NIHR HTA Programme

18 December 2012

The NIHR Evaluation, Trials and Studies Coordinating Centre (NETSCC), based at the University of Southampton, manages evaluation research programmes and activities for the NIHR

Health Technology Assessment Programme
National Institute for Health Research
Evaluation, Trials and Studies Coordinating Centre
University of Southampton, Alpha House
Enterprise Road, Southampton, SO16 7NS

tel: +44(0)23 8059 5586
fax: +44(0)23 8059 5639

email: hta@hta.ac.uk
web: www.hta.ac.uk

Study Protocol

TOMMY trial: A comparison of TOMosynthesis with digital MammographY in the UK NHS Breast Screening Programme

Short Title: TOMMY trial: Comparison of TOMosynthesis with digital MammographY

Sponsor	University of Cambridge
Funder	National Institute for Health Research Health Technology Assessment Programme
Funding Reference Number	09/22/182
Chief Investigator	Professor Fiona J Gilbert
REC Number	10/MRE00/39
ISRCTN Number	
Version Number and Date	Version 3.0 18th September 2012

Chief Investigator	Professor Fiona Gilbert
Trial Office	Mrs Louise Harlow University of Cambridge School of Clinical Medicine University of Cambridge Box 218 Addenbrooke's Hospital Cambridge CB2 0QQ Tel: 01223 746 443 lh458@medschl.cam.ac.uk
Statistical Analysis Centre	Professor Stephen W Duffy Wolfson Institute of Preventive Medicine Barts & The London School of Medicine and Dentistry Queen Mary University of London London EC1M 6BQ email: s.w.duffy@qmul.ac.uk

ABBREVIATIONS

AE	Adverse event
AR	Adverse reaction
CI	Chief Investigator
DBT	Digital Breast Tomosynthesis
FFDM	Full Field Digital Mammography
GCP	Good Clinical Practice
GP	General Practitioner
ICF	Informed Consent Form
ICH	International Conference of Harmonisation
IRAS	Integrated Research Application System
ISF	Investigator Site File
NHSBSP	National Health Service Breast Screening Programme
PI	Principal Investigator
PIL	Participant Information Leaflet
R&D	NHS Trust R&D Department
REC	Research Ethics Committee
SAE	Serious Adverse Event
SAR	Serious Adverse Reaction
SOP	Standard Operating Procedure
TMF	Trial Master File

SUMMARY OF THE TRIAL

Breast screening is recognised as the most effective method of detecting early stage breast cancer and reducing breast cancer mortality. However, 15-30% of cancers are not detected by standard mammography screening and this percentage is higher in dense breasts and in women under 50 years. One of the major limitations of standard mammography is that overlapping dense fibroglandular tissue in normal breast tissue can decrease the visibility of malignant abnormalities or mimic abnormalities. This results in some cancers being missed or unnecessary recalls, assessments and psychological stress. 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 and facilitating differentiation between malignant and non malignant features.

The aim of this study is to assess whether DBT could improve upon digital mammography (FFDM) as a screening tool, particularly in certain groups of women e.g. those with a family history of breast cancer, or women with dense breasts.

Six UK National Health Service Breast Screening Programme (NHSBSP) centres will participate in the study and recruit a total of 9000 women over a 15-18 month period. Women (aged 47-73) who have been recalled to an assessment clinic following abnormal screening mammography and women (aged 40-49) with a family history of breast cancer attending annual mammography will undergo both a standard digital mammogram and DBT imaging. The diagnostic accuracy of the two sets of images will be compared in an independent retrospective reading study.

INTRODUCTION

In standard 2D mammography, overlapping dense fibroglandular tissue within the breast can decrease the visibility of malignant abnormalities, or simulate the appearance of an abnormality, creating a false positive result that leads to unnecessary recalls, biopsies and related psychological stress(1). Women with dense breasts are at higher risk of developing breast cancer (2) and the reduced sensitivity of mammography in women with increased breast density is recognised as a fundamental limitation of screening(3). This is of particular concern for the NHSBSP as it is extending the age range of screening to include younger, pre- or perimenopausal women who have a higher proportion of dense breast tissue (4) (5) It is also potentially problematic for women aged 40-49 at moderate or high risk of developing familial breast cancer who are recommended to have annual mammography(6).

Digital Breast Tomosynthesis

Although the fundamentals of tomographic imaging were established in the 1930s clinical applications of tomosynthesis 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 (7). 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 exposure used for each projection image is relatively small so that the overall mean glandular dose for DBT is comparable to that of conventional 2D imaging. The projection images acquired by the detector are processed by algebraic 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. Image quality of DBT is highly dependent on system geometry and the choice of optimal image acquisition, reconstruction and display parameters (8-10). A viewing workstation presents the reader with tools that enable them to scroll vertically through the tomographic images as well as to compare them with the corresponding 2D images if desired.

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 differentiation of superimposed breast tissue from malignant lesion, DBT could also decrease the number of false positive recalls and associated healthcare costs and patient anxiety(11).

A recent technology assessment review (12) of DBT as a screening tool concluded there was currently insufficient clinical evidence but recommended that the development of this emerging technology should be monitored. Rigorous clinical trials of the clinical utility and

cost effectiveness of DBT are required to establish its optimum role in breast imaging (13) (14).

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 (15-17). Improved lesion visibility and classification compared to standard film or digital mammography were reported in several studies (18-21) and superior estimation of lesion classification and size of subtle cancers compared to FFDM or ultrasound has also been reported (22,23). A study of 200 diagnostic cases using ROC analysis, reported that the clinical performance of one-view DBT was comparable (non-inferior) to that of two- view FFDM(24) and a significant improvement in diagnostic accuracy of one-view DBT over two-view FFDM was observed in a study of 200 subtle cases(25). In a study of 310 mixed cases, one or two-view DBT in combination with FFDM demonstrated superior ROC performance in women with dense breasts (26) and studies of parenchymal texture analysis in correlation with breast density suggested that DBT may be more discriminative than FFDM in estimating breast cancer risk(27,28).

Cancer detection and recall rate

A 30% reduction in false positive recall rate using DBT in combination with FFDM compared to FFDM alone and a smaller, non significant, decrease (10%) using DBT alone was reported in a retrospective multireader observer study of 125 cases including 35 cancer cases (29). An American screening population study involving 1957 women reported a reduction in recall rate from 7.5% to 4.3% (30)and a blinded reader study of 60 cases (31)reported an 83% reduction in false positive recall rate using DBT with no significant decrease in cancer detection rate compared to conventional 2D mammography. A blinded assessment of 100 cases (60 screening cases, 25 diagnostic cases and 10 subjects undergoing biopsy) reported 40% improvement in sensitivity and a 20% reduction in recall rate (32). However, an independent review of FFDM images, DBT images and reconstructed DBT images in a cancer enriched dataset of 30 cases reported no statistically significant differences in cancer detection rate or recall rate but conclusions were limited by the small sample size and study design (33). Similar sensitivity of DBT compared to FFDM was also reported in a study of 513 women with abnormal screening mammograms or clinical symptoms (34) and a trial of 2764 women attending screening mammography reported a 42% reduction in recall rate using one-view DBT compared to conventional mammography(35). Higher sensitivity with DBT compared to either single view or two-view

digital mammography was reported in a non blinded, retrospective, consensus study of 36 cases with subtle suspicious findings from screening mammography (36).

Reader performance

Multi-reader studies have demonstrated improved performance by radiologists with a range of experience, as measured by reduction in recall rate and ROC analysis, using DBT in combination with 2D mammography compared to 2D mammography alone.(37-40)

AIMS and OBJECTIVES

3.1 Research objectives

1. To compare the diagnostic accuracy of breast DBT and FFDM in women aged 47-73 years.
2. To determine whether the addition of DBT to FFDM improves the accuracy of detection of small or subtle breast cancers.
3. To determine the diagnostic accuracy in women with dense breasts as these are the most challenging in which to detect cancers in a screening programme.

Secondary objectives are to compare the accuracy of DBT vs FFDM in (a) women 40-49 years with a moderate or high risk of breast cancer attending annual screening mammography (b) lesions presenting as soft tissue masses (c) lesions presenting as microcalcifications.

STUDY DESIGN

The proposed trial is a multicentre retrospective matched comparison of the diagnostic performance of DBT and FFDM. The overall study design is shown in Figure 1

STUDY SETTING

Participants will be recruited from six NHSBSP centres in the UK**.**

STUDY POPULATION

A total of 9000 participants will be recruited over a 15-18 month period from women aged 47-73 recalled to an assessment clinic for a mammographic abnormality detected at routine breast screening. In addition, women under 50 years of age with a family history of breast cancer attending annual mammography screening (6) will also be included. This will accrue a dataset comprising a large number of abnormal cases with a relatively high proportion of cancer cases (approximately 18% cancer cases and 50% cases with overlapping tissues on standard mammography that simulate suspicious features which are actually normal breast tissue).

Inclusion Criteria

- Women 47-73 years attending routine breast screening (either film or digital mammography) and recalled for further assessment.
- Women 40-49 with family history 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 who has breast implants
- Any woman who is pregnant

RECRUITMENT

A participant information leaflet outlining the potential benefits and risks of the study will be sent to women who would be suitable for inclusion in the study and informed consent will be obtained on attendance at the assessment clinic or surveillance mammography appointment. Women who are unable to give consent will be excluded from the study.

Women aged 47-73 years who have been recalled to an assessment clinic

Women who have been recalled following abnormal film or digital mammography screening will be mailed a trial information leaflet and consent form in advance of their appointment. At the assessment clinic appointment a healthcare professional or member of the research team will discuss the content of the trial information sheet and answer any questions relating to trial participation prior to seeking written informed consent.

High/medium risk women attending annual mammography

Women with a family history of breast cancer who are invited to attend annual screening mammography will be mailed a trial information leaflet and consent form in advance of their screening appointment. At the time of the screening appointment, a healthcare professional or member of the research team will discuss the content of the trial information sheet and answer any questions relating to trial participation prior to seeking written informed consent.

INFORMED CONSENT

The person taking informed consent must be appropriately qualified, trained and authorised to do so by the CI/PI. The person taking consent must have a comprehensive understanding of the study. They will be responsible for ensuring that the subject fully understands the information contained in the Participant Information Sheet and answer any questions relating to study participation prior to seeking written consent.

Participants will be asked sign and date the consent form. The consent form should also be signed and dated by the person taking consent.
The original consent form should be filed in the ISF
One copy should be given to the participant
One copy filed with the participant's screening centre records

END OF TRIAL
The end of trial will be one year after the last participant has been recruited into the trial.

STUDY SITE STAFF
The PI at each centre must be familiar with the trial protocol and the study requirements and staff assisting in the study must be adequately informed about the trial protocol and their trial related duties. Delegation of responsibility for aspects of the informed consent process will be documented and filed in the ISF. The CI, with the agreement of the Sponsor, will ensure all other documents required for compliance with the principles of GCP are retained in the TMF and that appropriate documentation is available in local ISFs.

STUDY INTERVENTIONS
All participants will undergo standard two-view (MLO and CC) FFDM of both breasts and two-view (MLO and CC) DBT imaging. For participants recruited in assessment clinics, imaging will be conducted prior to any additional investigations deemed necessary in the assessment clinic. The imaging examinations will be conducted by NHSBSP staff with appropriate DBT training.
Following FFDM and DBT imaging, women will resume the normal pathway through the assessment or screening clinic. On completion of the FFDM and DBT imaging, a participant's involvement in the trial will be complete. Any subsequent management will be in accordance with standard assessment clinic or screening centre procedures.

IMAGE ACQUISITION
Both the standard FFDM and the DBT imaging will be performed as a single procedure under the same degree of breast compression on a Hologic Selenia 2D Dimensions Digital Mammography Unit.
Radiographers will be experienced specialist mammography radiographers, fully trained in accordance with NHSBSP standards, and who have had additional specific training on the DBT equipment used in the study.

READERS

A maximum of three readers from each centre will be trained to undertake this study. The readers will be a mixture of radiologists, advanced practitioner radiographers, radiologists and breast clinicians, representative of current reading practice in the NHSBSP. All readers will have 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 PERFORMS self assessment test
- Attendance at assessment clinics and multidisciplinary team meetings

READER TRAINING

Reader training will consist of two days of applications training from the DBT system manufacturer and one day of test set reading. Over the first 12 months of the study recruitment period readers will gain experience of DBT by reviewing the DBT and FFDM images of at least 500 cases (including approximately 100 cancer cases) acquired at their own site.

WORKSTATIONS

All readers will use Hologic SecurView DW workstations, optimised to read both FFDM and DBT images. Every reader will be blinded to the outcome status of each case and will read cases independently of all other readers.

IMAGE REVIEW and PROSPECTIVE DATA COLLECTION

1.Participants recruited at assessment clinics

The FFDM and DBT images will be reviewed in the assessment clinic by one reader and used to inform subsequent patient management. Data will be collected prospectively for each case using proforma data collection sheets. Since approximately 30% of recalled cases are judged to be normal after repeat mammography a case hanging protocol will be established to alternate the viewing sequence of cases (2D then 3D images or vice versa). For each imaging modality the case will be scored on a standard 5 point scale (1=normal, 2=benign, 3= probably benign, 4=suspicious, 5=malignant). The location, size, type and features of any abnormalities will be recorded. Lesion conspicuity will be recorded on a five point scale (1=no visible finding, 2=low conspicuity, 3= medium conspicuity, 4= high conspicuity, 5=very high conspicuity). Readers will also grade their confidence that the lesion is malignant on a 0-100% scale and assess breast density on a 10cm VAS scale. The outcome of other assessment clinic procedures e.g.ultrasound and biopsy will also be

recorded. Histopathology from core biopsy or surgical excision will be used as the gold standard to confirm the presence of a cancer and this information will be subsequently collated with the proforma data for each case to generate a ground truth database.

2. High/moderate risk participants recruited at screening mammography

The FFDM and DBT images will be reviewed by two readers as standard in the NHSBSP and data recorded on proforma data collection and clinical outcomes sheets.

READER ONE

The first reader will review the FFDM images then the DBT images.

On reviewing the FFDM images, each case will be scored on a standard 5 point scale (1=normal, 2=benign, 3= probably benign, 4=suspicious, 5=malignant). The location, size, type and features of any abnormalities will be recorded. Lesion conspicuity will be recorded on a five point scale (1=no visible finding, 2=low conspicuity, 3= medium conspicuity, 4= high conspicuity, 5=very high conspicuity). Readers will also grade their diagnostic confidence on a 0-100% scale and assess breast density on a 10cm VAS scale.

The first reader will then access the DBT images and score the case, as above, on a standard 5 point scale (1=normal, 2=benign, 3= probably benign, 4=suspicious, 5=malignant). The location (slice range), type and features of any abnormalities will be recorded. Lesion conspicuity will be recorded on a five point scale (1=no visible finding, 2=low conspicuity, 3= medium conspicuity, 4= high conspicuity, 5=very high conspicuity). An overall decision of recall/no recall will be recorded. The reader will record their confidence that a lesion is malignant on a 0-100% scale.

SECOND READER

The second reader (blinded to the results of the first reader) will review the DBT images then the FFDM images.

On reviewing the DBT images, each case will be scored on a standard 5 point scale (1=normal, 2=benign, 3= probably benign, 4=suspicious, 5=malignant). The location (slice range), size, type and features of any abnormalities will be recorded. Lesion conspicuity will be recorded on a five point scale (1=no visible finding, 2=low conspicuity, 3= medium conspicuity, 4= high conspicuity, 5=very high conspicuity).

The second reader will then access the FFDM images and score the case, as above, on a standard 5 point scale (1=normal, 2=benign, 3= probably benign, 4=suspicious, 5=malignant). The location, size, type and features of any abnormalities will be recorded.

Lesion conspicuity will be recorded on a five point scale (1=no visible finding, 2=low conspicuity, 3= medium conspicuity, 4= high conspicuity, 5=very high conspicuity).
An overall decision of recall/no recall will be recorded. The reader will record their confidence that a lesion is malignant on a 0-100% scale.

Breast density

Qualitative or quantitative measurement of breast density from 2D mammograms methods is highly subjective and variable (41)(Yaffe 2008). DBT has the potential to enable direct measurement of volumetric radiological density (42). Each reader will record a rating of breast density on a 10cm visual analogue scale from the mammographic views and from the DBT images. These will be converted to percentages and compared to the automatic measure of breast density from the digital mammographic views of each breast obtained from the Hologic Quantra software package.

RETROSPECTIVE READING STUDY

It has been shown that microcalcification clusters are not as easily detected on DBT images alone and therefore it has been suggested that a standard view digital mammogram (FFDM) should be acquired along with two-view DBT for optimal microcalcification assessment. Software has also become available that creates a "synthetic" 2D image from a single DBT scan, simulating a conventional FFDM slice. The combination of DBT with FFDM requires approximately doubling the radiation dose to the breast being imaged. If it can be demonstrated that 2Dsynthetic images are satisfactory and comparable to FFDM, double exposure could potentially be eliminated. Therefore the diagnostic performance of three imaging regimes will be compared:

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

IMAGE REVIEW and DATA COLLECTION (RETROSPECTIVE)

Cases will be read on a workstation without access to the original screening mammograms or prior examinations. The location of any abnormality will be recorded and cases will be scored on a standard 5 point scale (1=normal, 2=benign, 3= probably benign, 4=suspicious, 5=malignant). Readers will make a decision to recall or not recall each case.

STUDY DATASET

The local data manager will collate copies of the pseudoanonymised image files (i.e. FFDM and DBT images) and clinical outcomes for participants from each centre and send these to

the Trial Data Manager in Cambridge. The Trial Data Manager and research radiographer will be responsible for:

- Confirming receipt of the image files from participating centres
- Checking image quality of files for data integrity
- Collating and storing the files prior to distribution for independent reading
- Entering data from prospective reading study and ground truth information into a study database
- Distribution of cases to readers at other centres for the retrospective reading study
- Collation of data from retrospective reading study

The research radiographer and CI in the trial office will check the information from data collection sheets and ground truth data for each case. Cases will be allocated to one of three categories (Cancer/Suspicious/Normal). From the study image database, equal numbers of cases will be randomly selected from each category and distributed between readers and centres to minimise bias. Each reading set will comprise approximately 40 cases (per week). The database will track the dates on which images are sent to readers at each centre and the dates on which completed data collection forms are returned by the readers. Readers will review either (a) FFDM images or (b) FFDM+DBT or (c) DBT+2D Synthetic images for any one case and will not review any cases from their own centre (see Figure 2)

OUTCOME MEASURES

The primary outcome measures from the retrospective independent reading study will be the relative sensitivity and specificity of FFDM and DBT in the detection of early stage cancers (<15mm), subtle lesions (e.g. lesions that were detected by only one reader at the time of screening mammography), and in women with dense breasts. Secondary outcome measures will include an evaluation of the visibility of multifocal lesions and microcalcification detection.

Histopathology will be used as the gold standard to confirm the presence of a cancer.

STATISTICS AND DATA ANALYSIS

SAMPLE SIZE

The power calculations below assume that for any given cancer case, at least one of the imaging modalities gives the ‘correct’ answer (malignant or not). This is generally

conservative. Assuming that some cases will be wrongly classified by both, 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 wish to compare FFDM and two-view DBT and to detect as statistically significant any improvement of sensitivity or specificity conferred by the latter. DBT could prove to be particularly useful for a number of subgroups:

- Women 40-49 years with moderate or high risk of familial breast cancer
- Detection of grade 3 tumours of size <15mm
- Women with >70% beast density

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

Sensitivity: The smallest anticipated subgroup of cancers is likely to comprise around 15% of the total tumour population. In any given subgroup, we postulate a sensitivity for FFDM of 85% and of two-view DBT of 95%. Assuming that both detect a cancer in 80% of cases, (i.e. discordance between the two modalities of 20%), we would expect the following percentages to be observed:

Table 1 Potential differential sensitivity of FFDM and DBT
Anticipated percentage distribution of cancers identified by each modality

Detected by DBT	Detected by FFDM		
	No	Yes	Total
No	0	5	5
Yes	15	80	95
Total	15	85	100

With a 5% significance level and 2-sided testing, to have 90% power to detect the above difference (5% missed by DBT and 15% missed by FFDM) as significant, requires at least 38 cancers with discordant findings(43). Thus, 190 cancers (38/0.2) are needed in the subgroup. As stated above, the smallest subgroup is likely to be approximately 15% of the total, therefore a total of 1,267 cancer cases is 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 will have at least 90% power for any subgroup comprising at least 15% of the total study and 80% for any subgroup comprising at least 11%. We anticipate that the difference between FFDM and FFDM plus

DBT would be larger than that between FFDM and DBT alone, and therefore these comparisons would also be sufficiently powered.

***Specificity*:** It might be reasonable to anticipate that the specificity of FFDM would be 93% and that DBT might improve this to 97%. Assuming 90% agreement between the two modalities, we would have the following table for negative assessment outcomes.

Table 2 Potential differential specificity Anticipated percentage of non-cancers ruled out by each modality			
Ruled out by DBT	Ruled out by FFDM		
	No	Yes	Total
No	0	3	3
Yes	7	90	97
Total	7	93	100

For 90% power to detect this as significant, we require 62 discordant negative cases in any given subgroup, i.e. 620 negative cases in total in any given subgroup. Since the subgroups of interest are all anticipated to be at least 15% of the total study size, we expect 1,050 (15% of 7,000) subjects in each subgroup of whom 861 (82%) will be negative. Thus, there will be > 90% power for the postulated difference in specificity. Again, larger differences between FFDM and FFDM plus DBT would be expected, and so these are also sufficiently powered.

PROPOSED ANALYSES

In view of the matched nature of the data, analysis of binary outcomes (e.g. presence or absence of a specific feature) will be by McNemar methods and conditional logistic regression (44). Typical data for such analysis can be tabulated as follows:

Table 1 Cancers diagnosed at assessment tabulated by detection method			
Detected by DBT	Detected by FFDM		
	No	Yes	Total
No	a	b	a+b
Yes	c	d	c+d
Total	a+c	b+d	a+b+c+d

Cancer cases diagnosed after completion of assessment tests

The value a represents the cancers which were not seen by either FFDM or DBT (for example, seen only on ultrasound), b the tumours seen only on FFDM, not on DBT, c the tumours seen only on DBT, and d the tumours seen on both. The formal comparison of

sensitivity of the two imaging modalities depends only on the discordant observations, b and c. If both are equally sensitive, b and c will be approximately equal, i.e. 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. Both the McNemar and the logistic regression inference depend on the difference between these two discordant totals, b and c. Design and sample size aids are available in text form (43)and from statistical software (45). A similar comparison of discordant totals among the subjects with a non-cancer outcome of the complete assessment episode will be made to assess the significance of the difference in specificities.

Results for binary variables will be presented as comparisons of sensitivities and specificities. Analysis of ordinal variables such as scores of suspicion of malignancy will include trend tests in conditional logistic regression and calculation of ROC curves. Statistical analysis will be performed using Stata Version 10.0(45).

Cancers missed at assessment arising as interval cancers

It is possible that, after follow-up of those negatively assessed, a subsequent number of cancers were diagnosed, and after complete review it was decided that a number, say e, of these constituted cancers which were truly present at the assessment and were missed by the full combination of diagnostic techniques applied. These cases would be added to the number a in Table 1 since they would, by definition, have been detected by neither FFDM nor DBT. However, they would ***not*** affect the comparison of sensitivities of the two imaging modalities, as this comparison only depends on b and c. Thus, the inference on any difference in sensitivities can be performed without the additional expense of flagging. This was a successful strategy in the CADET II study of computer aided detection in mammography (46).

The estimated sensitivity of DBT from Table 1 above would be

$$S_1 = \frac{(c+d)}{(a+b+c+d)}$$

For FFDM, the estimated sensitivity would be

$$S_2 = \frac{(b+d)}{(a+b+c+d)}$$

These will be slight overestimates as the number *e* should be added to the denominator. However, the underestimation is likely to be very small. Although cancers can occur in those with negative assessments, those which were judged on audit to have been missed at the

assessment were rare even when core biopsy was rarely practised (47). The statistical comparison of the sensitivities is unaffected by the presence or absence of the number *e*. The outcomes from digital mammography and DBT will be compared to the gold standard of the final histopathological verification of the presence of benign or malignant disease. If a woman is returned to routine screening this will be deemed a normal case.

ETHICS

Ethical approval will be sought from using the Integrated Research Application System (IRAS). The primary ethical issues in the trial are the additional radiation dose involved for two-view DBT and the possibility of increased cancer detection during the retrospective reading study.

Radiation dose

All the radiation dose for DBT is additional to normal procedures and some of the dose from the FFDM imaging may be additional if local protocol is to take fewer initial images at assessment (some of this dose may be offset by not having to acquire magnification images).

Procedure	Estimated mean glandular dose for typical breast (50 to 60mm thick)	Diagnostic reference level (DRL)*
Two-view FFDM	3 mGy	7 mGy
Two-view DBT	4 mGy	Not available
Total study dose	7mGy	Not applicable

*The DRL is 3.5 mGy for one oblique view; this has been doubled for two-views

The additional lifetime risk of inducing a breast cancer due to a single two-view mammography examination is estimated to be approximately 1 in 20,000 between the ages of 50 and 70 (48)(49). For this trial protocol, the total mean glandular dose is estimated as 7 mGy, 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 be received during normal assessment procedures (estimated at 1.5 to 3 mGy depending on local practice), therefore the additional dose ranges from 4 to 5.5 mGy. The total dose for the trial falls just within the DRL for standard two-view FFDM. Some of the trial participants will be women aged 40-49 years with a family history of breast cancer who are attending annual surveillance mammography (6). 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 (48). In the trial, the total dose including DBT will be approximately 7 mGy. Overall, the additional radiation dose involved is very low and within the range currently accepted for routine screening examinations.

Additional cancers detected

Approximately 18% of women recalled to assessment clinics for further investigation after routine screening have a cancer diagnosis. However, 8% of cancers subsequently diagnosed as interval cancers are cases that have been missed at the assessment clinic. Some of these interval cancers may be identified during the retrospective reading study. These cases will be referred back to the original screening centre for independent re-evaluation. These cases will also be recorded as adverse events.

QUALITY CONTROL AND QUALITY ASSURANCE

The study will be conducted in accordance with the current approved trial protocol, ICH GCP, relevant regulations and trial specific standard operating procedures (SOPs).

DBT QUALITY CONTROL

The DBT system to be used in the study will be tested prior to the start of the trial by the National Coordinating Centre for the Physics of Mammography (NCCPM) to ensure that the 2D imaging performance meets the minimum standards required by the NHSBSP and to establish baseline 2D and 3D performance (50). Each system will also be tested by physicists on installation, prior to clinical use, and every six months for the duration of the trial. Physicists from NCCPM will work with the local physics service at each participating centre to establish QC procedures and act as the coordinating centre for data collection. The 2D performance of the systems will be measured according to current NHSBSP guidelines (50). A specific protocol for testing 3D performance will also be provided by NCCPM. This will include imaging of a routine QC DBT phantom at each centre by radiographic staff on a weekly basis. These images will be reviewed centrally at NCCPM.

PROCEDURE FOR RECORDING AND REPORTING ADVERSE EVENTS

Any additional cancer detected during the retrospective reading study is an expected occurrence but will still be recorded as an adverse event.

A serious adverse event (SAE) is defined as an untoward occurrence that:

(a) results in death, (b) is life-threatening,(c) requires hospitalisation or prolongation of existing hospitalisation,(d) results in persistent or significant disability or incapacity,

(e) consists of a congenital anomaly or birth defect, or (f) is otherwise considered medically significant by the PI.
Any SAE experienced by a study participant will be reported to the main REC using the NRES report of serious adverse event form, v3. (See guidance: http://www.nres.npsa.nhs.uk/applications/after-ethical-review/safetyreports/safety-reports-for-all-other-research/#safetynonCTIMPreportingSAEs) where, in the opinion of the CI, the event was 'related i.e. resulted from administration of the study intervention; and 'unexpected' i.e. the type of event is not listed in the protocol as an expected occurrence.

DATA HANDLING

Images will be stored locally and/or in each centre's PACS system as part of the screening centre's records. Copies of the FFDM and DBT image files for each participant will be pseudoanonymised prior to transfer to the Trial office.

Archiving will begin immediately after the end of the trial or following the processing of all the images collected for research, whichever is the later.

With participants' consent, anonymised data and images collected during the study will be transferred to researchers outwith the EEA to improve the technology used in this study to analyse mammographic images.

Consent forms will be archived in the ISF at each centre. The trial dataset of FFDM and DBT image files will be archived by the Trial office at the end of the retrospective reading study and retained for 15 years on a secure network server at University of Aberdeen. The Chief Investigator will act as custodian for the trial data.

Trial data will be stored securely in locked cabinets and/or password protected computers and only accessible by trial staff and authorised personnel. Personal data will be regarded as strictly confidential and will be handled and stored in accordance with the Data Protection Act (1998). In electronic study databases, trial participants will be identified by a Study ID number.

FINANCIAL SUPPORT

The TOMMY trial is funded by a grant (09/22/182) from the National Institute of Health Research Health Technology Assessment Programme part of the Department of Health (www.hta.nhs.uk).

ETHICAL CONDUCT OF THE STUDY

The study will be conducted in compliance with the principles of the Declaration of Helsinki (1964;2008 Amendment), Good Clinical Practice (ICH Guidelines for Good Clinical Practice (CPMP/ICH/135/95, 1996), and the applicable regulatory requirements of the Research Governance Framework (2nd edition).

The University of Aberdeen will act as Sponsor for the study and will be responsible for the governance of the trial.

TRIAL MANAGEMENT GROUP

The trial will be coordinated by a Trial Management Group (TMG) consisting of the grantholders, the trial manager and data manager, and local data managers. The TMG meetings will be conducted by teleconference on a monthly basis to monitor progress and identify and address any key problems in the conduct of the study.

TRIAL CO-ORDINATION OFFICE

The Trial Office based in University of Cambridge will be responsible for the day to day management of the trial under the supervision of Professor Fiona Gilbert. The Trial Office will provide support to each site and will be responsible for collection of data in collaboration with the local data managers. Publication and dissemination of the study results at conferences and in peer-reviewed journals will be coordinated by the Trial Office in collaboration with the CI and PIs.

TRIAL STEERING COMMITTEE/ DATA MONITORING COMMITTEE

A Trial Steering Committee/Data Monitoring Committee (TMC/DMC) independent of the TMG, will be established to oversee the conduct and progress of the trial.

TRIAL MONITORING

The trial manager will visit each participating centre prior to the start of the study and during the course of the study be to ensure that the trial is conducted, documented and reported in compliance with the protocol, GCP and the applicable regulatory requirements Trial data will be evaluated for compliance with the trial protocol and accuracy in relation to source documents.

INSPECTION OF RECORDS

PIs and institutions involved in the study will permit trial related monitoring, audits, REC review, and regulatory inspection(s). In the event of an audit, the PI agrees to allow the Sponsor, representatives of the Sponsor or regulatory authorities direct access to all study records and source documentation.

PUBLICATION POLICY

The results of the study will be disseminated at international conferences, in peer-reviewed journals and at the participating NHSBSP centres.

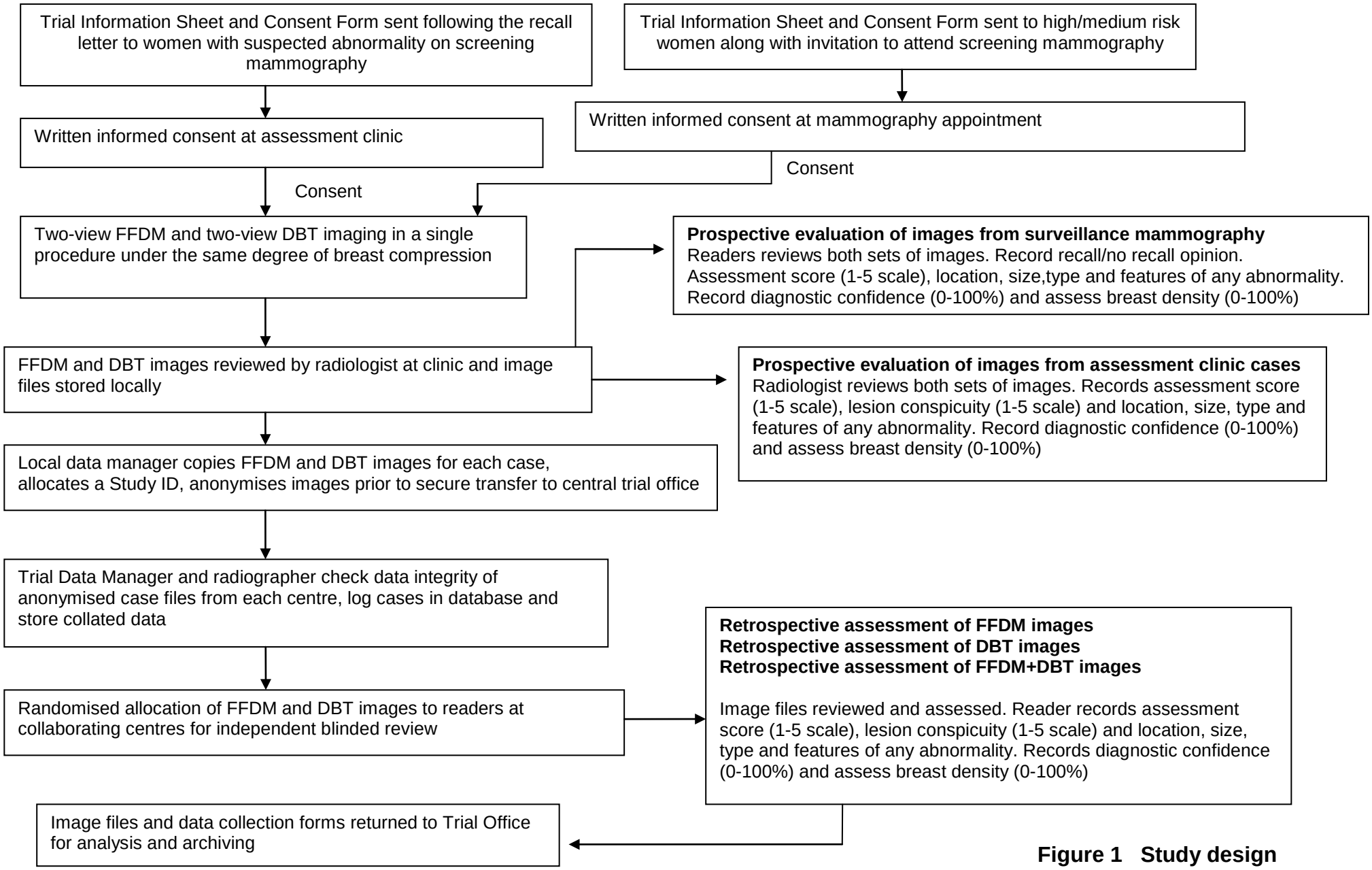

Figure 1 Study design

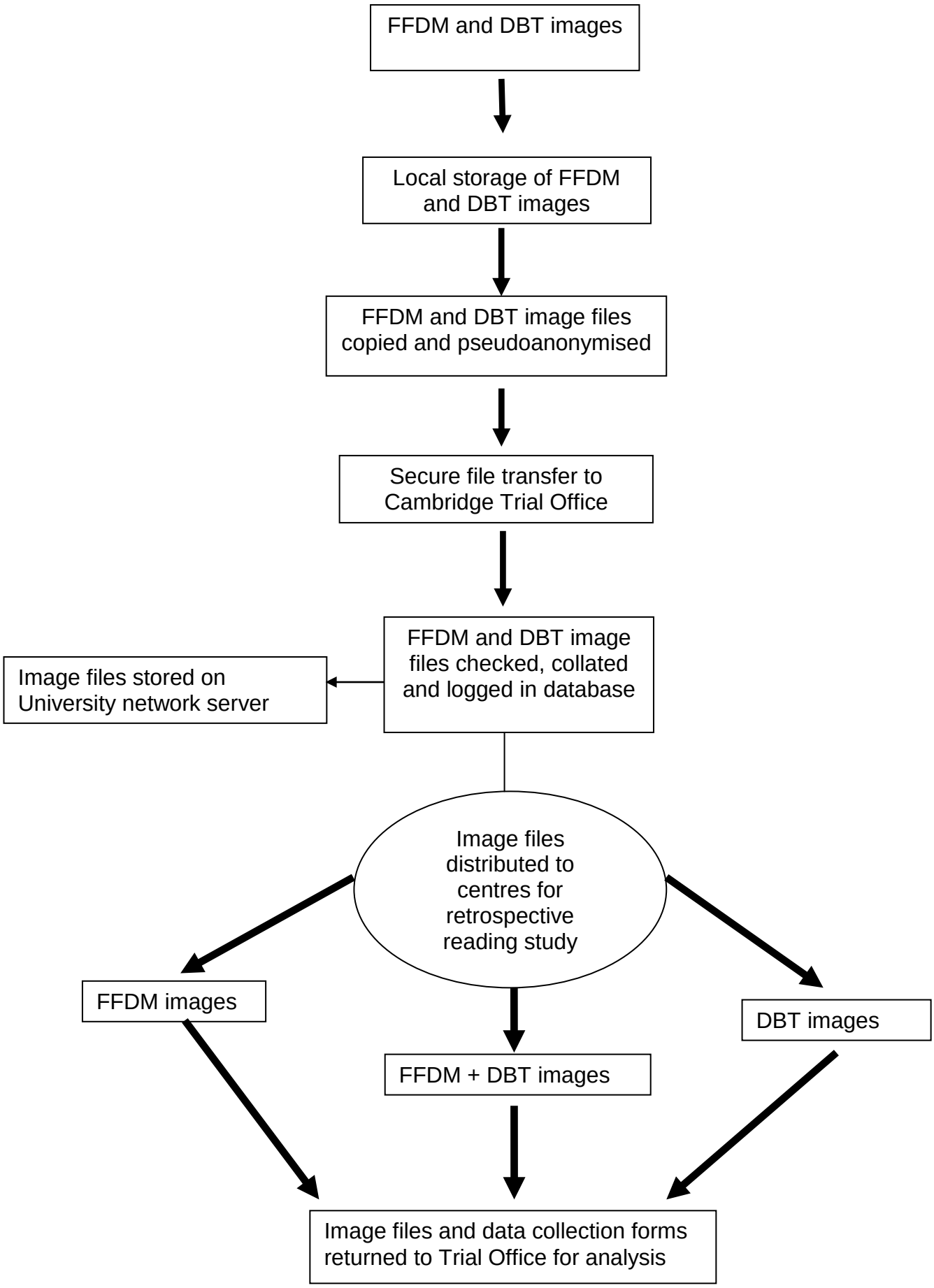

Figure 2 Data management

References

(1) Brett J, Bankhead C, Henderson B, Watson E, Austoker J. The psychological impact of mammographic screening. A systematic review. Psychooncology 2005;14 (11):917-938.

(2) Norman F. Boyd NF, Guo H, et al. Mammographic density and the risk and detection of breast cancer. NEJM 2007;356(3):227-236.

(3) Mandelson MT, Oestreicher N, Porter PL, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. Journal of the National Cancer Institute 2000 ; 92(13):1081-1087.

(4) Fletcher SW, Elmore JG. Clinical practice. Mammographic screening for breast cancer. NEJM 2003;348(17):1672-1680.

(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 ; 25(1):165-175.

(6) National Collaborating Centre for Primary Care. Familial breast cancer: the classification and care of women at risk of familial breast cancer in primary, secondary and tertiary care (partial update of CG14). 2006:1-75.

(7) Dobbins JT,3rd, Godfrey DJ. Digital x-ray tomosynthesis: current state of the art and clinical potential. Physics in Medicine & Biology 2003 ;48(19):R65-106.

(8) Smith A. Full-field breast tomosynthesis. Radiology Management 2005 ;27(5):25-31.

(9) Dobbins JT. Tomosynthesis imaging:At a translational crossroads. Medical Physics 2009;36(6):1956-1967.

(10) Chan H, Wei J, Zhang Y, et al. Digital breast tomosynthesis (DBT) mammography: Effect of number of projection views on computersied mass detection using 2D and 3D approaches. RSNA 2008.

(11) McCann J., Stockton D., Goward S. Impact of false-posittive mammography on subsequent screening attendance and risk of cancer. Breast Cancer Research 2002;4: R11.

(12) Mundy L, Hiller JE. Breast tomosynthesis: a breast cancer screening tool. Australia and New Zealand Horizon Scanning Network. Prioritising Summary. 2008; Vol 21 August 2008.

(13) Berg WA. Tailored Supplemental Screening for Breast Cancer: What Now and What Next? AJR 2009 ;192(2):390-399.

(14) Gur D. Tomosynthesis: potential clinical role in breast imaging. AJR 2007;189(3):614-615.

(15) Niklason LT, Christian BT, Niklason LE, Kopans DB, Castleberry DE, Opsahl-Ong BH, et al. Digital tomosynthesis in breast imaging. Radiology 1997 205(2):399-406.

(16) Suryanarayanan S, Karellas A, Vedantham S, et al. Comparison of tomosynthesis methods used with digital mammography, Academic Radiology 2000 12;7(12):1085-1097.

(17) Wu T, Stewart A, Stanton M, et al. Tomographic mammography using a limited number of low-dose cone-beam projection images. Medical Physics 2003 30(3):365-380.

(18) Rafferty E.A., Georgian-Smith D., Kopans D. B., et al. Comparison of Full-Field Digital Tomosynthesiswith Two View Conventional Film Screen Mammography in the Prediction of Lesion Malignancy. RSNA 2002

(19) Rafferty E.A., Kopans D. B., Georgian-Smith D., et al. Comparison of Full-Field Digital Tomosynthesis and Conventional Two View Screen Film Mammography in Lesion Detection and Assessment of Lesion Conspicuity. RSNA 2003

(20) Poplack SP, Tosteson TD, Kogel CA, Nagy HM. Digital breast tomosynthesis: initial experience in 98 women with abnormal digital screening mammography. American Journal of Roentgenology 2007;189(3):616-623.

(21) Helvie M, Hadjiiski L, Zhang Y, Sahiner B, Schmitz A, Chan H. Digital breast tomosynthesis mammography:initial assessment of non-palpable microcalcifications. RSNA 2008 .

(22) Förnvik 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 Radiologica 2010 51(3) 240-7

(23) Tingberg A, Andersson I, Ikeda DM, Ruschin M, Svahn T, Timberg P editors. BIRADS classification of breast tomosynthesis compared to mammography and ultrasonography. Proceedings IWDM 2008 5116:67-73.

(24) Gennaro G, Toledano A, di Maggio C, et al. Digital breast tomosynthesis versus digital mammography: a clinical performance study. European Radiology 2009.

(25) Svahn T., Andersson I., Chakraborty D., et al. The Diagnostic Accuracy of Breast Tomosynthesis: A Free-Response Observer Performance Study. RSNA2009.

(26) Rafferty E.A., Smith A. P., Niklason L. Assesssing Radiologist Performance in Dense Versus Fatty Breasts using Combined Full-Field Digital Mammography and Breast Tomosynthesis Compared to Full-Field Digital Mammography Alone. RSNA 2009

(27) Kontos D, Bakic PR, Carton A, Tet al. Parenchymal Texture Analysis in Digital Breast Tomosynthesis for Breast Cancer Risk Estimation: A Preliminary Study. Academic Radiology 2009 ;16(3):283-298.

(28) Kontos D., Ikejimba L., Bakic PR et al. Digital Breast Tomosynthesis Parenchymal Texture Analysis for Breast Cancer Estimation: Results from a Screening Trial. IWDM 2008

(29) Gur D, Abrams GS, Chough DM, et al. Digital breast tomosynthesis: observer performance study. AJR. 2009 ;193(2):586-591.

(30) Moore R, Kopans DB, Rafferty EA, et al. Initial callback rates for conventional and digital breast tomosynthesis mammography in the screening setting. RSNA 2007.

(31) Rafferty E.A.,Kopans D. B., Georgian-Smith D. et al. Evaluation of Call-Back Rate for Screening Mammography using Full-Field Digital Tomosynthesis Versus Conventional Film Screen Mammography. RSNA 2003.

(32) Lo J., Baker J., Orman J., Mertelmeier T., Singh S. Breast Tomosynthesis:Initial Clinical Experience with 100 Human Subjects. RSNA 2006

(33) Good WF, Abrams GS, Catullo VJ, et al. Digital breast tomosynthesis: a pilot observer study. American Journal of Roentgenology 2008;190:865-869.

(34) Teertstra HJ, Lo CE, van den Bosch MA,et al. Breast tomosynthesis in clinical practice: initial results. European Radiology 2010;20:16-24.

(35) Kopans D, Moore R. Digital Breast Tomosynthesis (DBT) NCI 3000 Women Trial. RSNA 2009

(36) Andersson I, Ikeda DM, Zackrisson S, Ruschin M, 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. European Radiology 2008;18(12):2817-2825.

(37) Rafferty E.A., Niklason L., Halpern E., Sumkin J., et al. Assessing Radiologist Performance using Combined Full-Field Digital Mammography and Breast Tomosynthesis Versus Full-Field Digital Mammography Alone: Results of a Multi-Centre, Multi-Reader Trial. RSNA 2007

(38) Smith AP, Rafferty EA, Niklason L. Clinical Performance of Breast Tomosynthesis as a Function of Radiologist Experience Level. Lecture Notes in Computer Science 2008; 5116:61-66.

(39) Rafferty EA. Inter-reader variability: Comparing breast tomosynthesis combined with FFDM to FFDM alone. ECR 2009.

(40) Gennaro G., Toledano A., Baldan E., Bezzon E.,et al. Inter-Reader Variability: Tomosynthesis Versus Digital Mammography. RSNA 2008

(41) Kopans DB. Basic physics and doubts about relationship between mammographically determined tissue density and breast cancer risk. Radiology 2008;246(2):348-353.

(42) Aitken Z, McCormack VA, Highnam RP, Martin L, 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 Epidemiology Biomarkers & Prevention 2010;19 (2):418-428.

(43) Machin D, Campbell M, Fayers P, Pinol A. Sample size Tables for Clinical Studies. Oxford: Blackwell; 1997.

(44) Clayton D, Hills M. Statistical Models in Epidemiology. Oxford: Oxford University Press; 1993.

(45) StataCorp. Stata Statistical Software Release 10.0 College Station TX: Stata Corporation.

(46) Gilbert FJ, Astley SM, Gillan MG, Agbaje OF, et al. Single reading with computer-aided detection for screening mammography. NEJM 2008; 359(16):1675-1684.

(47) Warren R, Allgood P, Hunnan G, Godward S, Duffy SW. An audit of breast assessment procedures in women who later develop cancer after a negative result. Journal of Medical Screening 2004;11:180-186.

(48) Law J, Faulkner K, Young KC. Risk factors for induction of breast cancer by x−rays and their implications for breast screening. British Journal of Radiology 2007;80:261-266.

(49) Young KC, Faulkner K, Wall B, Muirhead C. Review of radiation risk in breast screening. February 2003; NHSBSP Publication No 54.

(50) Workman A, Castallano I, Kulama L, C.P., Marshall N, Young KC. Commissioning and routine testing of full field digital mammography systems. 2009; NHSBSP Equipment Report 0604 version 3.