digital breast tomosynthesis is comparable to mammographic spot views for mass characterization. *Radiology* 2012;**262**:61–8. <http://dx.doi.org/10.1148/radiol.11101763>
43. Poplack SP, Tosteson TD, Kogel CA, Nagy HM. Digital breast tomosynthesis: initial experience in 98 women with abnormal digital screening mammography. *AJR Am J Roentgenol* 2007;**189**:616–23. <http://dx.doi.org/10.2214/AJR.07.2231>
44. Spangler ML, Zuley ML, Sumkin JH, Abrams G, Ganott MA, Hakim C, *et al.* Detection and classification of calcifications on digital breast tomosynthesis and 2D digital mammography: a comparison. *AJR Am J Roentgenol* 2011;**196**:320–4. <http://dx.doi.org/10.2214/AJR.10.4656>



45. Michell MJ, Iqbal A, Wasan RK, Evans DR, Peacock C, Lawinski CP, *et al.* A comparison of the accuracy of film-screen mammography, full-field digital mammography, and digital breast tomosynthesis. *Clin Radiol* 2012;**67**:976–81. <http://dx.doi.org/10.1016/j.crad.2012.03.009>
46. Wallis MG, Moa E, Zanca F, Leifland K, Danielsson M. Two-view and single-view tomosynthesis versus full-field digital mammography: high-resolution X-ray imaging observer study. *Radiology* 2012;**262**:788–96. <http://dx.doi.org/10.1148/radiol.11103514>
47. Teertstra HJ, Loo CE, van den Bosch MA, van Tinteren H, Rutgers EJ, Muller SH, *et al.* Breast tomosynthesis in clinical practice: initial results. *Eur Radiol* 2010;**20**:16–24. <http://dx.doi.org/10.1007/s00330-009-1523-2>
48. Andersson I, Ikeda DM, Zackrisson S, Ruschin M, Svahn T, Timberg P, *et al.* Breast tomosynthesis and digital mammography: a comparison of breast cancer visibility and BIRADS classification in a population of cancers with subtle mammographic findings. *Eur Radiol* 2008;**18**:2817–25. <http://dx.doi.org/10.1007/s00330-008-1076-9>
49. Kopans D, Gavenonis S, Halpern E, Moore R. Calcifications in the breast and digital breast tomosynthesis. *Breast J* 2011;**17**:638–44. <http://dx.doi.org/10.1111/j.1524-4741.2011.01152.x>
50. Reiser I, Nishikawa RM, Edwards AV, Kopans DB, Schmidt RA, Papaioannou J, *et al.* Automated detection of microcalcification clusters for digital breast tomosynthesis using projection data only: a preliminary study. *Med Phys* 2008;**35**:1486–93. <http://dx.doi.org/10.1118/1.2885366>
51. Timberg P, Lang K, Nystrom M, Holmqvist K, Wagner P, Fornvik D, *et al.* Investigation of viewing procedures for interpretation of breast tomosynthesis image volumes: a detection-task study with eye tracking. *Eur Radiol* 2013;**23**:997–1005. <http://dx.doi.org/10.1007/s00330-012-2675-z>
52. Das M, Gifford HC, O'Connor JM, Glick SJ. Evaluation of a variable dose acquisition technique for microcalcification and mass detection in digital breast tomosynthesis. *Med Phys* 2009;**36**:1976–84. <http://dx.doi.org/10.1118/1.3116902>
53. Nishikawa RM, Reiser I, Seifi P. A new approach to digital breast tomosynthesis for breast cancer screening. *Proc SPIE Med Imaging* 2007;**6510**:65103C–65108.
54. Tingberg A, Andersson I, Ikeda D, Ruschin M, Svahn T, Timberg P. BIRADS classification of breast tomosynthesis compared to mammography and ultrasonography. In Krupinski EA, editor. *Digital Mammography: 9th International Workshop on Digital Mammography*. Tucson, AZ: Heidelberg Springer; 2008. pp. 67–73. [http://dx.doi.org/10.1007/978-3-540-70538-3\\_10](http://dx.doi.org/10.1007/978-3-540-70538-3_10)
55. Svahn TM, Chakraborty DP, Ikeda D, Zackrisson S, Do Y, Mattsson S, *et al.* Breast tomosynthesis and digital mammography: a comparison of diagnostic accuracy. *Br J Radiol* 2012;**85**:1019, e1074–82. <http://dx.doi.org/10.1259/bjr/53282892>
56. Thibault F, Dromain C, Breucq C, Balleyguier CS, Malhaire C, Steyaert L, *et al.* Digital breast tomosynthesis versus mammography and breast ultrasound: a multireader performance study. *Eur Radiol* 2013;**23**:2441–9. <http://dx.doi.org/10.1007/s00330-013-2863-5>
57. Gennaro G, Toledano A, di Maggio C, Baldan E, Bezzon E, La Grassa M, *et al.* Digital breast tomosynthesis versus digital mammography: a clinical performance study. *Eur Radiol* 2010;**20**:1545–53. <http://dx.doi.org/10.1007/s00330-009-1699-5>
58. Smith AP, Rafferty EA, Niklason L. Clinical performance of breast tomosynthesis as a function of radiologist experience level. In Kuprinski EA, editor. *Digital Mammography: 9th International Workshop on Digital Mammography*. Tucson, AZ: Heidelberg Springer; 2008. pp. 61–66. [http://dx.doi.org/10.1007/978-3-540-70538-3\\_9](http://dx.doi.org/10.1007/978-3-540-70538-3_9)
59. Haas B, Kalra V, Geisel J, Raghu M, Durand M, Philpotts L. Comparison of tomosynthesis plus digital mammography and digital alone for breast cancer screening. *Radiology* 2013;**269**:694–700. <http://dx.doi.org/10.1148/radiol.13130307>



60. Rose SL, Tidwell AL, Bujnoch LJ, Kushwaha AC, Nordmann AS, Sexton R Jr, Implementation of breast tomosynthesis in a routine screening practice: an observational study. *AJR Am J Roentgenol* 2013;**200**:1401–8. <http://dx.doi.org/10.2214/AJR.12.9672>
61. Diekmann F, Bick U. Breast tomosynthesis. *Semin Ultrasound CT MR* 2011;**32**:281–7. <http://dx.doi.org/10.1053/j.sult.2011.03.002>
62. Tingberg A, Fornvik D, Mattsson S, Svahn T, Timberg P, Zackrisson S. Breast cancer screening with tomosynthesis: initial experiences. *Radiat Prot Dosimetry* 2011;**147**:180–3. <http://dx.doi.org/10.1093/rpd/ncr296>
63. Alakhras M, Bourne R, Rickard M, Ng KH, Pietrzyk M, Brennan PC. Digital tomosynthesis: a new future for breast imaging? *Clin Radiol* 2013;**68**:225–36. <http://dx.doi.org/10.1016/j.crad.2013.01.007>
64. Houssami N, Skaane P. Overview of the evidence on digital breast tomosynthesis in breast cancer detection. *Breast* 2013;**22**:101–8. <http://dx.doi.org/10.1016/j.breast.2013.01.017>
65. Evaluation of BreastScreen Australia Program. *Digital Breast Tomosynthesis: Overview of the Evidence and Issues for Its Use in Screening for Breast Cancer*. Australia: Department of Health and Ageing, Australian Government; 2013.
66. Ciatto S, Houssami N, Bernardi D, Caumo F, Pellegrini M, Brunelli S, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *Lancet Oncol* 2013;**14**:583–9. [http://dx.doi.org/10.1016/S1470-2045\(13\)70134-7](http://dx.doi.org/10.1016/S1470-2045(13)70134-7)
67. Gur D. Tomosynthesis: potential clinical role in breast imaging. *AJR Am J Roentgenol* 2007;**189**:614–15. <http://dx.doi.org/10.2214/AJR.07.2588>
68. Good WF, Abrams GS, Catullo VJ, Chough DM, Ganott MA, Hakim CM, et al. Digital breast tomosynthesis: a pilot observer study. *AJR Am J Roentgenol* 2008;**190**:865–9. <http://dx.doi.org/10.2214/AJR.07.2841>
69. Kopans DB. Mammography: yet another challenge. *Radiology* 2009;**253**:587–9. <http://dx.doi.org/10.1148/radiol.2533091517>
70. Gur D, Zuley ML, Anello MI, Rathfon GY, Chough DM, Ganott MA, et al. Dose reduction in digital breast tomosynthesis (DBT) screening using synthetically reconstructed projection images: an observer performance study. *Acad Radiol* 2012;**19**:166–71. <http://dx.doi.org/10.1016/j.acra.2011.10.003>
71. Lee CI, Lehman CD. Digital breast tomosynthesis and the challenges of implementing an emerging breast cancer screening technology into clinical practice. *J Am Coll Radiol* 2013;**10**:913–17. <http://dx.doi.org/10.1016/j.jacr.2013.09.010>
72. Helvie MA. Digital mammography imaging: breast tomosynthesis and advanced applications. *Radiol Clin North Am* 2010;**48**:917–29. <http://dx.doi.org/10.1016/j.rcl.2010.06.009>
73. Baker JA, Lo JY. Breast tomosynthesis: state-of-the-art and review of the literature. *Acad Radiol* 2011;**18**:298–310. <http://dx.doi.org/10.1016/j.acra.2011.06.011>
74. AETNA. *Clinical Policy Bulletin: Mammography*. URL: [www.aetna.com/cpb/medical/data/500\\_599/0584.html](http://www.aetna.com/cpb/medical/data/500_599/0584.html) (accessed 15 December 2013).
75. Mundy L, Liufu V, Braunack-Mayer A, Merlin T, Hiller JE. *New and Emerging Technologies for Breast Cancer Detection: Australia and New Zealand Horizon Scanning Network. Emerging Technology Bulletin*. Canberra, ACT: Australian Government; 2009.
76. Technology Assessment No. 9: Digital breast tomosynthesis. *Obstet Gynecol* 2013;**121**:1415–17. <http://dx.doi.org/10.1097/01.AOG.0000431055.71711.dc>

77. Oslo University Hospital. *Tomosynthesis in the Oslo Breast Cancer Screening Program (DBT)*. URL: <http://clinicaltrials.gov/show/NCT01248546> (accessed 4 November 2013).
78. Skaane P, Bandos AI, Eben EB, Jepsen IN, Krager M, Haakenaasen U, et al. Two-view digital breast tomosynthesis screening with synthetically reconstructed projection images: comparison with digital breast tomosynthesis with full-field digital mammographic images. *Radiology* 2014;**271**:655–63. <http://dx.doi.org/10.1148/radiol.13131391>
79. Berg WA. Tailored supplemental screening for breast cancer: what now and what next? *AJR Am J Roentgenol* 2009;**192**:390–9. <http://dx.doi.org/10.2214/AJR.08.1706>
80. Gilbert FJ, Young KC, Astley SM, Whelehan P, Gillan MGC. *Digital Breast Tomosynthesis*. NHSBSP Publication No. 69. Sheffield: NHS Cancer Screening Programmes; 2010.
81. National Institute for Health and Care Excellence (NICE). *Familial Breast Cancer: Classification and Care of People at Risk of Familial Breast Cancer and Management of Breast Cancer and Related Risks in People with a Family History of Breast Cancer*. London, UK: NICE; 2013. [www.nice.org.uk/guidance/cg41](http://www.nice.org.uk/guidance/cg41) (accessed August 2013).
82. Law J, Faulkner K, Young KC. Risk factors for induction of breast cancer by X-rays and their implications for breast screening. *Br J Radiol* 2007;**80**:261–6. <http://dx.doi.org/10.1259/bjr/20496795>
83. Young KC, Faulkner K, Wall B, Muirhead C. *Review of Radiation Risk in Breast Screening*. NHSBSP Publication No. 54. Sheffield: NHS Cancer Screening Programmes; 2003.
84. Kulama E, Burch A, Castellano I, Lawinski CP, Marshall NW, Young KC. *Commissioning and Routine Testing of Full Field Digital Mammography Systems*. NHSBSP Equipment Report 0604 Version 3. Sheffield: NHS Cancer Screening Programmes; 2009.
85. Maxwell AJ, Ridley NT, Rubin G, Wallis MG, Gilbert FJ, Michell MJ. The Royal College of Radiologists Breast Group breast imaging classification. *Clin Radiol* 2009;**64**:624–7. <http://dx.doi.org/10.1016/j.crad.2009.01.010>
86. Kopans DB. Basic physics and doubts about relationship between mammographically determined tissue density and breast cancer risk. *Radiology* 2008;**246**:348–53. <http://dx.doi.org/10.1148/radiol.2461070309>
87. Aitken Z, McCormack VA, Highnam RP, Martin L, Gunasekara A, Melnichouk O, et al. Screen-film mammographic density and breast cancer risk: a comparison of the volumetric standard mammogram form and the interactive threshold measurement methods. *Cancer Epidemiol Biomarkers Prev* 2010;**19**:418–28. <http://dx.doi.org/10.1158/1055-9965.EPI-09-1059>
88. Uematsu T. The emerging role of breast tomosynthesis. *Breast Cancer* 2013;**20**:204–12. <http://dx.doi.org/10.1007/s12282-013-0456-4>
89. Machin D, Campbell MJ, Tan SB, Tan SH. *Sample Size Tables for Clinical Studies*. 2nd edn. Oxford: Blackwell Science; 1997.
90. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 1947;**12**:153–7. <http://dx.doi.org/10.1007/BF02295996>
91. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;**44**:837–45. <http://dx.doi.org/10.2307/2531595>
92. Astley SM, Connor S, Lim Y, Tate C, Entwistle H, Morris J, et al. A comparison of image interpretation times in full field digital mammography and digital breast tomosynthesis. *Proc SPIE 8637 Medical Imaging 2013: Image Perception, Observer Performance, and Technology Assessment*. <http://dx.doi.org/10.1117/12.2006039>

93. Assi V, Warwick J, Cuzick J, Duffy SW. Clinical and epidemiological issues in mammographic density. *Nat Rev Clin Oncol* 2012;**9**:33–40. <http://dx.doi.org/10.1038/nrclinonc.2011.173>
94. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2006;**15**:1159–69. <http://dx.doi.org/10.1158/1055-9965.EPI-06-0034>
95. Boyd NF, Byng JW, Jong RA, Fishell EK, Little LE, Miller AB, *et al*. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. *J Natl Canc Inst* 1995;**87**:670–5. <http://dx.doi.org/10.1093/jnci/87.9.670>
96. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, *et al*. Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 2007;**356**:227–36. <http://dx.doi.org/10.1056/NEJMoa062790>
97. Boyd N, Martin L, Gunasekara A, Melnichouk O, Maudsley G, Peressotti C, *et al*. Mammographic density and breast cancer risk: evaluation of a novel method of measuring breast tissue volumes. *Cancer Epidemiol Biomarkers Prev* 2009;**18**:1754–62. <http://dx.doi.org/10.1158/1055-9965.EPI-09-0107>
98. Boyd N, Martin L, Chavez S, Gunasekara A, Salleh A, Melnichouk O, *et al*. Breast-tissue composition and other risk factors for breast cancer in young women: a cross-sectional study. *Lancet Oncol* 2009;**10**:569–80. [http://dx.doi.org/10.1016/S1470-2045\(09\)70078-6](http://dx.doi.org/10.1016/S1470-2045(09)70078-6)
99. Boyd NF, Martin LJ, Yaffe MJ, Minkin S. Mammographic density and breast cancer risk: current understanding and future prospects. *Breast Cancer Res* 2011;**13**:223. <http://dx.doi.org/10.1186/bcr2942>
100. Cuzick J, Forbes JF, Howell A. Re: Tamoxifen for the prevention of breast cancer: current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study. *J Natl Canc Inst* 2006;**98**:643; author reply–4.
101. Vachon CM, van Gils CH, Sellers TA, Ghosh K, Pruthi S, Brandt KR, *et al*. Mammographic density, breast cancer risk and risk prediction. *Breast Cancer Res* 2007;**9**:217. <http://dx.doi.org/10.1186/bcr1829>
102. Ursin G, Ma H, Wu AH, Bernstein L, Salane M, Parisky YR, *et al*. Mammographic density and breast cancer in three ethnic groups. *Cancer Epidemiol Biomarkers Prev* 2003;**12**:332–8.
103. Byrne C, Schairer C, Wolfe J, Parekh N, Salane M, Brinton LA, *et al*. Mammographic features and breast cancer risk: effects with time, age, and menopause status. *J Natl Canc Inst* 1995;**87**:1622–9. <http://dx.doi.org/10.1093/jnci/87.21.1622>
104. Torres-Mejia G, De Stavola B, Allen DS, Perez-Gavilan JJ, Ferreira JM, Fentiman IS, *et al*. Mammographic features and subsequent risk of breast cancer: a comparison of qualitative and quantitative evaluations in the Guernsey prospective studies. *Cancer Epidemiol Biomarkers Prev* 2005;**14**:1052–9. <http://dx.doi.org/10.1158/1055-9965.EPI-04-0717>
105. Ziv E, Shepherd J, Smith-Bindman R, Kerlikowske K. Mammographic breast density and family history of breast cancer. *J Natl Canc Inst* 2003;**95**:556–8. <http://dx.doi.org/10.1093/jnci/95.7.556>
106. Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammography. *JAMA* 1996;**276**:33–8. <http://dx.doi.org/10.1001/jama.1996.03540010035027>
107. Chiarelli AM, Kirsh VA, Klar NS, Shumak R, Jong R, Fishell E, *et al*. Influence of patterns of hormone replacement therapy use and mammographic density on breast cancer detection. *Cancer Epidemiol Biomarkers Prev* 2006;**15**:1856–62. <http://dx.doi.org/10.1158/1055-9965.EPI-06-0290>

108. Ghosh K, Brandt KR, Sellers TA, Reynolds C, Scott CG, Maloney SD, *et al.* Association of mammographic density with the pathology of subsequent breast cancer among postmenopausal women. *Cancer Epidemiol Biomarkers Prev* 2008;**17**:872–9. <http://dx.doi.org/10.1158/1055-9965.EPI-07-0559>
109. Aiello EJ, Buist DS, White E, Porter PL. Association between mammographic breast density and breast cancer tumor characteristics. *Cancer Epidemiol Biomarkers Prev* 2005;**14**:662–8. <http://dx.doi.org/10.1158/1055-9965.EPI-04-0327>
110. Yaghjian L, Colditz GA, Collins LC, Schnitt SJ, Rosner B, Vachon C, *et al.* Mammographic breast density and subsequent risk of breast cancer in postmenopausal women according to tumor characteristics. *J Natl Canc Inst* 2011;**103**:1179–89. <http://dx.doi.org/10.1093/jnci/djr225>
111. Sala E, Solomon L, Warren R, McCann J, Duffy S, Luben R, *et al.* Size, node status and grade of breast tumours: association with mammographic parenchymal patterns. *Eur Radiol* 2000;**10**:157–61. <http://dx.doi.org/10.1007/s003300050025>
112. Roubidoux MA, Bailey JE, Wray LA, Helvie MA. Invasive cancers detected after breast cancer screening yielded a negative result: relationship of mammographic density to tumor prognostic factors. *Radiology* 2004;**230**:42–8. <http://dx.doi.org/10.1148/radiol.2301020589>
113. Kerlikowske K, Zhu W, Hubbard RA, Geller B, Dittus K, Braithwaite D, *et al.* Outcomes of screening mammography by frequency, breast density, and postmenopausal hormone therapy. *JAMA* 2013;**173**:807–16.
114. van Gils CH, Otten JD, Verbeek AL, Hendriks JH. Mammographic breast density and risk of breast cancer: masking bias or causality? *Eur J Epidemiol* 1998;**14**:315–20. <http://dx.doi.org/10.1023/A:1007423824675>
115. Fletcher SW, Elmore JG. Clinical practice: mammographic screening for breast cancer. *N Engl J Med* 2003;**348**:1672–80. <http://dx.doi.org/10.1056/NEJMcp021804>
116. Corsetti V, Houssami N, Ghirardi M, Ferrari A, Speziani M, Bellarosa S, *et al.* Evidence of the effect of adjunct ultrasound screening in women with mammography-negative dense breasts: interval breast cancers at 1 year follow-up. *Eur J Cancer* 2011;**47**:1021–6. <http://dx.doi.org/10.1016/j.ejca.2010.12.002>
117. Berg WA, Zhang Z, Lehrer D, Jong RA, Pisano ED, Barr RG, *et al.* Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. *JAMA* 2012;**307**:1394–404. <http://dx.doi.org/10.1001/jama.2012.388>
118. Nohacker M, Duda V, Hahn M, Warm M, Degenhardt F, Madjar H, *et al.* Early detection of breast cancer: benefits and risks of supplemental breast ultrasound in asymptomatic women with mammographically dense breast tissue. A systematic review. *BMC Cancer* 2009;**9**:335. <http://dx.doi.org/10.1186/1471-2407-9-335>
119. Houssami N, Kerlikowske K. The impact of breast density on breast cancer risk and breast screening. *Curr Breast Cancer Rep* 2012;**4**:161–8. <http://dx.doi.org/10.1007/s12609-012-0070-z>
120. Are You Dense, Inc. *Are You Dense? Exposing the Best-Kept Secret®*; 2012. URL: [www.areyoudense.org](http://www.areyoudense.org) (accessed 29 December 2014).
121. Wang J, Azziz A, Fan B, Malkov S, Klifa C, Newitt D, *et al.* Agreement of mammographic measures of volumetric breast density to MRI. *PLOS ONE* 2013;**8**:e81653. <http://dx.doi.org/10.1371/journal.pone.0081653>
122. Nicholson BT, LoRusso AP, Smolkin M, Bovbjerg VE, Petroni GR, Harvey JA. Accuracy of assigned BI-RADS breast density category definitions. *Acad Radiol* 2006;**13**:1143–9. <http://dx.doi.org/10.1016/j.acra.2006.06.005>

123. Martin KE, Helvie MA, Zhou C, Roubidoux MA, Bailey JE, Paramagul C, *et al.* Mammographic density measured with quantitative computer-aided method: comparison with radiologists' estimates and BI-RADS categories. *Radiology* 2006;**240**:656–65. <http://dx.doi.org/10.1148/radiol.2402041947>
124. Wang XH, Good WF, Chapman BE, Chang YH, Poller WR, Chang TS, *et al.* Automated assessment of the composition of breast tissue revealed on tissue-thickness-corrected mammography. *AJR Am J Roentgenol* 2003;**180**:257–62. <http://dx.doi.org/10.2214/ajr.180.1.1800257>
125. Byng JW, Yaffe MJ, Jong RA, Shumak RS, Lockwood GA, Tritchler DL, *et al.* Analysis of mammographic density and breast cancer risk from digitized mammograms. *Radiographics* 1998;**18**:1587–98. <http://dx.doi.org/10.1148/radiographics.18.6.9821201>
126. Conant EF, Li D, Gavenonis S, Bakic PR, Carton AK, Zang C, *et al.* A comparative study of the inter-reader variability of breast percent density estimation in digital mammography: potential effect of readers training and clinical experience. In J Marti, Oliver A, Freixenet J, editors. *Digital Mammography: 10th International Workshop on Digital Mammography*. Girona, Catalonia, Spain: Heidelberg Springer; 2010. pp. 114–120.
127. Yaffe MJ. *Advanced Applications of Digital Mammography*. Pisani ED, Yaffe MJ, Kuzmiak C, editors. Philadelphia: Lippincott, Williams & Wilkins; 2004.
128. Hartman K, Highnam R, Warren R, Jackson V. Volumetric assessment of breast tissue composition from FFDM images. In Krupinski EA, editors. *Digital Mammography: 9th International Workshop on Digital Mammography*. Tucson, AZ: Heidelberg Springer; 2008. pp. 33–9.
129. HTA Programme. *TOMMY Trial: A Comparison of TOMosynthesis with Digital Mammography in the UK NHS Breast Screening Programme* (26 August 2013). URL: [www.nets.nihr.ac.uk/projects/hta/0922182](http://www.nets.nihr.ac.uk/projects/hta/0922182) (accessed 11 November 2014).
130. Highnam R, Brady MS, Yaffe M, Karssemeijer N, Harvey J. Robust breast composition: Volpara™. In Marti J, Oliver A, Freixenet J, Marti R, editors. *Digital Mammography: 10th International Workshop on Digital Mammography*. Girona, Catalonia, Spain: Heidelberg Springer; 2010. pp. 342–9.
131. Sergeant JC, Warwick J, Evans DG, Howell A, Berks M, Stavrinou P, *et al.* Volumetric and area-based breast density measurement in the Predicting Risk of Cancer at Screening (PROCAS) study. In Maidment AA, Bakic P, Gavenonis S, editors. *Breast Imaging: 11th International Workshop on Digital Mammography*. Philadelphia, PA: Heidelberg Springer; 2012. pp. 228–35.
132. Makaronidis J, Berks M, Sergeant J, Morris J, Boggis C, Wilson M, *et al.* Assessment of breast density: reader performance using synthetic mammographic images. In Manning DJ, Abbey CK, editors. *Medical Imaging 2011: Image Perception, Observer Performance and Technology Assessment*. Lake Buena Vista: Florida; 2011. pp. 796603.
133. Sperrin M, Bardwell L, Sergeant JC, Astley S, Buchan I. Correcting for rater bias in scores on a continuous scale, with application to breast density. *Stat Med* 2013;**32**:4666–78. <http://dx.doi.org/10.1002/sim.5848>
134. McCormack VA, Highnam R, Perry N, dos Santos Silva I. Comparison of a new and existing method of mammographic density measurement: intramethod reliability and associations with known risk factors. *Cancer Epidemiol Biomarkers Prev* 2007;**16**:1148–54. <http://dx.doi.org/10.1158/1055-9965.EPI-07-0085>
135. Sukha A, Berks M, Morris J, Boggis C, Wilson M, Barr N, *et al.* Visual assessment of density in digital mammograms. *10th International Workshop on Digital Mammography*, Girona, Catalonia, Spain: Heidelberg Springer; 2010. pp. 414–20.



136. Shepherd JA, Kerlikowske K, Ma L, Duewer F, Fan B, Wang J, *et al.* Volume of mammographic density and risk of breast cancer. *Cancer Epidemiol Biomarkers Prev* 2011;**20**:1473–82. <http://dx.doi.org/10.1158/1055-9965.EPI-10-1150>
137. Duffy SW, Rohan TE, Altman DG. A method for combining matched and unmatched binary data: application to randomized, controlled trials of photocoagulation in the treatment of diabetic retinopathy. *Am J Epidemiol* 1989;**130**:371–8.
138. Barry-Brooks MA, Lourenco AP, Mainiero MB. Breast cancer screening pre and post-tomosynthesis: comparison of recall rate, biopsy positive predictive value and cancer detection rate. Paper presented at the Radiological Society of North America 2013 Scientific Assembly and Annual Meeting, Chicago, November 2013.
139. Conant EF. Clinical implication of digital breast tomosynthesis. *Radiol Clin N Am* 2014;**52**:499–518. <http://dx.doi.org/10.1016/j.rcl.2013.11.013>
140. Friedewald SM, Rafferty EA, Rose SL, Durand MA, Plecha DM, Greenberg JS, *et al.* Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA* 2014;**311**:2499–507. <http://dx.doi.org/10.1001/jama.2014.6095>
141. Gur D, Bandos AI, Rockette HE, Zuley ML, Sumkin JH, Chough DM, *et al.* Localized detection and classification of abnormalities on FFDM and tomosynthesis examinations rated under an FROC paradigm. *AJR Am J Roentgenol* 2011;**196**:737–41. <http://dx.doi.org/10.2214/AJR.10.4760>
142. Kopans DB. Digital breast tomosynthesis from concept to clinical care. *Am J Roentgenol* 2014;**202**:299–308. <http://dx.doi.org/10.2214/AJR.13.11520>
143. Zuley ML, Guo B, Catullo UJ, Chough DM, Kelly AE, Lu AH, *et al.* Comparison of two dimensional synthesised mammograms versus original digital mammograms alone and in combination with tomosynthesis images. *Radiology* 2014;**271**: 664–71. <http://dx.doi.org/10.1148/radiol.13131530>
144. Locatelli M, Tonutti M, Trianni A. *First Experience with the new Generation Low-dose Digital Breast Tomosynthesis: can 2D Synthetic Image Replace Digital Mammography in Combination with Digital Breast Tomosynthesis?* Paper presented at the European Congress of Radiology, Vienna, 2014.
145. American College of Radiology. *ACR statement on Breast Tomosynthesis*. 2014. URL: [www.acr.org/About-Us/Media-Center/Press-Releases/2014-Press-Releases/20140722-ACR-Statement-on-Breast-Tomosynthesis](http://www.acr.org/About-Us/Media-Center/Press-Releases/2014-Press-Releases/20140722-ACR-Statement-on-Breast-Tomosynthesis) (accessed 4 November 2014).
146. Young K. *OPTIMAM 2: Optimisation of Breast Cancer Detection Using Digital X-Ray Technology 2*. National Coordinating Centre for the Physics of Mammography. Cancer Research UK (CRUK) grant no. 17321.
147. Altman DG. Prognostic models: a methodological framework and review of models for breast cancer. *Cancer Investigation* 2009;**27**:235–43. <http://dx.doi.org/10.1080/07357900802572110>
148. Cush S, Burch A, Whelehan P, Young K. *Routine Quality Control Tests for Full Field Digital Mammography Systems*. NHSBSP Equipment Report 0702 Version 1. Sheffield: NHS Cancer Screening Programmes; 2007.
149. Strudley CJ, Young KC, Oduko JM, Looney P, Barnard A, Gilbert FJ. Development of a quality control protocol for digital breast tomosynthesis systems in the TOMMY trial. In Maidment DA, Bakic PR, Gavenonis S, editors. *Breast Imaging: 11th International Workshop on Digital Mammography*. Philadelphia, PA: Heidelberg Springer; 2012. pp. 330–7.
150. Dance DR, Strudley CJ, Young KC, Oduko JM, Whelehan PJ, Mungutroy EHL. Comparison of breast doses for digital tomosynthesis estimated from patient exposures and using PMMA breast phantoms. In Maidment ADA, Bakic PR, Gavenonis S, editors. *Breast Imaging: 11th International Workshop on Digital Mammography*. Philadelphia, PA: Heidelberg Springer; 2012. pp. 316–21.

151. Strudley CJ, Oduko JM, Looney P, Young KC, Gilbert FJ. *Physics Tomosynthesis Quality Control Tests for the TOMMY Trial*. Poster at Symposium Mammographicum 2012.
152. Strudley CJ, Looney P, Young KC, Gilbert FJ. *Technical Performance of the Seven Hologic Dimensions Tomosynthesis Systems used in the TOMMY Trial*. Poster at Symposium Mammographicum 2012.
153. Strudley CJ, Oduko JM, Barnard A, Young KC, Gilbert FJ. *Radiographer Tomosynthesis Quality Control Tests for the TOMMY Trial*. Poster at Symposium Mammographicum 2012.





# Appendix 1 Participant information sheet for women recalled for assessment

## A Comparison of TOMosynthesis with Digital Mammography in the UK Breast Screening Programme (the TOMMY trial)

You are being invited to take part in our research study. It is important for you to understand why the research is being done and what it will involve. Please take time to read this information sheet carefully and decide whether or not you wish to take part. One of our research team will answer any questions you may have about the study.

### *What is the purpose of the study and why have I been chosen?*

Mammography (breast X-ray) is undertaken to detect breast cancer. Tomosynthesis is a new method of obtaining high-quality images that uses digital X-rays and a computer to generate three-dimensional (3D) images of the breast.

In standard two-dimensional (2D) breast X-rays, some abnormalities may be missed at screening because they are hidden by overlapping normal breast tissue. In addition, overlapping normal tissues of the breast may show changes that raise concern. This leads to recall for assessment and further investigation, and can generate unnecessary anxiety and stress. Preliminary studies have suggested that tomosynthesis has advantages over standard 2D mammography by decreasing the problem of overlapping tissues. This enables the structures of the breast to be seen more clearly.

You have been invited to take part in this research because you have been recalled for further assessment following routine breast screening. The purpose of this study is to compare the accuracy of tomosynthesis and standard mammography in the diagnosis of breast abnormalities found at breast screening.

### *Do I have to take part?*

No, it is up to you to decide to participate in the study. If you do decide to take part you will be asked to sign a consent form. You will be given this information sheet and a copy of the consent form to keep. You are free to withdraw at any time, without giving a reason. This would not affect the standard of care you receive.

### *What are the alternatives for diagnosis or treatment?*

The standard procedure for patients who, like yourself, are recalled for further assessments is to have additional mammographic imaging, ultrasound and needle biopsy if needed. You will still undergo this assessment if you decide to participate in this study.

### *What will happen to me if I take part?*

If you agree to take part in the study, you will have a standard 2D mammogram (the same as your breast screening mammogram) and tomosynthesis imaging in addition to any other routine test like ultrasound or biopsy during your visit to the clinic. Taking part in the study will add five to ten minutes of time to your clinic visit.

### *Design of the study*

We aim to conduct the study in six NHS Breast Screening centres in order to recruit a total of 7000 women. This will provide sufficient data to compare the accuracy of the two imaging techniques.

***What will happen to the additional images obtained during the study?***

If tomosynthesis provides any additional information, this will be used to decide your treatment. The tomosynthesis images will be kept on a computer system along with the standard mammogram as part of your medical record. Copies of the tomosynthesis and standard mammography images from each centre will be sent to researchers at a different study centre for independent comparison of the two types of images. Any information about you which is sent to another centre will have your name and address removed and replaced by a unique study number known only to the study research team.

***What do I have to do?***

If you wish to take part in the study we will ask you to sign a consent form when you attend your clinic appointment. The radiographer will explain the tomosynthesis and the standard mammogram examinations. This means you will have a further chance to change your mind if you then find you do not wish to proceed with the tomosynthesis.

***Will my taking part in this study be kept confidential?***

All information which is collected about you during the course of the research will be kept strictly confidential and in accordance with the Data Protection Act 1998. All patient information will be stored securely on password protected computer databases or in locked filing cabinets. Information held by the NHS and records maintained by the NHS Information Centre and the NHS Central Register may be used to help contact you and follow-up your health status. Any information about you which leaves the hospital will be identified by a unique study number.

If you join the study, some parts of your medical records and the data collected for the study may be looked at by representatives of regulatory authorities to check that the study is being carried out correctly. All staff will have a duty of confidentiality to you as a research participant. Nothing that could reveal your identity will be disclosed outside the research site.

We will include information regarding your participation in the study in our letter to your general practitioner following your clinic visit unless you instruct us otherwise.

***What will happen to my data?***

With your permission, images and data from this study will be used by other researchers to improve the technology used to process and read mammograms and help us improve our understanding of breast cancer. This information may be sent to a country outwith the European Economic Area. The information will be kept strictly confidential. Your personal information will be removed and no one will be able to connect you to the data.

***Expenses***

There are no payments or reimbursements for participating in this study as you will not be required to make any additional visits.

***Possible disadvantages and risks of taking part in the study***

Most women experience minor or moderate discomfort during an X-ray mammography examination. The tomosynthesis examination will be taken at the same time as the standard mammogram so there should be no greater discomfort.

Women who are recalled to the clinic often have additional X-ray mammography. The tomosynthesis examination involves an additional dose of radiation to the breast tissue. Participation in the study would be equivalent to having two mammograms in a row. This has been considered by radiation protection advisors and is not considered to be a significant risk to health. (The radiation dose of a mammogram is equivalent to a few months exposure to natural background radiation).

### **Side effects of tomosynthesis**

The additional radiation dose has been discussed above (see *Possible disadvantages and risks of taking part in the study*). No other potential side effects are known.

### **What are the possible benefits of taking part?**

The tomosynthesis examination may provide more accurate judgement on whether or not an abnormality is a cancer and it may be better at detecting very small or subtle cancers.

In the future, fewer women may need to be recalled for further tests due to 'false alarms'.

Women in the study will benefit by having their mammograms looked at by additional specialists at another breast screening centre. There is a small chance (1%) that they could detect an abnormality that was missed at the assessment clinic. If this happens you would be invited back to your screening centre for re-assessment.

We hope that the study will provide useful information for the NHS Breast Screening Programme.

### **What happens if there is a problem?**

If you have a concern about any aspect of taking part in this study, you should contact the trial coordination office (xxxxx xxxxxx) or the Patient Advisory Liaison Service (PALS) on xxx xxxx xxxx.

If you are harmed as a result of your participation in the study, due to someone's negligence, then you may have grounds for legal action but you may have to pay for it.

### **What if relevant new information becomes available?**

Sometimes during a research project, new information becomes available about the diagnostic equipment that is being studied. If this happens, the research team will consider whether any changes should be made to the research method.

### **What happens when the research study stops?**

In line with Good Clinical Practice guidelines, at the end of the study, your data will be securely archived for a minimum of 15 years. Arrangements for confidential destruction will then be made.

### **What will happen to the results of the research project?**

The results of the study will be published in medical journals and will be available at your local breast screening centre.

### **Who is organising and funding the research?**

The TOMMY trial is funded through the Health Technology Assessment (HTA) Programme, part of the Department of Health ([www.hta.nhs.uk](http://www.hta.nhs.uk)). The Department of Radiology, University of Cambridge will undertake the day-to-day running of the trial, under the supervision of Professor Fiona Gilbert. The University of Cambridge will act as Sponsor for the study and will be responsible for the governance of the trial.

This study has been reviewed and given favourable opinion by the Scotland A Research Ethics Committee.

**Thank you, we hope you will agree to take part in this study.**

**If you have any questions or would like any more information please contact xxxxx xxxxx by phone xxx xxxxxxxx or contact the TOMMY trial office (Tel: xxxxx xxxxxxx; email: xxxxxxxxxxxxxxxxxxxxxxxxx).**



## Appendix 2 Participant information sheet for moderate-/high-risk women as a result of family history

### A Comparison of TOMosynthesis with Digital MammographY in the UK Breast Screening Programme (TOMMY trial)

You are being invited to take part in our research study. It is important for you to understand why the research is being done and what it will involve. Please take time to read this information sheet carefully and decide whether or not you wish to take part. One of our research team will answer any questions you may have about the study.

#### *What is the purpose of the study and why have I been chosen?*

Mammography (breast X-ray) is undertaken to detect breast cancer. Tomosynthesis is a new method of obtaining high-quality images that uses digital X-rays and a computer to generate three-dimensional (3D) images of the breast.

In standard two-dimensional (2D) breast X-rays, some abnormalities may be missed because they are hidden by overlapping normal breast tissue. This is a particular problem in younger women because their breast tissue is more dense. In addition, overlapping normal tissues of the breast may show changes that raise concern as they can mimic a cancer. This leads to further investigation, and can generate unnecessary anxiety and stress. Preliminary studies have suggested that tomosynthesis may enable the structures of the breast to be seen more clearly, even in dense breasts.

You have been invited to take part in this research because you are attending annual mammography because of your family history. The purpose of this study is to compare the accuracy of tomosynthesis and standard 2D mammography in the diagnosis of breast abnormalities.

#### *Do I have to take part?*

No, it is up to you to decide to participate in the study. If you do decide to take part you will be asked to sign a consent form. You will be given this information sheet and a copy of the consent form to keep. You are free to withdraw at any time, without giving a reason. This would not affect the standard of care you receive.

#### *What are the alternatives for diagnosis or treatment?*

The standard procedure for patients who, like yourself, are at higher risk of developing breast cancer is to have a standard 2D mammogram. You will still have a standard mammogram if you decide to participate in this study.

#### *What will happen to me if I take part?*

If you agree to take part in the study, you will have tomosynthesis imaging in addition to a standard 2D mammogram. Taking part in the study will add five to ten minutes of time to your visit to the screening centre.

#### *Design of the study*

We aim to conduct the study in six NHS Breast screening centres in the UK. In addition to recruiting women with a family history of breast cancer we will also be inviting women who have had an abnormality detected after routine breast screening to participate in the study. We aim to recruit a total of 7000 women. This should provide sufficient data to compare the accuracy of the two imaging techniques.

***What will happen to the additional images obtained during the study?***

If tomosynthesis provides any additional information, this will be used to decide if you need any further tests. The tomosynthesis images will be kept on a hospital computer system along with your standard mammogram as part of your medical record. Copies of the tomosynthesis and standard mammography images from each centre will be sent to researchers at a different study centre for an independent comparison of the two types of images. Any information about you which leaves the breast centre will have your name and address removed so that you cannot be recognised from it. Participants in the study will be allocated a study number known only to the research team.

***What do I have to do?***

If you wish to take part in the study we will ask you to sign a consent form when you attend your clinic appointment. The radiographer will explain the tomosynthesis and mammogram examinations. This means you will have a further chance to change your mind if you then find you do not wish to proceed with the tomosynthesis.

***Will my taking part in this study be kept confidential?***

All information which is collected about you during the course of the research will be kept strictly confidential and in accordance with the Data Protection Act 1998. All patient information will be stored securely on password protected computer databases or in locked filing cabinets. Information held by the NHS and records maintained by the NHS Information Centre and the NHS Central Register may be used to follow-up your health status. Any information about you which leaves the hospital will be identified by a unique study number.

If you join the study, some parts of your medical records and the data collected for the study may be looked at by representatives of regulatory authorities to check that the study is being carried out correctly. All staff will have a duty of confidentiality to you as a research participant. Nothing that could reveal your identity will be disclosed outside the research site.

We will inform your general practitioner that you are participating in the study unless you instruct us otherwise.

***What will happen to my data?***

With your permission, images and data from this study will be used by other researchers to improve the technology used to process and read mammograms and help us improve our understanding of breast cancer. This information may be sent to a country outwith the European Economic Area. The information will be kept strictly confidential. Your personal information will be removed and no one will be able to connect you to the data.

***Expenses***

There are no payments or reimbursements for participating in this study as you will not be required to make any additional visits.

***Possible disadvantages and risks of taking part in the study***

Most women experience minor or moderate discomfort during an X-ray mammography examination. The tomosynthesis examination will be taken at the same time as the standard mammogram so there should be no greater discomfort.

The tomosynthesis examination involves an additional dose of radiation equal to that of a standard X-ray mammogram. Participation in the study would be equivalent to having two mammograms in a row. This has been considered by radiation protection advisors and is not considered to be a significant risk to health. (The radiation dose of a mammogram is equivalent to a few months exposure to natural background radiation).

### ***Side effects of tomosynthesis***

The additional radiation dose has been discussed above (see *Possible disadvantages and risks of taking part in the study*). No other potential side effects are known.

### ***What are the possible benefits of taking part?***

The tomosynthesis examination may provide more accurate judgement on whether or not an abnormality is a cancer and may be better at detecting very small or subtle cancers.

In the future, fewer women may need to be recalled for further tests due to 'false alarms'.

Women in the study will benefit by having their mammograms and tomosynthesis images looked at by additional specialists at another breast screening centre at a later date. There is a small chance (1%) that they could detect an abnormality that was missed at the assessment clinic. If this happens you would be invited back to your screening centre for re-assessment.

We hope that the study will provide useful information for the NHS Breast Screening Programme.

### ***What happens if there is a problem?***

If you have any concerns about any aspect of taking part in this study, you should contact the trial coordination office (xxxxx xxxxxx) or the Patient Advisory Liaison Service (PALS) on xxxxx xxxxxx.

If you are harmed as a result of your participation in the study, due to someone's negligence, then you may have grounds for legal action but you may have to pay for it.

### ***What if relevant new information becomes available?***

Sometimes during a research project, new information becomes available about the diagnostic equipment that is being studied. If this happens, the research team will consider whether any changes should be made to the research method.

### ***What happens when the research study stops?***

If the research study stops, this will not affect your attendance for annual breast screening.

In line with Good Clinical Practice guidelines, at the end of the study, your data will be securely archived for a minimum of 15 years. Arrangements for confidential destruction will then be made.

### ***What will happen to the results of the research project?***

The results of the study will be published in medical journals and will be available at your local breast centre.

***Who is organising and funding the research?***

The TOMMY trial is funded through the Health Technology Assessment (HTA) Programme, part of the Department of Health ([www.hta.nhs.uk](http://www.hta.nhs.uk)).

The Department of Radiology, University of Cambridge, will undertake the day to day running of the trial, under the supervision of Professor Fiona Gilbert. The University of Cambridge, will act as Sponsor for the study and will be responsible for the governance of the trial.

This study has been reviewed and given favourable opinion by the Scotland A Research Ethics Committee.

**Thank you, we hope you will agree to take part in this study.**

**If you have any questions or would like any more information please contact xxxxxxxx xxxxx by phone xxxx xxxxx or contact the TOMMY trial office (Tel: xxxxx xxxxxx; email: xxxxxxxxxxxxxxxxxxxxxxxxx).**



## Appendix 3 Consent form

### CONSENT FORM

#### A Comparison of TOMosynthesis with Digital MammographY in the UK Breast Screening Programme (TOMMY Trial)

#### Please initial box

1. I have read and understand the Information Sheet (Version 2.0 Dated 7<sup>th</sup> June 2010 ) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason. I understand that withdrawal would not affect my medical care or my legal rights.
3. I give permission for authorised researchers and responsible individuals from regulatory authorities or from the NHS Trust or Board or University of Aberdeen to look at relevant sections of my medical notes if required.
4. I agree to the transfer of my data and images out with the European Economic Area for use in future research.
5. I understand that information held by the NHS, NHS Information Centre and NHS Central Register may be used to provide information about my health status
6. I agree to my GP being informed of my participation in the study.
7. I agree to take part in the above study

Name of participant	Date:	Signature:
Name of Person taking Consent: (if different from researcher)	Date:	Signature:
Researcher:	Date:	Signature:

White: Site File; Blue: Participant; Pink: Breast Screening Records



# Appendix 4 Image management report

## Design of the data transfer process

Early in the design of the trial it was decided to create a CDR to act as a hub for redistribution of the images for the retrospective study and to provide a batch processing facility for the density analysis. The CDR would also provide an archive for the total data set.

The CDR was designed and created at the University of Aberdeen on a HP DL385G7 running Red Hat Enterprise Linux, Red Hat Inc, Raleigh, NC, USA and attached to network storage.

Hologic provided an existing anonymisation system (the Windows-based 'HICS' workstation) which accompanied the commercial imaging equipment, and was installed at each site. Network transmission was chosen to be the method of transfer for the anonymised images from each HICS to a central repository. This satisfied the criteria of security, speed, flexibility, ease of use, reliability, data integrity and robustness. Customised user software was written for the project to invoke the data transfer applications.

### Network protocol

The protocol chosen was rsync-over-SSH (remote synchronisation over secure shell), with SSH providing the security framework and rsync providing the data transfer framework; rsync is designed specifically for mirroring large data stores over a network, and is highly optimised for such use. Both the protocols themselves and, equally important, easily available software implementations, are mature, well proven and highly regarded in the industry.

### Ease of use

Ease of use was essential to managing the transfers at the sites, and this was achieved by arranging that anonymised data sets simply needed to be deposited in a designated 'Outgoing' folder in order to enable fully automatic upload, overnight, at regular prescheduled times. If the central data manager required distributing images back to the sites, this was configured in a similar way on the central server. During each scheduled 'transfer slot', following the upload, the images configured for sending were downloaded automatically to the 'Incoming' folder at each site. All transfers were initiated from the NHS-located sites, thereby avoiding any requirement to establish incoming connections to the New National Network (N3).

### Robustness and data integrity

The rsync protocol intrinsically provides both the ability to resume interrupted transfers and a guarantee of end-to-end data integrity following a successfully completed transfer, owing to its use of rolling checksums. In order to take full advantage of this, the wrapper application was designed to employ a delay-and-retry algorithm if a network interruption is encountered, allowing transfers to resume automatically following brief network downtime, while correctly terminating the transfer if downtime exceeds a configurable threshold. This feature has proved invaluable, since the network path is long and complex, and overall network reliability is not guaranteed, particularly when transferring large numbers of data.

When a data transfer commenced, all data sets found in the 'Outgoing' folder were immediately moved to a dated folder in the 'Sending' folder, and were not moved to the 'Sent' folder until all transfers had completed successfully. This meant that if transfers were interrupted for any reason, when the application was next run, the previous set of transfers ran to completion before any new transfers were started and also ran to completion. These combined features provided a high degree of robustness and flexibility.

### **Monitoring and accounting**

Since data transfers necessarily took place automatically, overnight (to avoid network congestion during working hours since large numbers of data are being transferred), it was essential to monitor transfers and maintain accounting records. The wrapper transfer application therefore generated log files for every transfer, and once the main transmitting and receiving transfers were completed, the log files also were transferred to the central server, so that the logs might be examined both by the local staff at the sites, and also by the central data manager. All transfers and logs were dated for ease of scrutiny.

### **Automated access credentials**

In order to make automated access possible without human intervention (for overnight transfers), automated management of credentials was required. The ability to do this is a standard feature of SSH, and is based on public key cryptography. A unique key was installed at each site during installation of the transfer application, thereby allowing each site to identify and authenticate itself when opening a connection to start data transfer. On the CDR, SSH keys were stored securely on the local disk (so access could not be interrupted by network problems), while the data folders were mounted from the central storage resource. The SSH connections for the site accounts were 'locked down' so that connections using any protocol other than rsync were refused by the CDR.

### **Main data store file system hierarchy**

One issue which caused a certain amount of confusion within the project was the distinction between the naming of the transmitted data set files and the internal R2ID numbers. The Hologic HICS anonymiser replaces both the image accession numbers and the patient names and IDs with numeric IDs which simply increment for successive anonymisations. However, the data set folder names (and individual filenames) are based on the R2NUM (the anonymised accession number tag), not the R2ID (anonymised patient ID tag). Since an accession is always unique, but a single patient may be scanned on a return visit, the R2NUM and R2ID do not necessarily match (although for initial data sets they are likely to be the same). When each image in a single data set is anonymised, the accession number is replaced by the generated R2NUM, and both the Patient Name and the Patient ID are replaced by the generated R2ID. For this reason, it was decided to add an additional level when creating the central data storage file system hierarchy (since the internal R2ID is otherwise hidden). The hierarchy was SITENUM/R2ID/R2NUM.

## **Operation of the data transfer process**

### **Relocation**

Following the relocation of the chief investigator to Cambridge in 2011, it was agreed that image transfer and storage of the images should be based in Cambridge.

In June 2012, Aberdeen distributed a new client software configuration that pointed the data collection sites image transfers to Cambridge.

### **Storage and management of data**

The CDR was reimplemented to the original design from the University of Aberdeen on the University of Cambridge Clinical School of Computing Services virtual infrastructure [Ubuntu 12.04 LTS (Canonical, London, UK) on VmWare (Palo Alto, CA, USA)] and network storage [(Network File System, Oracle America, Inc., Santa Clara, California, USA) on NetApp filer with snap-mirroring on a 7-daily, 3-weekly rotation]. All trial image data (1.5 TB) acquired prior to the relocation were copied to the new site on Linux Unified Key Setup encrypted external disk and verified against Message Digest 5 checksums.

### **Local transfer**

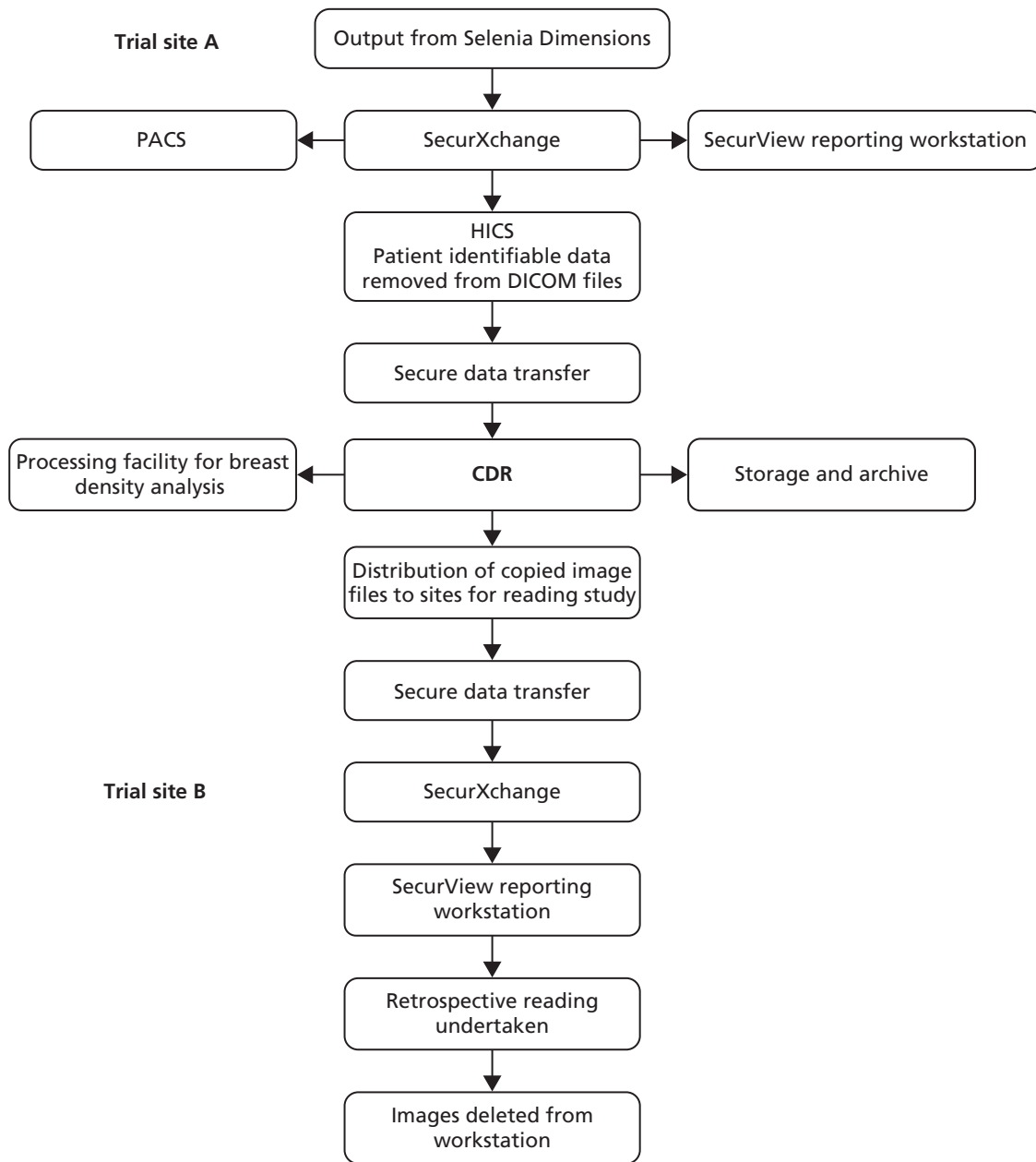
Initially images were transferred for quality assurance (QA) to the Cambridge Hologic SecurView workstation in batches using the software store from the DCMTK DICOM toolset. Later, DICOM server software (DCMTK dcmqrscp) was used to provide a 'PACS query' connection between the Cambridge Hologic workstation and the CDR system for ad-hoc querying. The local network speed was 100 Mb/second and was found to be inadequate for interactive case retrieval of the large data sets.

### **Issues involved in network connections**

Transfers between Glasgow and Cambridge were hindered by the slow connection speed, around 87 kB/second up (Glasgow–Cambridge) and 750 kB/second down (Cambridge–Glasgow), about 20 times and two times slower respectively than the other sites, which regularly achieved 1.5 MB–5 MB/second. Most sites completed their transfers in a few hours in the middle of the night, but Glasgow had to start immediately after office hours and spread the transfers out over the week. If for any reason a Glasgow return transfer was missed or incomplete, it was difficult to catch back up again in time for the next.



## Appendix 5 Flow diagram indicating management of image files



**FIGURE 29** Flow diagram indicating the management of image files.





## Appendix 6 Quality control report on imaging equipment

The NCCPM, which provides scientific and technical advice to the NHSBSP, was tasked with assuring the performance of the imaging equipment involved in the trial. The purpose of the testing carried out was to ensure that the dose and image quality on the seven systems employed for the trial complied with existing standards<sup>84,148</sup> for 2D mammography, to ensure that the imaging performances of all systems in both 2D and tomosynthesis modes were well matched and to monitor the performance of the equipment during the course of the trial. As well as physical testing of the imaging equipment, a survey was carried out of clinical radiation doses delivered by each of the seven systems used in the trial. The results of the QC testing showed that all imaging equipment met the relevant standards, the imaging performances of the systems used in the trial were well matched and that they remained so during the course of the trial.

The knowledge and experience gained in carrying out the tomosynthesis QC testing for the trial has been published<sup>149–153</sup> and utilised in the preparation of UK and European DBT QC protocols.

### Methods

The imaging equipment used for the trial consisted of seven Hologic Dimensions Tomosynthesis mammography X-ray systems and the associated Hologic SecurView image display workstations, of which there were eight at the start of the trial and a further two were added during the course of the trial.

For 2D imaging performance, the existing UK NHSBSP QC protocols<sup>84,148</sup> were applied. In the absence of any established protocols for tomosynthesis QC, it was necessary to devise suitable tests and phantoms for the purpose, some of which were an extension of the 2D tests, and others novel tests to check aspects of the quality of the reconstructed tomosynthesis images. The NCCPM provided test phantoms which were specially manufactured for use by radiographers and physicists at each centre to ensure the comparability of QC results between centres. The NCCPM prepared and provided QC protocols (see *Appendices 15* and *16*) to be used by radiographers and physicists at all centres during the trial. These protocols referred to existing NHSBSP test protocols for 2D testing and included additional tests of tomosynthesis performance.

The NCCPM carried out rigorous initial testing at all six centres during the period July to December 2011, before each of the seven X-ray systems commenced use in the trial. At this stage, efforts were made, with the assistance of the installing X-ray engineers, to minimise differences between systems to ensure that radiation dose and image quality was closely matched. Subsequent 6-monthly testing was carried out by physicists contracted by NCCPM, who, in addition to providing their usual reports on the standard 2D tests, returned the required 2D and tomosynthesis data and images to NCCPM for further analysis and reporting. Radiographers at all centres carried out routine QC on a daily, weekly and monthly basis, and returned data and images, on a weekly basis, to NCCPM for collation and analysis.

The QC testing carried out included tests of the image display equipment used in the retrospective reading phase of the trial, as well as the X-ray equipment used to acquire the patient images.

In accordance with NHSBSP 0604, 84 clinical breast doses were calculated for a sample of 50 or more patients imaged on each system.

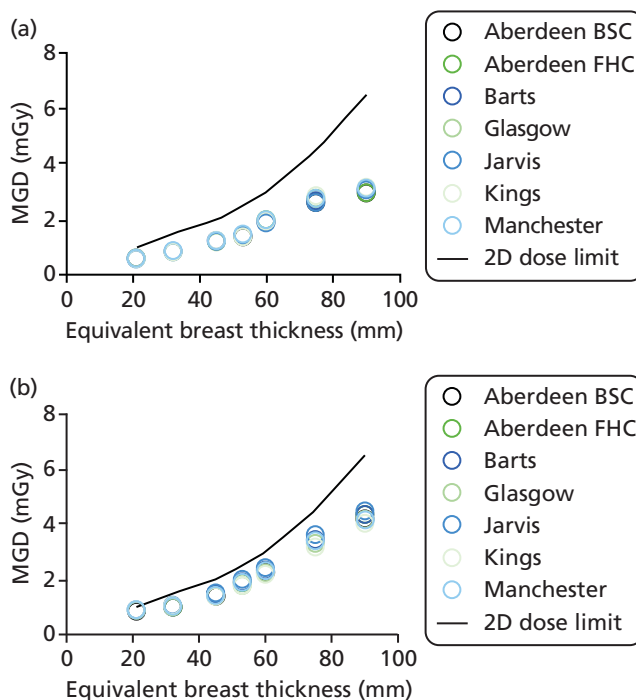
Data from key measurements comparing the performance of systems at the start of the trial, and subsequent 6-monthly testing, are presented, as well as an overview of the routine QC carried out by radiographers and the results of the patient dose survey.

## Results

### Dose

The MGD to the standard breast model, simulated with polymethyl methacrylate (PMMA) phantoms, was assessed for conventional 2D imaging and tomosynthesis for a range of equivalent breast thicknesses. The initial results for the seven systems are shown in *Figure 30* and *Table 26*. The NHSBSP dose limits for conventional digital mammography are also shown in *Figure 30*.

At 6-month intervals the MGD measurements were repeated. *Table 27* summarises the overall averages and ranges of values for all systems throughout the trial.



**FIGURE 30** Mean glandular dose to the standard breast, for (a) conventional exposures and (b) tomosynthesis exposures. BSC, Breast Screening Centre; FHC, Family History Clinic.

**TABLE 26** Mean glandular dose to the standard breast and contrast-to-noise ratio averaged over the seven systems and coefficient of variation, for conventional and tomosynthesis images

Equivalent breast thickness (mm)	MGD for 2D images		MGD for tomosynthesis		CNR for 2D images		CNR for tomosynthesis images	
	Mean (mGy)	CoV (%)	Mean (mGy)	CoV (%)	Mean	CoV (%)	Mean	CoV (%)
21	0.60	3.4	0.89	3.1	10.9	2.9	29.1	2.5
32	0.84	2.4	1.04	3.5	9.9	2.3	22.2	2.8
45	1.17	2.4	1.43	3.9	9.0	3.5	18.9	2.8
53	1.41	3.2	1.87	4.2	8.4	3.4	18.5	3.5
60	1.98	1.6	2.29	3.8	8.6	3.3	17.2	3.7
75	2.69	2.1	3.39	4.3	8.3	3.0	14.8	4.7
90	3.03	2.5	4.25	3.6	6.7	4.7	11.3	4.3

CNR, contrast-to-noise ratio; CoV, coefficient of variation.

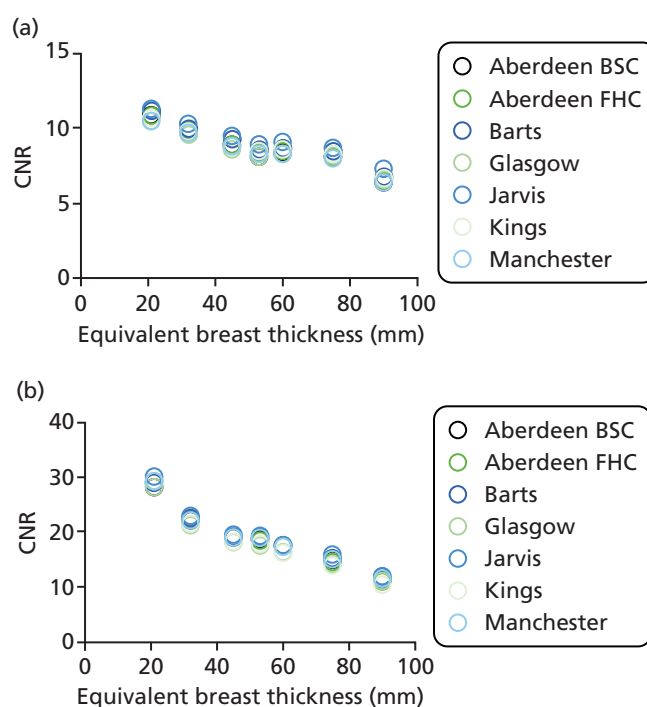
**TABLE 27** Mean glandular dose measurements for all seven systems during the trial

Equivalent breast thickness (mm)	Average 2D MGD (mGy)	Minimum 2D MGD (mGy)	Maximum 2D MGD (mGy)	Average tomosynthesis MGD (mGy)	Minimum tomosynthesis MGD (mGy)	Maximum tomosynthesis MGD (mGy)
21	0.62	0.58	0.69	0.88	0.82	0.94
32	0.86	0.81	0.94	1.02	0.94	1.08
45	1.18	1.11	1.24	1.42	1.34	1.52
53	1.43	1.36	1.56	1.85	1.73	2.00
60	1.98	1.86	2.15	2.27	2.11	2.42
75	2.66	2.50	2.86	3.36	3.11	3.36
90	2.95	2.76	3.16	4.23	3.91	4.48

The data in *Table 27* show that the maximum deviation from the mean for 2D or tomosynthesis dose measurements is 11%. To put these deviations into context, the remedial level for dose measurements (NHSBSP 0604<sup>84</sup>) is a change of 25% relative to baseline dose levels.

### Contrast-to-noise ratio

Average contrast-to-noise ratio (CNR) measurements in conventional and tomosynthesis modes for the seven systems involved in the trial are shown in *Figure 31* and *Table 26*. Average CNR measurements in 2D and tomosynthesis modes for the seven systems throughout the trial are shown in *Table 28*.



**FIGURE 31** Contrast-to-noise ratio for (a) conventional exposures and (b) DBT exposures. BSC, Breast Screening Centre; FHC, Family History Clinic

The data in *Table 28* show a maximum deviation from the mean of 13% for 2D and 18% for tomosynthesis, compared with a maximum deviation from the mean for the seven systems of 8% at the initial testing. This is an acceptable spread in CNR measurements, given that the remedial level for a single system is a 10% change from baseline measurements.

## Tomosynthesis geometric distortion and interplane resolution

### Tomosynthesis geometric distortion

For all systems, the height of best focus assessed at multiple positions within the image deviated by no more than 1 mm from that assessed at the centre of the chest wall edge. No distortion was observed in the x and y directions. There was a 4% scaling error for all systems, but this had no impact, as measurements of distance did not feature in the trial. There was no change in tomosynthesis geometrical distortion measurements during the course of the trial.

### Tomosynthesis z-resolution

The z-resolution for 1-mm aluminium balls was found to range between 9.8 mm and 12.0 mm, with some dependence on position within the image. The mean value was 10.8 mm. No significant variation between the systems or over time during the course of the trial was seen.

### Detector modulation transfer function

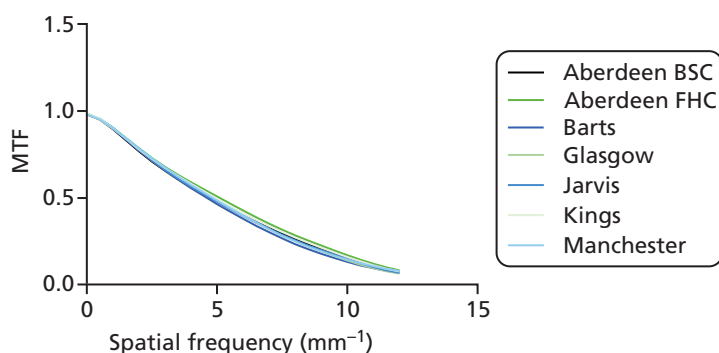
Measurements of modulation transfer function (MTF) were made at the start of the trial. Only small differences were seen between the seven systems. The MTF measurements are shown in *Figure 32* and *Table 29*.

During the course of the trial, detector resolution for the seven systems was shown to have remained unchanged by measurement of the threshold contrast detail detection, which included assessment of small details.

**TABLE 28** Contrast-to-noise ratio measurements for all seven systems during the trial

Equivalent breast thickness (mm)	Average 2D CNR	Minimum 2D CNR	Maximum 2D CNR	Average tomosynthesis CNR	Minimum tomosynthesis CNR	Maximum tomosynthesis CNR
21	10.9	10.4	11.7	29.1	27.4	30.4
32	9.8	9.4	10.6	22.1	20.7	23.2
45	8.9	8.3	9.5	18.9	18.1	20.2
53	8.4	7.8	9.0	18.7	17.6	21.2 <sup>a</sup>
60	8.4	8.0	9.1	17.1	16.0	20.3 <sup>a</sup>
75	8.0	7.5	8.7	14.8	13.9	17.0 <sup>a</sup>
90	6.4	6.0	7.3	10.9	10.4	12.0

<sup>a</sup> The higher maximum figures for tomosynthesis CNR are a result of some relatively high measurements from one centre (Jarvis) during the second half of the trial, where installation of a software upgrade altered the apparent noise in the reconstructed tomosynthesis QC images. Hologic provided assurance that this fault affected QC images only.



**FIGURE 32** Modulation transfer function measurements for the seven systems. BSC, Breast Screening Centre; FHC, Family History Clinic.

**TABLE 29** Mean modulation transfer function measurements for the seven systems

Spatial frequency (mm <sup>-1</sup> )	Mean MTF	Coefficient of variation of MTF (%)
2	0.779	0.7
4	0.573	1.7
6	0.402	2.6
8	0.308	3.3
10	0.259	4.8
12	0.074	9.7

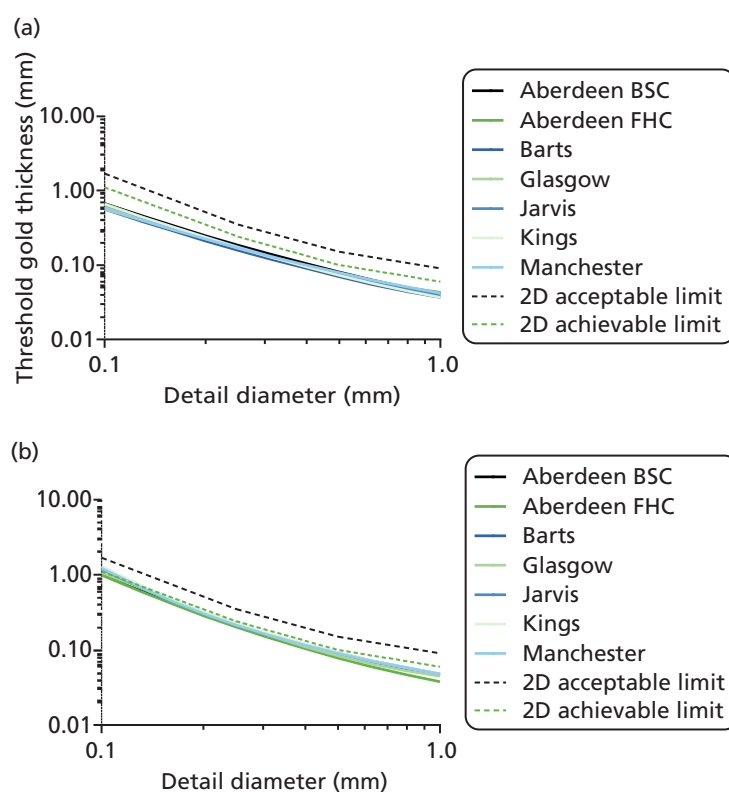
## Threshold contrast detail detectability

Figure 33 and Table 30 show the threshold gold thickness results for a range of detail diameters for conventional and tomosynthesis images of a CDMAM test object (version 3.4, UMC St. Radboud, Nijmegen University, the Netherlands). The results from initial testing on all systems are shown and compared against the minimum acceptable and achievable values for conventional mammography as defined in the NHSBSP protocol.<sup>84</sup>

Repeat measurements of image quality using a CDMAM test object on all systems every 6 months demonstrated that there has been no significant change in image quality in either 2D or tomosynthesis modes and that all systems continue to exceed the achievable image quality standard for 2D mammography.

## Image display equipment

Testing of the performance of image display equipment was carried out at all centres by physicists and radiographers as prescribed by NHSBSP protocols.<sup>84,148</sup> Two of the workstations tested did fail the DICOM greyscale standard as detailed in NHSBSP 0604,<sup>84</sup> but these systems were recalibrated and reassessed to ensure that they met the standard before commencement of the retrospective reading phase of the trial.



**FIGURE 33** Contrast detail curves from each of the systems for conventional (a) and tomosynthesis (b) images. BSC, Breast Screening Centre; FHC, Family History Clinic.

**TABLE 30** Average threshold gold thickness and coefficient of variation for 2D and tomosynthesis CDMAM images

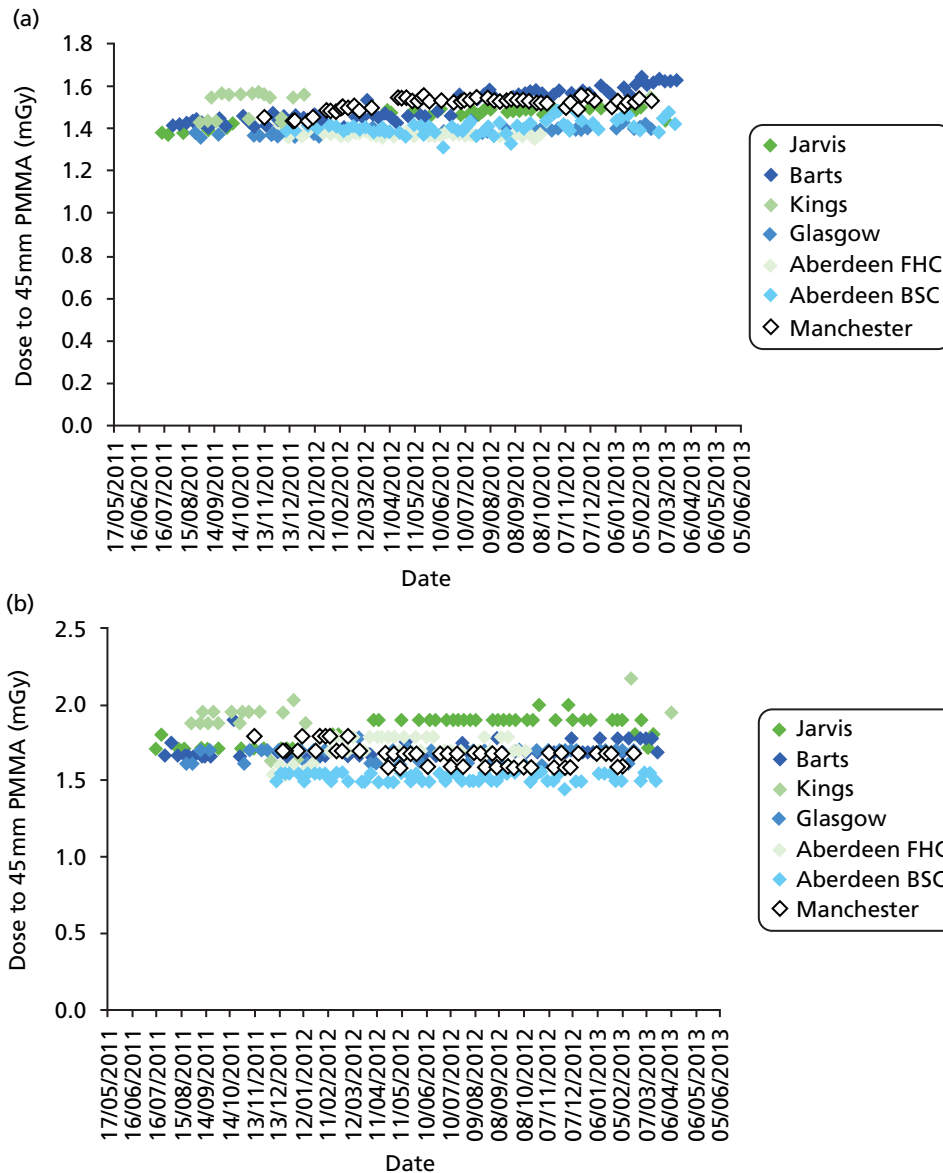
Detail size (mm)	Threshold gold thickness ( $\mu\text{m}$ )	CoV (%)	Thickness gold ( $\mu\text{m}$ )	CoV (%)	Minimum acceptable	Achievable
0.10	0.655	8.3	1.277	6.7	1.680	1.100
0.25	0.175	4.0	0.218	2.3	0.352	0.244
0.50	0.075	6.8	0.095	2.3	0.150	0.103
1.00	0.040	12.3	0.043	8.9	0.091	

CoV, coefficient of variation.

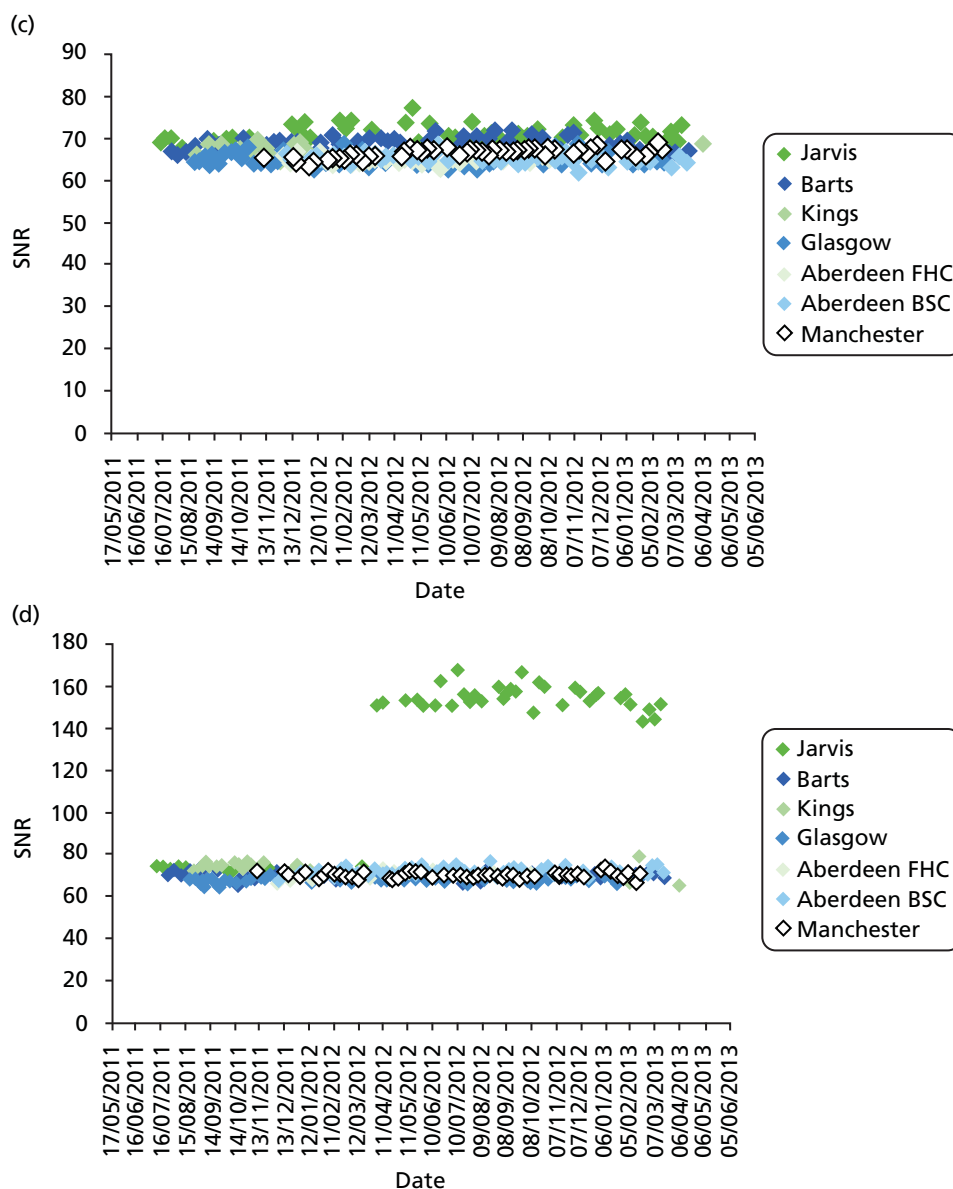
### Quality control by radiographers

Daily, weekly and monthly tests of equipment performance were carried out by radiographers on the seven systems used in the trial, in accordance with the trial protocol for radiographer QC (see *Appendix 17*). All QC results were recorded electronically using a spreadsheet or database, and copies were sent to NCCPM for checking. By carrying out daily QC, radiographers were able to check that X-ray systems were functioning properly prior to use on patients and that no significant changes in dose or image quality had occurred. As well as checking that no significant abnormal artefacts were evident in images of plain PMMA, radiographers also examined reconstructed tomosynthesis images of a 1-mm-diameter aluminium ball, to ensure that there had been no change in its appearance, which might

suggest a failure in the tomosynthesis image reconstruction. Weekly QC images sent to NCCPM for detailed analysis showed that dose and signal-to-noise ratio (SNR) in both conventional and tomosynthesis images have remained stable and no clinically significant artefacts have been seen. The variation in dose and SNR measurements for each system over time are shown in *Figure 34*. [The tomosynthesis SNR results for one centre (Jarvis) doubled half way through the trial owing to a software upgrade necessitated by an equipment breakdown. This led to a change in the calibration of the pixel values and hence SNR, which has since remained stable. Subsequent checks have confirmed the manufacturer's assertion that the change is seen in QC images only, and that clinical image quality will not have been affected. This change also affected the physicist's measurements of CNR as described in pages 99–100.]



**FIGURE 34** Variation in 2D mammography and tomosynthesis dose and SNR from analysis of weekly QC images of 45-mm PMMA. (a) Weekly 2D mammography dose, (b) weekly Tomo dose, (c) weekly 2D mammography SNR and (d) weekly Tomo SNR. BSC, Breast Screening Centre; FHC, Family History Clinic. (*continued*)



**FIGURE 34** Variation in 2D mammography and tomosynthesis dose and SNR from analysis of weekly QC images of 45-mm PMMA. (a) Weekly 2D mammography dose, (b) weekly Tomo dose, (c) weekly 2D mammography SNR and (d) weekly Tomo SNR. BSC, Breast Screening Centre; FHC, Family History Clinic.

Radiographers were also responsible for regular checks on the SecurView image display workstations and reported to NCCPM that these checks were continued through to the end of the retrospective reading phase of the trial.

### Patient dose survey

Exposure data from each centre were analysed in order to calculate the estimated patient doses. A summary of the data and calculated patient doses is presented. These results demonstrate that the seven systems employed for the trial were set up to give doses that varied by less than 10% from the mean in 2D and tomosynthesis modes. Although measurements of MGD to the standard breast model are one-third higher for tomosynthesis than for 2D, estimates of clinical doses show that tomosynthesis doses are on average only about 13% higher than 2D doses.



A summary of the patient data gathered at each centre is shown in *Table 31*.

Average MGD to 50- to 60-mm breasts for each unit are shown in *Table 32*, together with the MGD to a 53-mm standard breast, simulated with 45-mm PMMA.

The patient dose measurements from all sites are shown in *Figure 35*. Also shown are the average measurements from all sites of MGD to the standard breast model, as listed in *Table 29*.

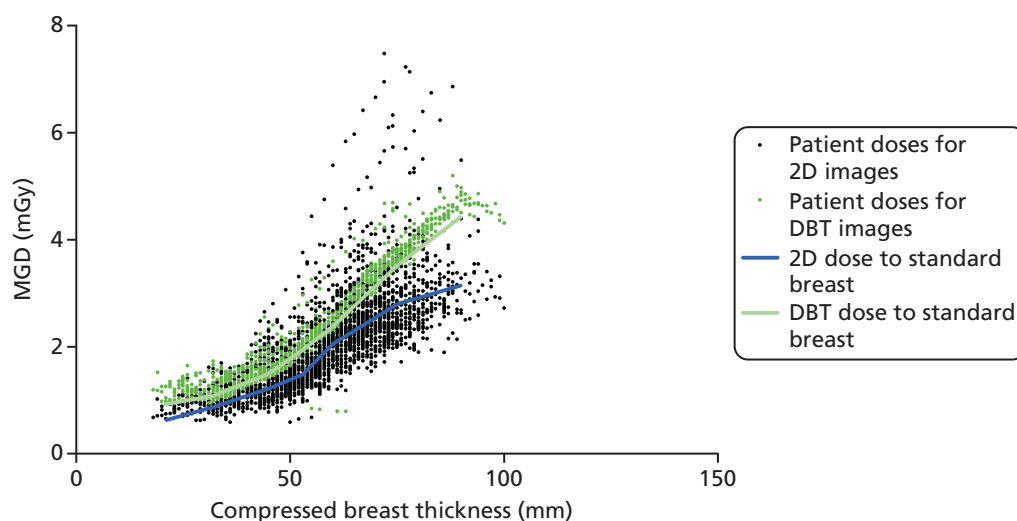
A summary giving the range of clinical patient doses (minimum, maximum and mean MGD measurements) for 2D and tomosynthesis views from patient data for all breast thicknesses at all sites is shown in *Table 33*.

**TABLE 31** Summary of patient dose data gathered from each centre

Centre	Number of patients for whom dose data was gathered	Period over which patient dose data was collected		Mean age of patient (years)	Mean compression force (N)
		From	To		
Aberdeen (FH)	56	15 December 2011	19 March 2012	45	94
Aberdeen (screening)	106	19 December 2011	21 March 2012	57	109
Barts	74	12 July 2011	20 January 2012	55	80
Glasgow	99	30 August 2011	06 October 2011	57	119
Guildford	89	11 July 2011	25 August 2011	57	102
King's College	105	6 September 2011	9 February 2012	57	91
Manchester	178	15 March 2012	30 April 2012	55	119

**TABLE 32** Mean glandular dose to 53-mm standard breast and average MGD for oblique views of 50- to 60-mm breasts

Centre	2D		Tomosynthesis	
	MGD to 53-mm standard breast (mGy)	Average MGD to 50- to 60-mm breasts (mGy)	MGD to 53-mm standard breast (mGy)	Average MGD to 50- to 60-mm breasts (mGy)
Aberdeen (FH)	1.41	2.02	1.88	2.13
Aberdeen (screening)	1.36	1.87	1.80	2.04
Barts	1.39	1.62	1.92	2.11
Glasgow	1.41	2.16	1.79	2.10
Guildford	1.38	1.93	2.00	2.23
King's College	1.47	2.15	1.80	2.02
Manchester	1.48	1.60	1.89	2.14
Average for all sites	1.41	1.87	1.87	2.11



**FIGURE 35** Patient dose data from all sites (all views). Also shown are lines representing the average MGD to the standard breast measured using PMMA for 2D and tomosynthesis exposures.

**TABLE 33** Summary of patient dose measurements at all sites, all views, all breast thicknesses

Mammographic views	Minimum MGD (mGy)	Maximum MGD (mGy)	Mean MGD (mGy)
All 2D views	0.59	7.48	2.12
All tomosynthesis views	0.79	5.20	2.57

## Discussion

Performance testing of the imaging equipment employed for the trial ensured that the seven X-ray systems were set up to give exposures that were closely matched, thus ensuring that there was no significant variation in image quality between the systems. Thorough testing by physicists on behalf of NCCPM at 6-month intervals ensured that there were no significant variations in imaging performance during the course of the trial. More frequent tests carried out by radiographers at daily, weekly and monthly intervals ensured that all systems were closely monitored during the trial so that any problems that could have arisen were dealt with promptly. Tests carried out on image display equipment used to view clinical images were tested regularly in accordance with NHSBSP protocols. The QC testing demonstrated that all of the imaging equipment met, or exceeded, existing UK standards for 2D mammography, and that the 2D and tomosynthesis imaging performance of all systems were well matched.

The survey of patient radiation doses from each centre confirmed the findings of the physical QC measurements, indicating that radiation exposures employed by the seven systems were well matched.

## Conclusions

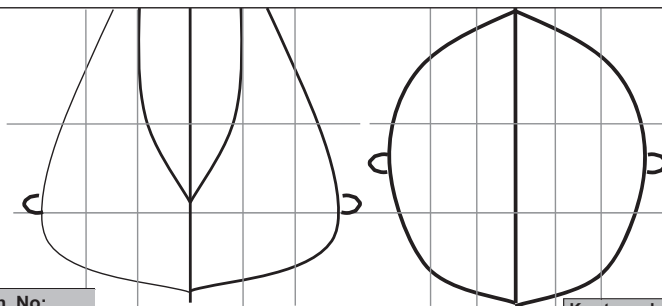
The QC program employed for the trial ensured that the imaging performance of all equipment employed for the trial was well matched and met current applicable UK standards for 2D digital mammography, both at the start and during the course of the trial. The experience gained has contributed to the development of QC protocols for tomosynthesis in the UK and Europe.

# Appendix 7 Prospective data collection form: assessment

Affix Patient  Label Name:  Date of Birth Sx /CHI Number	Case recalled by: <i>Please tick</i>	Screen films: <i>Please circle</i>
	First Reader	Analogue (1)   Digital
	Second Reader	(2) Assessment
	Reader initials:	Trial ID:

Date:

Mark ALL lesions on diagram with a number (1,2,3) Primary lesion, that initiated recall, is 1



Please complete separate form for EACH lesion and staple the forms together

Are you completing this form with knowledge of pathology from biopsy?  
YES (1) NO (2)

Lesion No:	Sign 0 - 5	Suspicion 1 - 5	Conspicuity 0-3	
Screen Films MLO				
CC				
Assess. Clinic 2D MLO				
2D CC				Slice No ↓
DBT MLO				
DBT CC				

## Key to code numbers

Sign	0	None
	1	Circumscribed Mass
	2	Spiculated Mass
	3	Micro-calcification
	4	Distortion
	5	ASD
Suspicion	1	Normal
	2	Benign
	3	Probably Benign
	4	Suspicious
	5	Malignant
Conspicuity	0	Not visible
	1	Barely visible
	2	Visible, not well seen
	3	Clearly visible

DBT vs 2D Please tick	
-2	DBT much poorer than 2D
-1	DBT slightly poorer than 2D
0	DBT equal to 2D
+1	DBT slightly better than 2D
+2	DBT much better than 2D

## Additional information from DBT Tick all that apply

1	Lesion visibility
2	Lesion extent
3	Lesion characteristics
4	Discrimination benign vs malignant
5	Other (please write comment)

DENSITY: please mark visual estimate of breast density on the line

0 \_\_\_\_\_ 100%

Additional comments:

Biopsy (1)/ No biopsy(0) (please circle) Biopsy score (if known)

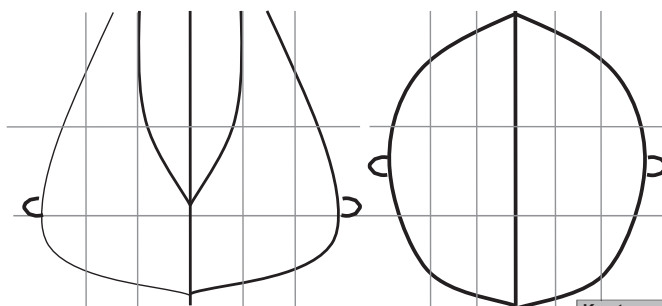


## Appendix 8 Prospective data collection form: family history

Affix Patient Label			
Name			
Screening /CHI Number			
	Reader 1		
	Reader 2		
	Arbitration		
	Reader initials:	Trial ID:	Screening Date:

Prior films:  Please circle  
Analogue (1) Digital (2) None (3)  
Prior films:  Please circle

Mark ALL lesions on diagram with a number (1,2,3). Primary lesion, that initiates recall, is 1



Please complete a separate form for EACH lesion and staple together

Lesion No:	Sign 0 - 5	Suspicion 1 - 5	Conspicuity 0-3	
2D MLO				
2D CC				Slice No ↓
DBT MLO				
DBT CC				

Key to code numbers		
Sign	0	None
	1	Circumscribed Mass
	2	Spiculated Mass
	3	Micro-calcification
	4	Distortion
	5	ASD
Suspicion	1	Normal
	2	Benign
	3	Probably Benign
	4	Suspicious
	5	Malignant
Conspicuity	0	Not visible
	1	Barely visible
	2	Visible, not well seen
	3	Clearly visible

DBT vs 2D Please tick	
<input type="checkbox"/>	-2 DBT much poorer than 2D
<input type="checkbox"/>	-1 DBT slightly poorer than 2D
<input type="checkbox"/>	0 DBT equal to 2D
<input type="checkbox"/>	+1 DBT slightly better than 2D
<input type="checkbox"/>	+2 DBT much better than 2D

Additional comments:

Additional information from DBT (tick all that apply)	
<input type="checkbox"/>	1 Lesion visibility
<input type="checkbox"/>	2 Lesion extent
<input type="checkbox"/>	3 Lesion characterisation
<input type="checkbox"/>	4 Discrimination benign vs malignant
<input type="checkbox"/>	5 Other (please write comment)

Density Please mark visual estimate of breast density on the line  
0 \_\_\_\_\_ 100%

Please circle:

RECALL (1) / NO RECALL (2)



## Appendix 9 Prospective data collection form: pathology

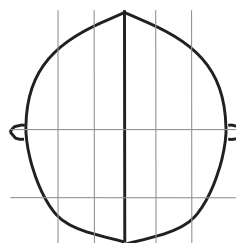
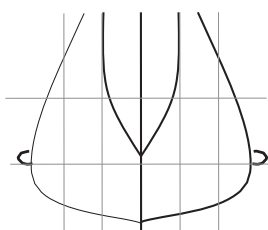
<b>Trial ID:</b>
<b>D.O.B:</b>

<b>Assessment/FH case (please circle)</b>
---

Lesion number (from prospective data collection proforma):
--

If there is more than one lesion, please complete a separate form for EACH malignant lesion and staple the forms together
---

Lesion location



Assessment Biopsy Score (please tick)	
1	Normal
2	Benign
3	Benign equivocal
4	Suspicious of malignancy
5a	Malignant non invasive /in situ
5b	Malignant invasive
0	Not performed

Tumour Type (please tick)			
1	Invasive Ductal	3e	Metastases
2	Invasive Lobular	3f	Lymphoma
3a	Medullary	3g	Other
3b	Mucinous	4	Lobular In Situ
3c	Tubular	5	Ductal In Situ
3d	Intracystic papillary		

Tumour Grade (please tick)	
1	I
2	II
3	III
4	Unknown

Lymph Node Status (please tick)	
1	Normal
2	< 4 nodes positive
3	>4 nodes positive
4	Not known

Hormone/HER-2 status	Quick Score (0-8)	Score	Code*
ER status	Quick Score (0-8)		
PgR status	Quick Score (0-8)		
C-erb-B2/HER-2 Status	Grade (0-+3)		

*Coding for Hormones and HER status	
1	Positive
2	Negative
3	Unknown
0	Not assessed

Tumour Size (mm)	
DCIS only size (no invasive component)	
Invasive size	
Whole tumour size	
Baseline image measurement if neo-adjuvant chemotherapy	

Comments:

*Please complete form from final histology*





## Appendix 10 Randomisation process for retrospective reading study

### RANDOMISATION PROCESS FOR RETROSPECTIVE READING STUDY

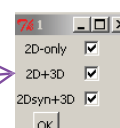
The cases from the TOMMY trial will be randomised using a programme called R.



Click on the main icon (  ) to start the tool.



Select which types of image(s) you wish to allocate to the available readers by unticking the boxes.



Choose the input file. This is a spreadsheet with 3 columns, R2NUM, R2ID, Site Name. This details the patients whose images are to be allocated to a reader.

Next choose the readers file. This is a spreadsheet with 4 columns: SiteName, SiteCode, Reader, Capacity. The capacity is the number of images each reader can read per week.

(You will have the opportunity later to say how many weeks this allocation is to be read over.)

**The format of these files must be strictly adhered to.**



Now select how many weeks this allocation will be read over. The total number of images that can be read in theory is the capacity times the number of weeks, added up over all the sites and all the readers.



Choose the subdirectory folder called "Log" that contains all the output files that were previously generated.



The tool will create an output text file that gives the allocations. If this file is used, then a copy needs to be stored in the subdirectory of this folder called "Log". The tool will use any files in this directory to check if the images for each case have previously been read and apply the randomisation rules as appropriate. **No other files may be stored in this log.** A copy of this output file should then be emailed to Richard Black for the distribution of the correct images to the correct site. The data managers at each site will then allocate these cases to the specified reader.



Copies of the input/output files created by the tool will be stored inside Z:\Cambridge Clinical Trials Unit\11. STUDY MANAGEMENT FILES\CCTU0091 - TOMMY\Randomisation and are archived in the relevant folders each time a new one is used. Such copies should be named and version controlled in the correct manner to enable the creation of an audit trail. The reader.csv file is expected to change less often than the Batch.csv file, which will update on a weekly basis/each time a new randomisation is required.



Details of the allocation of each image will be captured in MACRO for each subject and an excel sheet completed to log how many cases have been randomised.



## Appendix 11 Retrospective study data collection form: two-dimensional

<b>R2 ID:</b>
<b>Reading site ID:</b>
<b>Reader initials/ID:</b>

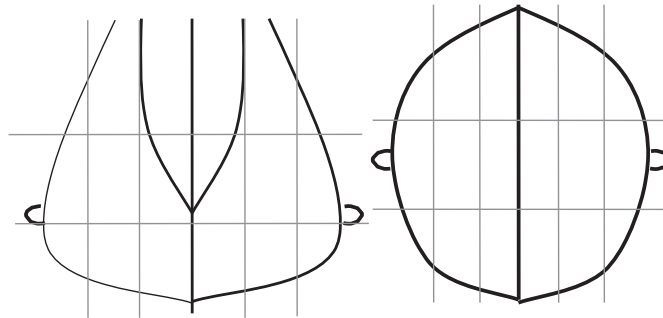
Reading Protocol for this case:  <b>2D only</b>
---

<b>Opinion (please circle)</b>
<b>RECALL</b>
<b>NO RECALL</b>

Please mark most suspicious lesion with X (i.e. lesion responsible for recall)

Subsequent suspicious lesions for recall should be recorded on 'Additional Lesion' form and numbered as 2, 3 etc. Additional Lesion form completed? (Please circle): YES / NO

Reading Protocol for this case:



Please score most suspicious lesion by ticking appropriate box:

Suspicion	1	Normal	
	2	Benign	
	3	Probably benign	
	4	Suspicious	
	5	Malignant	



## Appendix 12 Retrospective study data collection form: two-dimensional and three-dimensional

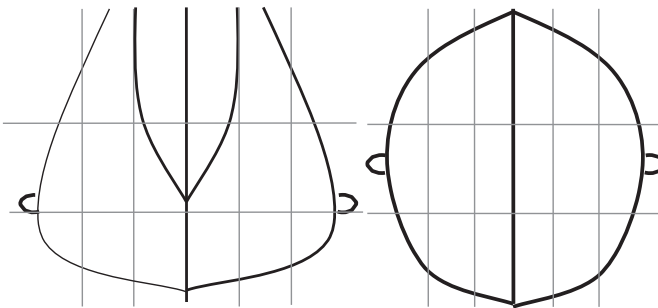
R2 ID:  Reading site ID:	Reading Protocol for this case (please circle)
--------------------------------	--

2D+3D / 2Dsynth+3D

Opinion (please circle)

2D alone: RECALL / NO RECALL    3D alone: RECALL / NO RECALL    Combined: RECALL / NO RECALL

Please mark **most suspicious lesion** with X (i.e. lesion responsible for recall IN COMBINED MODE ONLY) and fill in DBT slices below.



Additional lesion form completed?	YES	NO
If YES how many additional forms completed?		

DBT MLO Slice no.	
DBT CC Slice no.	

Please score items 1 to 5 by ticking the appropriate box:

1.Suspicion	1	Normal	
	2	Benign	
	3	Probably Benign	
	4	Suspicious	
	5	Malignant	
2.Lesion visibility	-2	Only seen on 2D	
	-1	DBT poorer than 2D	
	0	DBT equal to 2D	
	+1	DBT better than 2D	
	+2	Only seen on DBT	
3.Lesion extent	-2	DBT much poorer than 2D	
	-1	DBT slightly poorer than 2D	
	0	DBT equal to 2D	
	+1	DBT slightly better than 2D	
	+2	DBT much better than 2D	

4.Discrimination	-2	DBT much poorer than 2D	
	-1	DBT slightly poorer than 2D	
	0	DBT equal to 2D	
	+1	DBT slightly better than 2D	
	+2	DBT much better than 2D	
5.Overall opinion	-2	DBT much poorer than 2D	
	-1	DBT slightly poorer than 2D	
	0	DBT equal to 2D	
	+1	DBT slightly better than 2D	
	+2	DBT much better than 2D	



## Appendix 13 Invitation letter for assessment women

*(To be printed on local headed paper)*

Dear

You have been requested to attend an appointment at <<local centre>> for further investigation of an abnormality detected on your recent breast screening X-ray examination (mammogram).

I am writing to let you know that <<local centre>> is currently taking part in a research study to evaluate whether a new imaging technique could be more accurate than standard screening X-rays in detecting breast abnormalities. We would like to give you the opportunity to participate in this study.

Information about the research study is included with this letter. If you agree to participate in the study you will be asked to have an extra imaging examination at the assessment clinic, in addition to any other routine diagnostic tests.

If you have any questions about the study please call <<local contact name>> at <<local centre>> on telephone number <<xxxxx>> or contact the trial office on xxxxx xxxxxx.

Thank you for taking the time to read this information.

Yours sincerely,

Name of Local PI





## Appendix 14 Invitation letter for moderate- or high-risk women as a result of family history

*(To be printed on local headed paper)*

Dear

You have been invited to attend your annual breast mammogram at <<local centre>>.

I am writing to let you know that <<local centre> is currently taking part in a research study to evaluate whether a new imaging technique could be more accurate than standard screening X-rays in detecting breast abnormalities. We would like to give you the opportunity to participate in this study.

Information about the research study is included with this letter. If you agree to participate in the study you will be asked to have an extra imaging examination. Your results will be sent to you as normal.

If you have any questions about the study please call the research team at the <<local contact name>> at <<local centre >> on telephone number <<xxxxx>> or the Trial Office on xxxxx xxxxxx.

Thank you for taking the time to read this information.

Yours sincerely,

Name of local PI



## Appendix 15 Letter to general practitioner advising of trial participation

Dear Doctor Patient name(s)

The above patient(s) has consented to participate in the TOMMY trial at <<local centre>>.

The TOMMY (TOMosynthesis and MammographY) Trial is a comparison of Digital Breast Tomosynthesis (DBT) and standard digital mammography in the NHS Breast Screening Programme. The study has been funded by the Health Technology Assessment Programme and has received favourable ethical opinion from the Scotland A Research Ethics Committee.

The purpose of the study is to compare the performance of DBT and standard mammography in diagnosis of breast abnormalities. We aim to recruit 1000–1200 women locally over a 15–18-month period.

Women (age 47–73 years) who have been recalled for assessment after abnormal routine screening mammography will be invited to participate. Women (age 40–49 years) with a family history of breast cancer who are attending annual screening mammography will also be eligible.

For women giving written informed consent, a standard digital mammogram and DBT imaging examination will be performed. Both sets of images will be reviewed independently to evaluate the relative sensitivity and specificity of the two imaging techniques.

For further information please contact << local contact details>> or the Trial Coordination Office (Tel: xxxxx xxxxxx or email: xxxxxxxxxxxxxxxxxxxxxxxxx).

Yours sincerely,

Local PI



## Appendix 16 National Co-ordinating Centre for the Physics of Mammography tomosynthesis commissioning protocol for the TOMMY trial

*The mammography equipment to be used in the TOMMY trial will be commissioned according to IPEM 89 and NHSBSP Equipment Report 0604, which cover the testing of conventional digital mammography X-ray equipment and image display equipment. This protocol details the additional tests which will be carried out, including newly developed tests for digital breast tomosynthesis systems.*

For all tests the 'Std Acq' mode should be selected – not the enhanced mode as this increases the doses delivered. Use fixed rather than flexible paddle.

### Compressed breast thickness indication

Use blocks of PMMA of known thickness. Position on breast support table overhanging the chest wall edge such that the back edge of the top block is 8 cm from the chest wall edge (thus allowing the paddle to tilt slightly as it would during a clinical mammogram). Compress to 100 N and record the actual thickness and the indicated thickness. Repeat for thicknesses of approximately 2 cm, 5 cm and 7 cm or 9 cm.

### Reproducibility and homogeneity

Make 5 flatfield tomo exposures of large (24 × 30) 45-mm PMMA block with automatic exposure control (AEC) in Autofilter mode at the beginning of testing, with the paddle and using the same compressed breast thickness for all exposures to ensure selection of the same kV. Make a further exposure a few hours later and/or at the end of testing. These are used to check stability of mAs and SNR (though a valid SNR cannot be calculated for the Hologic Dimensions system due to the scaled pixel values), and also to check for homogeneity and artefacts.

Also make flatfield tomo exposures as above in AutokV and Autotime modes (with kV manually selected to match that for the Autofilter exposures) to ensure that the mAs at a given kV does not vary between AEC modes. Also try a combo exposure in each AEC mode and check that the exposure factors for the 2D and tomo exposures are not different from those obtained when making individual exposures.

During these scans measure tomo and scan times with a stopwatch. Tomo time begins when the expose button is pressed and ends at decompression. Scan time begins at start of first X-ray exposure to end of last X-ray exposure (measure this in Autofilter/Autotime AEC mode and in manual mode).

### Polymethyl methacrylate dose and thickness compensation

The Hologic system chooses beam quality based purely on indicated compressed breast thickness which is unfortunately in error when large PMMA blocks are compressed due to the paddle being prevented from tilting as it normally would in clinical situation. PMMA doses and CNR measurements should therefore be measured without spacers and the paddle positioned at such a height as to give the desired compressed breast thickness (i.e. equal to the equivalent breast thickness for the thickness of PMMA on the table). For 2 to 4-cm PMMA the top block may need to be moved forwards in order to allow the paddle to come down far enough. PMMA dose measurements are combined with the thickness compensation measurements.

**Tomo**

Position ball phantom (15-mm PMMA containing 10-mm 100% glandularity ball) on top of a 5-mm sheet of PMMA on the breast support table making up a total thickness of 20 mm. Image using a flatfield tomo scan in Autofilter AEC mode and record the kV and mAs. Repeat with additional thicknesses of PMMA placed on top of the phantom and for each thickness position the paddle to give the appropriate indicated breast thickness.

**Conventional**

Use 0.2-mm aluminium square on the midline 60 mm from chest wall edge, sandwiched between two 10-mm plates PMMA. Image using flatfield conventional in Autofilter AEC mode. Repeat with additional thicknesses of PMMA placed on top of the phantom and for each thickness position the paddle to give the appropriate indicated breast thickness.

Repeat using local physics aluminium square and PMMA for comparison.

**Tube output**

Half-value layers (HVLs) and tube outputs are measured in conventional and tomosynthesis modes with the paddle in the beam for the purpose of MGD calculations. The Monte Carlo simulations on which MGD calculations are based were actually carried out with the paddle in contact with the ion chamber, which increases the dose measured by a few per cent. For the TOMMY trial we have decided to use doses measured with the paddle in contact with the ion chamber for both tomosynthesis and conventional MGD so that the measurements are comparable. The dose with the paddle not in contact with the ion chamber will also be measured so that the MGDs for conventional mammograms can be compared to those from other systems measured in the usual way with the paddle not in contact according to the UK protocol.

Place steel or lead plate on breast support table to protect detector (should be a piece of 'heavy wood' labelled MIS in the room).

Position ion chamber at standard position on midline 4 cm from chest wall edge 10 cm above the table. Position the paddle immediately above and in contact with the ion chamber. (It may be easier to put the paddle in position first.) Position a collimator above the paddle close to the tube port such that the light beam is collimated to the chamber. The HVL filters will be placed on top of this collimator.

Select 50 mAs and measure the output and HVL for the clinical range of beam qualities in conventional and tomosynthesis modes. The HVL should be measured using thicknesses of aluminium filters just below and above the actual HVL.

For the tomosynthesis exposures use the 'zero degree tomo' setting (a series of short exposures without rotation, dosimeter will need to be set to accumulation mode).

Record actual mAs (likely to differ slightly from setting) and dose reading, focus-to-chamber distance, focus-to-table distance.

## Detector response, modulation transfer function, noise power spectrum

These tests are to be carried out in conventional 2D mode only.

### *Detector response*

Collimate beam to 10 × 10 cm centred on the midline 6 cm from the chest wall edge. Attach 2-mm aluminium to the tube port. Remove grid and protect detector. Place ion chamber 6 cm from the chest wall edge on the midline. Use 29WRh (same kV, target, filter as was selected automatically for 45-mm PMMA). Measure doses for 6 dose levels ranging from 4 to 125 mAs. Repeat measurements for lowest dose to ensure repeatability. Remove ion chamber and protector and image using conventional flatfield exposures for the same mAs values.

### *Modulation transfer function*

Position MTF tool on table so that the middle of its edge passes through the midline 6 cm from the chest wall edge at a slight angle (1.5–3 degrees) to a line either parallel or perpendicular to the chest wall edge. Take image using 29 WRh 100 mAs for each orientation take two images with the MTF tool either side of the line, giving four images in total.

### *Noise power spectrum*

Take 4 images at each of 5 dose levels using 29 WRh to give 0.25, 0.5, 1, 2, and 4 times a typical detector dose (approximately 7, 14, 28, 56 and 112 mAs).

## Alignment

### *Alignment of X-ray beam with imaged plane at detector surface*

Place sheet of white paper or card on table for better visibility and reduced slipperiness. Position rulers (+ outwards) with edge of light beam aligned with zero. Two rulers at front and one on each lateral side. At the back edge there is not enough space for a ruler so use another marker. An 18 × 24 sheet of PMMA (maximum 5-mm thick) or something else not too attenuating is helpful to place on top to hold rulers in place. Place strips of Gafchromic on top of rulers with line across middle aligned with zero on the rulers. Expose using flatfield tomo, manual, 32 kV, 140 mAs (two exposures to get enough blackening), no paddle. To reduce size of image keep paddle holder as low as possible without casting shadow on back part of image.

### *Missed tissue at chest wall edge*

Position a 6-cm stack of PMMA aligned with front edge of table. Insert X-ray rulers at 0, 3, 6 cm above table with zero aligned with front edge of PMMA/table. Expose with flatfield tomo under AEC control.

## Geometric distortion

Place the geometric distortion grid on the breast support table with large slabs of PMMA on top to make a total thickness of around 60 mm. Bring paddle down to rest on top of the PMMA and image using flatfield tomo Autofilter AEC setting. Repeat with the geometric distortion phantom in the middle of and on top of the stack of PMMA. (If there is evidence of distortion in the images then further investigation may be required at additional heights within the PMMA stack.)

## Z-resolution

Place large 5-mm sheet of PMMA on breast support table with Z-resolution phantom (5-mm PMMA containing 6 × 1-mm aluminium balls) on top. Add a further 40-mm PMMA to make up the total thickness to 50 mm. Bring paddle down to rest on top of the PMMA and image using flatfield tomo Autofilter AEC setting. Repeat with the Z-resolution phantom positioned approximately in the middle of the stack, and again 5 mm from the top of the stack.

## Image quality

Position CDMAM aligned with the chest wall edge of the breast support table sandwiched between 2- × 2-cm PMMA and bring paddle down to rest on top of the PMMA. Check the image to ensure that the edges of the CDMAM are within the image. Repeat to give a total of 16 images, moving the CDMAM very slightly between exposures.

### *Tomo*

Use flatfield tomo and manual setting to give the same kV and mAs (or as close as possible using either Manual or Autotime AEC setting) as was selected automatically for tomo exposures of 5-cm PMMA with the paddle at 6 cm.

Using the same settings also take 16 images of local CDMAM for comparison.

### *Conventional*

Use flatfield conventional and manual setting to give the same kV and filter (31 WRh) and mAs (or as close as possible using either Manual or Autotime AEC setting) as was selected automatically for conventional exposures of 5-cm PMMA with the paddle at 6 cm.



## Appendix 17 Radiographer quality control for the TOMMY trial

The radiographer QC required for the TOMMY trial will be the NHSBSP-recommended tests for digital mammography with four additional tests for tomosynthesis. The tests are listed below.

### List of required tests

#### Monitors

Daily checks on acquisition and reporting monitors (NHSBSP 0702 section 2.1). Monthly test of reporting monitors (NHSBSP 0702 section 2.2).

#### X-ray unit

Daily system check in conventional 2D mode (NHSBSP 0702 section 3.1).

\*Daily system check in tomosynthesis mode.

Weekly CNR check in conventional 2D mode (NHSBSP 0702 section 3.2). Weekly uniformity check in conventional 2D mode (NHSBSP 0702 section 3.3). Weekly image quality test in conventional 2D mode (NHSBSP 0702 section 4).

\*Weekly remote QC images in 2D and tomosynthesis modes.

Weekly detector flat field calibration (according to manufacturer's instructions). Monthly thickness check in conventional 2D mode (NHSBSP 0702 section 3.4).

\*Monthly thickness check in tomosynthesis mode.

Monthly mechanical and safety function tests (NHSBSP 0702 section 7).

\*Six-monthly geometry calibration for tomosynthesis (according to manufacturer's instructions).

Repeat analysis (NHSBSP 0702 section 8).

(\* denotes additional QC for TOMMY trial.)

### Instructions for quality control tests

For the conventional 2D digital mammography tests, the instructions given in NHSBSP 0702 should be followed. Further notes on some of these tests are included later in this document.

There are two additional tests required by Hologic which should be carried out according to their instructions given in the Hologic Selenia Dimensions QC Manual. One of these is the weekly flat field calibration for conventional 2D mammography, the other is the geometry calibration for tomosynthesis systems.

Instructions for the other three additional tomosynthesis tests for the TOMMY trial are included later in this document.

## Quality control equipment

It is expected that mammography centres taking part in the trial will already have the required equipment for testing conventional digital mammography systems according to NHSBSP 0702.

Equipment for the manufacturer's required QC (flat field and geometry calibrations) are supplied with the system.

National Co-ordinating Centre for the Physics of Mammography supply a test phantom with which to carry out the additional weekly remote QC tests, the images from which are sent to NCCPM.

## Recording of quality control results

Conventional 2D QC should be recorded in the normal way as per local procedure. A weekly summary of 2D local QC is also to be entered onto the spreadsheet provided and sent to NCCPM each week. The spreadsheet also contains pages for the recording of the additional tomosynthesis tests. These pages can be printed out and used as paper forms to collect the data on a daily basis, to then be copied into the spreadsheet and sent to NCCPM each week. Where there is an existing digital QC spreadsheet in use, the TOMMY spreadsheet may be grafted onto it to avoid unnecessary duplication in data entry.

## Physics support

National Co-ordinating Centre for the Physics of Mammography intend to appoint a local mammography physicist for each site to provide physics cover for the equipment used for the TOMMY trial. If it is not possible to appoint a local TOMMY physicist then NCCPM will provide the required physics cover. Where QC results fail remedial levels, and the problem is not resolved, then the TOMMY physicist should be asked for advice before continuing to use the equipment for the trial. If the local TOMMY physicist is not available then NCCPM should be consulted instead. In any event, NCCPM should be kept informed of any problems that occur.

## Quality control baselines and remedial levels

Baselines for conventional 2D tests will be set in the usual manner according to local protocol and remedial levels will be set following recommendations in NHSBSP 0702. Baselines and remedial levels for the tomosynthesis tests should be set in a similar manner using the provisional remedial levels quoted in the table summarising remedial levels for all tests appended to this document. The TOMMY physicist should be consulted if assistance is required in setting baselines and remedial levels. Once the baselines and remedial levels have been set they should be sent to the TOMMY physicist and to NCCPM, so that they can be checked and compared with those from other sites in the trial. If there is ever any need to alter any baselines or remedial levels, the TOMMY physicist and NCCPM must be consulted. As experience is gained in the testing of tomosynthesis equipment, NCCPM may decide to make changes to test methods or remedial levels, and will inform all concerned.

## Reporting of quality control to the National Co-ordinating Centre for the Physics of Mammography

Each week the following should be sent to NCCPM:

- The completed QC spreadsheet containing a summary of the 2D QC results and the tomo test records.
- The weekly remote QC images (not from SecurView or PACS):
  - The tomosynthesis image of aluminium ball in PMMA.
  - The tomosynthesis image of plain PMMA.
  - The 2D image of plain PMMA.

The weekly QC summary should be sent by e-mail to the e-mail address given below.

Quality control images need to be written to a CD/DVD directly from the acquisition workstation and sent to NCCPM. We require raw tomosynthesis images which are not transferred to the PACS network or SecurView workstation. These can only be written to CD or DVD using the acquisition workstation in the X-ray room.

Instructions for downloading QC images to CD/DVD:

Open the weekly QC study and check that it contains the required QC images and not patient images. (Obviously it is very important that no patient image is included as this would contravene data protection laws). Put a blank CD or DVD in the drive and click on the 'Export' icon. Select destination 'E:' and start. Only one study can be written onto each disc.

When the facility becomes available it is planned that QC images will be transferred from the acquisition workstation to NCCPM via secure internet link.

### Contact details for the National Co-ordinating Centre for the Physics of Mammography:

Celia Strudley is the main contact for TOMMY QC, but if not available other staff should be able to help.

Address: NCCPM (Medical Physics)

Level B, St Luke's Wing Royal Surrey County Hospital Guildford

GU2 7XX

Telephone: NCCPM office: xxxxx xxxxxx

Celia direct: xxxxx xxxxxx ext xxxx E-mail: xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

(This e-mail address will be monitored by NCCPM staff dealing with QC issues during the trial)

## Tomosynthesis quality control instructions

### *Daily system check in tomosynthesis mode*

This test is an extension of the existing 2D daily system test, and the two would be most easily carried out together.

Use the same PMMA block and compression paddle as are used for the daily 2D system test.

Place the PMMA on the breast support table.

This test should always be carried out with the paddle in the fixed position or carried out with the paddle in the flexible position.

Compress to 100 N (or as near to as possible, 95–105 N is acceptable) Set the AEC sensor to position 2.

In addition to the conventional 2D Autofilter flatfield image, take a tomosynthesis Autofilter image of the block. These two images may be taken as a combo exposure.

Record the following:

- compressed breast thickness
- kV
- mAs.

View the reconstructed image slices (not the projections) on the X-ray unit. While scrolling through the tomosynthesis slices, look for any abnormal artefacts in the image, and record whether or not any abnormality has been observed.

Enter all the recorded information onto the form/spreadsheet provided, and compare the mAs against the remedial level. Where any remedial levels are exceeded, repeat the test to ensure that a mistake has not been made.

If remedial levels are exceeded or abnormal artefacts are observed then contact the TOMMY physicist for advice before continuing to use the equipment for patients in the trial. Any image containing abnormal artefacts must be saved for investigation and a copy should also be sent to NCCPM.

(It is not possible to measure pixel values within a region of interest [ROI] in a reconstructed tomosynthesis slice on the Dimensions AWS, so the calculation SNR or CNR values for tomosynthesis images is not required for Radiographer QC.)

### *Weekly remote quality control images*

#### **Image of test object containing 1-mm aluminium ball in tomosynthesis mode**

Place the 5-mm sheet of PMMA containing the 1-mm aluminium ball between the two 22.5-mm thick sheets of PMMA on the breast support table. The aluminium ball is located on the midline 6 cm from one of the long edges of the sheet – this long edge should be aligned with the chest wall edge of the table.

Use the large 24 × 29 cm compression paddle (in fixed mode). Compress to 100 N (or as near as possible, 95–100 N is acceptable) Set the AEC sensor to position 2.

Take a flatfield tomosynthesis image with the AEC in Autofilter mode.

View the reconstructed image slices (not the projections) on the acquisition workstation in the X-ray room. Press the 'Actual pixels' button to bring the image into full resolution, and using the wheel scroll through the image. The aluminium ball should be visible as a spot somewhere between slice numbers 20 and 30.

Find the slice at which the image of the ball appears sharpest and least distorted and record that slice number. The ball may appear fairly sharp over a range of 3 to 5 slices – record a slice number in the middle of the range.

Check that the ball appears sharp and circular in this slice. Record whether or not this is so.

Above and below the slice in which the ball appears sharpest the ball will appear to become elongated in the vertical direction (parallel to chest wall edge). Check that the appearance of this line is as normal with no broadening, twisting, or otherwise abnormal appearance. Record whether or not the appearance is as normal.

If any abnormalities are seen in the appearance of the ball in the reconstructed tomo image, contact the TOMMY physicist for advice before continuing to use the equipment for patients in the trial.

Two dimensional and tomosynthesis images of large plain PMMA block.

Place the two 22.5-mm thick sheets of PMMA on the breast support table. Use the large 24 × 29 compression paddle (in flexible mode).

Compress to 100 N (or as near as possible, 95–100 N is acceptable) Set the AEC sensor to position 2.

Take a flatfield combo image with the AEC in Autofilter mode. Display the 2D image on the monitor and check carefully for artefacts.

Display the reconstructed tomo image on the monitor and scroll through the slices, checking carefully for artefacts.

Record whether or not any abnormalities are seen. If any abnormalities are seen in either the 2D or the tomo image, contact the TOMMY physicist for advice before continuing to use the equipment for patients in the trial.

Send the three weekly images and QA spreadsheet to NCCPM.

Every week, send a copy of the aluminium ball tomosynthesis image, and the 2D and tomosynthesis images from the Combo exposure to NCCPM for further analysis.

Also send a copy of the Radiographer QA spreadsheet to NCCPM by e-mail.

### **Monthly thickness check in tomosynthesis mode**

This test is an extension of the existing 2D monthly thickness test, and the two would be most easily carried out together.

Use the same PMMA blocks, compression paddle and thicknesses as are used for the 2D thickness test. (Record the details of the method used when setting the baseline so that it can be reproduced for future measurements.) In addition to the conventional 2D Autofilter image, take a tomosynthesis Autofilter image for each thickness.

Record the kV and mAs for the tomosynthesis exposure on the form/spreadsheet provided and compare against remedial levels. Where any remedial levels are exceeded, repeat the test to ensure that a mistake has not been made.

If remedial levels are exceeded or abnormal artefacts are observed then contact the TOMMY physicist for advice before continuing to use the equipment for patients in the trial.

## Notes on quality control tests

Acquisition of patient images for the trial may only take 12 to 18 months, after which time, the TOMMY trial requirements for QC on the X-ray units may cease, but QC will need to be continued for the reporting monitors until reading of the images for the trial has finished at the centre. When patient image acquisition has finished, weekly reporting of QC results to NCCPM will not be required and monthly reports of monitor QC to NCCPM will suffice.

### Monitors

#### Daily checks on acquisition and reporting monitors (NHSBSP 0702 section 2.1)

#### Monthly test of reporting monitors (NHSBSP 0702 section 2.2)

NHSBSP 0702 appears a little ambiguous regarding the need for the QC radiographer to have a lightmeter to measure ambient light levels at the reporting monitor. If a lightmeter is not available and not normally used for this test, then the QC radiographer needs to be aware of the ambient light levels measured by the physicist, and the conditions required to keep the ambient light below the 10 lux limit, for example the closing of blinds and doors, etc. The QC radiographer needs to check that these conditions are being maintained and take action to reinstate the required conditions, such as closing blinds and reminding users of the requirements. Where there is any doubt as to whether or not ambient light levels remain within limits, the physicist should be asked to repeat the measurement with a lightmeter.

### X-ray unit

#### Daily system check in conventional two-dimensional mode (NHSBSP 0702 section 3.1)

The conventional 2D block test may be carried out using the local PMMA usually used for this test or alternatively the PMMA provided for the tomosynthesis test (without the layer containing the aluminium ball) may be used if preferred, in order that the 2D and tomo test images may be conveniently carried out sequentially as a combo exposure.

Ensure that the mean pixel value taken from the ROI for SNR calculation is the corrected value from which the offset has been subtracted.

The daily 2D image should also be checked for artefacts. Press the 'Actual pixels' button to bring the image into full resolution.

*\*Daily system check in tomosynthesis mode.*

*Weekly CNR check in conventional 2D mode (NHSBSP 0702 section 3.2). Weekly uniformity check (NHSBSP 0702 section 3.3).*

Results need not be reported to NCCPM as we will use the 2D image sent to us each week to assess homogeneity.

*Weekly image quality test (NHSBSP 0702 section 4).*

Either Tormam or Tormax may be used according to local procedure. Results need not be reported to NCCPM.

*\*Weekly remote QC images.*

Three images to be sent to NCCPM each week for further analysis.

*Weekly detector flat field calibration (according to manufacturer's instructions).*

The manufacturer's special flat field phantom must be used for this. In order to keep the flatfield phantom in pristine condition it should not be used for any other tests.

If the calibration fails the advice given in the manufacturer's QC manual should be followed and, if not resolved, use of the equipment should be suspended, an engineer called in, and the TOMMY physicist and NCCPM informed.

Monthly thickness check in conventional 2D mode (NHSBSP 0702 section 3.4).

*\*Monthly thickness check in tomosynthesis mode.*

Combine with 2D monthly thickness checks and for maximum efficiency use combo mode to acquire the 2D and tomo images.

The only remedial level for this test is for the mAs, which needs to be within 10% of the baseline assuming that the baseline was set using the same kV. If the kV differs by only 1kV from the baseline, repeat the exposure using the kV used for baseline. If the result is now satisfactory, record the new kV and mAs as a supplementary baseline.

*Monthly mechanical and safety function tests (NHSBSP 0702 section 7).*

According to local procedure. No need to report this to NCCPM.

*\*Six-monthly geometry calibration for tomosynthesis (according to manufacturer's instructions).*

The Dimensions system should automatically alert the user when this is due.

If the calibration fails the advice given in the manufacturer's QC manual should be followed and, if not resolved, use of the equipment for tomosynthesis should be suspended, an engineer called in, and the TOMMY physicist and NCCPM informed.

*Repeat analysis (NHSBSP 0702 section 8).*

According to local procedure. No need to report this to NCCPM.

TABLE 34 Summary of remedial levels for radiographer QC tests

QC test	Remedial level
Daily checks on acquisition and reporting monitors (NHSBSP 0702 section 2.1)	Monitor fails any of the checks
Monthly test of reporting monitors (NHSBSP 0702 section 2.2)	Monitor fails any of the checks
Daily system check in conventional 2D mode (NHSBSP 0702 section 3.1)	mAs baseline $\pm$ 10% (provided kV & filter match baseline) DDI baseline $\pm$ 10% SNR baseline $\pm$ 20%
Daily system check in tomosynthesis mode <sup>a</sup>	mAs baseline $\pm$ 10% (provided kV & filter match baseline) Any significant abnormal artefacts
Weekly CNR check in conventional 2D mode (NHSBSP 0702 section 3.2)	CNR $\pm$ 20%
Weekly uniformity check (NHSBSP 0702 section 3.3)	> 10% deviation from mean PV at centre Any significant artefacts
Weekly image quality test (NHSBSP 0702 section 4)	Any significant change from baseline image
Weekly image of test object containing 1mm aluminium ball in tomosynthesis mode <sup>a</sup>	Any significant change in the appearance of the aluminium ball and the reconstruction artefacts above and below it
Weekly detector flat field calibration (according to manufacturer's instructions)	Calibration failure
Monthly thickness check in conventional 2D mode (NHSBSP 0702 section 3.4)	For each thickness: DDI baseline $\pm$ 10% SNR baseline $\pm$ 20% CNR baseline $\pm$ 20%
Monthly thickness check in tomosynthesis mode <sup>a</sup>	For each thickness: mAs $\pm$ 10% (provided kV matches baseline)
Monthly mechanical and safety function tests (NHSBSP 0702 section 7)	Equipment fails any of the checks
Six-monthly geometry calibration for tomosynthesis (according to manufacturer's instructions) <sup>a</sup>	Calibration failure

DDI, detector dose indicator.  
a Denotes tomosynthesis QC test.





A decorative graphic consisting of numerous thin, parallel green lines that curve from the left side of the page towards the right, creating a sense of movement and depth.

**EME  
HS&DR  
HTA  
PGfAR  
PHR**

Part of the NIHR Journals Library  
[www.journalslibrary.nihr.ac.uk](http://www.journalslibrary.nihr.ac.uk)

*This report presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health*

***Published by the NIHR Journals Library***