Breast cancer risk associated with changes in mammographic density

Emma Atakpa

Centre for Cancer Prevention
Wolfson Institute of Preventive Medicine
Barts and The London School of Medicine and Dentistry
Queen Mary University of London

Submitted in partial fulfilment of the requirements of the Degree of Doctor of Philosophy

August 2019

Statement of originality

I, Emma Atakpa, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature:

Date: 09/08/19

Details of collaboration and publications

This thesis was undertaken at the Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, by the thesis author (Emma Atakpa) under the supervision of Professor Jack Cuzick and Dr Adam Brentnall.

- Collaboration with colleagues at the University of Manchester and the University of Cambridge for Chapter 2. Michelle Harvie and Anthony Howell provided the data, Ruth Warren and Caroline Boggis assessed area-based density measures, Susan Astley developed the Stepwedge volumetric density measurement tool, and D. Gareth Evans provided clinical guidance. Collaborators also provided feedback on the analysis.

- Collaboration with the Kaiser Permanente Washington Health Research Institute and Breast Cancer Surveillance Consortium. Diana Buist and Erin Bowles provided the data and gave feedback on the analysis.

- Collaboration with the Cochrane Breast Cancer Group. Melina Willson assisted with the systematic review editorial process, Anne Parkhill ran the searches, Mangesh Thorat provided clinical guidance, and the paper reviewers, Richard Riley, Peggy Devine, Gretchen Gierach and a clinical expert who wished to remain anonymous, provided feedback on the study protocol.

- Collaboration with the International Breast Cancer Intervention Study-II. The bio-specimen assistants and bio-specimen managers assisted with the collection and processing of mammograms for the trial, Ivana Sestak ran the Stata code, Linda Metaxa visually assessed density measures, and the Trial Steering Committee approved the analysis plans.

Chapter 4 was partially published as:

Atakpa EC, Thorat MA, Cuzick J, Brentnall AR. Mammographic density, endocrine therapy and breast cancer risk: a prognostic and predictive biomarker review. Cochrane Database of Systematic Reviews 2018, Issue 8. Art. No.: CD013091. DOI: 10.1002/14651858.CD013091.

Funding

This work was supported by Cancer Research UK [grant number: C569/A16891].

Acknowledgements

First and foremost, I would like to express my sincere appreciation and gratitude to my supervisors, Professor Jack Cuzick and Dr Adam Brentnall, for their continual support, guidance, encouragement and patience throughout my PhD. I have been privileged to work with them and to learn from their immense knowledge and wisdom. I am forever grateful for everything that they have done for me and for the PhD.

I would also like to thank Professor Louise Jones and Dr Tamara Suaris who examined my 9 and 18 month progressions and whose valuable advice and suggestions helped to shape the PhD.

I am grateful to the staff at the Centre for Cancer Prevention and the Wolfson Institute of Preventive Medicine who have helped me to grow on both a professional and personal level. I am especially grateful to my fellow PhD students and ground floor colleagues for sharing this journey with me and for providing great company, conversation and friendship along the way.

Thank you also to everyone involved in the Lifestyle study, the Kaiser Permanente Washington Breast Cancer Surveillance Consortium, the Cochrane Breast Cancer Group and the International Breast Cancer Intervention Study-II who allowed their data to be used for this thesis. A particular thank you to the women enrolled on these studies, without whom none of this would be possible.

Furthermore, I would like to thank Cancer Research UK for supporting me financially with the PhD studentship. The work they do is amazing, and I am honoured and proud to have been a part of it.

Last, but not least, a huge thank you to my family and friends who have been so supportive and patient over these past four years. My wonderful mum, dad, partner, sister, grandparents and friends have shown me an incredible amount of love and support that has been the driving force for me to keep going and to see the PhD through until the end. A very special thank you to my incredible mum and partner, Tom, for reading the thesis, and more generally, for always being there for me, supporting me through the ups and downs and for keeping a smile on my face.

Thank you.

Abstract

Breast cancer is the most common cancer in the UK, and mammographic density ('density') is one of its strongest known risk factors. At present, most research focuses on static measures of density to determine population effects. The central hypothesis of this thesis is that repeated measures of density are more valuable for personalised breast cancer prevention. This hypothesis was tested through the following research.

Study-I investigated within-women associations between body mass index (BMI) and density, to assess whether density (visual/Cumulus/volumetric 'Stepwedge') acts as a mediator for breast cancer risk reduction during a premenopausal weight-loss intervention (n=65). Study-II evaluated the benefit of using a woman's longitudinal history of (BI-RADS) density to improve breast cancer risk estimation (n=132,439). Study-III was a Cochrane systematic review investigating the association between endocrine therapy-induced density reduction and breast cancer risk and mortality. Studies-IV and V (n=575) evaluated visually-assessed density reduction with prophylactic anastrozole during the International Breast Cancer Intervention Study-II, and its use as a biomarker for concurrent breast cancer risk reduction, respectively.

In Study-I, change in BMI was associated with change in breast fat but not dense tissue, negating density reduction as a biomarker for risk reduction with weight-loss. In Study-II, longitudinal density provided approximately a quarter more statistical information than most recent density and improved discriminatory accuracy. Study-III found evidence that density reduction may be a biomarker for reduction in risk and mortality with tamoxifen, but the level of evidence was limited by some study quality issues. Study-IV indicated that preventive anastrozole might marginally reduce density, but statistical significance was not obtained. In Study-V, sample size was too small to draw definitive conclusions.

Overall, changes in density were useful for the study of breast cancer risk and should be considered for personalised breast cancer prevention strategies.

Table of Contents

List of Figures

List of Tables

Commonly used abbreviations

AIs	Aromatase Inhibitors
BCSC	Breast Cancer Surveillance Consortium
BI-RADS	Breast Imaging Reporting and Data System
BMI	Body Mass Index
CC	Cranio-Caudal
DA	Dense Area
DCIS	Ductal Carcinoma In Situ
DV	Dense Volume
ER+	Oestrogen Receptor-Positive
ER-	Oestrogen Receptor- Negative
FA	Fat Area
FFDM	Full-Field Digital Mammography
FV	Fat Volume
HRT	Hormone Replacement Therapy
IBIS-I	International Breast Cancer Intervention Study-I
IBIS-II	International Breast Cancer Intervention Study-II
MLO	Medio-Lateral Oblique
MRI	Magnetic Resonance Imaging
PDA	Percent Dense Area
PDV	Percent Dense Volume
SERMs	Selective Oestrogen Receptor Modulators
TA	Total Area
TV	Total Volume
VAS	Visual Assessment Score

Chapter 1: Thesis background, rationale and aims

1.1 Thesis background and rationale

1.1.1 Introduction to breast cancer

Breast cancer is the most common cancer in the UK, with approximately 55,000 new cases of invasive breast cancer diagnosed every year (1). An estimated 1 in 7 women in the UK will be diagnosed with the disease in their lifetime (based on multiple primary incidences) and around 11,000 women will die from breast cancer in the UK each year; that's 31 women every day (1). Worldwide, breast cancer is the second most common cancer, with an estimated 2.1 million new cases and 600,000 deaths recorded in 2018 (2).

There is higher incidence of the disease in the Western world, which is thought to be a result of lifestyle factors such as delayed age at first full term birth, nulliparity, use of exogenous hormones and obesity (3-8). Implementation of routine mammographic screening programmes has also been linked to an increase in annual breast cancer incidence due to greater and earlier detection of tumours (6-8). Furthermore, early detection programmes allow treatment to be administered at an early stage of progression when it is most effective, thereby improving survival and reducing mortality as a result of the disease (6-10). Other factors, such as wider use of chemotherapy and adjuvant hormone therapy, have also contributed to reduced breast cancer mortality rates (6-8, 11).

1.1.2 Anatomy of the breast

The breast is part of the female reproductive system, and its main function is to produce and secrete milk. It consists of fatty adipose tissue, stroma (fibrous connective tissue), lobules (mammary glands), milk ducts, the lymphatic network, the nipple, the areola, blood vessels and Copper's ligaments. Individual structures, known as terminal ductal lobular units (TDLUs), consist of a lobule and an extra-lobular duct connected to a larger milk duct. Milk production occurs in TDLUs, which is then transported through the milk ducts to the nipple. TDLUs are also the site of hormonal exchange, hence their number and size changes in response to hormone fluctuations during different reproductive stages such as menstruation, pregnancy, lactation and menopause. Most breast cancers develop in the epithelial cells lining TDLUs (12). Cancers developing in the lobules are known as lobular breast cancers, and cancers developing in the extra-lobular ducts are known as ductal breast cancers.

1.1.3 Principles of mammography

In the UK, women aged 50-70yr (or 47-73yr in some areas) are invited to attend a mammography screening appointment every three years. At their first screen, mammograms are used to assess the presence of an already existent prevalent breast cancer, and at their subsequent screens, mammograms are used to detect incident breast cancers that have developed since the last screening appointment. Breast cancers which develop or are detected in between screens are known as interval breast cancers, which usually present as a result of symptoms such as a breast lump, skin dimpling or thickening, breast or nipple pain, and nipple retraction or discharge. Diagnostic mammograms are used to image the breast after a screening mammogram with suspicious results or as a result of the patient reporting symptoms that the physician believes warrants further investigation.

The mammogram examination is an x-ray of the breast. The breast is placed between two compression paddles and an x-ray beam is used to penetrate the breast, which is then detected at the other side of the breast by a receptor (film-screen or digital detector). During screening mammography, women undergo two-view assessment: the cranio-caudal (CC) view and the medio-lateral oblique (MLO) view. These two complementary views capture the breast from different angles to image as much tissue as possible.

Various imaging factors affect the quality of the image produced during a mammogram. The optimal mammogram is a trade-off between factors such as contrast (x-ray tube kilovoltage peak, kVp), exposure (milliampere-seconds, mAs) and breast compression. A lower level of kilovoltage peak produces a higher contrast (desirable), and a lower level of exposure produces a lower radiation dose (desirable) but grainy images (undesirable). However, a lower kilovoltage peak does not penetrate through thick or dense breast tissue, and requires higher exposure to produce the necessary dose for penetration through the breast to the receptor below. Therefore, exposure and dose generally increase with increased breast size and density. Breast compression can be described by the force (Newton, N) or pressure (kiloPascal, kPa); the latter being a measure of the force divided by the contact area (13). Contact area depends on the breast size and deformation of breast tissue under pressure. A higher breast compression leads to a smaller breast thickness (millimetres, mm) and increased spread of tissue which reduces the required dose whilst producing a higher quality image. These factors work in tandem and can be largely dependent on the patient and radiographer's technique at the time of mammography examination. However, most (digital) mammography machines nowadays use calibration techniques to automate the process and determine the optimal factors for each examination, negating the need for radiographers to subjectively determine the imaging factors.

Mammography can either be film-screen or full-field digital mammography (FFDM). FFDM generates a 'raw' ('for processing') mammogram whereby the pixel values are linearly related to the exposure. These raw mammograms then undergo image processing to make the mammogram visible to the naked eye, which varies depending on the mammography machine's manufacturer. The applied algorithms are usually unspecified and irreversible, so the appearance of 'for presentation' mammograms can differ from one mammography machine to another.

1.1.4 Mammographic density – literature review

1.1.4.1 Mammographic density

Mammographic density is also referred to as 'density' or 'breast density'.

Broadly, the breast has two main components when assessed via mammography: dense and fatty tissue. The dense tissue are stroma (fibrous connective tissue) and duct epithelium (parenchyma or glandular tissue) within the breast; collectively known as the fibroglandular tissue. This fibroglandular tissue appears as white, radio-dense material on an x-ray, whereas the fatty adipose tissue is dark and translucent.

The degree of density involvement within the breast can be described in a variety of ways. It is often defined as the area of fibroglandular tissue, known as the absolute dense area (DA); or as the percent dense area (PDA), which is a measure of the DA over the total breast area (TA). This TA is comprised of the DA and the fatty adipose tissue within the breast, also known as the absolute non-dense or fat area (FA). With the introduction of volumetric methods, density has also recently been described in terms of absolute dense volume (DV), percent dense volume (PDV), total breast volume (TV) and non-dense or fat volume (FV). Measurement of these area-based and volumetric density measures are described in detail in the following section.

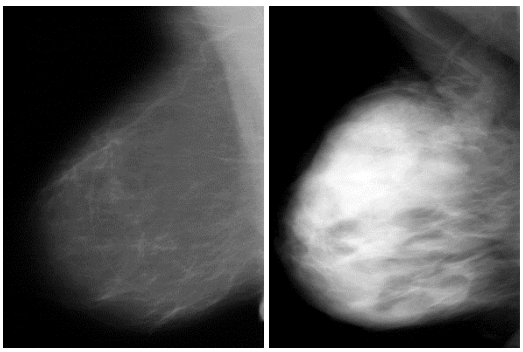


Figure 1.1 Mammograms depicting low (left) and high (right) density. Images from the International Breast Cancer Intervention Study-I.

1.1.4.2 Measurement of mammographic density

The first reference to density as a risk factor for breast cancer was made by Dr John Wolfe in 1969 (14) who categorised density based on variations in the appearance of mammographic parenchymal patterns as well as the approximate proportion of the breast occupied by these patterns. These 'Wolfe grades' were described as: N1 (predominantly fat), P1 (ductal prominence in <25% of the breast), P2 (ductal prominence in >25% of the breast) and DY (extensive dysplasia). Wolfe observed that women with extensive dysplasia (DY) had an incidence of breast cancer that was 22 times higher than those with predominantly fatty breasts (N1).

A similar method of assessment, known as the Tabar classification, was developed by Gram et al. in 1997 (15). This measurement of density is categorised into: I (scalloped contours and Cooper's ligaments, evenly scattered terminal ductal lobular units, 1-2 mm nodular densities and oval-shaped lucent areas corresponding to fatty replacement), II (complete fatty replacement), III (retro-areolar prominent duct pattern due to fatty involution), IV (extensive nodular and linear densities) and V (homogenous ground glass like, structure-less fibrosis with convex contour). The major difference between Wolfe and Tabar categorisations is the closer representation of premenopausal density seen with the Tabar grades. Since premenopausal women have denser breasts than postmenopausal women, Tabar I is thought to represent the high density patterns commonly observed in premenopausal women, whilst Tabar II, III, IV and V would represent Wolfe grades N1, P1, P2 and DY, respectively (16).

Another assessment of density, the Breast Imaging Reporting and Data System (BI-RADS) lexicon, was first proposed by the American College of Radiology (ACR) in 1993 (17). In the third edition of the lexicon, BI-RADS density was defined qualitatively: BI-RADS I (predominantly fat), BI-RADS II (scattered fibroglandular density), BI-RADS III (heterogeneously dense) and BI-RADS IV (extremely dense), but in the fourth edition, quantitative descriptions of PDA were also included: BI-RADS I (<25%, predominantly fat), BI-RADS II (25%-50%, scattered fibroglandular density), BI-RADS III (50%-75%, heterogeneously dense) and BI-RADS IV (>75%, extremely dense) (18). These have since been dropped in the fifth edition which is again purely qualitative: BI-RADS I (predominantly fat), BI-RADS II (scattered fibroglandular density), BI-RADS III (heterogeneously dense breasts which may obscure small masses) and BI-RADS IV (extremely dense breasts which lower the sensitivity of mammography) (17).

Since Wolfe, Tabar and BI-RADS grades assess density with a maximum of five categories; substantial risk information could be lost due to the grouping of density. An alternative measurement method, known as the visual assessment score (VAS), was therefore developed to

aid in the assessment of PDA on a continuous and quantitative scale. Visual assessment scoring ranks PDA from 0% to 100%; although PDA is sometimes also described in 5% incremental scores, creating a semi-continuous 21 point scale (19).

Although the methods outlined so far are thought to do well in describing density, a major limitation is their subjective nature; basing measurements on a radiologist's visual interpretation of density. Only moderate inter-observer (kappa score=0.54) and intra-observer (kappa score=0.71) agreements have been reported with BI-RADS assessment (20), and studies on visual assessment have reported differences as high as 30% and 35% for intra- and inter-observer scores, respectively (21). This means that adequate reader training is essential in order to reduce the heterogeneity of density scores assessed subjectively (22).

As well as purely visual assessments, computer-assisted methods of density measurement have also been introduced. In 1994, a computer-assisted thresholding technique, known as 'Cumulus', was introduced to improve on the subjective nature of scoring (23). Cumulus is an interactive thresholding technique which requires its users to set pixel thresholds on a digitised (23) or digital (24, 25) mammogram. The user sets an initial grey-level threshold, i_{edge}, to separate the breast edge from the background, for which any pixel with grey-level higher than this threshold is classed as the breast and any pixel with a lower grey-level is considered background. Another threshold, i_{DY}, is then determined to separate the dense tissue (pixels higher than i_{DY}) from the fatty tissue (pixels lower than i_{DY}) within the breast. The user can also mask out the pectoral muscle (particularly visible in MLO views). Cumulus then sums the number of pixels categorised as breast, dense and non-dense tissue to provide information on compartmentalised breast composition i.e. TA, DA and FA, respectively. PDA can then be calculated by dividing the DA over the TA. Before Cumulus, a similar method of assessment called 'planimetry' was used to provide information on separate areas of breast tissue. This worked by tracing around the breast edge and regions of dense tissue on an acetate overlay and measuring TA and DA using an outlining tool (26). However, this method is somewhat cumbersome and requires greater user involvement than Cumulus, so is rarely used.

The semi-automated, semi-subjective and quantitative nature of Cumulus provides a density score that has shown strong associations with breast cancer risk (23, 27, 28); produces comprehensive information about separate breast tissue components; and reports high levels of agreement within and between adequately trained observers (intraclass correlation coefficient between observers >0.9 in Byng et al. (23)). In addition to the continuous Cumulus score, Boyd et al. introduced a categorisation of Cumulus, known as the Boyd classification, which sees density grouped into: none, <10%, 10 to <25%, 25 to <50%, 50 to <75% and ≥75% (27). Other

fully-automated or semi-automated area-based segmentation methods exist, including AutoDensity (29), fuzzy c-means-based methods (30) and an ImageJ-based method (31).

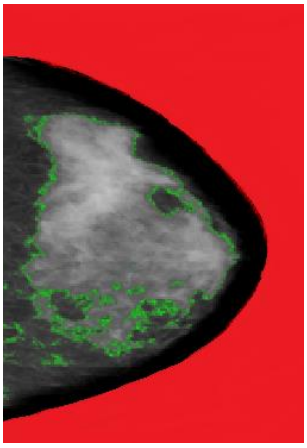


Figure 1.2 Cumulus thresholds (23) applied to a mammogram. Image from the International Breast Cancer Intervention Study-I.

Red pixels indicate the image background (excluded from area calculations); green pixels indicate the outline of dense tissue such that pixels inside the green boundary are dense tissue and pixels outside of the green boundary are non-dense breast fat.

The methods introduced so far have their limitations. Manual (or semi-manual) density readings would be too labour-intensive and time-consuming to be incorporated into the UK National Health Service (NHS) breast screening programme (NHSBSP), which requires high-throughput mammography for approximately two million women per year (32); and subjective (or semi-subjective) density readings are only practical with sufficient reader training (22). Incorporation of these density methods into the NHSBSP would place a time and cost burden on health services in order to cover extra staffing and training costs. Furthermore, the insufficient reproducibility and large intra- and inter-reader variability that occur with subjective density measures (21, 22) might make these methods of assessment unreliable in guiding clinical decisions about a woman's healthcare. In addition to this, these measurement methods are based on a 2-dimensional projection of the breast, which undoubtedly loses information regarding the anatomical breast structure. Even with sufficient compression, superimposition can occur during mammography, creating an overlap of structures which may distort area-based interpretations.

With the aim of resolving many of these issues, various volumetric methods based on two-class tissue models and FFDM have now been developed. These volumetric methods are intended to give a more realistic representation of dense tissue in the breast and provide objective, automated measures of absolute fibroglandular and non-dense tissue (33-36). The breast is assumed to contain only 2 mediums, dense and fatty tissue, which have separate x-ray

attenuation coefficients. Greyscale levels at each pixel in the mammogram are modelled as a function of these x-ray attenuations and the initial x-ray beam. This function is then used to produce estimates of the different amounts of dense and fatty tissue within each 'stack' of breast tissue at each pixel. Summing over the pixels in the whole breast gives the total volume of breast tissue (33). The most widely used and commercially available volumetric software is Volpara (36), but alternative approaches also exist (37-43). Volpara is a development of the standard mammographic form (SMF) (33) that produces estimates of TV, DV, FV and PDV.

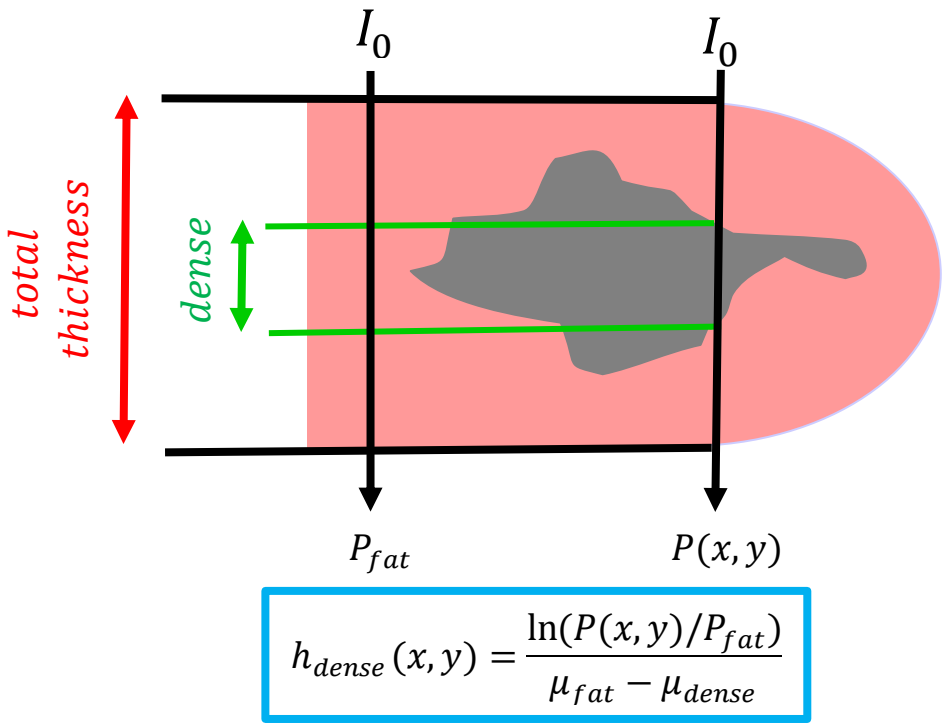


Figure 1.3: Representation of Volpara volumetric density physics model to estimate dense volume.

Total breast thickness at $P(x,y)$ (i.e. pixel value of pixel (x,y)) is made up of the thickness of dense and fatty tissue. The thickness of dense tissue at $P(x,y)$ is $h_{dense}(x,y)$. The energy penetrating the breast is I_0, and the energy reaching the detector is assumed to be linearly related to $P(x,y)$. The pixel value of pure fat, P_{fat}, is determined by finding the pixel value with the least x-ray attenuation (using an iterative approach to find the fatty, uncompressed breast edge). The linear x-ray attenuation coefficients for dense and fatty tissue using a particular filter, tube voltage and breast thickness are μ_{dense} and μ_{fat}, respectively. Values for $P(x,y)$ and P_{fat} are measured on the 'raw' mammogram, and μ_{dense} and μ_{fat} are assumed from reported data. Summing values of $h_{dense}(x,y)$ over all pixels of the breast gives an estimate of the volume of dense tissue in the breast. Derived from the physics model based on (44).

An advantage of volumetric methods is that they offer high levels of agreement with magnetic resonance imaging (MRI)-assessed breast density (often seen as the ground truth for volumetric tissue distributions) (40, 44, 45). A study by Gilhuijs et al. found Pearson correlations of 0.93,

0.97 and 0.85 for PDV, TV and DV, respectively, when comparing Volpara-assessed density to density measured using segmentation techniques in MRI scans (45). With such good representation of breast tissue and high reproducibility (46), it was hypothesised that volumetric density would also improve breast cancer risk prediction. However, studies have indicated a somewhat similar performance to area-based methods (41, 47-51).

Some researchers have also hypothesised that density may be acting as a risk factor not purely through the amount of fibroglandular tissue it encompasses, but also through its distribution and pattern complexity (52, 53). This idea has led to a focus on structural components and spatial distributions ('texture features') of mammographic images, to assess whether they could have an effect on breast cancer risk, independent of the amount of density (52, 54-62). Some studies suggest that these textural features could add to or even outweigh the predictive power of density (52, 54-56, 61, 62), but these are less well validated than volumetric or visual density.

Even though most measures of density involve assessment of mammograms, the breast may be assessed by other modalities. Density can be read from computed tomography (CT) (63-65), ultrasound (66, 67), MRI (68, 69) and dual energy x-ray (70) images. A relatively new approach to breast imaging comes in the form of digital breast tomosynthesis (DBT). DBT is an x-ray imaging modality which produces pseudo-3D images of the breast, using multiple low-dose exposures taken at different angles through 30° and reconstructed to produce image 'slices' through the breast (71). This provides a novel approach to density assessment, allowing cross-sectional visualization through the breast to reduce the effect of superimposition of overlapping tissue. Since DBT is still relatively new, only a few studies have looked into its use in density assessment. Methods of density assessment using DBT include the application of Volpara (72) or the fully-automated volumetric software, Quantra, (73) to raw projections; a Cumulus-like threshold to reconstructed slices (74); Cumulus to a central projection (55); and integral curves (75), maximum entropy (75) or BI-RADS (73, 75, 76) to a set of projections. Density can then be calculated as the average density from all raw projections (72, 73, 75) or slices (74), or it can be calculated as the total number of voxels identified as dense over the total number of voxels in the DBT 3D reconstruction (74). Cumulus-assessed percent density from DBT (central projection, mean of all projections, or mean of reconstructed slices) has shown high correlation with Cumulus-assessed percent density from 2D mammography, with Pearson correlations ranging between 0.76 and 0.97 (77). Similarly, Volpara-assessed density from DBT (one projection) has shown high correlation with Volpara-assessed density from 2D mammography (Pearson correlation 0.903) (77).

Table 1.1: Overview of density measurement techniques.

Wolfe	Tabár	BI-RADS	VAS	Planimetry	Semi-automated area-based e.g. Cumulus	Boyd	Fully-automated volumetric e.g. Volpara	Techniques applied to Tomosynthesis images
Subjective	Subjective	Subjective	Subjective	Subjective	Semi-subjective	Semi-subjective	Objective	Semi-subjective/ Objective
Qualitative and quantitative *categorised*	Qualitative *categorised*	Qualitative *categorised*	Quantitative *continuous*	Quantitative *continuous*	Quantitative *continuous*	Quantitative *categorised*	Quantitative *continuous*	Quantitative *continuous*
2D	2D	2D	2D	2D	2D	2D	Volumetric	Volumetric/ Pseudo-3D
Manual	Manual	Manual	Manual	Manual	Semi-automated	Semi-automated	Automatic	Semi-automated/ Automatic

Favourable density measurement technique characteristics:

- *Objective–consistent and reproducible scores, intra- and inter-reader variability eliminated, does not require user training, ideal for use in clinical and screening environments.*
- *Quantitative–larger gradients in risk than qualitative measures (78); fine scale and continuous description of density makes it easier to distinguish small differences in density (28).*
- *Volumetric–more realistic representation of the 3D breast structure, takes into account the thickness of the breast.*
- *Automatic–fast results, ideal for use in clinical and screening environments, not very labour or time-intensive.*

1.1.4.3 Mammographic density as a breast cancer risk factor

Mammographic density is one of the strongest known independent breast cancer risk factors; women with mostly dense breasts are at a 4 to 6-fold increased risk relative to women with fatty breasts (79). The density-risk association is seen with both qualitative and quantitative density measurements, however many studies show a better risk prediction for quantitative density (78-81). In a meta-analysis of over 14,000 cases and 226,000 controls from 42 studies, McCormack and dos Santos Silva reported a pooled relative risk (RR) estimate of 3.98 (95% CI, 2.53 to 6.27) for Wolfe grade DY relative to N1; 4.08 (95% CI, 2.96 to 5.63) for extremely dense relative to fatty BI-RADS scores; and 4.64 (95% CI, 3.64 to 5.91) for VAS PDA ≥75% relative to <5% (79). As for computerised methods, a study by Boyd et al. showed a fairly high relative risk of 4.04 (95% CI, 2.12 to 7.69) for Cumulus ≥75% density relative to no density (27).

Risk associations with volumetric modalities have been somewhat mixed (41, 47-51, 82-84), with some studies suggesting a superior performance over area-based methods (82-84), but many reporting similar risk associations (41, 47-51). In a recent case-control study, the screen-detected cancer odds ratios (ORs) for the highest quintile percent density (relative to the lowest quintile percent density) were 2.42 (95% CI, 1.56 to 3.78), 2.12 (95% CI, 1.30 to 3.45) and 2.17 (95% CI, 1.41 to 3.33) for Volpara, Cumulus and the fully-automated area-based measure, Densitas, respectively. VAS was the strongest predictor with an OR of 4.37 (95% CI, 2.72 to 7.03), whilst the fully-automated volumetric method, Quantra, had no significant association with risk, with an OR of 1.02 (95% CI, 0.67 to 1.54) (50). Another recent study reported similar risk associations for BI-RADS and Volpara (fractioned into categories analogous to BI-RADS categories) (85). It has also been suggested that a combination measure of Volpara DV and BI-RADS may improve breast cancer risk estimation beyond using only one of the measures (86).

As well as looking into the effects of qualitative and quantitative density measures on breast cancer risk, many studies have investigated whether the chosen description of density can vary the extent of risk. Both percent and absolute density are strongly and positively associated with breast cancer risk (87-90), but most literature suggests that PDA has stronger risk associations than DA (88-90).

Other examinations of density as a risk factor for breast cancer suggest that risk is not specific to breast side or location of the eventual cancer (89), and risk associations are similar for both MLO and CC mammographic views and right or left side (91). Of note, even though risk profiles are similar between views, PDA does tend to differ between views, with CC views regularly reporting higher PDA estimates than MLO views (92). The reason for this is quite intuitive. Subcutaneous fat is more visible in MLO views than CC views, which leads to larger

TA values in MLO images, and subsequently lower PDA (93). It has been suggested, therefore, that both MLO and CC views should be used for density assessment, which is supported by evidence of a more consistent risk score when assessing density from two view mammography (93).

With the advent of FFDM, differences in risk estimates may also exist depending on the type of mammogram used for density assessment. Density tends to appear darker and therefore less dense in FFDM images compared with film images (94), which can be partly explained by better recognition of the skin line on digital mammograms (94). There is also an increased variability in density measured using FFDM on different machine types. Each mammography machine uses its own processing algorithm to display mammograms 'for presentation', which can affect the perceived relationship between the input x-ray and received image signal (35). Volumetric measures use raw, pre-processed mammograms in their estimation of density (35), because these images represent the x-ray attenuation of the breast tissue directly. However, raw, processed and film images have different appearances, and adjustments may be required to calibrate their density-breast cancer risk associations.

1.1.4.4 Mammographic density in screening

There is some concern regarding density in screening; mainly due to the masking effect caused by dense breasts. Since dense tissue has a similar attenuation coefficient to many types of tumours, high density can cause an inability to detect breast cancers (95), resulting in higher levels of missed prevalent cancers at first screen in women with dense breasts. Because of this masking effect, many advocacy groups in the US, such as the 'Are You Dense?' campaign (96), have been arguing for access to information regarding density scores. In 2009, the first breast density legislation was subsequently passed in Connecticut, mandating the disclosure of BI-RADS density to screened women (97). As of July 2019, a total of 38 US states now require some level of breast density notification (98). The ACR also addressed the issue of masking in their fifth edition BI-RADS lexicon. This latest fifth edition lexicon now includes guidance on grading breasts with high density behind the nipple as BI-RADS III or IV (99). This emphasis on masking effects has, however, also caused some concern that healthcare practitioners may be more likely to grade women as having dense breasts (100), perhaps as a safeguard for missing prevalent cancers at the time of screening.

It was previously thought that this masking effect could introduce bias into the density-risk relationship (101). Masking bias arises from lower mammographic sensitivity in dense breasts compared with fatty breasts, creating higher rates of false negative screens. Prevalent, but missed, cancers might be detected after a negative screen, creating a high level of interval

cancers and low level of screen-detected cancers (95, 102, 103) (as shown by the 20-30% stronger density-risk association seen in studies of incident cancer (cancer diagnosed after a negative screen) compared with prevalent cancer (density assessment and diagnosis occurring at the same time) in McCormack and dos Santos Silva's meta-analysis (79)). Supporting evidence for this masking effect comes from Boyd et al. (95) who found an increased probability of interval cancer detection within 12 months after a negative screen (OR=17.8, 95% CI, 4.8 to 65.9) for women with >75% visually-assessed density relative to women with <10% visually-assessed density, suggesting that these cancers were likely already prevalent, but missed, at screening. It has also been hypothesised that masking bias could cause an overestimation of incident cancers in cohort studies, and an underestimation in case-control studies (22). The reason for the overestimation in cohort studies would be due to missed prevalent cancers in dense breasts at study entry being revealed during follow-up, whilst the underestimation in case-control studies is thought to be due to the error in categorisation of women with dense breasts. Prevalent cancers in dense breasts could be missed and therefore misplaced as 'healthy', hence underrepresenting dense breasts amongst cases and over-representing dense breasts amongst controls. This would create an artificially low relationship between high density and the subsequent cancers detected. However, another study by Boyd et al. also reported that both cohort and case-control studies carried similar risk associations (22), which would disprove the idea that masking bias was an influence on these study designs. In addition to this, studies by Byrne et al. and Rebolj et al. have found that the risk effect of density exists for at least 10 years post mammogram examination (104, 105). Harvey et al. estimated that if masking bias did have a substantial effect on the density-risk relationship, about 75% of prevalent cancers in dense breasts of women in Byrne et al.'s study would have been missed and subsequently diagnosed 10 or more years later, which is a highly unlikely scenario (106). It is therefore now commonly accepted that the effect of density on breast cancer risk is not a result of masking (79).

A further issue regarding density and screening is the debate surrounding the treatment of women who present dense breasts at screening. Due to the masking effect, dense breasts tend to decrease both sensitivity and specificity of breast cancer detection during mammography (107, 108). This effect is especially apparent in younger women (for example, <50 years old) who are likely to have denser breasts than older screening ages (108). Sensitivity has been shown to improve with digital mammography in pre- and perimenopausal women <50 years old with dense breasts (109, 110) but with film mammography in postmenopausal women >65 years old with fatty breasts (109). Specificity is somewhat different between digital and film mammography (109), if not slightly improved with film mammography in younger women (aged 40-49) (110). The lower levels of sensitivity and specificity seen in dense breasts assessed by mammography have led many to advocate the need for stratified surveillance dependent on density.

One proposed method of stratified screening in women with dense breasts is the use of supplemental or adjunct imaging. These may be particularly useful in young, high-risk or symptomatic women requiring frequent examinations throughout their lifetime. Whilst mammography in these women would increase lifetime exposure to x-ray ionisation, the use of MRI or ultrasound (which do not use ionising radiation) would lower this lifetime x-ray exposure. Adjunct automated whole-breast ultrasound (111-115) and MRI (113, 114, 116) have the potential to improve sensitivity compared with mammography alone in women with dense breasts. However, compared with mammography alone, specificity has been shown to fall with these modalities, which could increase the number of false positives (111, 114). Supplemental screening using ultrasound or MRI in women with dense breasts given the 'all-clear' after a negative mammogram has also been suggested (117). A simulation study by Sprague et al. (118) found that supplemental screening ultrasound after a negative mammogram in women with BI-RADS III or IV would avert 0.36 additional breast cancer deaths and gain 1.7 quality-adjusted life years per 1000 women. However, they also found that there would be an additional 354 false-positive biopsy recommendations per 1000 women with supplemental ultrasound screening. With an estimated cost-benefit ratio of $325,000 per quality-adjusted life years gained, this study suggested that supplemental ultrasound screening for women with heterogeneously or extremely dense breasts could greatly increase costs while producing relatively small benefits in comparison.

Another increasingly popular imaging modality is DBT. Although relatively new, this imaging modality is progressively replacing FFDM in the US and it is quickly gathering more interest in other countries. DBT has been shown to lower the effects of masking and provide better imaging for the detection of tumours within dense breasts (119-121). It is also possible to capture a DBT image in the same compression as FFDM imaging with little added dose and little extra resource (117). Since MRI is a relatively expensive modality requiring a large amount of user training, and both MRI and ultrasound still produce relatively high false positive rates in comparison to tomosynthesis (122), DBT could be a prime candidate for adjunct or supplemental screening, particularly since its U.S. Food and Drug Administration approval in 2011. However, adjunct screening tomosynthesis may increase the time required to interpret a mammographic examination (123); but one could argue that with the increased specificity of tomosynthesis (120), this additional reading time would be balanced by the reduced number of non-cancers recalled for diagnostic tests. It is also possible to produce a synthesised 2D image with DBT that replicates projections captured with FFDM. Studies of DBT + synthesised 2D mammography have shown similar results to DBT + FFDM (124); and with no additional x-ray dose than DBT alone, DBT + synthesised 2D mammography could negate the need for both FFDM and DBT.

Whilst supplemental screening in women with dense breasts is promising, recent evidence suggests that this implementation may be too premature (125). For supplemental screening to be beneficial, it must improve sensitivity as well as reducing the number of interval cancers, advanced-stage disease or breast cancer-specific mortality (125). However, so far, studies have shown only a small reduction in breast cancer deaths (118) and many have been underpowered to show an effect on women with advanced-stage disease (126). There is also limited evidence for the effects of supplementary screening on interval cancers in women with dense breasts. A large randomised control trial (RCT), DENSE, is currently underway to assess whether supplementary MRI can reduce rates of interval cancer amongst women with dense breasts (127).

Not only does high density increase the risk of an incorrect 'all-clear' mammographic examination, but density has also been shown to increase the risk of more aggressive (107, 128, 129) or larger (128, 130) cancers due to its masking effect. In the UK, women between 50-70 years (or 47-73 years in some areas) are currently invited for screening every three years regardless of their level of density (131). However, shorter screening intervals in women with dense breasts may be more beneficial. More frequent screening in women with dense breasts could increase the likelihood of catching rapidly progressing cancers and existent cancers missed by masking on previous screens, at an earlier stage. One must note, however, that basing surveillance frequency on density alone would ignore women who are at an increased risk of more severe tumours but who do not necessarily have dense breasts. It may therefore be better to base stratified surveillance on risk assessment from established risk models rather than density alone, to ensure that those at the highest risk of breast cancer are given frequent examinations, even if they perhaps have low density. This idea has already been implemented in some healthcare services, such as the NHS, who now offer more frequent surveillance to women at a high risk of breast cancer based on family history (132, 133).

Whilst stratified screening could potentially improve efficiency and provide more targeted surveillance for the most at-risk women (131, 134), there are still questions surrounding the cost-effectiveness of supplemental screening as well as the advantages of notifying women of their density (135). For example, a study by Hooley et al. found that only 45% of women in Connecticut referred for supplemental ultrasound screening actually received it (136). This raises concerns over the inconsistency of implementing breast density legislation. It also creates a problem whereby women may be informed of their dense breasts but are not offered suitable treatment to decrease their risk or to reduce the effects of masking. Whether this lack of follow-through is due to personal choice, inefficiencies in healthcare services or possibly insufficient medical insurance to cover further examinations, appropriate care should be available to women

with dense breasts if they are to be notified of their high density and increased risk of incidence and masking.

1.1.4.5 Mammographic density and breast cancer risk factors

Mammographic density is associated with many other breast cancer risk factors including age, body mass index and reproductive factors such as parity and menopause. These relationships are discussed in more detail below.

1.1.4.5.1 Age

Density decreases with age (137-141). Adjustments are therefore necessary in order to counteract the negative confounding of age on the density-risk relationship (27, 104). If no adjustment is made for age, the effect of density on risk will be underestimated (142). This creates a contradiction between density, age and breast cancer risk, since age and density are positively associated with risk, but inversely associated with each other. To help to understand this inconsistency, one could consider the cumulative rate of 'breast tissue aging' (i.e. cumulative rate of exposure to hormones) rather than chronological age, as suggested in Pike's model (143, 144). Risk from density might reflect the breast tissue response to lifetime exposure of reproductive hormones (such as oestrogen) and growth factors (such as insulin-like growth factor (IGF-I) or prolactin) which stimulate epithelial and stromal cell division in the breast (78, 145-149). According to Pike's model, the rate of breast tissue aging is most rapid at the time of menarche, slows with each pregnancy, slows further in the perimenopausal period, and is lowest after the menopause. This implies that an earlier age at menarche, nulliparity, later age at first birth and later age at menopause will increase cumulative exposure to hormones. Later menarche, parity, earlier age at first birth and earlier age at menopause are suggested to decrease cumulative exposure to hormones (27, 143, 144). It has been hypothesised that the higher the cumulative exposure to hormones, the higher the density. Hormonal exposures in early life might therefore be the most important predictors in the development of density, since this stage in life sees the highest rates of breast tissue aging.

Furthermore, the density and breast cancer risk association can be seen in both younger and older women (27, 79, 104). However, one must bear in mind that density estimates in women younger than the screening age may not be fully applicable to general populations of women. Since these younger women do not undergo routine mammography, density measures taken from this age group may be skewed by the potentially symptomatic or high-risk populations of women examined.

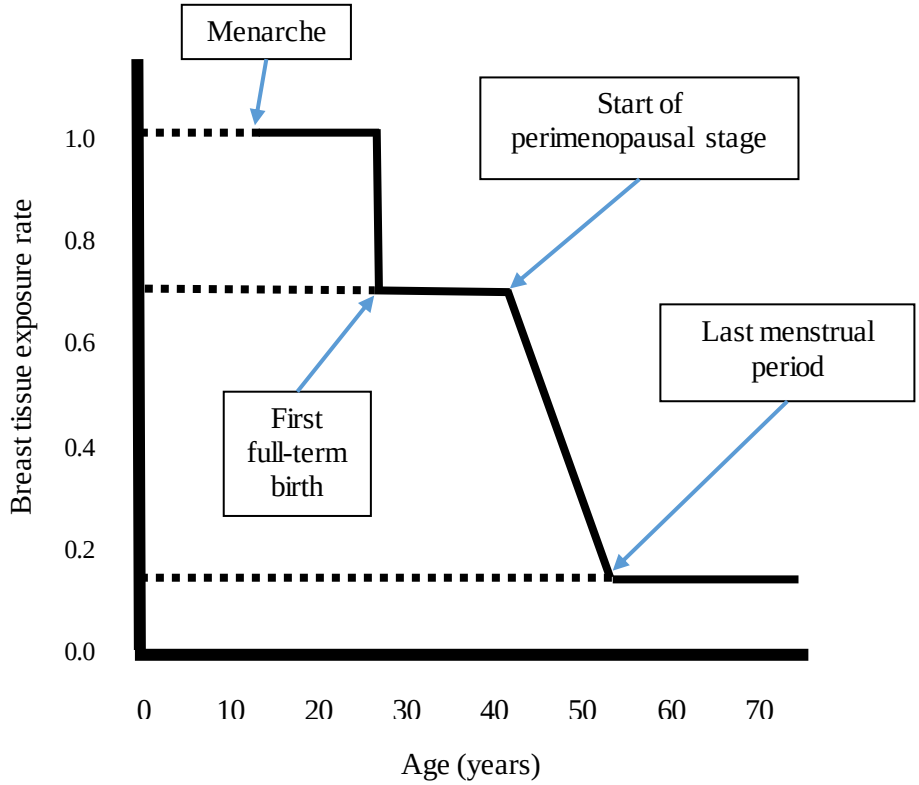


Figure 1.4: Pike's model showing rates of breast tissue aging with chronological age.

The rate of breast tissue aging is greatest after menarche, declines with successive pregnancies and in the perimenopausal period, and is lowest after the last menstrual period i.e. post-menopause. This model is used as a theory to explain the increasing incidence rate of breast cancer with increasing age. Derived from Boyd et al. (78) which uses data from Pike et al. (143) and Rosner and Colditz (144).

1.1.4.5.2 Menarche

There is only a small amount of literature regarding the association between age at menarche and breast density, but according to Pike's model, one would expect there to be a correlation between early menarche and both increased density and increased breast cancer risk. There is some evidence to suggest that density is higher in women with early menarche (150), but this is not always the case (138). On the other hand, breast cancer risk increases with earlier menarche, but the effects are only marginal (151).

1.1.4.5.3 Parity

Density reduces with a first full-term pregnancy (152, 153), and reduces even further with each subsequent pregnancy (81, 153). This is thought to be due to the lower levels of reproductive hormones circulating post-pregnancy. Lower density is therefore associated with parity (81, 138, 152, 153) and earlier age at first full-term birth (152, 153).

1.1.4.5.4 Menopause

Density reductions occur over the menopause (140, 154). This effect is thought to reflect the decrease in circulating reproductive hormones and increase in breast tissue involution that occur at this stage of female reproduction (155). Density is also positively associated with age at menopause, with lower densities existing in women who begin menopause at an earlier age (154). The density and breast cancer risk association is not limited to a particular menopausal status, with both premenopausal and postmenopausal women seeing higher risk with increasing density (104).

1.1.4.5.5 Body Mass Index

As well as age, systemic adiposity (commonly measured as body mass index (BMI)) is one of the strongest confounders of density (138, 156). Percent density, whether measured as PDA or PDV, is negatively associated with BMI (88, 157-165); as women with higher BMI are more likely to have higher non-dense tissue and total breast tissue (157, 166, 167) which will lead to lower percentage estimates of density. Absolute dense area has a less consistent relationship with BMI (87, 157-160, 168, 169). DV, on the other hand, has shown positive associations with BMI (161-165). These relationships are further complicated by the fact that density and excessive postmenopausal BMI are both positively associated with risk of breast cancer (170-173), but (percent) density reduces with increasing BMI (88, 157-165). Relationships involving (percent) density as a risk factor therefore require an adjustment for BMI, otherwise breast cancer risk will be underestimated (174).

Some studies have also suggested a protective effect of BMI on breast cancer risk in premenopausal women (172, 175, 176). However, others have argued that these contrary findings are a result of negative confounding by density (166, 169).

To understand the contradictory relationships between adiposity, density and breast cancer risk, it helps to first understand the biological mechanisms behind the positive effect of BMI on breast cancer risk in postmenopausal women. This can largely be explained by the high levels of oestrogen present in overweight or obese postmenopausal women, as a result of the aromatase enzyme converting androgens to oestrogen in peripheral adipose tissue. This process of aromatisation acts as the main source of oestrogen in postmenopausal women whose hormonal production in the ovaries has ceased (177). Elevated levels of oestrogen act as a risk factor by binding to oestrogen receptor (ER)-positive tumour cells and stimulating their growth and proliferation (177).

1.1.4.5.6 Endogenous hormones

Circulating endogenous oestrogen levels have been shown to influence the growth of density (146, 147, 178), making density a potential mediator for the effects of reproductive hormones on breast cancer risk. It has been theorised that density is a reflection of the breast tissue response to lifetime hormone exposure, as outlined in Pike's model (143, 144). According to this theory, variations in density would mirror different levels of cumulative hormonal stimulation. However, this relationship between systemic hormones and density might not be applicable to all women, since little to no association between density and blood serum oestrogen has been seen in postmenopausal women (179-181). Some studies have investigated the role of local environments surrounding density, suggesting that certain hormones produced locally at these sites, may be stimulating the proliferation of epithelial cells. Most evidence suggests that a relationship exists between circulating oestrogens and density in premenopausal women, but not in postmenopausal women, with local breast tissue perhaps acting as the main oestrogen source in postmenopausal women (180-182).

Other hormones that are known to influence density include IGF-I and prolactin. High levels of serum IGF-I in premenopausal women and prolactin in postmenopausal women have shown significantly positive associations with density (145, 183, 184). Recent evidence to support this suggests that the breast cancer risk associations of plasma prolactin and mammographic density are independent in premenopausal women (185). It has also been suggested that consideration of both density and endogenous hormones (such as prolactin, circulating testosterone and estrone sulphate) may add to current breast cancer risk prediction models (185, 186).

1.1.4.5.7 Exogenous hormones

Hormone replacement therapy (HRT) increases both risk of breast cancer (142, 187, 188) and density (189-194). Specifically, combined oestrogen and progesterone HRT has greater associations with density than oestrogen only HRT (192-194), and continuous use of combined HRT is also associated with higher density than cycled HRT use (191, 194). However, the effects of HRT are only short-term, with a decrease in density visible just 4 weeks after HRT cessation (195, 196) and a decrease in risk (to the level of a non-HRT user) is apparent within a few years of stopping treatment (197).

Whilst HRT increases both risk of breast cancer and density, selective oestrogen receptor modulators (SERMs) have been shown to decrease risk (198-202), and certain SERMs, such as tamoxifen, can also reduce density (203-206). A study from the International Breast Cancer Intervention Study-I (IBIS-I) found that visually-assessed density reductions of 10% or more after 12-18 months of tamoxifen treatment were associated with an approximately 63%

reduction in breast cancer risk (OR=0.37, 95% CI, 0.20 to 0.69), p=0.002) compared with placebo, whereas smaller reductions or increases in density on tamoxifen had the same association with risk as placebo (OR=1.13, 95% CI, 0.72 to 1.77), p=0.60) (19). This suggests that change in density could be used as a biomarker to measure the efficacy of tamoxifen for prevention. Aromatase inhibitors (AIs), such as anastrozole, can also be used to treat and prevent breast cancer (207, 208). However, the effect of AIs on density is less clear, and it is yet unknown whether change in density can also be used as a biomarker for response to treatment with these drugs.

1.1.4.5.8 Heritability

Family history and heritability can also influence mammographic density. Twin studies have shown a 60% correlation between Cumulus-assessed PDA in monozygotic twins compared with 30% in dizygotic twins (209), and findings suggest that heritability can explain around two thirds of the residual variance seen in Cumulus-assessed PDA (209).

There has also been interest in the links between women with *BRCA1/2* mutations and density (210, 211). Weak evidence suggests that *BRCA1/2* carriers have higher density that is lower in contrast and coarser than low-risk women without these genetic mutations (210). However, studies also show similar relative risks of developing breast cancer for high density relative to low density amongst carriers of the *BRCA1/2* mutation and non-carriers (211), proposing no effect modification by genetic mutations.

In addition to this, a series of genome-wide association studies (GWAS) have so far identified at least 100 Single Nucleotide Polymorphisms (SNPs) that are thought to be associated with breast cancer susceptibility (212, 213). Each of these SNPs has been shown to slightly modify an individual woman's risk, but together, they could provide a significant amount of information regarding a woman's risk. Attention has therefore turned towards these SNPs and investigations are on-going to ascertain their effect on density (212, 214-220). Candidate SNPs suggested to have an association between both breast cancer risk and density include SNPs located within the *HSD17B1*, *CYP1B1* and *COMT* oestrogen-related genes (218, 220), *rs6220* (*IGF-1*) (217), *rs3817198* (*LSP1*) (214), *rs13281615* (*8q*) (214) and *rs10509168* (*ZNF365*) (216), but the effects of these latter SNPs on density require validation. A particular SNP, *rs10995190* (*ZNF365*), has shown significance in more than one study (215, 216), suggesting a promising locus for further genetic evaluation. However, not all studies report an association between density and those SNPs identified in GWAS (219), and some studies suggest that independent information can be gained from the two risk factors (221); hence the relationship between SNPs, density and breast cancer risk remains an active area of research.

1.1.4.6 Mammographic density and breast cancer risk models

Various breast cancer risk models aim to distinguish between women at different levels of risk. These include the Gail (222), Tyrer-Cuzick (223), BCSC (224), BOADICEA (225), Claus (226) and Ford (227) models. The Tyrer-Cuzick model is used in the UK and US, whilst the Gail model is more commonly used in the US. The Gail model includes age, age at menarche, age at first full-term birth, family history, number of biopsies, presence of atypical hyperplasia and ethnicity. A major limitation of the Gail model, however, is that it includes only first-degree relatives, which can result in underestimation of risk in women with a familial risk of cancer from the paternal side, for example. The Tyrer-Cuzick model expands on this by incorporating extensive family history information. The Tyrer-Cuzick model also includes age, age at menarche, BMI, age at first full-term birth, menopausal status, age at menopause, benign breast disease, presence of atypical hyperplasia, HRT use and length, Ashkenazi Jewish heritage, and genetic mutations in the *BRCA1/2* genes, as well as a lower-penetrance 'unknown' *BRCAX* gene which may increase susceptibility to hereditary breast cancer, and is used to account for residual familial clustering.

The Gail model has shown good calibration between predicted and observed numbers of breast cancer (228), but there is some evidence that including more risk factors provides better discriminatory accuracy (229-234).

Breast cancer risk models are continually being updated, and attempts to include density into the models have shown promising results (235-244). One early study investigating the inclusion of density in a breast cancer risk model was conducted by Tice et al., who compared a risk model containing age and ethnicity-adjusted BI-RADS density only, the Gail model, and a combined version of the two. The age and ethnicity-adjusted BI-RADS model was shown to perform just as well as the Gail model. However, adding density to the Gail model modestly but significantly increased the discriminatory accuracy of the Gail model (237). Further studies by Chen et al. (241) and Barlow et al. (240) have also found slight improvements in discriminatory power after adding visually-assessed (241) or BI-RADS (240) density to the Gail model, and a recent study by Rice et al. found an improvement in the Rosner-Colditz breast cancer risk model when percent density was included (245). In 2014, Warwick et al. showed that density significantly added to the Tyrer-Cuzick model in a case–control analysis of high-risk women from the placebo arm of the IBIS-I trial (242). In this study, the area under the receiver operator characteristic curve (AUC) increased from 0.51 in the Tyrer-Cuzick model to 0.62 in the model containing both Tyrer-Cuzick risk and a density residual ($p=0.002$). More recently, Brentnall et al. showed that visually-assessed density improved the Tyrer-Cuzick model (235). Using data

from the PROCAS cohort study of around 55,000 women attending screening, they found that AUC (0.57) and IQR-OR (an odds ratio for the 75th vs. 25th percentile; 1.36, 95% CI, 1.25 to 1.48) for the Tyrer-Cuzick model increased with the addition of density (AUC=0.61; IQR-OR=1.47, 95% CI, 1.33 to 1.62). In another study, Brentnall et al. also showed that more high-risk screened women can be identified when using the Tyrer-Cuzick model with BI-RADS or volumetric density than without (without: 4.8% identified, with BI-RADS: 7.1% identified, with volumetric density: 6.8% identified) (246). The Tyrer-Cuzick model now includes VAS, BI-RADS and Volpara-assessed density in its latest version of risk calculation (243, 244).

1.1.5 Changes in mammographic density – literature review

Mammographic density is a promising tool, with great potential for breast cancer prevention. However, most research has so far focused on static measures of density, giving insight into population-based relationships. Density is a dynamic phenotype, so repeated measures of density may be more informative for predicting individual breast cancer risk and for developing personalised breast cancer prevention strategies. Assessing individual women's repeated measures could help to reveal within-women relationships between density and other breast cancer risk factors, to help to understand the aetiology of breast cancer development and the interacting influences of different risk factors. It may also provide information on risk of breast cancer for individual women and hence be useful for personalised breast cancer risk estimation. Consideration of changes in density may also be useful for indicating a woman's response to breast cancer treatment, such as endocrine therapy. If reductions in risk are mirrored by reductions in density, change in density could be used as a potential biomarker for decrease in risk as a result of the drug. It is therefore hypothesised that changes in density may be of greater use in breast cancer prevention than fixed density measures.

Several studies have previously looked into the benefit of using repeated measures of density for breast cancer risk and prevention. A review of key studies is outlined below, along with further research ideas arising from the studies that formed the rationale for this thesis.

1.1.5.1 Repeated measures of mammographic density and other breast cancer risk factors (body mass index)

An important breast cancer risk factor and confounder of density is BMI. BMI is a well-established risk factor for postmenopausal women (170, 172, 173), but weight gain across premenopausal years has also been linked to an increased risk of postmenopausal breast cancer (173, 247). However, this can be reversible with short-term weight-loss through dietary (248) or surgical (249) means. For example, the Iowa Women's Health Study showed a 25-40% decrease in postmenopausal breast cancer risk in women who sustained a 5% loss of body weight

compared with women who continued to gain weight at different periods of time between 18 years of age and menopause (250). However, the effects of short-term weight-loss on density are less well understood.

There have been few studies assessing the effect of short-term weight change on density, particularly over the premenopausal years when a loss in weight is most effective. A dietary intervention study by Boyd et al. assessed the effect of a two year low-fat, high-carbohydrate diet on density, and found that women on the weight-loss intervention saw a reduction in Cumulus TA (2.4% reduction), whereas the control group had increased TA (0.3% increase), and DA decreased in the intervention group more so than the control group (6.1% reduction vs. 2.1% reduction, respectively). The reduction in dense area was particularly apparent in women who transitioned from pre- to postmenopausal or who remained premenopausal during the study (251). Other studies exploring the effect of more drastic weight-loss after bariatric surgery on premenopausal dense tissue have not shown any consistent effect of the weight-loss intervention on dense tissue (252, 253).

If weight-loss-induced reductions in risk are shown to be mediated by density, a reduction in density could act as a possible biomarker for risk reduction as a result of weight-loss and lifestyle interventions. However, with only one known study published to have previously assessed dietary-based weight-loss on density in premenopausal women, more studies are required to assess this mediating pathway, and to test this possible risk reduction biomarker.

1.1.5.2 Repeated measures of mammographic density for breast cancer risk estimation

Several studies have made use of repeated measures of density to predict breast cancer risk in populations of women attending screening (254-261). These have mainly focused on change in density between two serial mammograms and its effect on breast cancer risk. For instance, in a case-control study of 85 breast cancer cases and 85 matched controls in the Women at Risk (WAR) Columbia University study, Work et al. reported that Cumulus-assessed density between two pre-diagnostic mammograms (median 4 years apart) decreased in time with controls ($p=0.004$), but not with cases ($p=0.6$) (259). This suggested that a lack of density reduction over time may be indicative of a future risk of breast cancer. Another study from the Breast Cancer Surveillance Consortium (BCSC) tested whether changes in density between current and previous mammograms (average 3 years apart) were associated with risk of breast cancer. This study involved a large cohort of over 300,000 women screened at various US registries, with around 2600 subsequent breast cancers diagnosed during follow-up. Here, Kerlikowske et al. found that within-women changes in BI-RADS categories were associated with risk in women with previous BI-RADS categories I, II and III, but not for women with

previous BI-RADS category IV (258), suggesting a potential residual effect of high density. However, these interpretations were limited by the small number of women in the most extreme categories (for instance, only 0.1% of controls and 0.2% of cases moved from BI-RADS IV to I). Furthermore, no adjustments could be made for BMI which may have introduced negative confounding to the density-risk association.

The null effect seen in Kerlikowske et al.'s study in women with initially high density, was also reported in a study by vans Gils et al. (257). Fully computerised methods were used to measure density change over a 10 year period in over 100 postmenopausal breast cancer patients and 400 matched controls. This study found that women who started the study with high density (>25%) which decreased over time, experienced the same risk as women who had prolonged levels of high density. However, similar to Kerlikowske et al., very few women moved between the extreme density categories (only 12 women had initial density >25% which reduced to <5% during the study). Another key finding suggested that women whose density decreased from moderate (5-25%) to low (<5%), had (non-significantly) higher risk than women who had consistently low density (OR=1.9, 95% CI, 0.6 to 6.1). Compared with the consistently low group, women with consistently moderate density had an OR of 5.7 (95% CI, 2.2 to 15.2), and women whose density increased from moderate to high had an OR of 6.9 (95% CI, 2.1 to 22.9).

However, not all studies show an effect of change in serial density measurements on breast cancer risk. Longitudinal studies by Maskarinec et al. and Vachon et al. showed that changes in Cumulus percent density did not differ between women with and without breast cancer (255, 256). Nonetheless, both studies were limited by their collection of BMI information. Maskarinec et al. reported that many of their mammograms did not have corresponding BMI measurements taken at the same time as mammography, and Vachon et al. also reported differences in the timings of BMI assessments, with 17% of women having BMI data extracted over a year after their mammogram. BMI is not a static measurement and may have changed between the time of mammography and BMI assessment, potentially affecting the results.

Whilst, changes between two measures of density may have an effect on breast cancer risk, little is known as to whether repeated measures of density add information to risk estimation beyond what's already explained by a woman's current density. Only one other known study has evaluated this by assessing the predictive ability of using two density measures. Kerlikowske et al., again using data from the BCSC, found that the BCSC 5-year risk model better discriminated between cases and controls with a two-measure density predictor than with a one-measure density predictor (AUCs 0.640 vs. 0.635, respectively) (262). However, no studies have evaluated the benefit of including more than two density measures; particularly an

unlimited number of mammograms taken at arbitrary points in time, as would be seen in a screening environment.

1.1.5.3 Repeated measures of mammographic density for breast cancer risk with endocrine therapy interventions

IBIS-I was the first trial to show that change in density could reflect the beneficial effect of tamoxifen in the primary prevention of breast cancer. A nested case-control study within the trial assessed 123 breast cancer cases and 942 controls to test whether density reduction on tamoxifen was associated with risk of developing breast cancer. Cuzick et al. found that women who had at least a 10% reduction in VAS density in the first 12-18 months after the start of tamoxifen had an approximately 63% reduction in breast cancer risk compared with women on placebo, whilst women who experienced <10% density reduction on tamoxifen had no difference in risk compared with women on placebo (19). This result suggested that density change could be used as an early biomarker to assess the efficacy of prophylactic tamoxifen in order to predict a woman's response to treatment. With the help of this biomarker, healthcare practitioners may advise women who see at least a 10% density reduction after 12-18 months of treatment to continue with their 5 year course of chemoprevention, whereas those who see a more modest reduction or increase in density might not be responding to treatment and would perhaps benefit from alternatives such as lifestyle interventions or chemoprevention with other SERMs or AIs (19).

Other studies have since tested the biomarker in the adjuvant setting for breast cancer patients on endocrine therapy for treatment of the disease. Some studies have suggested that a reduction in density may be used as a biomarker for breast cancer recurrence on tamoxifen (263, 264) and AIs (264), and others have suggested its use for predicting a reduction in mortality for tamoxifen treatment (265, 266). However, there are currently no systematic reviews focussing on the evidence to suggest that mammographic density reduction in women receiving endocrine therapy is a biomarker for breast cancer outcomes such as reduction in risk, recurrence, mortality and incidence of contralateral breast cancer. A review of this sort is essential to determine the strength of certainty for this biomarker before it can be implemented into clinical practice.

There is also very little evidence for the mammographic density biomarker in women treated with AIs, and there are no known studies in women on preventive AI therapy. The IBIS-II trial showed that the AI, anastrozole, reduced the risk of ER+ breast cancer in high-risk postmenopausal women by 60% (208), and it is a good resource to test this biomarker for preventive anastrozole therapy.

Previous studies assessing the effect of AIs on density have reported only modest (and often underpowered) results (267-270). In the preventive setting, the NCIC CTG MAP.1 prevention trial of letrozole vs. placebo found that 12 and 24 month changes in Cumulus-assessed PDA were small and similar between arms (12 months: mean PDA change -1.74 on letrozole, -0.24 on placebo (adjusted p=0.61); 24 months: mean PDA change -0.01 on letrozole, -1.32 on placebo (adjusted p=0.61)) (268). Vachon et al. also found similar results in a study of over 100 postmenopausal women (adjusted mean PDA change -1.0% on letrozole vs. -0.3% on placebo (p=0.58)) (270). The NCIC CTG MAP.2 prevention trial found similar results for exemestane (mean 12 month Cumulus-assessed PDA change: 0.56 on exemestane and 0.58 on placebo (adjusted mean difference between arms p=0.96), mean 24 month PDA change: -0.17 on exemestane and -2.93 on placebo (adjusted mean difference between arms p=0.52)) (269). Studies in the adjuvant setting have shown similar results (267), but there has been some suggestion of a small effect of AIs on volumetric density with a larger sample size (271). However, there are currently no studies testing the effect of AIs on density in the preventive setting with a similarly sufficient sample size. The IBIS-II trial could be an important resource for testing the effect of preventive AIs on density with the potential to provide an adequately sized sample of women.

1.2 Aims and thesis outline

1.2.1 Repeated measures of mammographic density and other breast cancer risk factors (body mass index)

Chapter two aimed to assess the dynamic relationship between BMI and density during a dietary-based weight-loss intervention in premenopausal women to help assess whether weight-loss-induced reductions in risk are potentially mediated by reductions in density. Repeated measures data on density (visual, Cumulus and a 'Stepwedge' volumetric method) and BMI were collected over 2 years during the weight-loss intervention (Lifestyle study) in Manchester, UK (n=65). The intention of this intervention was to reduce postmenopausal breast cancer risk in susceptible premenopausal women who had gained weight since the age of 20 years through improvements in diet and exercise. Each woman's measure of BMI varied across the study as she actively lost weight, and density was measured at the same time as BMI in order to assess concurrent changes in density. These within-women associations were tested using repeated measures correlation coefficients and a linear mixed model for short-term BMI change on density.

1.2.2 Repeated measures of mammographic density for breast cancer risk estimation

Chapter three aimed to develop a longitudinal density measure that accounted for individual women's changing density values, and to assess the benefit of using this longitudinal density measure for breast cancer risk assessment. Repeated measures data on density and BMI were collected as part of the Kaiser Permanente Washington Breast Cancer Surveillance Consortium breast imaging registry, taken over a period of up to 20 years in an at-risk screening population in Washington State, USA. Substantial follow-up and linkage to cancer registries recording breast cancer incidence allowed for long-term assessment of density over time and its effect on breast cancer risk. The longitudinal density measure was developed using a linear mixed model for age and BMI on density (BI-RADS). The benefit of using longitudinal density for breast cancer risk assessment was tested using likelihood ratio statistics to assess the predictive ability of proportional-hazards Cox models for time to breast cancer diagnosis with baseline, most recent or longitudinal density. Discriminatory accuracy was also tested using a yearly at-risk concordance index.

1.2.3 Repeated measures of mammographic density for breast cancer risk with endocrine therapy interventions

Most breast cancer risk factors are difficult to change, for example: age, female sex, family history, genetics and endogenous hormone levels; or can jeopardise a woman's integrity and significant life-choices, such as reproductive events. However, density is a dynamic trait and it is modifiable. Density has also been shown to decrease in response to risk-reducing therapy by tamoxifen (19, 203). This is a promising result since an endocrine therapy-induced reduction in density that is concomitant with a reduction in risk could be used as a potential biomarker for monitoring the efficacy of risk-reducing endocrine therapies. Use of this biomarker would be more beneficial than the current "wait-and-see" approach, and because mammography is less invasive than the alternative tissue and blood sample biomarkers, change in density is a particularly appealing tool for indicating risk.

Chapter four is a Cochrane systematic review of the published evidence to suggest that endocrine therapy-induced reduction in density can be used as a biomarker to predict breast cancer risk and mortality. This biomarker may exist in both the preventive and adjuvant settings, hence both risk and mortality outcomes were considered.

Chapter five aimed to assess whether reductions in density occur with preventive anastrozole therapy. This is because it is important to first establish whether anastrozole has the potential to change density before assessing the biomarker's association with breast cancer risk. This study was nested within the IBIS-II international, double-blind, randomised placebo-controlled

prevention trial of anastrozole vs. placebo in high-risk postmenopausal women. Change in density at approximately 2 and 5 years after initiation of treatment was compared between women on anastrozole or placebo, to determine whether preventive anastrozole treatment reduces density more than the natural decline that occurs with age.

Chapter six aimed to assess whether preventive anastrozole-induced density reduction at approximately 2 years after the start of therapy is associated with a reduction in breast cancer risk. This study used the same data as Chapter 5, but with density change and treatment as predictors and breast cancer incidence as the outcome.

Chapter seven concludes the thesis with a discussion of findings and future direction for research on changes in mammographic density and breast cancer risk.

Chapter 2: The relationship between body mass index and mammographic density during a premenopausal weight-loss intervention study

2.1 Introduction

As discussed in Chapter 1, mammographic density is one of the strongest risk factors for breast cancer. Percent breast density is measured as the relative proportion of dense tissue in the breast, either in terms of area or volume depending on the measurement method. Visual assessment measures the percent density with respect to the TA; whilst automated and semi-automated methods can also measure the extent of dense and fatty tissue separately. Both DA and PDA are positively associated with risk of premenopausal (and postmenopausal) breast cancer (27, 104, 272), and absolute DV and PDV have also shown positive relationships (82, 246). Associations of FA and FV with breast cancer risk are unclear, although there is some suggestion of an inverse relationship with premenopausal breast cancer risk (246, 272).

In postmenopausal women, higher attained BMI is associated with a higher risk of breast cancer (170, 172, 173), with an estimated 40% increase in risk for every 10kg/m^2 of BMI in never users of hormone replacement therapy (172). This increase in risk is partly explained by increased aromatisation of androgens to oestrogen in peripheral adipose tissue, which promotes cell proliferation (273, 274), carcinogenesis (273, 274) and insulin resistance (275). Whilst BMI is a widely accepted risk factor for breast cancer in postmenopausal women, there may be an inverse relationship in premenopausal women (175). However, this is not always consistent (276). For example, a 5kg/m^2 increase is sometimes associated with a reduction in risk of premenopausal breast cancer risk amongst Caucasian and African women, but an increase amongst Asian women (276).

Weight gain across the premenopausal years has also been linked to an increased risk of postmenopausal breast cancer. Every 5kg of adult weight gain is associated with an approximate 10% increase in risk amongst never users (or low dose users) of hormone replacement therapy (173, 247). However, a number of studies (as summarised by Hardfeldt et al. (248)) suggest that these effects are reversible with efficient weight-loss, whereby short-term weight-loss is associated with an overall 20% breast cancer risk reduction (248). A reduction in risk of approximately 40% can also be seen with large weight-losses as a result of bariatric surgery in populations of pre- and postmenopausal women (249).

The effects of short-term weight change on breast density are less well understood, particularly those as a result of dietary weight-loss. Mammographic density is a dynamic phenotype, and has the potential to respond to short-term weight changes, making density reduction a possible

biomarker for reduction in risk as a result of weight-loss. This study aims to explore the effect of short-term dietary weight change on density using both area-based and volumetric methods in a cohort of premenopausal women, to ascertain whether the relationship between weight-loss and reduced postmenopausal breast cancer risk could, in part, be mediated by reductions in mammographic tissue.

2.2 Methods

2.2.1 Study design and participants

The Lifestyle study (277-279) was a 1 year diet and exercise weight-loss intervention study amongst 79 high-risk premenopausal women who attended annual screening within the Breast Cancer Family History clinic at the Prevent Breast Cancer research unit at the Manchester University Hospital Foundation NHS Trust between 2002 and 2004. Women were required to be aged 35-45 years, premenopausal with regular menstrual cycles, non-smokers, have a self-reported adult weight gain ≥7kg, and a sedentary lifestyle (<40 minutes moderate physical activity per week). All women had a family history of breast cancer (with lifetime risk 16–40% as assessed by the Tyrer-Cuzick model (223, 243)), but were excluded if they had a known BRCA1/2 mutation or a previous history of cancer. Women were also excluded if they were already successfully dieting or losing weight, were pregnant or planning to become pregnant over the next year, had used hormonal oral contraceptives in the last 6 months, or had psychiatric or physical co-morbidities that could affect their ability to take part in a diet and physical activity weight-loss programme. In the intervention group, 40 women were assigned to a 12 month intensive supervised weight-loss programme which involved a 25% energy-restricted Mediterranean type diet and an individualised physical activity program (150 minutes moderate intensity physical activity and 40 minutes of resistance exercise per week). A further 39 women were separately recruited to a limited intervention control group who received standard written advice about diet and physical activity but no additional support for weight-loss.

The objective of this analysis was to assess the relationship between BMI and density in a cohort of women with changing BMI measures, regardless of their method of weight-loss. Since women from both the intervention and control groups were given lifestyle advice to lose weight (although less so for the control group), all women had within-women variation in BMI. To increase power, the analysis combines both intervention arms. Additionally, to limit the effect of women contributing observations to an area-based measure or volumetric measure only, the cohort is restricted to those with both an area and volumetric density measurement at any one or more time points (n=65, 82% of the cohort).

2.2.2 Mammographic density

Mammographic films were digitised using a Kodak LS85 digitiser at a pixel size of 50μm and with 12-bits (4096 grey levels) pixel depth. Mammograms were analysed using three different methods: (1) a visual assessment score of percentage density read to the nearest 5% by two experienced readers (Dr Ruth Warren, Caroline Boggis), expressed as an average of the two scores to calculate PDA, (2) a semi-automated area-based measure based on computer-assisted thresholding (Cumulus, Sunnybrook health sciences centre, Toronto, Canada, (23)) (Dr Ruth Warren), and (3) an automated volumetric 'Stepwedge' method developed at Manchester University (280) (Dr Sue Astley). The Manchester Stepwedge method used markers on the compression paddle to determine breast thickness, and a calibration device (Stepwedge) to match each pixel density in the mammogram with the equivalent density in the Stepwedge. This method therefore required availability of the Stepwedge so that it could be imaged alongside the breast at the time of mammography. The Manchester Stepwedge method calculated TV, DV, FV and PDV and Cumulus was used to calculate TA, DA, FA and PDA. Density assessments were made at 3 time points: baseline, 1yr follow-up (at the end of the intervention) and 1yr after the end of the intervention. Baseline mammograms were taken at the point of entry to the study; for those women with a mammogram performed within one year of entry, their most recent mammogram within the last 12 months was used. Each woman had four mammographic views taken at each time point: left CC, right CC, left MLO and right MLO, and a final mammographic score at each time point was calculated using an average of the four views. The primary analysis refers to Cumulus-assessed DA, FA and PDA, and Stepwedge-assessed DV, FV and PDV to assess the effect of BMI on dense and non-dense tissue separately. Visually-assessed density had similar results to Cumulus-assessed PDA so is included as a secondary density measure only. Results for TA and TV are also reported as secondary density measures in the results tables.

2.2.3 Body weight and body composition

Weight, BMI and a variety of different measures of body composition were assessed at baseline, 1yr and 2yr after the start of the intervention. Weight (kg) and height at baseline (m) were determined using a calibrated beam balance and stadiometer, and used to calculate BMI (kg/m^2). Other body composition assessments were also made i.e. waist circumference (measured by a trained research nurse using a measuring tape); total body fat, fat free mass and % body fat (assessed using a DXA whole body scanner (Hologic Inc., Bedford, MA, USA) and bioelectrical impedance (Tanita TBF-300A, Tanita Europe B.V., Hoogoorddreef 56E, 1101 BE Amsterdam, The Netherlands)); and intra-abdominal and abdominal subcutaneous area (assessed using an MRI scan with a single transverse scan taken at the level of the intervertebral disc between the L2 and L3 vertebrae). Weight, BMI, waist circumference, and total body fat,

fat free mass and % body fat (impedance) were recorded at all 3 time points. Intra-abdominal area, abdominal subcutaneous area, and total body fat, fat free mass and % body fat (DXA) were only measured at baseline and at 1yr. Weight at age 20yr was self-reported via questionnaire, and BMI at age 20yr was calculated using weight at age 20yr and height at study entry. Long-term BMI gain was calculated as the difference between baseline BMI and BMI at age 20yr. BMI is discussed as the primary measure of body weight throughout the analysis because BMI is a commonly used adjustment for density and it is a well-established risk factor for breast cancer. BMI also provided the most longitudinal information because it was only missing for 1 observation at 2yr, whereas other measures had more missing data (such as impedance which was missing for 19 observations). Other body composition measures gave similar correlations with density to those of BMI and were highly correlated with BMI. Therefore, other body composition measures are included as secondary analyses. Weight gain during the intervention was defined as ≥+3% of baseline weight, weight-loss was defined as ≤ -3% of baseline weight, and a weight change >-3% to <+3% of baseline weight was defined as a stable weight (281).

2.2.4 Statistical methods

This analysis used the statistical software, R (282). All tests were two-sided and considered significant at the 5% level.

2.2.4.1 Repeated measures correlation coefficients (primary analysis)

Correlation (*r*) between BMI and mammographic density was assessed on a cross-sectional basis (between-women), and within-women as their short-term BMI changed, using repeated-measures methods as described by Bland and Altman (283, 284). These correlations used all of the available data together to get an overall statistic across repeated measures. Between-women correlations were used to evaluate the relationship between breast density measures and BMI cross-sectionally across the group of women, for example, whether heavier women were more likely to have dense breasts. Within-women correlations were used to assess whether breast density changed for an individual woman in line with their changing BMI. If there was little relationship seen between-women, then it was unlikely that there would be a relationship within-women.

A Pearson correlation coefficient could have been used to calculate between-women correlation, however this does not take into account the different number of observations contributed by each woman. Repeated measures correlation coefficients overcome this issue by calculating a

weighted Pearson correlation coefficient. With summations for woman $i = 1 \ldots n$, the weighted Pearson correlation coefficient is defined as:

$$\frac{\sum m_i \bar{x}_i \bar{y}_i - \frac{\sum m_i \bar{x}_i \sum m_i \bar{y}_i}{\sum m_i}}{\sqrt{\left(\sum m_i \bar{x}_i^2 - \frac{(\sum m_i \bar{x}_i)^2}{\sum m_i}\right) - \left(\sum m_i \bar{y}_i^2 - \frac{(\sum m_i \bar{y}_i)^2}{\sum m_i}\right)}};$$

where m_i is the number of observations for woman i; and $\bar{x}_i$ and $\bar{y}_i$ are the mean BMI and density measures for woman i, respectively (284). Missing pairs of density and BMI were excluded.

The within-women correlation coefficients effectively remove the differences between subjects to assess the changes within subjects only. The measure is based on the decomposition of sums of squares from an Analysis of variance (ANOVA). A linear model was first fit with a factor for each woman $i = 1 \ldots n$ at each time point $j = 1, 2, 3$ so that:

$$y_{ij} = \beta_0 + \psi_i + \beta_1 x_{ij} + e_{ij};$$

where y_{ij} is the density measure for woman i at time point j; x_{ij} is the BMI for woman i at time point j; β_1 is the parameter for BMI; β_0 is an overall intercept; ψ_i is a categorical factor variable with n factors (dummy variables); and e_{ij} is the random error for each observation with mean zero and unknown variance. Missing pairs of density and BMI were excluded.

The ANOVA table for this linear regression model is:

Source of variation	Degrees of freedom	Sum of squares	Mean square error	F-ratio
Women	$n - 1$	SS_{women}	$\frac{SS_{women}}{n-1}$	$\frac{MSE_{BMI}}{MSE_{error}}$
BMI	1	SS_{BMI}	$\frac{SS_{BMI}}{1}$	
Residual	$n(k-1) - 1$	SS_{error}	$\frac{SS_{error}}{n(k-1)-1}$	
Total	$(nxk) - 1$	SS_{total}	$\frac{SS_{total}}{(nxk)-1}$	

where n is the sample size; and k is the total number of observations divided by n i.e. mean number of observations per woman. Proof of the number of degrees of freedom has been described elsewhere (285).

ANOVA assessed the variability in density partitioned into components based on the source of the variation. Removing the variation from women (SS_{woman}), the within-women correlation coefficient was obtained:

$$(\pm)\sqrt{\frac{SS_{BMI}}{SS_{BMI}+SS_{error}}}\ ;$$

where the sign is equal to that of β_1 (283).

As a secondary analysis, these repeated measures correlation coefficients were also completed for all mammographic density measures against each other and for all adiposity measures against each other. As an exploratory analysis, DXA bone density was also measured and added to the list of adiposity measures.

2.2.4.2 Tadpole plots

Density data were also plotted against BMI using 'tadpole plots' to show simultaneous between- and within-women associations graphically. In these scatter plots, each tadpole represents a woman: the head of the tadpole represents a woman's density and BMI at their last mammogram, and the tail shows the same but for previous follow-ups (if density was available). This way, the reader may assess a woman's joint between- and within-women effects over the 2-year period, for example, assess whether the tadpole tails (within-women effects) followed the pattern seen by comparing the tadpole heads (between-women effects). If there was no relationship between-women, the heads would be horizontal, and if density did not change as a woman lost weight, the tails would be horizontal.

2.2.4.3 Empirical bootstrap

95% confidence intervals for correlation coefficients were estimated using an empirical bootstrap with 10,000 resamples. The idea behind this non-parametric method is to generate a random bootstrap sample from the original dataset sample using 'with replacement' selection so that each unit of the original sample may be selected more than once. The bootstrap sample is selected so that it is of the same size as the original dataset sample. The statistic of interest can then be calculated (here, this was the correlation coefficient). This is repeated a number of times (for example, 10,000 times) to get a distribution of bootstrapped statistics. From this, the 2.5th percentile and 97.5th percentile can be obtained to give the empirical lower and upper bounds for the 95% confidence interval, respectively (286).

2.2.4.4 Linear mixed models (1)

The ANOVA method for within-women correlations does not account for unbalanced data. Namely, some women were missing a density-BMI pair at one or more of their time points, therefore each woman contributed a different number of repeated measurements. Additionally, including a factor for each woman in the linear model for the ANOVA method might be affected by overfitting. Therefore, to check the robustness of repeated measure correlation coefficients, a multivariable linear mixed model was also fitted (287). Linear mixed models are robust to unbalanced data and reduce overfitting by including random effects per woman, as opposed to factors for each woman. The linear mixed model also allowed for assessment of the simultaneous association of between- and within-women correlations, alongside the adjustment for age.

Linear mixed models are frequently used to model repeated measures. Repeated measures give rise to clustered data where data points within the same group more closely resemble each other than data points in other groups. Clustered or grouped measurements, for example, within-person, tend to have high correlation with each other which results in lower variance than independent measurements. The linear mixed model overcomes this by introducing random effects in the linear regression model alongside the usual fixed effects representing population predictors. Estimated fixed effects are shared across all observations, whereas random effects vary across clusters.

Random effects are comprised of two or more levels. The lowest base level represents each outcome measurement and the level above represents the grouping of repeated measures, for example, each individual woman (as seen in this study). If the data structure were to involve additional grouping such as hospitals or regions, a higher level could have been included to account for further clustering. These random effects are used to model deviations of each level's groups about population mean effects (after accounting for the deviations for the levels below). Each level above the base level therefore separates the variance into two components: one that comprises the unobservable variance from that level and one comprising the variance from the level(s) below. Random effects can be modelled using random intercepts and random slopes. A random intercept allows for deviations about the mean fixed intercept and a random slope allows for deviations about a mean fixed effect slope so that each woman's slope is not necessarily parallel to that of the population.

In this chapter, the hierarchical structure of the linear mixed model includes two levels: the base level for each density measure at each time point, and the second level representing each woman. A diagram of this model is represented in Figure 2.1; which is also depicted as an equation (Equation 2.1).

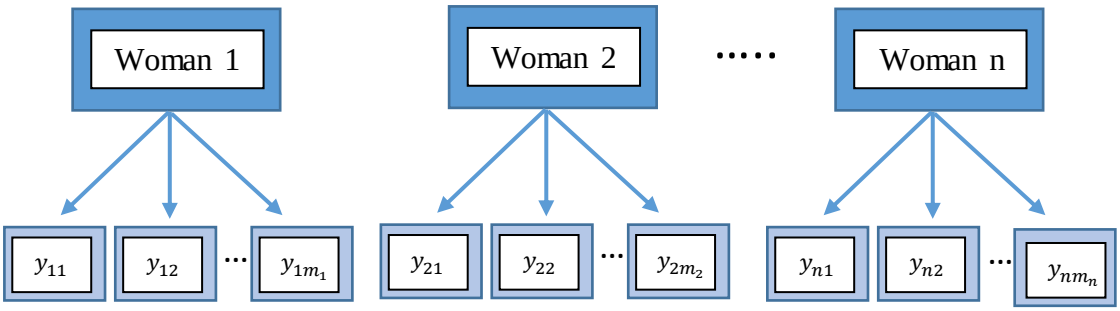


Figure 2.1: Representation of the 2-level linear mixed model.

Equation 2.1: The univariate linear mixed model.

Level 1: $$y_{ij} = \beta_{0_i} + \beta_{1_i} x_{ij} + e_{ij}$$

Level 2: $$\beta_{0_i} = \zeta_0 + u_{0_i}$$

$$\beta_{1_i} = \zeta_1 + u_{1_i}$$

where y_{ij} is the outcome for woman $i = 1, \dots, n$ at time point $j = 1, \dots, \mathrm{m}_i$, ζ_0 is the population fixed intercept, ζ_1 is the population fixed parameter for variable x_{ij}, e_{ij} is the conditional random error (residual), u_{0_i} is the random intercept for woman i, and u_{1_i} is the random slope for woman i.

The basic model assumptions used in the analysis are:

- The random effects $\boldsymbol{u}_i = (u_{0_i}, u_{1_i})$ are normally-distributed, such that $\boldsymbol{u}_i \sim N(\boldsymbol{\mu}_u, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu}_u = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\boldsymbol{\Sigma}$ is the 2x2 square, symmetric, and positive semi-definite variance-covariance matrix, defined as: $\boldsymbol{\Sigma} = \begin{pmatrix} \tau_{00}^2 & \tau_{01}^2 \\ \tau_{10}^2 & \tau_{11}^2 \end{pmatrix}$, where τ_{00}^2 is the variance of the random intercept, τ_{11}^2 is the variance of the random slope, and τ_{01}^2 and τ_{10}^2 represent the covariance of the random intercept and slope. Different structures can be assumed for $\boldsymbol{\Sigma}$. The two main structures mentioned in this thesis are: independent (covariance elements, τ_{01}^2 and τ_{10}^2, are constrained to be zero, hence random effects are uncorrelated) which is the simplest assumed structure, and unstructured (all elements are estimated and each of the random effects is allowed to be correlated with each other) which is commonly used for repeated measures data.
- The random errors (residuals) $\boldsymbol{e}_i = (e_{i1}, \dots, e_{i\mathrm{m}_i})$ are normally-distributed, such that $\boldsymbol{e}_i \sim N(\boldsymbol{\mu}_e, \boldsymbol{E}_i)$ where $\boldsymbol{\mu}_{e_{m_i x 1}} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ and $\boldsymbol{E}_i = \sigma^2 \boldsymbol{I}_{m_i}$ with σ^2 being the sample residual

variance and $\boldsymbol{I}_{m_i}$ is the m_i x m_i identity matrix. This is the most common structure assumption and the only assumption mentioned in this thesis, although other structures, such as compound symmetry or autoregressive, can be assumed.

- For the same woman:
  - All residuals are independent.
  - All random effects are (conditionally) independent of all residuals.
- For different women:
  - All random effects for one woman are independent of all random effects for another woman.
  - All residuals for one woman are independent of all residuals for another woman.
  - All random effects for one woman are independent of all residuals for another woman.

The linear mixed model for this study is described below.

Breast density y_{ij} for woman $i = 1, \ldots, n$ at time $j = 1, 2, 3$ was modelled as:

Equation 2.2

$$y_{ij} = \alpha + \beta age_{ij} + \gamma \bar{x}_{i.} + \delta\left(x_{ij} - \bar{x}_{i.}\right) + u_{0i} + e_{ij};$$

where α is an overall intercept, age_{ij} is the age for woman i at time j, β is the parameter for age, $\bar{x}_{i.}$ is mean BMI for woman i, γ is the between-women BMI parameter, x_{ij} is the BMI of woman i at time j, δ is the within-women parameter, and e_{ij} is an independent random error. The term that allowed for differences between women in their overall density level is the independent random intercept u_{0i} for woman i. The model is completed by assuming normal distributions for u_{0i} and e_{ij}, with zero mean, unknown variances and: zero covariance between e_{ij} of the same woman or different women, zero covariance between u_{0i} and e_{ij} of the same woman or different women, and zero covariance between u_{0i} of different women. Missing pairs of density and BMI were excluded. The model was fitted by maximum likelihood (2.2.4.7). To aid interpretation of the estimates across different measures of density, the density values were first standardised (2.2.4.6). To test γ=0 (between-women correlation) and δ=0 (within-women correlation) a Wald test was applied (2.2.4.9).

A secondary analysis was also undertaken to assess the effect of adding BMI gain since 20yr of age to the model. The model was extended to consider BMI gain from age 20yr:

Equation 2.3

$$y_{ij} = \alpha + \beta age_{ij} + \gamma \bar{x}_{i.} + \delta\left(x_{ij} - \bar{x}_{i.}\right) + u_{0i} + \varepsilon z_i + e_{ij};$$

where z_i is the BMI gain since age 20yr for woman i: calculated as the difference between baseline BMI for woman i and BMI at age 20yr for woman i, and ε is the parameter for BMI gain since age 20yr. To test ε=0 a Wald test was applied (2.2.4.9).

In the linear mixed model, BMI was modelled as both a between-women and within-women effect to mirror the repeated measures correlation coefficients. Using the mean BMI for the between-women effect ensured that all of the data across the intervention was used for each woman, whilst determining a stable reference point from which a relative within-women change measure could be calculated. As this model contained repeated measures, random variation about the overall population mean density was allowed by including random intercepts.

Likelihood ratio tests were conducted to test for inclusion of random slopes which would allow individual-woman deviation about the within-women population effect, δ. These likelihood ratio tests are explained in more detail in section 2.2.4.8. No interactions were considered so that effects from the linear mixed model were the same as the repeated measures correlation coefficients.

2.2.4.5 Transformations

Diagnostic Q-Q plots were used to check the normality assumption of residual errors and random effects i.e. $\boldsymbol{e}_i \sim N(\boldsymbol{\mu}_e, \boldsymbol{E}_i)$ and $\boldsymbol{u}_i \sim N(\boldsymbol{\mu}_u, \boldsymbol{\Sigma})$. Quantiles of the estimated residual errors, $\hat{\boldsymbol{e}}$, and predicted random effects, $\hat{\boldsymbol{u}}$, were plotted against theoretical quantiles from a standard normal distribution, to visually assess whether plots formed a straight line and were thus normally distributed (288). To make density measures more symmetric and approximately normal-distributed they were transformed: a square root transformation for area measures and a cube root transformation for volumetric measures.

2.2.4.6 Standardisation

To help with comparisons across different measures of breast density, the breast density values were first standardised for woman $i = 1, \ldots, n$ at time point $j = 1, 2, 3$ using:

an overall mean: $\bar{x} = \frac{\sum_{i=1}^{n} \bar{x}_i}{n}$,

and variance: $\sigma^2 = \frac{\sum_{i=1}^{n} (\bar{x}_i - \bar{x})^2}{n-1}$,

to get a standardised density measure: $\varphi_{ij} = \frac{d_{ij} - \bar{x}}{\sigma}$;

where $\bar{x}_i$ is the mean density for woman i and d_{ij} is the density measure for woman i at time point j.

2.2.4.7 Maximum likelihood

The likelihood of a parameter, θ, given the observed data, $x_1, \ldots, x_n = \boldsymbol{x}$, is $\mathcal{L}(\theta|\boldsymbol{x})$. The aim is to find the value for θ that maximises the likelihood function by taking the supremum ('sup') of $\mathcal{L}(\theta|\boldsymbol{x})$, also known as the maximum likelihood estimator (MLE). The natural logarithm of the likelihood (denoted with a lower case symbol, $\ell(\theta|\boldsymbol{x})$) is also often used. An estimated MLE ($\hat{\theta}$) can be found by taking the derivative of $\mathcal{L}(\theta|\boldsymbol{x})$ or $\ell(\theta|\boldsymbol{x})$ with respect to θ and equating this to zero to find the global maximum and then solving the resulting equation.

One problem that arises when using maximum likelihood (ML) for linear mixed models is that variance component estimators, $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\sigma}^2$, tend to be negatively biased because fixed coefficients are assumed to be known without uncertainty. As a solution, unbiased restricted maximum likelihood estimation (REML) can be used instead (289). Briefly, maximum likelihood is applied to the residuals from the fixed part of the model therefore estimation of the variance components is independent of the fixed effects coefficients. A limitation of REML is that it is biased when comparing nested models that differ in their fixed effects. On the other hand, ML is biased when comparing nested models that differ in their random effects, however this bias decreases as the sample size increases (290).

2.2.4.8 Likelihood ratio tests

For a statistical model with parameter space, Ω: the null hypothesis (H_0) states that parameter $\theta \in \Omega_0$, where Ω_0 is a subset of Ω, and the alternative hypothesis (H_1) states that $\theta \in \Omega_0^C$, where Ω_0^C is the complement of Ω_0 (291).

The likelihood ratio statistic for testing H_0 vs. H_1 is defined as:

$$\lambda(\boldsymbol{x}) = \frac{\sup(\mathcal{L}(\theta|\boldsymbol{x}) : \theta \in \Omega_0)}{\sup\,(\mathcal{L}(\theta|\boldsymbol{x}) : \theta \in \Omega)}$$

Assuming the null hypothesis is true, Wilks' theorem (292) can be used to conduct tests on whether to reject H_0 in nested models. Asymptotically (as the sample size $n \to \infty$), the statistic $-2\log(\lambda)$ follows a chi-squared distribution i.e. $-2\log(\lambda) \sim \chi^2$ with degrees of freedom equal to: dimensionality (Ω) - dimensionality(Ω_0).

The natural logarithm of the likelihood, $\ell(\theta|\boldsymbol{x})$, is often used because with logarithms, products become summations and division becomes subtraction: $log(ab) = log(a) + log(b)$; $log\left(\frac{a}{b}\right) = log(a) - log(b)$.

In this case: $$-2\log(\lambda) = -2\log\left(\frac{\sup(\mathcal{L}(\theta|\boldsymbol{x}):\theta\in\Omega_0)}{\sup(\mathcal{L}(\theta|\boldsymbol{x}):\theta\in\Omega)}\right)$$
$$= 2\big(\log\big(\sup(\mathcal{L}(\theta|\boldsymbol{x}) : \theta \in \Omega)\big) - \log\big(\sup(\mathcal{L}(\theta|\boldsymbol{x}) : \theta \in \Omega_0)\big)\big);$$

And since the logarithm function on the set of positive real numbers is a monotonically increasing function, $\log\Big(\sup\big(\mathcal{L}(\theta|\boldsymbol{x})\big)\Big) = \sup\Big(\log\big(\mathcal{L}(\theta|\boldsymbol{x})\big)\Big)$.

Therefore $-2\log(\lambda)$ can be rewritten as:

$$-2\log(\lambda) = 2(\sup((\ell(\theta|\boldsymbol{x}) : \theta \in \Omega)) - \sup((\ell(\theta|\boldsymbol{x}) : \theta \in \Omega_0)))$$

The significance of a model compared with its nested model can then be tested by assessing the statistic $-2\log(\lambda)$ using a χ^2 distribution.

2.2.4.9 *<u>Wald tests</u>*

A Wald test is used to assess how far an estimated parameter is from 0 (the value under the null hypothesis) in terms of its standard error. A Wald test that fails to reject the null hypothesis suggests that the estimated parameter is very small relative to its standard error and that removing this coefficient from the model will not harm the model fit.

2.3 Results

2.3.1 Baseline characteristics

Baseline characteristics of the cohort are shown in Table 2.1. Median age was 41yr (interquartile range (IQR), 38-43), and the majority of women were Caucasian (n=60, 92%) and parous (n=55, 85%). At baseline, 27 women (42%) were classified as overweight (BMI ≥25 kg/m^2 and <30 kg/m^2), 20 (31%) were obese (BMI ≥30 kg/m^2) and 18 (28%) were in the normal BMI range (BMI ≥18.5 kg/m^2 and <25 kg/m^2). All women gained at least 7kg of weight from the age of 20yr. By the end of the study, 16 women (25%) had gained weight, 22 (34%) had lost weight and 26 (41%) maintained their original weight.

2.3.2 Mammographic density measurements

Median PDA, DA and FA of each woman's average density measure over the intervention were 37.1% (IQR, 2.5%-71.3%), 59.9cm^2 (IQR, 5.8cm^2-158.4cm^2) and 107.3cm^2 (IQR, 23.6cm^2-405.1cm^2), respectively. For Stepwedge measures, PDV, DV and FV were 22.7% (IQR, 6.7%-69.4%), 191.5cm^3 (IQR, 56.7cm^3-710.4cm^3) and 573.0cm^3 (IQR, 72.8cm^3-1992.1cm^3), respectively. There was a large amount of missing data for volumetric density at baseline because of unavailability of the Stepwedge calibration tool, therefore volumetric density was only available for 36 women at baseline. Non-missing data was adequate for Stepwedge measurements at the other time points (n at 1yr and 2yr = 60 and 61, respectively) and for Cumulus measurements at all time points (n at baseline, 1yr and 2yr = 61, 64 and 55, respectively).

2.3.3 Repeated measures correlation coefficients – body mass index and mammographic density

The estimated repeated measure correlations are shown in Table 2.3. DV was positively correlated with BMI between-women (r=0.41, 95% CI, 0.17 to 0.61) but less so within-women (r=0.08, 95% CI, -0.16 to 0.28). There was little association between DA and BMI (between-women r=-0.12, 95% CI, -0.38 to 0.16; within-women r=0.01, 95% CI, -0.24 to 0.25). PDV was inversely associated with BMI between- and within-women (between r=-0.48, 95% CI, -0.64 to -0.33; within r=-0.36, 95% CI, -0.54 to -0.12), and PDA was inversely associated with BMI between-women (r=-0.58, 95% CI, -0.72 to -0.42), but less so within-women (r=-0.22, 95% CI, -0.44 to 0.01). FV and FA were positively correlated with BMI between- and within-women (volume: between r=0.77, 95% CI, 0.69 to 0.84, within r=0.58, 95% CI, 0.36 to 0.75; area: between r=0.74, 95% CI, 0.63 to 0.82, within r=0.45, 95% CI, 0.23 to 0.63). The magnitude and significance of correlations were weaker within-women than between-women.

Table 2.1: Participant characteristics at baseline.

Factor	Summary
Age (years)	41 (38-43)*
Baseline BMI (kg/m^2)	27.1 (24.7-33.4)*
Normal (≥18.5 to <25)	18 (28%)
Overweight (≥25 to <30)	27 (42%)
Obese (≥30)	20 (31%)
Height (m)	1.64 (1.60-1.68)*
Age at menarche (years)	12 (12-13)*
Number of live births	
Nulliparous	10 (15%)
1-2	41 (63%)
3-4	12 (18%)
≥5	2 (3%)
Age first live birth (years)	27 (22-29)*
Ethnicity (% Caucasian)	60 (92%)
Previous smoker	
Never	54 (83%)
Ever	11 (17%)
Previous oral contraception use	
Never	5 (8%)
Ever	58 (89%)
Missing	2 (3%)
Breastfed	
Never	22 (34%)
Ever	41 (63%)
Missing	2 (3%)
10 year Tyrer-Cuzick risk (%)	4.0 (3.0-5.0)*
Alcohol intake (units per week)	11 (3-24)*
Physical activity (kJ/kg per week)	974 (945-999)*

**Median (interquartile range); Body mass index (BMI).*

Table 2.2: Mammographic density measure counts during the intervention.

Factor	N (%)
N women with VAS density at 3 time points	53 (82%)
N women with VAS density at 2 time points	7 (11%)
N women with VAS density at 1 time point	5 (8%)
N women with Cumulus density at 3 time points	51 (78%)
N women with Cumulus density at 2 time points	13 (20%)
N women with Cumulus density at 1 time point	1 (2%)
N women with Stepwedge density at 3 time points	31 (48%)
N women with Stepwedge density at 2 time points	29 (45%)
N women with Stepwedge density at 1 time point	5 (8%)

Visual assessment score (VAS).

Table 2.3: Repeated-measures between-women and within-women correlations (95% confidence intervals) for density and body mass index.

Field	VAS (95% CI) [sqrt%]	PDA (95% CI) [sqrt%]	PDV (95% CI) [cbrt%]
Cross-sectional BMI (between-women)	-0.62 (-0.74 to -0.47)	-0.58 (-0.72 to -0.42)	-0.48 (-0.64 to -0.33)
Short-term BMI change (within-women)	-0.27 (-0.48 to -0.05)	-0.22 (-0.44 to 0.01)	-0.36 (-0.54 to -0.12)

Field	FA (95% CI) [sqrt]	FV (95% CI) [cbrt]	DA (95% CI) [sqrt]	DV (95% CI) [cbrt]
Cross-sectional BMI (between-women)	0.74 (0.63 to 0.82)	0.77 (0.69 to 0.84)	-0.12 (-0.38 to 0.16)	0.41 (0.17 to 0.61)
Short-term BMI change (within-women)	0.45 (0.23 to 0.63)	0.58 (0.36 to 0.75)	0.01 (-0.24 to 0.25)	0.08 (-0.16 to 0.28)

Visual assessment score (VAS), percent dense area (PDA), percent dense volume (PDV), fat area (FA), fat volume (FV), dense area (DA), dense volume (DV), square root transformed (sqrt), cube root transformed (cbrt), body mass index (BMI).

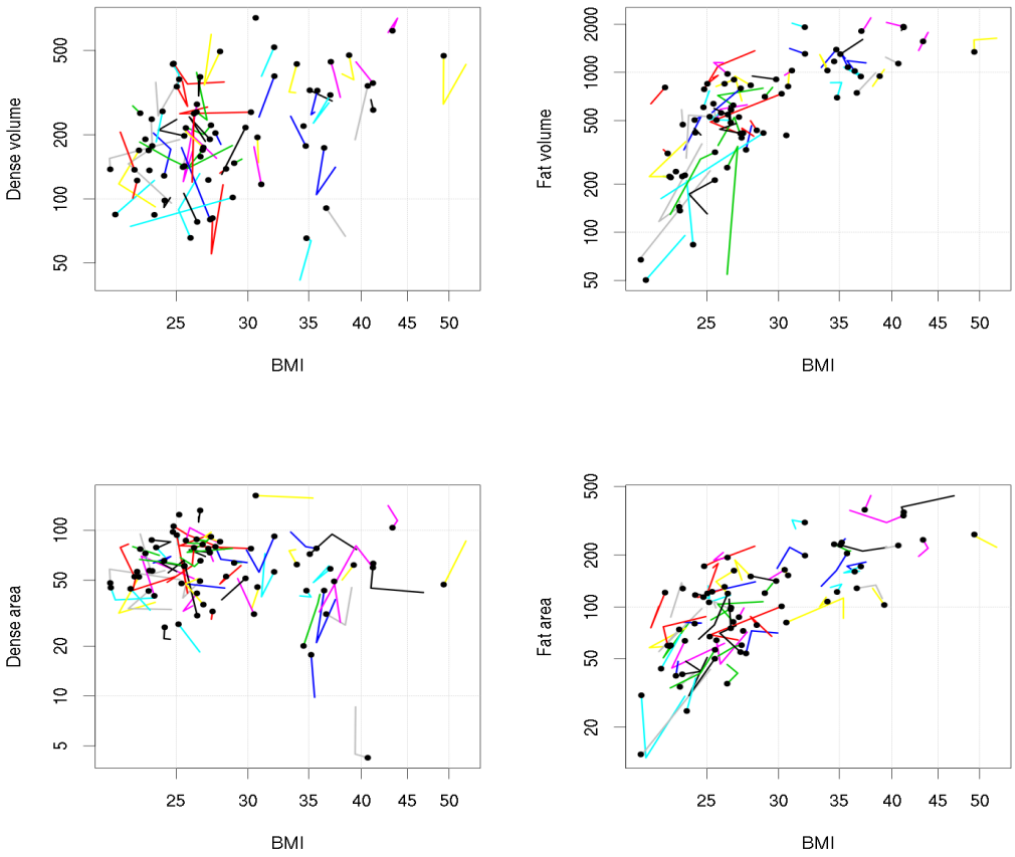


Figure 2.2: Tadpole plots showing density measures by body mass index during the intervention.

Each tadpole represents a woman: the head shows density and body mass index (BMI) at their last mammogram; the tail shows the same but for previous follow-ups (if density available).

2.3.4 Tadpole plots

Tadpole plots showed similar patterns to those seen with the repeated measures correlation coefficients (Figure 2.2). For Cumulus and Stepwedge, the heavier the woman was the fattier her breasts (between-women correlation of the tadpole heads). As women lost weight, their breast fat also decreased (within-women correlation depicted by the direction of tadpole tails). Tadpole heads for DA remained flat, suggesting that there was little association with BMI between-women; and lack of a discernible pattern for the tails indicated that there was little association within-women too. Tadpole heads for DV suggested that heavier women had higher dense volume (between-women), but the direction of tadpole tails (within-women associations) was less evident between BMI and DV. In general, the tadpole tails more-or-less followed the pattern for the tadpole heads, providing some evidence that the relationship between BMI and density reported in population studies can be applied to make predictions about the breast density of a woman as she diets.

2.3.5 Repeated measures correlation coefficients - other adiposity measures and mammographic density

There were similar associations between breast density and other body fat compositions as with BMI (Table 2.4, Table 2.5). There was a positive association for DV between-women of approximately 0.4 (although less so in DXA %fat, MRI subcutaneous and MRI abdominal fat), and little association within-women. There was also little association for DA and other body fat compositions between- or within-women. There was an inverse association for PDV between-women (approximately -0.5) and within-women (approximately -0.3), although within-women associations were less strong than between with only weight and impedance total fat showing significant effects. Similarly, there was an inverse association for PDA between-women (approximately -0.5), but less so within-women (only impedance %fat showed a significant effect of approximately -0.3). FV and FA were positively associated with other body fat compositions between-women (approximately 0.7) and within-women (approximately 0.4). However, the within-women correlations for FV and DXA %fat or MRI measures, and within-women correlations for FA and DXA lean mass or MRI total fat were not significant.

2.3.6 Repeated measures correlation coefficients – adiposity measures (between-women) and mammographic density measures (between-women)

Exploratory between-women correlations were also performed amongst the different body composition measures (Table 2.6) and amongst the different density measures (Table 2.7). As expected, associations between body composition measures were strong, albeit slightly weaker for measures of lean mass (DXA lean and impedance lean mass). Associations between different density measures were mostly expected for percent density where a strong positive association was seen between different percentage density methods and an inverse relationship was seen between percent density and both breast fat and total area or volume. However, a positive association was only seen between percent density methods and DA, but not DV (except for PDV which had a modest correlation with DV). Breast fat measures were strongly and positively correlated with each other and with total area or volume. Breast fat measures were moderately positively correlated with DV. There was some indication of an inverse relationship between FA and DA, but this was less so when assessing the association between FV and DA. Similarly, TA and TV were strongly positively correlated with each other, and both were moderately correlated with DV; but little association was seen with DA. A moderately positive association was seen between DA and DV.

Table 2.4: Complete results for repeated-measures between-women correlations (95% confidence intervals) for density and body composition measures.

	VAS (95% CI) [sqrt%]	PDA (95% CI) [sqrt%]	PDV (95% CI) [cbrt%]	FA (95% CI) [sqrt]	DA (95% CI) [sqrt]
Weight	-0.55 (-0.70 to -0.37)	-0.49 (-0.64 to -0.30)	-0.40 (-0.56 to -0.23)	0.69 (0.59 to 0.79)	0.00 (-0.28 to 0.27)
BMI	-0.62 (-0.74 to -0.47)	-0.58 (-0.72 to -0.42)	-0.48 (-0.64 to -0.33)	0.74 (0.63 to 0.82)	-0.12 (-0.38 to 0.16)
Waist	-0.63 (-0.76 to -0.46)	-0.59 (-0.73 to -0.41)	-0.54 (-0.67 to -0.38)	0.77 (0.67 to 0.85)	-0.09 (-0.34 to 0.18)
Imped(total fat)	-0.57 (-0.71 to -0.39)	-0.53 (-0.68 to -0.35)	-0.42 (-0.58 to -0.26)	0.71 (0.61 to 0.80)	-0.03 (-0.32 to 0.25)
Imped(% fat)	-0.57 (-0.71 to -0.40)	-0.55 (-0.69 to -0.38)	-0.49 (-0.66 to -0.30)	0.71 (0.60 to 0.80)	-0.06 (-0.33 to 0.22)
Imped(lean)	-0.40 (-0.58 to -0.17)	-0.34 (-0.55 to -0.13)	-0.26 (-0.45 to -0.07)	0.57 (0.44 to 0.69)	0.06 (-0.23 to 0.33)
DXA(total fat)	-0.55 (-0.69 to -0.36)	-0.53 (-0.68 to -0.35)	-0.54 (-0.68 to -0.39)	0.69 (0.58 to 0.79)	-0.02 (-0.32 to 0.26)
DXA(lean)	-0.36 (-0.56 to -0.15)	-0.30 (-0.52 to -0.07)	-0.23 (-0.41 to -0.04)	0.53 (0.39 to 0.66)	0.13 (-0.21 to 0.42)
DXA(% fat)	-0.53 (-0.69 to -0.31)	-0.55 (-0.70 to -0.36)	-0.64 (-0.77 to -0.49)	0.63 (0.48 to 0.76)	-0.09 (-0.35 to 0.17)
DXA(bone)	-0.23 (-0.47 to 0.03)	-0.08 (-0.31 to 0.17)	-0.03 (-0.31 to 0.25)	0.17 (-0.08 to 0.41)	0.11 (-0.14 to 0.37)
MRI(subcutaneous)	-0.64 (-0.77 to -0.46)	-0.62 (-0.76 to -0.46)	-0.60 (-0.73 to -0.48)	0.74 (0.64 to 0.83)	-0.13 (-0.42 to 0.15)
MRI(abdominal)	-0.65 (-0.76 to -0.48)	-0.65 (-0.77 to -0.49)	-0.60 (-0.73 to -0.48)	0.78 (0.68 to 0.86)	-0.15 (-0.42 to 0.10)
MRI(total fat)	-0.57 (-0.69 to -0.42)	-0.61 (-0.73 to -0.47)	-0.52 (-0.66 to -0.38)	0.76 (0.61 to 0.86)	-0.18 (-0.39 to 0.05)

	TA (95% CI) [sqrt]	FV (95% CI) [cbrt]	DV (95% CI) [cbrt]	TV (95% CI) [cbrt]
Weight	0.71 (0.61 to 0.80)	0.74 (0.65 to 0.82)	0.49 (0.28 to 0.66)	0.77 (0.69 to 0.85)
BMI	0.72 (0.59 to 0.82)	0.77 (0.69 to 0.84)	0.41 (0.17 to 0.61)	0.78 (0.70 to 0.86)
Waist	0.76 (0.64 to 0.85)	0.80 (0.72 to 0.87)	0.37 (0.12 to 0.59)	0.79 (0.69 to 0.87)
Imped(total fat)	0.72 (0.61 to 0.81)	0.76 (0.67 to 0.84)	0.48 (0.26 to 0.66)	0.79 (0.70 to 0.86)
Imped(% fat)	0.71 (0.59 to 0.81)	0.78 (0.68 to 0.86)	0.43 (0.20 to 0.62)	0.79 (0.69 to 0.87)
Imped(lean)	0.59 (0.47 to 0.71)	0.60 (0.46 to 0.71)	0.45 (0.25 to 0.62)	0.64 (0.51 to 0.74)
DXA(total fat)	0.69 (0.57 to 0.79)	0.76 (0.68 to 0.84)	0.35 (0.06 to 0.58)	0.77 (0.68 to 0.85)
DXA(lean)	0.59 (0.45 to 0.71)	0.58 (0.45 to 0.70)	0.52 (0.29 to 0.70)	0.64 (0.51 to 0.74)
DXA(% fat)	0.61 (0.45 to 0.74)	0.72 (0.60 to 0.81)	0.16 (-0.12 to 0.43)	0.69 (0.55 to 0.80)
DXA(bone)	0.22 (-0.03 to 0.47)	0.26 (0.00 to 0.50)	0.33 (0.09 to 0.54)	0.31 (0.06 to 0.54)
MRI(subcutaneous)	0.73 (0.62 to 0.81)	0.78 (0.70 to 0.85)	0.25 (-0.04 to 0.48)	0.77 (0.67 to 0.85)
MRI(abdominal)	0.76 (0.63 to 0.85)	0.80 (0.73 to 0.86)	0.28 (-0.01 to 0.51)	0.79 (0.70 to 0.86)
MRI(total fat)	0.73 (0.55 to 0.85)	0.75 (0.65 to 0.83)	0.34 (0.07 to 0.54)	0.75 (0.64 to 0.84)

Significant, non-significant, borderline significance, visual assessment score (VAS), percent dense area (PDA), percent dense volume (PDV), fat area (FA), fat volume (FV), dense area (DA), dense volume (DV), total area (TA), total volume (TV), square root transformed (sqrt), cube root transformed (cbrt), body mass index (BMI), impedance (Imped).

Table 2.5: Complete results for repeated-measures within-women correlations (95% confidence intervals) for density and body composition measures.

	VAS (95% CI) [sqrt%]	PDA (95% CI) [sqrt%]	PDV (95% CI) [cbrt%]	FA (95% CI) [sqrt]	DA (95% CI) [sqrt]
Weight	-0.27 (-0.49 to -0.05)	-0.22 (-0.44 to 0.02)	-0.37 (-0.55 to -0.13)	0.46 (0.24 to 0.65)	0.01 (-0.24 to 0.25)
BMI	-0.27 (-0.48 to -0.05)	-0.22 (-0.44 to 0.01)	-0.36 (-0.54 to -0.12)	0.45 (0.23 to 0.63)	0.01 (-0.24 to 0.25)
Waist	-0.17 (-0.38 to 0.06)	-0.10 (-0.31 to 0.12)	-0.23 (-0.42 to 0.00)	0.25 (0.03 to 0.47)	0.01 (-0.19 to 0.21)
Imped(total fat)	-0.22 (-0.44 to 0.03)	-0.24 (-0.46 to 0.01)	-0.32 (-0.52 to -0.09)	0.44 (0.22 to 0.63)	-0.07 (-0.31 to 0.17)
Imped(% fat)	-0.14 (-0.35 to 0.10)	-0.28 (-0.47 to -0.03)	-0.23 (-0.45 to 0.03)	0.44 (0.21 to 0.62)	-0.10 (-0.29 to 0.11)
Imped(lean)	-0.29 (-0.63 to 0.03)	-0.25 (-0.67 to 0.28)	-0.34 (-0.64 to 0.03)	0.51 (0.05 to 0.81)	-0.12 (-0.57 to 0.40)
DXA(total fat)	-0.08 (-0.51 to 0.34)	-0.24 (-0.56 to 0.16)	-0.27 (-0.61 to 0.24)	0.46 (0.12 to 0.71)	0.07 (-0.31 to 0.41)
DXA(lean)	-0.05 (-0.50 to 0.33)	-0.07 (-0.47 to 0.25)	-0.39 (-0.75 to 0.08)	0.27 (-0.02 to 0.63)	0.01 (-0.42 to 0.43)
DXA(% fat)	-0.09 (-0.48 to 0.32)	-0.32 (-0.61 to 0.06)	-0.26 (-0.61 to 0.28)	0.49 (0.15 to 0.74)	-0.02 (-0.29 to 0.23)
DXA(bone)	-0.07 (-0.48 to 0.30)	-0.01 (-0.33 to 0.34)	-0.07 (-0.52 to 0.43)	-0.21 (-0.52 to 0.14)	-0.14 (-0.45 to 0.20)
MRI(subcutaneous)	-0.24 (-0.59 to 0.18)	-0.38 (-0.68 to 0.01)	-0.38 (-0.71 to 0.19)	0.55 (0.22 to 0.78)	-0.01 (-0.38 to 0.32)
MRI(abdominal)	-0.23 (-0.57 to 0.19)	-0.33 (-0.65 to 0.08)	-0.28 (-0.63 to 0.28)	0.49 (0.09 to 0.77)	0.01 (-0.39 to 0.37)
MRI(total fat)	-0.17 (-0.53 to 0.22)	-0.21 (-0.61 to 0.20)	-0.08 (-0.50 to 0.42)	0.32 (-0.13 to 0.71)	0.04 (-0.44 to 0.41)

	TA (95% CI) [sqrt]	FV (95% CI) [cbrt]	DV (95% CI) [cbrt]	TV (95% CI) [cbrt]
Weight	0.54 (0.36 to 0.69)	0.59 (0.37 to 0.76)	0.07 (-0.17 to 0.28)	0.75 (0.55 to 0.85)
BMI	0.54 (0.35 to 0.69)	0.58 (0.36 to 0.75)	0.08 (-0.16 to 0.28)	0.74 (0.54 to 0.85)
Waist	0.32 (0.10 to 0.51)	0.37 (0.09 to 0.58)	0.05 (-0.15 to 0.24)	0.46 (0.14 to 0.68)
Imped(total fat)	0.47 (0.27 to 0.63)	0.47 (0.20 to 0.67)	0.01 (-0.24 to 0.24)	0.58 (0.29 to 0.76)
Imped(% fat)	0.43 (0.19 to 0.62)	0.36 (0.05 to 0.60)	0.09 (-0.15 to 0.29)	0.48 (0.15 to 0.70)
Imped(lean)	0.65 (0.26 to 0.85)	0.52 (0.06 to 0.79)	-0.03 (-0.46 to 0.38)	0.55 (-0.04 to 0.83)
DXA(total fat)	0.55 (0.27 to 0.75)	0.48 (0.03 to 0.77)	0.19 (-0.22 to 0.53)	0.65 (0.15 to 0.87)
DXA(lean)	0.31 (0.06 to 0.63)	0.53 (0.12 to 0.80)	-0.13 (-0.58 to 0.37)	0.48 (0.08 to 0.78)
DXA(% fat)	0.52 (0.20 to 0.74)	0.44 (-0.12 to 0.79)	0.20 (-0.21 to 0.48)	0.63 (-0.03 to 0.87)
DXA(bone)	-0.31 (-0.61 to 0.05)	-0.06 (-0.62 to 0.48)	-0.29 (-0.60 to 0.20)	-0.25 (-0.70 to 0.27)
MRI(subcutaneous)	0.64 (0.38 to 0.81)	0.48 (-0.13 to 0.80)	-0.01 (-0.47 to 0.47)	0.56 (-0.19 to 0.87)
MRI(abdominal)	0.57 (0.27 to 0.77)	0.42 (-0.16 to 0.77)	0.10 (-0.37 to 0.54)	0.53 (-0.21 to 0.83)
MRI(total fat)	0.37 (0.04 to 0.67)	0.24 (-0.25 to 0.66)	0.21 (-0.25 to 0.61)	0.38 (-0.23 to 0.75)

Significant, non-significant, borderline significance, visual assessment score (VAS), percent dense area (PDA), percent dense volume (PDV), fat area (FA), fat volume (FV), dense area (DA), dense volume (DV), total area (TA), total volume (TV), square root transformed (sqrt), cube root transformed (cbrt), body mass index (BMI), impedance (Imped).

Table 2.6: Complete results for repeated-measures between-women correlations (95% confidence intervals) for different body composition measures.

	Weight (95% CI)	BMI (95% CI)	Waist (95% CI)	Imped(total fat) (95% CI)	Imped(% fat) (95% CI)	Imped(lean) (95% CI)	DXA(total fat) (95% CI)
Weight	1.00 (1.00 to 1.00)	0.92 (0.88 to 0.95)	0.87 (0.81 to 0.91)	0.98 (0.98 to 0.99)	0.91 (0.87 to 0.94)	0.93 (0.88 to 0.96)	0.95 (0.93 to 0.97)
BMI	0.92 (0.88 to 0.95)	1.00 (1.00 to 1.00)	0.87 (0.81 to 0.92)	0.94 (0.91 to 0.96)	0.89 (0.85 to 0.94)	0.79 (0.65 to 0.88)	0.94 (0.91 to 0.97)
Waist	0.87 (0.81 to 0.91)	0.87 (0.81 to 0.92)	1.00 (1.00 to 1.00)	0.89 (0.84 to 0.93)	0.87 (0.82 to 0.91)	0.72 (0.61 to 0.83)	0.89 (0.83 to 0.93)
Imped(total fat)	0.98 (0.98 to 0.99)	0.94 (0.91 to 0.96)	0.89 (0.84 to 0.93)	1.00 (1.00 to 1.00)	0.96 (0.94 to 0.97)	0.85 (0.77 to 0.91)	0.97 (0.96 to 0.98)
Imped(% fat)	0.91 (0.87 to 0.94)	0.89 (0.85 to 0.94)	0.87 (0.82 to 0.91)	0.96 (0.94 to 0.97)	1.00 (1.00 to 1.00)	0.71 (0.58 to 0.81)	0.94 (0.92 to 0.97)
Imped(lean)	0.93 (0.88 to 0.96)	0.79 (0.65 to 0.88)	0.72 (0.61 to 0.82)	0.85 (0.77 to 0.91)	0.71 (0.58 to 0.81)	1.00 (1.00 to 1.00)	0.82 (0.70 to 0.90)
DXA(total fat)	0.95 (0.93 to 0.97)	0.94 (0.91 to 0.97)	0.89 (0.83 to 0.93)	0.97 (0.96 to 0.98)	0.94 (0.92 to 0.97)	0.82 (0.70 to 0.90)	1.00 (1.00 to 1.00)
DXA(lean)	0.88 (0.81 to 0.92)	0.72 (0.59 to 0.81)	0.64 (0.50 to 0.75)	0.82 (0.72 to 0.89)	0.70 (0.56 to 0.80)	0.93 (0.90 to 0.96)	0.70 (0.58 to 0.79)
DXA(% fat)	0.75 (0.65 to 0.83)	0.81 (0.75 to 0.87)	0.80 (0.71 to 0.88)	0.81 (0.73 to 0.88)	0.87 (0.81 to 0.91)	0.50 (0.30 to 0.67)	0.90 (0.86 to 0.94)
DXA(bone)	0.31 (0.11 to 0.50)	0.18 (-0.03 to 0.40)	0.11 (-0.11 to 0.34)	0.27 (0.07 to 0.47)	0.25 (0.02 to 0.46)	0.40 (0.18 to 0.61)	0.17 (-0.04 to 0.38)
MRI(subcutaneous)	0.86 (0.80 to 0.92)	0.90 (0.85 to 0.94)	0.89 (0.85 to 0.94)	0.88 (0.84 to 0.93)	0.87 (0.83 to 0.92)	0.72 (0.55 to 0.84)	0.93 (0.88 to 0.96)
MRI(abdominal)	0.87 (0.82 to 0.92)	0.92 (0.88 to 0.96)	0.93 (0.89 to 0.96)	0.90 (0.86 to 0.94)	0.89 (0.85 to 0.92)	0.73 (0.58 to 0.85)	0.93 (0.89 to 0.96)
MRI(total fat)	0.77 (0.69 to 0.88)	0.84 (0.77 to 0.90)	0.88 (0.80 to 0.93)	0.80 (0.72 to 0.90)	0.79 (0.71 to 0.87)	0.66 (0.51 to 0.84)	0.80 (0.71 to 0.90)

	DXA(lean) (95% CI)	DXA(% fat) (95% CI)	DXA(bone) (95% CI)	MRI(subcutaneous) (95% CI)	MRI(abdominal) (95% CI)	MRI(total fat) (95% CI)
Weight	0.88 (0.81 to 0.92)	0.75 (0.65 to 0.83)	0.31 (0.11 to 0.50)	0.86 (0.80 to 0.92)	0.87 (0.82 to 0.92)	0.77 (0.69 to 0.87)
BMI	0.72 (0.58 to 0.81)	0.81 (0.75 to 0.87)	0.18 (-0.03 to 0.39)	0.90 (0.85 to 0.94)	0.92 (0.88 to 0.96)	0.84 (0.77 to 0.90)
Waist	0.64 (0.50 to 0.75)	0.80 (0.71 to 0.88)	0.11 (-0.11 to 0.34)	0.89 (0.85 to 0.94)	0.93 (0.89 to 0.96)	0.88 (0.79 to 0.93)
Imped(total fat)	0.82 (0.72 to 0.89)	0.81 (0.73 to 0.88)	0.27 (0.07 to 0.47)	0.88 (0.84 to 0.93)	0.90 (0.86 to 0.94)	0.80 (0.72 to 0.90)
Imped(% fat)	0.70 (0.55 to 0.80)	0.87 (0.81 to 0.91)	0.25 (0.02 to 0.46)	0.87 (0.83 to 0.92)	0.89 (0.85 to 0.92)	0.79 (0.71 to 0.87)
Imped(lean)	0.93 (0.90 to 0.96)	0.50 (0.30 to 0.67)	0.40 (0.18 to 0.61)	0.72 (0.55 to 0.84)	0.73 (0.59 to 0.85)	0.66 (0.51 to 0.84)
DXA(total fat)	0.70 (0.58 to 0.80)	0.90 (0.86 to 0.94)	0.17 (-0.04 to 0.38)	0.93 (0.88 to 0.96)	0.93 (0.89 to 0.96)	0.80 (0.71 to 0.90)
DXA(lean)	1.00 (1.00 to 1.00)	0.36 (0.18 to 0.52)	0.47 (0.30 to 0.62)	0.58 (0.44 to 0.69)	0.61 (0.47 to 0.71)	0.57 (0.41 to 0.73)
DXA(% fat)	0.36 (0.18 to 0.53)	1.00 (1.00 to 1.00)	-0.03 (-0.27 to 0.21)	0.86 (0.79 to 0.91)	0.85 (0.78 to 0.90)	0.71 (0.60 to 0.82)
DXA(bone)	0.47 (0.30 to 0.62)	-0.03 (-0.27 to 0.21)	1.00 (1.00 to 1.00)	0.16 (-0.07 to 0.38)	0.13 (-0.09 to 0.37)	0.06 (-0.15 to 0.32)
MRI(subcutaneous)	0.58 (0.44 to 0.70)	0.86 (0.79 to 0.90)	0.16 (-0.07 to 0.39)	1.00 (1.00 to 1.00)	0.98 (0.97 to 0.99)	0.80 (0.72 to 0.89)
MRI(abdominal)	0.61 (0.48 to 0.71)	0.85 (0.78 to 0.90)	0.13 (-0.09 to 0.37)	0.98 (0.97 to 0.99)	1.00 (1.00 to 1.00)	0.90 (0.86 to 0.94)
MRI(total fat)	0.57 (0.42 to 0.73)	0.71 (0.60 to 0.82)	0.06 (-0.15 to 0.32)	0.80 (0.72 to 0.89)	0.90 (0.86 to 0.94)	1.00 (1.00 to 1.00)

Significant, non-significant, borderline significance, repeated, body mass index (BMI), impedance (Imped).

Table 2.7: Complete results for repeated-measures between-women correlations (95% confidence intervals) for different density measures.

	VAS (95% CI) [sqrt%]	PDA (95% CI) [sqrt%]	PDV (95% CI) [cbrt%]	FA (95% CI) [sqrt]	DA (95% CI) [sqrt]
VAS (sqrt%)	1.00 (1.00 to 1.00)	0.90 (0.83 to 0.95)	0.79 (0.70 to 0.86)	-0.83 (-0.89 to -0.77)	0.50 (0.29 to 0.67)
PDA (sqrt%)	0.90 (0.83 to 0.95)	1.00 (1.00 to 1.00)	0.78 (0.71 to 0.85)	-0.84 (-0.90 to -0.78)	0.68 (0.52 to 0.80)
PDV (cbrt%)	0.79 (0.70 to 0.86)	0.78 (0.71 to 0.85)	1.00 (1.00 to 1.00)	-0.67 (-0.77 to -0.56)	0.48 (0.29 to 0.64)
FA (sqrt)	-0.83 (-0.89 to -0.77)	-0.84 (-0.90 to -0.78)	-0.67 (-0.77 to -0.56)	1.00 (1.00 to 1.00)	-0.23 (-0.41 to -0.01)
DA (sqrt)	0.50 (0.30 to 0.67)	0.68 (0.51 to 0.80)	0.48 (0.29 to 0.64)	-0.23 (-0.42 to -0.01)	1.00 (1.00 to 1.00)
TA (sqrt)	-0.67 (-0.77 to -0.54)	-0.62 (-0.75 to -0.46)	-0.48 (-0.64 to -0.28)	0.94 (0.89 to 0.97)	0.11 (-0.10 to 0.34)
FV (cbrt)	-0.81 (-0.88 to -0.73)	-0.74 (-0.83 to -0.63)	-0.72 (-0.81 to -0.60)	0.94 (0.92 to 0.97)	-0.10 (-0.30 to 0.12)
DV (cbrt)	-0.09 (-0.31 to 0.14)	0.02 (-0.20 to 0.24)	0.30 (0.05 to 0.52)	0.42 (0.23 to 0.58)	0.55 (0.32 to 0.74)
TV (cbrt)	-0.71 (-0.81 to -0.59)	-0.63 (-0.75 to -0.49)	-0.55 (-0.69 to -0.37)	0.92 (0.87 to 0.95)	0.05 (-0.16 to 0.28)

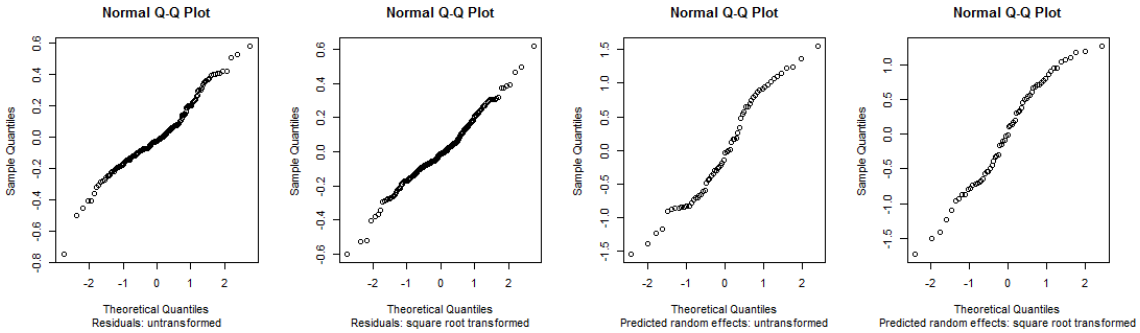
	TA (95% CI) [sqrt]	FV (95% CI) [cbrt]	DV (95% CI) [cbrt]	TV (95% CI) [cbrt]
VAS (sqrt%)	-0.67 (-0.77 to -0.53)	-0.81 (-0.88 to -0.72)	-0.09 (-0.31 to 0.13)	-0.71 (-0.81 to -0.59)
PDA (sqrt%)	-0.62 (-0.75 to -0.46)	-0.74 (-0.83 to -0.63)	0.02 (-0.20 to 0.23)	-0.63 (-0.75 to -0.49)
PDV (cbrt%)	-0.48 (-0.64 to -0.28)	-0.72 (-0.81 to -0.60)	0.30 (0.05 to 0.52)	-0.55 (-0.69 to -0.38)
FA (sqrt)	0.94 (0.89 to 0.97)	0.94 (0.92 to 0.97)	0.42 (0.23 to 0.58)	0.92 (0.87 to 0.95)
DA (sqrt)	0.11 (-0.10 to 0.34)	-0.10 (-0.30 to 0.12)	0.55 (0.32 to 0.74)	0.05 (-0.16 to 0.28)
TA (sqrt)	1.00 (1.00 to 1.00)	0.92 (0.87 to 0.95)	0.66 (0.54 to 0.76)	0.96 (0.93 to 0.98)
FV (cbrt)	0.92 (0.87 to 0.95)	1.00 (1.00 to 1.00)	0.43 (0.22 to 0.61)	0.97 (0.95 to 0.98)
DV (cbrt)	0.66 (0.54 to 0.75)	0.43 (0.23 to 0.62)	1.00 (1.00 to 1.00)	0.63 (0.49 to 0.75)
TV (cbrt)	0.96 (0.93 to 0.98)	0.97 (0.95 to 0.98)	0.63 (0.49 to 0.75)	1.00 (1.00 to 1.00)

Significant, non-significant, borderline significance, repeated, visual assessment score (VAS), percent dense area (PDA), percent dense volume (PDV), fat area (FA), fat volume (FV), dense area (DA), dense volume (DV), total area (TA), total volume (TV), square root transformed (sqrt), cube root transformed (cbrt).

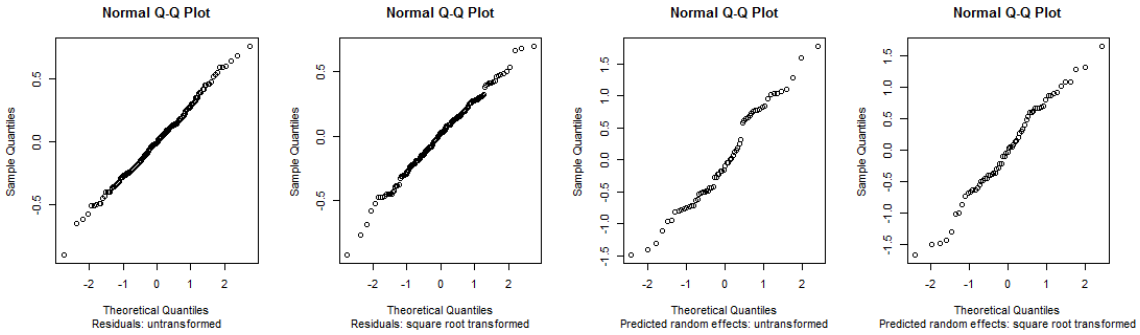
2.3.7 Linear mixed model

The between- and within-women associations for density and BMI measures were also estimated jointly in an age-adjusted linear mixed model. Q-Q plots for conditional residuals and predicted random effects showed a slight improvement when area density measures were square root transformed and volumetric density measures were cube root transformed since lines became straighter after these transformations (Figure 2.3).

VAS



PDA



PDV

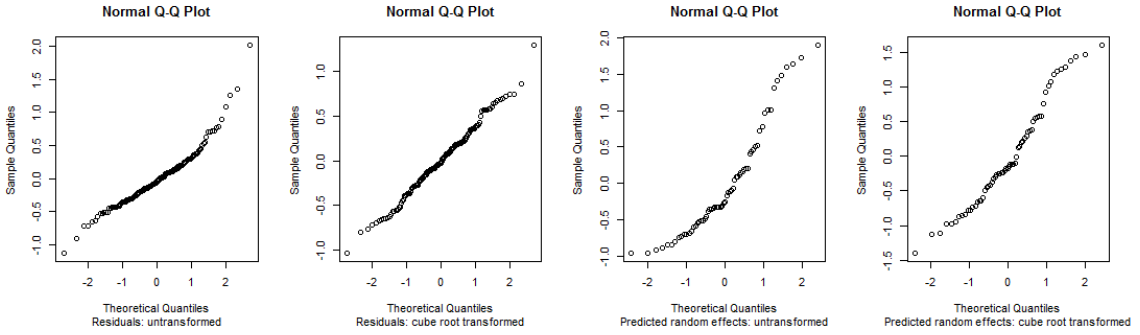
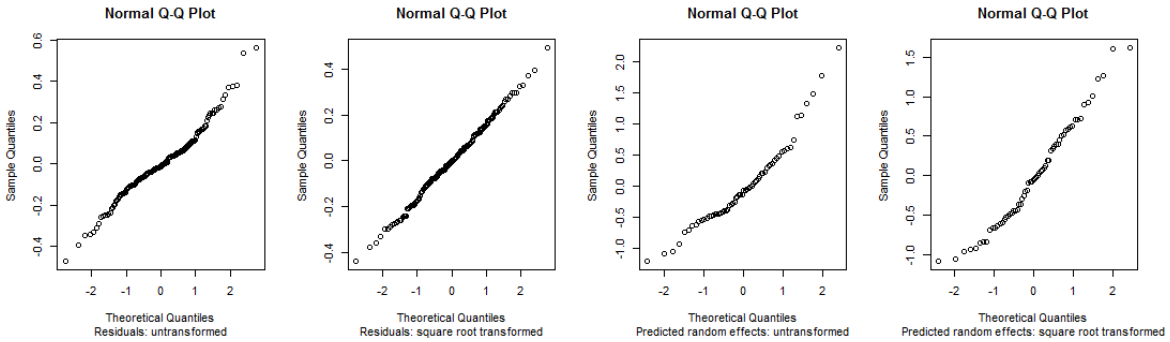


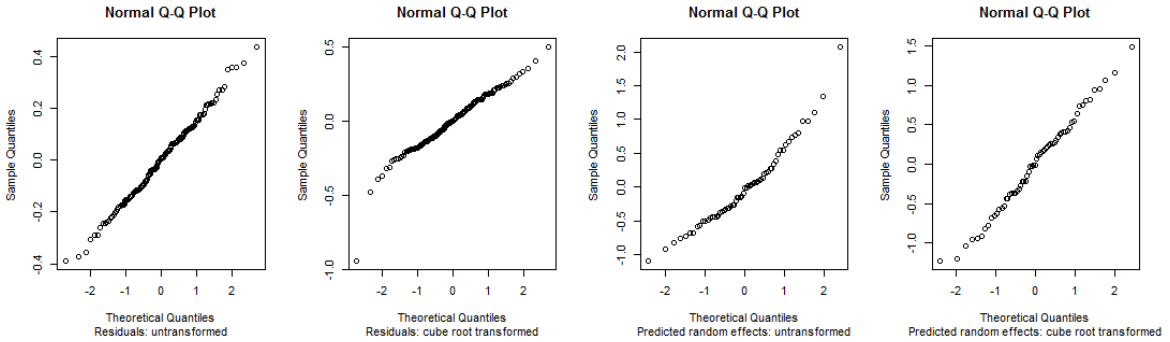
Figure 2.3: Q-Q plots for conditional residuals and predicted random effects.

Visual assessment score (VAS), percent dense area (PDA), percent dense volume (PDV), fat area (FA), fat volume (FV), dense area (DA), dense volume (DV).

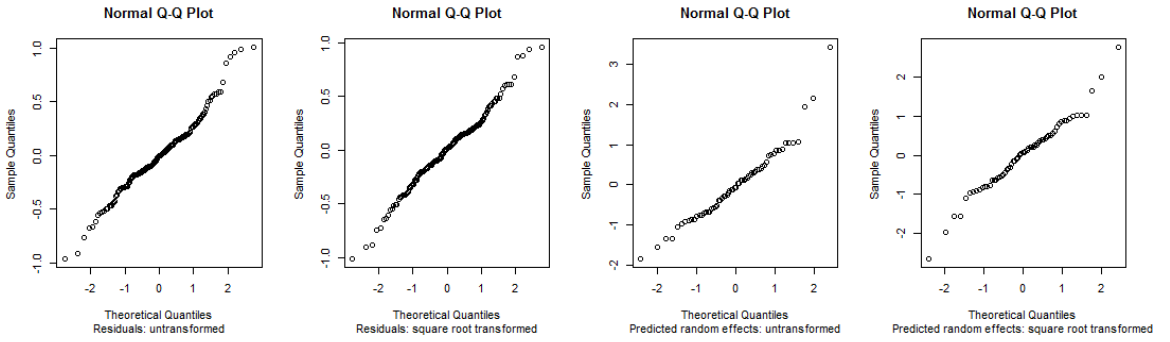
FA



FV



DA



DV

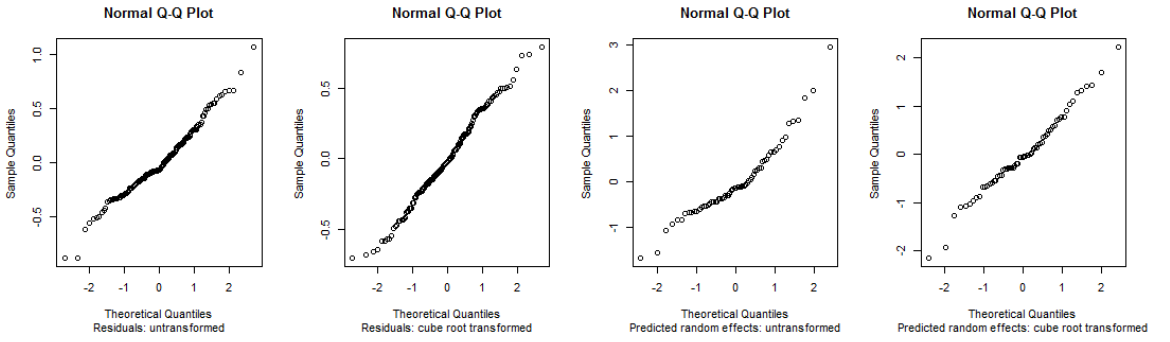


Figure 2.3 continued

Tests for random slopes in the linear mixed model for within-women BMI and age were not significant (Table 2.8, Table 2.9). The parameters, β and δ, from Equation 2.2 are therefore assumed to be the same for all women. No covariance structure was required because only one random effect was included (intercept).

Table 2.8: Linear mixed model likelihood ratio tests for within-women body mass index random slope.

	Likelihood ratio test	
Density outcome	ΔLR-χ^2	P-value
VAS (sqrt%)	0.86	0.65
PDA (sqrt%)	0.35	0.84
PDV (cbrt%)	1.71	0.43
FA (sqrt)	4.13	0.13
FV (cbrt)	1.23	0.54
DA (sqrt)	1.13	0.57
DV (cbrt)	0.00	1.00

ΔLR-χ^2 represents the difference in likelihood ratio for Equation 2.2 with and without a within-women BMI random slope (test with 2 degrees of freedom: random slope and covariance), visual assessment score (VAS), percent dense area (PDA), percent dense volume (PDV), fat area (FA), fat volume (FV), dense area (DA), dense volume (DV), square root transformed (sqrt), cube root transformed (cbrt).

Table 2.9: Linear mixed model likelihood ratio tests for age random slope.

	Likelihood ratio test	
Density outcome	ΔLR-χ^2	P-value
VAS (sqrt%)	0.00	1.00
PDA (sqrt%)	0.01	0.99
PDV (cbrt%)	0.00	1.00
FA (sqrt)	0.00	1.00
FV (cbrt)	0.00	1.00
DA (sqrt)	0.04	0.98
DV (cbrt)	0.11	0.95

ΔLR-χ^2 represents the difference in likelihood ratio for Equation 2.2 with and without an age random slope (test with 2 degrees of freedom: random slope and covariance), visual assessment score (VAS), percent dense area (PDA), percent dense volume (PDV), fat area (FA), fat volume (FV), dense area (DA), dense volume (DV), square root transformed (sqrt), cube root transformed (cbrt).

In a sensitivity analysis, the linear mixed model was fit using weight instead of BMI but it had a worse model fit for almost all density measures Table 2.10. All models (except DV) had a higher log-likelihood when fitting with BMI compared to weight.

Table 2.10: Multivariable linear mixed model fit results for Equation 2.2 using either body mass index or weight.

	Log-likelihood	
	BMI	Weight
VAS (sqrt%)	-105.9	-109.7
PDA (sqrt%)	-147.6	-151.8
PDV (cbrt%)	-171.3	-173.5
FA (sqrt)	-77.8	-81.4
FV (cbrt)	-88.4	-90.1
DA (sqrt)	-180.5	-180.9
DV (cbrt)	-156.3	-153.4

Model fit with body mass index (BMI) or weight. Model: density on age and BMI (between and within) or weight (between and within), with a random per-woman intercept; between-women BMI calculated as the mean BMI for each woman; within-women BMI calculated as the difference between each woman's BMI and her mean BMI; between-women weight calculated as the mean weight for each woman; within-women weight calculated as the difference between each woman's weight and her mean weight. Visual assessment score (VAS), percent dense area (PDA), percent dense volume (PDV), fat area (FA), fat volume (FV), dense area (DA), dense volume (DV), square root transformed (sqrt), cube root transformed (cbrt).

The jointly-fit and age-adjusted between- and within-women associations (Table 2.11) were very similar to those using repeated measures correlation coefficients (Table 2.3), showing the robustness of the estimates using either method.

When a term for BMI gain since age 20yr was added to the linear mixed model, the model fit improved for PDA, PDV, FV and DA (all ΔLR-χ^2 $p<0.05$) (Table 2.12). Within-women effects of BMI on density were almost unchanged when including BMI gain since age 20yr (Table 2.11, Table 2.12). After including BMI gain since age 20yr, between-women associations for BMI became more strongly inversely associated with percent density (coefficient approximately -0.5 to -0.8), and more strongly positively associated with breast fat (coefficient approximately 0.6 to 0.8). BMI became more strongly inversely associated with DA (coefficient -0.1 to -0.5) and less strongly positively associated with DV between-women (coefficient 0.4 to 0.2). BMI gain from age 20yr was positively associated with DA, PDA and PDV (5kg/m^2 increase in BMI

gain since age 20yr associated with 0.61 (95% CI, 0.12 to 1.09), 0.61 (95% CI, 0.21 to 1.02) and 0.47 (95% CI, 0.05 to 0.88) standard deviation increase in breast density (β), respectively), and inversely associated with FV (β=-0.31, 95% CI, -0.62 to 0.00), but less association was seen with DV (β=0.15, 95% CI, -0.29 to 0.59) and FA (β=-0.32, 95% CI, -0.67 to 0.03).

Table 2.11: Multivariable linear mixed model fit results (95% confidence intervals) for density on body mass index (between- and within-women), adjusted for age (Equation 2.2).

Density outcome	Intercept (95% CI)	Age (95% CI) [per 10yr]	BMI (95% CI) [between] [per 5kg/m^2]	BMI (95% CI) [within] [per 5kg/m^2]
VAS (sqrt%)	3.75 (1.88 to 5.61)	-0.19 (-0.56 to 0.19)	-0.51 (-0.68 to -0.35)	-0.27 (-0.44 to -0.10)
PDA (sqrt%)	2.87 (0.57 to 5.17)	-0.05 (-0.53 to 0.43)	-0.46 (-0.63 to -0.30)	-0.32 (-0.59 to -0.05)
PDV (cbrt%)	1.73 (-1.07 to 4.53)	0.12 (-0.48 to 0.71)	-0.39 (-0.57 to -0.21)	-0.85 (-1.32 to -0.39)
FA (sqrt)	-3.63 (-5.25 to -2.02)	0.04 (-0.28 to 0.36)	0.60 (0.46 to 0.74)	0.43 (0.27 to 0.58)
FV (cbrt)	-3.46 (-5.27 to -1.64)	-0.04 (-0.42 to 0.34)	0.63 (0.50 to 0.76)	0.79 (0.56 to 1.03)
DA (sqrt)	0.57 (-2.13 to 3.27)	-0.03 (-0.59 to 0.53)	-0.08 (-0.28 to 0.11)	0.01 (-0.30 to 0.33)
DV (cbrt)	-2.39 (-5.11 to 0.33)	0.09 (-0.48 to 0.66)	0.35 (0.16 to 0.53)	0.16 (-0.24 to 0.55)

Between-women body mass index (BMI) calculated as the mean BMI for each woman; within-women BMI calculated as the difference between each woman's BMI and her mean BMI; density measures are standardised (2.2.4.6); 1 woman with missing BMI at age 20yr excluded. Visual assessment score (VAS), percent dense area (PDA), percent dense volume (PDV), fat area (FA), fat volume (FV), dense area (DA), dense volume (DV), square root transformed (sqrt), cube root transformed (cbrt), 95% confidence interval (95% CI).

Table 2.12: Multivariable linear mixed model fit results (95% confidence intervals) for density on body mass index (between- and within-women) and body mass index gain since 20yr of age, adjusted for age (Equation 2.3).

Density outcome	Intercept (95% CI)	Age (95% CI) [per 10yr]	BMI (95% CI) [between] [per 5kg/m^2]	BMI (95% CI) [within] [per 5kg/m^2]	BMI gain since 20yr of age (95% CI) [per 5kg/m^2]	ΔLR-χ^2 p-value Equation 2.3 vs. Equation 2.2
VAS (sqrt%)	5.47 (3.34 to 7.60)	-0.25 (-0.61 to 0.12)	-0.92 (-1.23 to -0.62)	-0.27 (-0.45 to -0.10)	0.59 (0.20 to 0.97)	0.0031
PDA (sqrt%)	4.90 (2.34 to 7.46)	-0.16 (-0.63 to 0.31)	-0.89 (-1.22 to -0.57)	-0.32 (-0.59 to -0.06)	0.61 (0.21 to 1.02)	0.0033
PDV (cbrt%)	3.35 (0.30 to 6.40)	0.01 (-0.57 to 0.60)	-0.71 (-1.05 to -0.38)	-0.85 (-1.32 to -0.39)	0.47 (0.05 to 0.88)	0.0267
FA (sqrt)	-4.59 (-6.49 to -2.69)	0.08 (-0.24 to 0.40)	0.82 (0.54 to 1.10)	0.43 (0.28 to 0.59)	-0.32 (-0.67 to 0.03)	0.0704
FV (cbrt)	-4.42 (-6.44 to -2.40)	0.01 (-0.37 to 0.38)	0.84 (0.59 to 1.09)	0.79 (0.56 to 1.03)	-0.31 (-0.62 to 0.00)	0.0476
DA (sqrt)	2.58 (-0.48 to 5.64)	-0.14 (-0.70 to 0.41)	-0.51 (-0.90 to -0.12)	0.01 (-0.31 to 0.32)	0.61 (0.12 to 1.09)	0.0145
DV (cbrt)	-1.90 (-4.96 to 1.15)	0.06 (-0.51 to 0.64)	0.24 (-0.12 to 0.60)	0.16 (-0.24 to 0.55)	0.15 (-0.29 to 0.59)	0.4967

Between-women body mass index (BMI) calculated as the mean BMI for each woman; within-women BMI calculated as the difference between each woman's BMI and her mean BMI; BMI gain from age 20yr calculated as the difference between each woman's BMI at baseline and her BMI at age 20yr; density measures are standardised (2.2.4.6); 1 woman with missing BMI at age 20yr excluded. Visual assessment score (VAS), percent dense area (PDA), percent dense volume (PDV), fat area (FA), fat volume (FV), dense area (DA), dense volume (DV), square root transformed (sqrt), cube root transformed (cbrt), 95% confidence interval (95% CI). ΔLR-χ^2 represents the difference in likelihood ratio for Equation 2.3 vs. Equation 2.2.

2.3.8 Exploratory analysis

Finally, in tests of association between breast and bone density, there was little correlation within-women, but there was some indication of a positive between-women correlation for bone density and FV (r=0.26, 95% CI, 0.00 to 0.50), DV (r=0.33, 95% CI, 0.09 to 0.54) and TV (r=0.31, 95% CI, 0.06 to 0.54) (Table 2.4, Table 2.5). Correlations between DXA bone and other body composition measures (between-women) were weak to moderate and only significant for weight, DXA lean mass and impedance measures.

2.4 Discussion

In this dietary weight-loss intervention study amongst premenopausal women, changes in breast fat were seen within-women as they lost weight, but little change was seen in dense tissue. Effective weight-loss during premenopausal years has been shown to reduce risk of postmenopausal breast cancer (248), but this study suggests that the effect is unlikely to be mediated by a reduction in dense breast tissue.

The between-women associations of attained premenopausal BMI and density observed in this study are consistent with the literature. Some of the effects of high attained BMI increasing breast cancer risk may be explained through its relationship with dense tissue. As suggested in this study (and others (161, 164, 165)), high BMI is associated with high DV in premenopausal women. The relationship between BMI and DA is less consistent. Some previous studies suggest an inverse association in premenopausal women (158, 169, 293) (which was also suggested in this study), but other studies have suggested a positive relationship (159). Since the breast is a deposit for adipose tissue, high attained BMI is strongly associated with high levels of FA (158, 159, 169, 293) and FV (161, 164), which in turn leads to an inverse association between BMI and both PDA (158, 159, 169, 293-295) and PDV (161-165). This is expected since increased breast fat contributes to an increased total breast area or volume, which is the denominator in percent density calculations.

A reduction in postmenopausal risk can be seen with effective premenopausal weight-loss (248-250). In a large cohort study of almost 34,000 women, the Iowa women's health study showed that weight gain from 18yr to 30yr followed by weight-loss from 30yr to menopause had a risk comparable to weight maintenance in both time periods (RR=0.61; 95% CI, 0.46 to 0.80 and RR=0.73; 95% CI, 0.64 to 0.84, respectively, relative to women who gained weight during both time periods). Premenopausal women were therefore the target for recruitment in this weight-loss intervention, because they were of an age that is thought to be the most important for reducing breast cancer risk later on in life.

There have been few studies assessing the effect of weight-loss on density. In this study, a positive within-women relationship was seen for short-term BMI change and breast fat, but no association was seen with dense tissue, resulting in an inverse association for percent density. Another dietary intervention trial showed reductions in TA with weight-loss, but unlike this study, they also found reductions in DA. Boyd et al. reported a 5.4% decrease in DA for premenopausal women on a 2 year low-fat, high-carbohydrate diet (n=249) compared with a 2.5% decrease in the control group (n=264) (251). Since dietary interventions similar to Boyd et al. have reported lower blood levels of estradiol and estrone (particularly amongst premenopausal women (296)), one theory for this reduction in dense tissue is that the dietary-induced reduction of oestrogen restricted fibroglandular tissue growth. Therefore, it is possible that a specific diet similar to that of Boyd et al.'s might be necessary to see an effect on dense tissue. Another intervention study in postmenopausal women showed a reduction in PDA after 2 years of dieting or physical exercise (297). Other studies have explored the effect of weight-loss after bariatric surgery on dense tissue. Some studies suggest a decrease in dense tissue with bariatric surgery in premenopausal women (252), whilst others suggest little effect (253). Weight-loss interventions can also be exercise-induced. An intervention study of one year of moderate exercise amongst postmenopausal women reported results similar to those found in this chapter, reporting a decline in FA and FV with moderate to high exercise duration but no significant effect on absolute density (298). The within-women associations of short-term BMI change and density in this chapter are consistent with previous evidence for breast fat, but the effect of dietary weight-loss on dense tissue, particularly in premenopausal women, is still unclear. It may be that specific diets or an extended period of intervention time are required to see an effect on dense tissue.

Adult weight gain over the premenopausal years is a risk factor for postmenopausal breast cancer (173, 247, 299-303). Some evidence in this chapter suggested that increased adult BMI gain was linked with higher dense tissue and percent density, which might partly explain an increase in risk with adult weight gain. In several studies, breast cancer risk from adult weight gain has been limited to (or has been stronger in) women who have never used HRT (299-301). Therefore, it has been suggested that this increase in risk may be oestrogen-related. HRT raises oestrogen levels; hence any breast cancer risk derived from elevated oestrogen would be attenuated in women with already high amounts of the hormone. Risk from adult weight gain may be mediated by higher amounts of dense tissue, which are thought to reflect cumulative lifetime exposure to oestrogen (143), and would explain the results seen in this study. Pollan et al. reported increased premenopausal PDA with adult weight gain (294); however, an inverse association was seen in a study by Samimi et al. (295). Tseng et al. found a positive association between adult weight gain and premenopausal DA, but very little association with PDA and FA

(159). Very few studies have assessed this relationship volumetrically. Alimujiang et al. suggested a positive association with premenopausal DV and FV, but an inverse association with PDV (164). Some of these conflicting results may be explained by the different adjustments used; for instance, some of the studies adjusted for current adiposity (294), some adjusted for adiposity at 18yr (164, 295), whilst others adjusted for both (159). To fully understand the long-term effects of weight on density it would be useful to assess life course effects in a large cohort of women. However, since mammography is not routinely conducted in young premenopausal women, alternative non-ionising methods of measuring density may prove to be more useful (304).

The associations between adipose tissue, dense tissue and breast cancer risk are somewhat contradictory. A high BMI represents elevated amounts of adipose tissue, which is a main site of aromatisation of androgens to oestrogen. Oestrogen is known to promote cell proliferation and carcinogenesis (273, 274), which may explain the positive relationship between BMI, dense volume and breast cancer risk in postmenopausal women whose hormonal production in the ovaries has ceased and whose main oestrogen source is adipose tissue. However, the effect of aromatisation is negligible in premenopausal women whose main source of oestrogen is the ovaries. Moreover, the association between circulating blood serum oestrogen and density is seen in premenopausal (178) but not postmenopausal (180) women. One suggestion for these differences is that systemic oestrogens transported in the blood have an effect on dense tissue growth in premenopausal women, but not postmenopausal women; and local oestrogen from aromatisation in breast fat affects dense tissue development in postmenopausal women but less so in premenopausal women (182). This leads to the idea that adipose tissue has differing effects on dense tissue (and breast cancer risk) whether measured during pre- or postmenopausal years and whether distributed systemically or locally within the breast. The idea of systemic and local breast fat operating through different mechanisms may also explain the contradiction seen in weight-loss studies where decreased BMI reduces risk but also elevates percent density, which is itself associated with an increased risk of breast cancer. Perhaps the reduction in adipose tissue elsewhere in the body offsets the increased risk from a reduction in breast fat.

There is some suggestion that BMI has a protective effect on dense tissue and breast cancer risk in premenopausal women (172, 175, 176). In this study, there was indication of an inverse effect of attained BMI and breast fat on DA, which has been seen previously in premenopausal women (attained BMI: (158, 169, 293); FA: (158, 305)). However, this effect is unclear since the study conversely found a positive effect of attained BMI and breast fat on DV. This positive association between attained BMI and DV has been seen previously in premenopausal women (161, 164, 165), but there have been few studies assessing the relationship between breast fat and DV. There is also some evidence to suggest a protective effect of breast fat on

premenopausal breast cancer risk (90, 175, 246, 272), however, some studies show that the protective effect of BMI on premenopausal risk is reversed after adjustment for percent density (166, 169).

An exploratory analysis tested the association between breast density and DXA bone density. Both are considered to be markers of the cumulative rate of exposure to oestrogen (143, 306) (which is related to breast cancer risk (307)), but their association with each other has shown inconsistent results (308-311). There was some indication of a positive between-women correlation for bone density and FV, DV and TV, but little correlation within-women (which is expected since bone density is unlikely to have changed over a 2 year period).

Strengths of this study include the use of Cumulus and the Stepwedge method which allowed for the assessment of dense and fatty tissue separately as well as volumetrically, which in theory should represent breast tissue more accurately than area-based methods by accounting for overlapping tissue. Additionally, this study used many different measures of body weight to assess adiposity overall and deposited in different parts of the body. All women were encouraged to lose weight (more so in the intervention arm), which provided data with large within-women variation in BMI, providing great potential to measure effects across the study. The intervention also took place at a time in a woman's life that is thought to be the most influential for breast cancer prevention and risk reduction. Additionally, the Lifestyle study provided a data source to assess premenopausal density associations, which is not available in studies of routine screening data. Moreover, the analysis utilised repeated measures correlation coefficients and linear mixed models which are robust techniques that used all of the data and assessed all of the time points simultaneously to provide an overall estimate of effects across the intervention as a whole.

Limitations of the study include the small sample size, which may have reduced power in the study. This may have been particularly relevant for volumetric measures which had a moderate amount of missing data at baseline. To increase statistical power, the two intervention arms were combined, but this limited the ability to determine the effects that were specific to the intervention. There may have also been methodological issues with these volumetric measures since a positive association was only seen between percent density measures and dense area, but not dense volume (except for PDV, which had a modest correlation with DV). Since both percent density and absolute density are risk factors for breast cancer, one would expect their measurements to be positively correlated. Volumetric measures are greatly influenced by breast thickness (312), which undoubtedly changed as BMI changed, and may have increased variation in serial measurements, reducing the accuracy of volumetric estimates. Different breast positioning between serial mammograms may have also introduced variation in both area-based

and volumetric measurements since radiographer technique can change from one examination to the next, causing different levels of breast compression, breast thickness and the amount of breast that is imaged. The latter is particularly relevant in MLO views since these views can capture subcutaneous fat (representing systemic BMI instead of breast fat) (93), which will be more prominent if the breast is adequately pulled onto the x-ray plate. This can make differentiation between breast adipose tissue and subcutaneous fat difficult, hence the effects of local and systemic adipose tissue can be hard to distinguish (313). At present, this is an unavoidable issue with subjective mammography techniques, and the need for image registration is essential for assessment of serial mammography. Another limitation of this study is the use of self-reported weight at age 20yr which may have suffered from recall bias. Nonetheless, a good correlation of 0.87 between recalled weight and actual weight in early adulthood has been reported previously in a similar cohort of young women, suggesting that this is a suitably robust measure (314). Finally, there was no adjustment for other lifestyle factors such as increased physical activity or reduced alcohol intake, which may have had independent effects on density beyond their indirect effect via weight-loss. However, these are unlikely to have confounded the results since no strong nor consistent effects of these variables have been reported previously, hence any influences are likely to be negligible (296).

2.5 Conclusion

This study suggests that premenopausal weight-loss reduced breast fat but did not reduce dense tissue. Short-term premenopausal weight-loss is likely to be linked to a lower postmenopausal breast cancer risk through reductions in adipose tissue but not fibroglandular tissue. This study suggests that density change is unlikely to be a useful biomarker for risk reduction associated with short-term weight-loss.

Chapter 3: Longitudinal modelling of mammographic density for accurate breast cancer risk estimation

3.1 Introduction

As described in Chapter 1, mammographic density is one of the strongest known risk factors for breast cancer (79) and inclusion of BI-RADS density in breast cancer risk models has been shown to improve the accuracy of assessments of individual risk (315). Providing accurate estimations of a woman's risk of breast cancer could help with decisions regarding supplemental screening, risk-reducing behaviour strategies and chemoprevention, as well as providing a prerequisite for risk-stratified screening.

Most studies assessing the effect of breast density on breast cancer risk are based on a density value at a single time point. However, mammographic density is a dynamic trait that decreases with increasing age and BMI, and changes in response to endocrine treatment and hormone replacement therapy. Using a woman's history of density might therefore be more informative for breast cancer risk estimation than density taken at a single point in time.

Several studies have explored the use of two serial breast density values in the assessment of breast cancer risk (254-261), including a recent large US cohort study of over 700,000 women, which showed a small improvement in the discriminatory accuracy of a breast cancer risk prediction model when two BI-RADS density values were used instead of one (AUC 0.640 vs. 0.635) (262). However, including information on the longitudinal history of density with more than two density values may improve risk prediction even further. Of particular importance is the ability to include information on an unlimited number of mammograms that are arbitrarily spaced through time, which reflects a screening environment in practice. No breast cancer risk model currently incorporates such information. The aims of this study are to develop a measure of density based on an individual woman's complete history of density taken at arbitrary time points (longitudinal density), and to assess how much more information this longitudinal density measure provides for risk assessment than a density value taken at a single time point.

3.2 Methods

3.2.1 Study design

3.2.1.1 Study population

This analysis is of a cohort of women from the Kaiser Permanente Washington Breast Cancer Surveillance Consortium breast imaging registry. All women were enrollees of Kaiser Permanente Washington, an integrated healthcare system that provides both insurance and healthcare in Washington State. The data were previously used to assess the long-term performance of breast cancer risk assessment with and without breast density (315). Women in the cohort attended screening from January 1, 1996, through December 31, 2013 (with follow-up from January 1, 1996, through December 31, 2014) with no prior diagnosis of invasive breast cancer or ductal carcinoma in situ (DCIS) at study entry, or lobular carcinoma in situ (LCIS) at baseline mammogram. To ensure that the included cohort represented the screening population, women aged <40 years or >73 years at baseline mammogram were excluded. To also ensure that there were no prevalent breast cancers at the start of the study, women who were diagnosed with DCIS or invasive breast cancer within 6 months after their baseline mammogram were excluded.

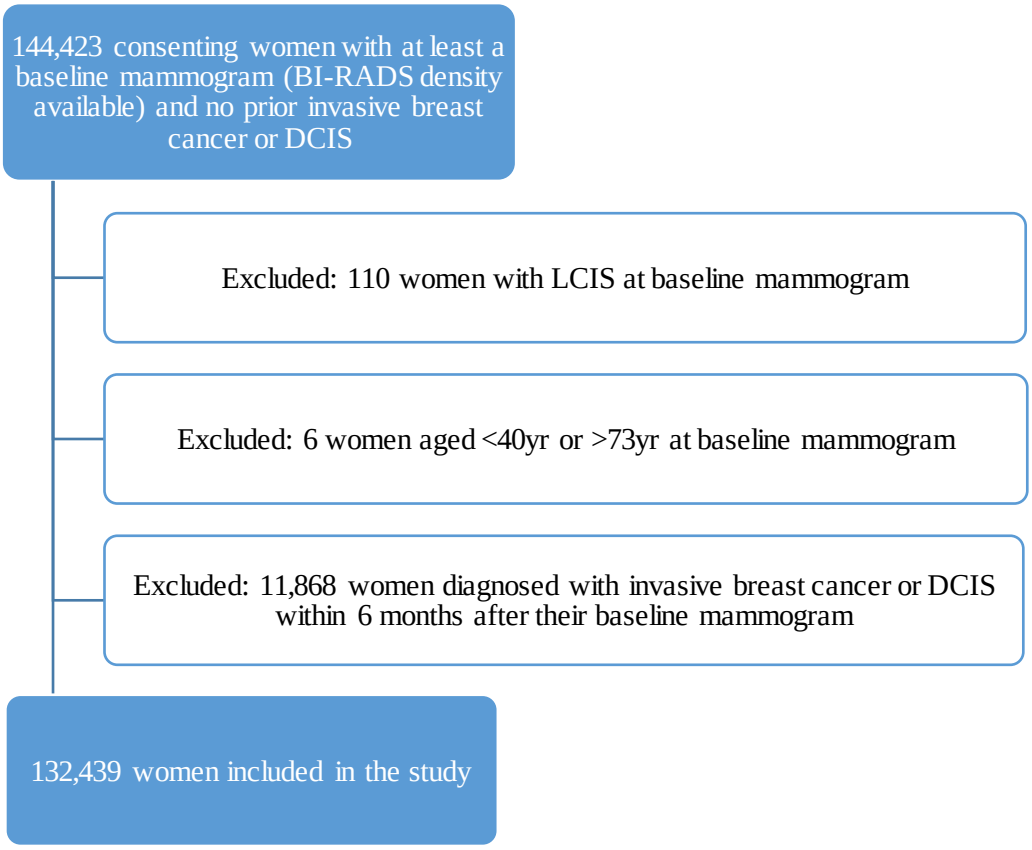


Figure 3.1: Flow diagram of the study population and reasons for exclusions.

3.2.1.2 Endpoints

The primary outcome was diagnosis of invasive breast cancer. Women were followed from their first mammogram with an available density assessment (baseline mammogram) until the earliest of: diagnosis of invasive breast cancer or censoring (at 75 years of age (the recommended end of screening age), December 31, 2014 (the end of calendar time follow-up), diagnosis of DCIS, death, or health plan disenrollment). Outcomes were obtained through linkage with the regional population-based Surveillance, Epidemiology, and End Results (SEER) tumour registry and pathology databases.

3.2.1.3 Exposure variables

Mammographic breast density was recorded at each screening mammogram by the interpreting radiologist using BI-RADS density categories (1=almost entirely fat, 2=scattered fibroglandular, 3=heterogeneously dense, or 4=extremely dense (18)). Only mammograms with a BI-RADS density were included. Self-reported height and weight were collected using a questionnaire completed at each screening mammogram. BMI was derived by dividing weight (kg) by height (m) squared. Values were also checked for validity at the time of scanning for research purposes. Approximately 5% of women who underwent screening opted out of having their questionnaire data used for research and were excluded. To enable a prognostic factor study design, any mammogram taken on the same date as a woman's breast cancer event was removed (no women were excluded because all women had a baseline mammogram at least 6 months before an event, by definition of the study design).

3.2.2 Statistical methods

Analysis was conducted using the statistical software packages Stata (316) and R (282). Statistical tests were two-sided with a significance level of 5%.

3.2.2.1 Missing data

Missing BMI values were imputed to allow for adjustment of density and to ensure that no data points were dropped from the analysis due to missing data. If BMI was unavailable at baseline mammogram, it was imputed using the sample mean BMI given age at the baseline mammogram; otherwise, by carrying forward the last recorded BMI. This was considered to be a robust method because the number of women requiring BMI imputation was small relative to the large sample size (6,047/132,439 women (5%)) (appendix A.I). Imputation was not required for age or density since all mammograms had a matching age and density value (only

mammograms with a BI-RADS density value were included in the analysis, as outlined in section 3.2.1.3). BMI was then winsorised for values below 15kg/m^2 and above 35kg/m^2; hence women who were morbidly obese were given the same risk for adiposity as women who were obese, and extremely underweight women were given a similar risk for adiposity as underweight women.

3.2.2.2 Linear mixed models (2)

A model for longitudinal density was developed by fitting a linear mixed model with BI-RADS density as the outcome (treated as an integer to approximate linear relationships with density). This model uses a similar two-level hierarchical structure as that in Chapter 2, whereby the base level is each density measure at each time point, and the second level is each woman (2.2.4.4).

The linear mixed model for this study is described below. Breast density y_{ij} for woman $i = 1, \ldots, n$ at time $j = 1, \ldots, \mathrm{m}_i$ was modelled as:

$$y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i})age_{ij} + \beta_2 age_{ij}^2 + \beta_3 age_{ij}^3 + \beta_4 age_{ij}^4 + \beta_5 BMI_{ij} + \beta_6 age_{ij} BMI_{ij} + e_{ij};$$

where β_0 is an overall intercept, age_{ij} is the age for woman i at time j, β_1 is the slope for age, β_2 is the slope for age-squared, β_3 is the slope for age-cubed, β_4 is the slope for age to the power of four, BMI_{ij} is the BMI for woman i at time j, β_5 is the slope for BMI, β_6 is the interaction effect for BMI and the linear age term, and e_{ij} is a random error. The term that allowed for differences between-women in their overall density level is the independent random intercept u_{0i} for woman i. The term that allowed for differences between-women in their age slope is the independent random slope u_{1i} for woman i. In other words, the random age slopes allowed each woman to have density trajectories that deviated from the average trajectory through time. Age had a non-linear relationship with density, as has been seen previously (141, 255, 317). The model is completed by assuming normal distributions for $\boldsymbol{u}_i = (u_{0\mathrm{i}}, u_{1\mathrm{i}})$ and e_{ij}, with zero mean, unknown variances and: zero covariance between e_{ij} of the same woman or different women, zero covariance between $\boldsymbol{u}_i$ and e_{ij} of the same woman or different women, zero covariance between $\boldsymbol{u}_i$ of different women, and unknown covariance between u_{0i} and u_{1i} of the same woman. The model was fitted by maximum likelihood (2.2.4.7). To test $\beta_k = 0$ for $k = 0, \ldots, 6$, Wald tests were applied (2.2.4.9).

The linear mixed model building strategy was based on a series of likelihood ratio tests to assess goodness of fit with various polynomial terms and interactions as well as visual assessment of

graphs plotting predicted density against age and BMI. Standard errors for the longitudinal model were calculated using robust sandwich estimators (318-320). These were calculated empirically without making any assumptions on the structure of heteroscedasticity (unequal variance across variable values) in the model.

From this linear mixed model, each woman's random effects, $\boldsymbol{u}_i = (u_{0\mathrm{i}}, u_{1\mathrm{i}})$, were then predicted using Empirical Bayes, as described below (287).

To better understand the Empirical Bayes prediction, it is useful to describe a linear mixed model in its matrix form, whereby, for woman $i = 1, \ldots, n$ with $j = 1, \ldots, \mathrm{m}_i$ time points:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{u}_i + \boldsymbol{e}_i \, ;$$

where:

- $\boldsymbol{y}_i = \left(y_{i1}, \ldots, y_{i\mathrm{m}_i}\right)^T$ is the m_i x 1 column vector of observed outcomes for woman i
- $\boldsymbol{X}_i = \begin{pmatrix} x_{i11} & \cdots & x_{i1a} \\ \vdots & \ddots & \vdots \\ x_{i\mathrm{m}_i1} & \cdots & x_{i\mathrm{m}_ia} \end{pmatrix}$ is the m_i x a design matrix of observed predictors for fixed effects $p = 1, \ldots, a$ for woman i
- $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{\mathrm{a}-1})^T$ is the a x 1 column vector of regression coefficients for fixed effects $p = 1, \ldots, a$
- $\boldsymbol{Z}_i = \begin{pmatrix} x_{i11} & \cdots & x_{i1b} \\ \vdots & \ddots & \vdots \\ x_{i\mathrm{m}_i1} & \cdots & x_{i\mathrm{m}_ib} \end{pmatrix}$ is the m_i x b design matrix of observed predictors for random effects $q = 1, \ldots, b$ for woman i
- $\boldsymbol{u}_i = (u_{0\mathrm{i}}, \ldots, u_{\mathrm{b}-1\,\mathrm{i}})^T$ is the b x 1 column vector of unobserved random effects $q = 1, \ldots, b$ for woman i
- $\boldsymbol{e}_i = \left(e_{i1}, \ldots, e_{i\mathrm{m}_i}\right)^T$ is the m_i x 1 column vector of unobserved random errors for woman i

Under the above model, density measures for woman i have mean = $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{u}_i$ and variance = $\boldsymbol{V}_i = \boldsymbol{Z}_i\boldsymbol{\Sigma}\boldsymbol{Z}_i^T + \sigma^2\boldsymbol{I}_{m_i}$.

Recalling from Chapter 2, $\boldsymbol{\Sigma}$ is the variance-covariance matrix, and $\sigma^2\boldsymbol{I}_{m_i} = \boldsymbol{E}_i$ which is the variance of the residuals for woman i, with σ^2 being the sample residual variance and $\boldsymbol{I}_{m_i}$ being the m_i x m_i identity matrix. The values for $\boldsymbol{y}_i$, $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are measured, and estimates of parameters $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$ and σ^2 are obtained by generalised least squares, which corresponds to maximum likelihood assuming normality of $\boldsymbol{u}_i$ and $\boldsymbol{e}_i$ (2.2.4.4).

Empirical Bayes can be used to predict the random effects, $\boldsymbol{u}_i$, by entering the observed values and estimated parameters into the following equation:

$$\mathbb{E}\left(\boldsymbol{u}_i \mid \boldsymbol{X}_i, \boldsymbol{Z}_i, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}, \widehat{\sigma}^2\right) = \widehat{\boldsymbol{\Sigma}} \boldsymbol{Z}_i^T \widehat{\boldsymbol{V}}_i^{-1} \left(\boldsymbol{y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\beta}}\right)$$

In this study, the above equation was used to predict the random intercept and random slope for each woman $i = 1, \ldots, n$ at each time point $j = 1, \ldots, \mathrm{m}_i$ using data from her most recent and previous observations only. Therefore, each woman's individual observations had a uniquely determined predicted random intercept and random slope.

A predicted density value was then calculated by entering the observed values, estimated parameters and predicted random effects into the equation:

$$\mathbb{E}\left(\boldsymbol{y}_i \mid \boldsymbol{X}_i, \boldsymbol{Z}_i, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}, \widehat{\sigma}^2\right) = \boldsymbol{X}_i \widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i \widehat{\boldsymbol{u}}_i$$

In the analysis, these Bayes predicted density measures are referred to as the longitudinal density measures.

At this point, each observation for each woman had a corresponding baseline density value (the starting value for woman *i*), most recent density value (the updated density value for woman i at time *j*), longitudinal density value (the updated Bayes predicted density for woman i at time *j*), age at baseline (the baseline age for woman *i*), baseline BMI (the baseline BMI for woman *i*), and most recent BMI (the updated BMI for woman i at time *j*).

3.2.2.3 Proportional-hazards Cox models for breast cancer risk

The primary analysis fitted proportional-hazards Cox models to assess the association between the survival time of women and density, age and BMI. Proportional-hazards Cox models were fit for an invasive breast cancer event using three different density measurements: baseline density (model 1), most recent density (model 2) and longitudinal density (model 3). Since all women had a value for baseline density, most recent density and longitudinal density as well as age and BMI at each of her observations, each woman contributed the same number of measurements to each model. Additionally, because the data consisted of repeated measures, the proportional-hazards Cox model equations used time-dependent covariates for density and BMI.

For model 1, the proportional-hazards Cox model equation for woman $i = 1, \dots, n$ is defined as:

$$\lambda_i(t) = \lambda_0(t) exp(\beta_1 age(t_0)_i + \beta_2 BMI(t_0)_i + \beta_3 density(t_0)_i) ;$$

where $\lambda_i(t)$ is the hazard function at time t for woman i; $\lambda_0(t)$ is the baseline hazard function; $t_0 =$ time 0 i.e. baseline; $age(t_0)_i$ is the age at baseline for woman i; $BMI(t_0)_i$ is the BMI at baseline for woman i; $density(t_0)_i$ is the BI-RADS density at baseline for woman i; β_1 is the effect of age at baseline; β_2 is the effect of BMI at baseline; and β_3 is the effect of density at baseline.

For model 2, the proportional-hazards Cox model equation for woman $i = 1, \dots, n$ is defined as:

$$\lambda_i(t) = \lambda_0(t) exp(\beta_1 age(t_0)_i + \beta_2 BMI(t)_i + \beta_3 density(t)_i) ;$$

where $\lambda_i(t)$ is the hazard function at time t for woman i; $\lambda_0(t)$ is the baseline hazard function; $t_0 =$ time 0 i.e. baseline; $age(t_0)_i$ is the age at baseline for woman i; $BMI(t)_i$ is the BMI at time t for woman i; $density(t)_i$ is the BI-RADS density at time t for woman i; β_1 is the effect of age at baseline; β_2 is the effect of BMI; and β_3 is the effect of density.

For model 3, the proportional-hazards Cox model equation for woman $i = 1, \dots, n$ is defined as:

$$\lambda_i(t) = \lambda_0(t) exp(\beta_1 age(t_0)_i + \beta_2 BMI(t)_i + \beta_3 long\_density(t)_i) ;$$

where $\lambda_i(t)$ is the hazard function at time t for woman i; $\lambda_0(t)$ is the baseline hazard function; $t_0 =$ time 0 i.e. baseline; $age(t_0)_i$ is the age at baseline for woman i; $BMI(t)_i$ is the BMI at time t for woman i; $long\_density(t)_i$ is the longitudinal density at time t for woman i; β_1 is the effect of age at baseline; β_2 is the effect of BMI; and β_3 is the effect of longitudinal density.

Each model was fitted by maximum partial likelihood (321). The coefficients β_1, β_2 and β_3 were different for each model and they were each tested using Wald tests (2.2.4.9).

All models were adjusted for age at baseline (per year; continuous). Model 1 was additionally adjusted for BMI at baseline (per kg/m^2; continuous) and models 2 and 3 were instead adjusted for most recent BMI (per kg/m^2; continuous). This was done so that BMI matched the corresponding density value.

3.2.2.4 Model building

To allow for a non-linear relationship between density and risk, density was included as a factor variable for models 1 and 2 (degrees of freedom (df)=3) (corresponding to the BI-RADS density value). Likelihood ratio tests were used to test for a non-linear fit with longitudinal density (2.2.4.8). The best fit for model 3 included longitudinal density modelled as a linear and quadratic variable (df=2). The rationale for including a quadratic longitudinal density term is described in section 3.3.4.

3.2.2.5 Measures of predictive ability (primary analysis)

To assess the predictive ability of each model, likelihood ratio statistics were estimated (2.2.4.8). Additionally, to account for the different number of parameters in each model, Akaike information criterion (AIC) were estimated. This is useful because the more parameters included in a model, the higher the likelihood. However, including too many parameters in a model can lead to overfitting which decreases the generalisability of results. There is a trade-off between developing a model that has both goodness of fit and parsimony. The AIC is a statistic that has been proposed to assess model fit by penalising overfitting. This is done by adding another term to the likelihood, namely:

$$AIC = -2\ell(\theta|\boldsymbol{x}) + 2k\ ;$$

where $\ell(\theta|\boldsymbol{x})$ is the log-likelihood and k is the number of parameters in the model (322).

3.2.2.6 Measures of discriminatory accuracy

To measure the discriminatory accuracy of longitudinal density, a yearly mean at-risk concordance index (yC) was estimated through time. This is a non-standard method that was developed for the purpose of this longitudinal study. The method for calculating the at-risk concordance index is described below.

Survival status for woman $i = 1, \ldots, n$ is denoted (t_i, δ_i) for the time t_i of breast cancer event $(\delta_i = 1)$ or censoring $(\delta_i = 0)$. A risk score, r_{ij}, is determined for each woman $i = 1, \ldots, n$ at each breast cancer event $j = 1, \ldots, m$ using the estimated hazard ratio (HR) from the proportional-hazards model. That is, $r = \exp(\boldsymbol{\beta x})$, where $\boldsymbol{x} = (x_1, x_2, x_3)$ includes age (x_1), BMI (x_2) and breast density (x_3), with corresponding parameters $\boldsymbol{\beta}^T = (\beta_1, \beta_2, \beta_3)$. BMI and breast density values are updated through time, hence x_2 and x_3 are time-varying covariates.

At each breast cancer event $j = 1, \dots, m$ (occurring at time, S_j), a concordance index, C_j, is defined as:

$$C_j = \frac{\sum_{\substack{i=1 \\ i \neq \omega_j}}^{n} I(t_i \geq S_j) \left\{ I(r_{ij} < \tilde{r}_j) + \frac{1}{2} I(r_{ij} = \tilde{r}_j) \right\}}{\sum_{\substack{i=1 \\ i \neq \omega_j}}^{n} I(t_i \geq S_j)} \quad ; \qquad (*)$$

where the risk score of the woman with the event $j = \tilde{r}_j$, and the index of the woman with the event $j = \omega_j$. That is, C_j is the proportion of women with a risk score less than or equal to the risk score of the woman generating the breast cancer event (out of the total number of women still at-risk at the time of the breast cancer event).

Generalising to include ties ($\geq$1 woman ($= \tilde{n}_j$) with an event at the same time, S_j), $(*)$ is calculated separately for each tied woman $k = 1, \dots, \tilde{n}_j$ at event j. The index, ω_j, is extended to be a vector of indices, $\boldsymbol{\omega_j} = (\omega_{j1}, \dots, \omega_{j\tilde{n}_j})$, and the risk score of the tied woman k at event $j = \tilde{r}_{jk}$. Hence:

$$C_{jk} = \frac{\sum_{\substack{i=1 \\ i \notin \boldsymbol{\omega_j}}}^{n} I(t_i \geq S_j) \left\{ I(r_{ij} < \tilde{r}_{jk}) + \frac{1}{2} I(r_{ij} = \tilde{r}_{jk}) \right\}}{\sum_{\substack{i=1 \\ i \notin \boldsymbol{\omega_j}}}^{n} I(t_i \geq S_j)}$$

In the results, a yearly mean concordance index, yC_z, is presented at each yearly interval $z = 1, \dots, 18$, starting at 0.5yr. So, for example, the yearly mean concordance index between 0.5yr and 1.5yr was defined as:

$$yC_1 = \frac{\sum_{j=1}^{m} \sum_{k=1}^{\tilde{n}_j} C_{jk} \left\{ I(0.5 \leq S_j < 1.5) \right\}}{\sum_{j=1}^{m} \tilde{n}_j \left\{ I(0.5 \leq S_j < 1.5) \right\}}$$

The maximum follow-up time was 19yr, hence the final yC_z ($= yC_{18}$) was calculated between 17.5yr and 18.5yr and the 4 women who developed breast cancer $\geq$18.5yr were excluded.

The standard error on each yC_z was calculated by estimating the variance about the mean, yC_z, (variance generated by the women with a breast cancer event, j, occurring in yearly interval, z).

So, for example, for $z = 1$ (between 0.5yr and 1.5yr), the variance about the mean, yC_1, is defined as:

$$y_1 = \frac{\sum_{j=1}^{m}\sum_{k=1}^{\tilde{n}_j}\left(C_{jk} - yC_1\right)^2 \left\{I\left(0.5 \le S_j < 1.5\right)\right\}}{\sum_{j=1}^{m}\tilde{n}_j \left\{I\left(0.5 \le S_j < 1.5\right)\right\}}$$

Hence, the standard error (SE) on yC_1 is calculated as:

$$SE_1 = \sqrt{\frac{y_1}{\sum_{j=1}^{m}\tilde{n}_j \left\{I\left(0.5 \le S_j < 1.5\right)\right\}}}$$

An overall concordance index (mean concordance index across the entire follow-up) was also calculated, and a 95% confidence interval was estimated using an empirical bootstrap of the mean with 10,000 resamples (2.2.4.3).

3.2.2.7 Assessment of risk stratification

To assess the effect of using longitudinal history of density on risk stratification, the distribution of observed risk based on the proportional-hazards model was assessed using histograms at 6 months, 5yr, 10yr and 15yr. Observed risk (i.e. relative hazard ratio (HR)) was generated by calculating each woman's risk score at each mammogram (defined as r_{ij} from section 3.2.2.6) relative to the average risk at 6 months. The proportion of lowest risk women (<1/2 relative HR) and highest risk women (≥2 relative HR) were plotted at 6 months, 5yr, 10yr and 15yr to assess the distribution of risk through time for the most extreme risk categories. The greater the spread of risk, the greater the ability for risk stratification. This analysis was also conducted in subgroups of women aged 40-49/50-59/≥60 years at baseline.

3.2.2.8 Secondary analyses

To assess whether predictive ability varied for different subgroups of women, a series of secondary analyses calculated likelihood ratio statistics throughout the follow-up for models 1-3 in:

- Women aged 40-44/45-49/50-54/55-59/60-64/≥65 years at baseline.
- Women with baseline mammogram before or after 2007 (as a proxy for film or digital mammography).
- Women with baseline mammogram before or after 2003 (as a proxy for the 3rd or 4th BI-RADS density lexicon (18).

- Premenopausal or postmenopausal women at baseline.
- Women <60 years old at baseline with baseline mammogram taken before the year 2000 (younger women starting the trial early enough to have a long follow-up before being censored at 75 years of age).
- Premenopausal women <60 years old at baseline with baseline mammogram taken before the year 2000 (premenopausal women whose long follow-up was likely to include their transition into postmenopausal status).

It was hypothesised that the random slopes (representing the likely future trajectory of density for each woman) would further improve the statistical output of the longitudinal model. Another secondary analysis tested the inclusion of the random slopes in model 3 by a likelihood ratio test (2.2.4.8). Furthermore, likelihood ratio tests were conducted to assess the benefit of an interaction between longitudinal density and age or BMI in model 3.

A final secondary analysis tested the predictive ability, discriminatory ability and capacity for risk stratification of longitudinal density in the subgroup of women with at least 3 mammograms (women with an adequate history of density), starting follow-up at their third mammogram.

3.2.2.9 Sensitivity analyses

To test the influence of BMI imputation on the results, predictive ability was also assessed in models 1-3 after removing mammograms with a missing corresponding BMI. Here, the start of follow-up began at the (potentially) new baseline mammogram for each woman. Additionally, to test the influence of screen-detected mammograms on results, predictive ability was also assessed in models 1-3 after removing mammograms taken within 6 months before a breast cancer event. Each woman had the same baseline mammogram (no breast cancer event occurred within 6 months after the baseline mammogram, by definition of the study design), so the start of follow-up remained the same as the primary analysis.

3.3 Results

Table 3.1: Univariate hazard ratios of age, body mass index and BI-RADS density baseline variables.

Baseline Variable	No. (%) of Women	Follow-up, 1000 Women-years	No. of Invasive Breast Cancer Cases	Incidence Rate per 1000 Women/yr	Univariate Hazard Ratio (95% CI)	LR-$\chi^2(1)$ Trend Test	P
Total	132,439 (100)	941	2,704	2.9	-	-	-
Age (yr)							
40-49	60,325 (45.6)	448	977	2.2	1 [Reference]	309.1	<0.001
50-59	43,878 (33.1)	339	1,057	3.1	1.41 (1.29-1.53)		
≥60	28,236 (21.3)	153	670	4.4	2.26 (2.04-2.50)		
BMI (kg/m^2)							
<18.5	1,657 (1.3)	11	25	2.3	0.88 (0.59-1.31)	5.5	0.019
≥18.5 to 25	48,713 (36.8)	355	931	2.6	1 [Reference]		
≥25 to 30	42,868 (32.4)	299	962	3.2	1.24 (1.13-1.36)		
≥30 to 35	20,791 (15.7)	146	415	2.8	1.09 (0.97-1.23)		
≥35	18,410 (13.9)	129	371	2.9	1.11 (0.99-1.25)		
BI-RADS density							
Fatty	10,387 (7.8)	66	107	1.6	0.70 (0.57-0.85)	91.8	<0.001
Scattered	46,206 (34.9)	332	786	2.4	1 [Reference]		
Heterogeneous	57,158 (43.2)	405	1,338	3.3	1.40 (1.28-1.53)		
Extremely dense	18,688 (14.1)	137	473	3.4	1.44 (1.29-1.62)		

Hazard Ratios from a Proportional-hazards Cox model; 95% confidence intervals (CIs) from Wald tests; LR-$\chi^2(1)$ trend test: represents the difference in likelihood ratio statistics (LR-χ^2) between the null model and a model fit to the covariate (age and body mass index (BMI) fit as continuous variables; BI-RADS density fit as an integer); P-value from LR-$\chi^2(1)$ trend test.

Table 3.2 Longitudinal model fit for BI-RADS density (integer) on age (continuous) and body mass index (continuous)

Fixed effects				
Variable	β-coefficient	Robust standard error**		
Intercept	2.9659	0.0044		
Age (per 5yr)	0.0491	0.0079		
Age^2 (per 5^2yr)	-0.0780	0.0048		
Age^3 (per 5^3yr)	0.0127	0.0011		
Age^4 (per 5^4yr)	-0.0006	0.0001		
BMI (per kg/m^2)	-0.0584	0.0005		
Age x BMI (per 5yr; per kg/m^2)	0.0033	0.0001		
Random effects				
Variable	Standard Deviation	Robust standard error**	Correlation	Robust standard error**
Intercept	0.5865	0.0027	-0.4590	0.0068
Age (per 5yr)	0.1107	0.0010		

***Standard errors calculated using robust sandwich estimators. Age from 40yr, body mass index (BMI) from 25kg/m^2.*

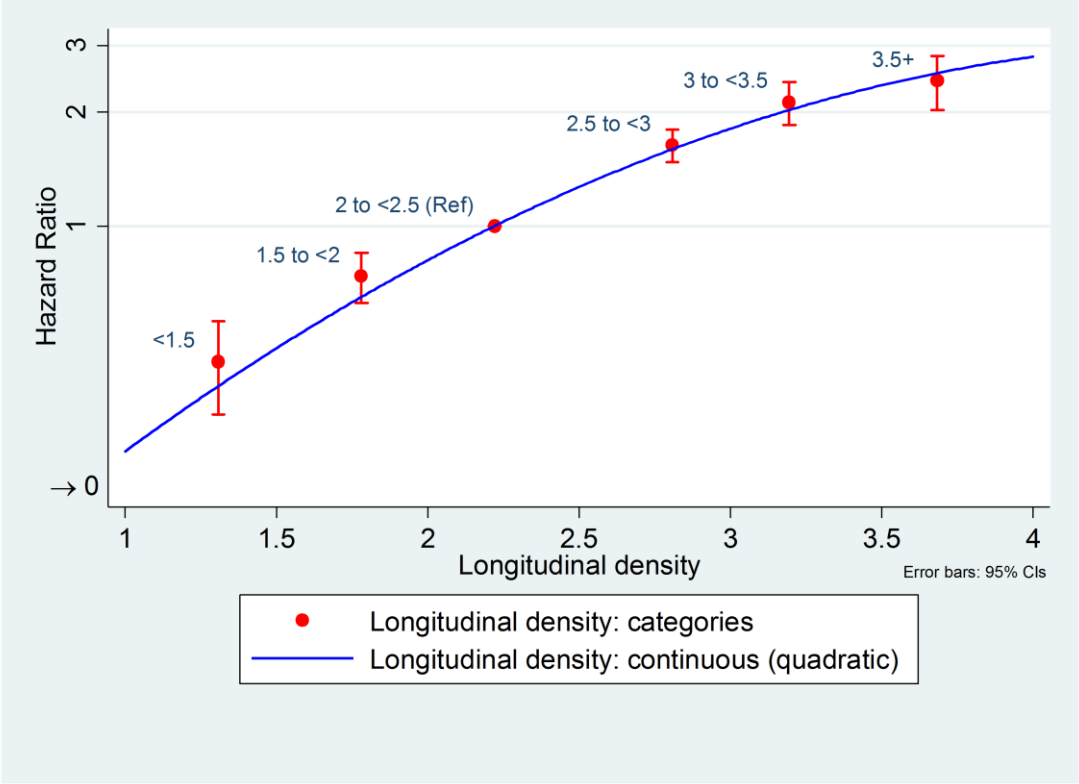


Figure 3.2: Adjusted Hazard Ratios for longitudinal density: continuous and categorical.

Hazard Ratios (HRs) from Proportional-hazards Cox models for longitudinal density: categorised into 6 arbitrary groups (relative to HR for longitudinal density group '2 to <2.5', plotted against mean longitudinal density in each group (x-axis)) and as a continuous variable (including a quadratic term, relative to HR for mean longitudinal density in group '2 to <2.5' (longitudinal density=2.22)); adjusted for age at baseline (continuous) and most recent body mass index (BMI) (continuous), centred at 40yr at baseline and BMI of 25kg/m^2; y-axis on a log-scale; 95% confidence intervals (CIs) from Wald tests.

3.3.1 Details of the cohort

In total, 132,439 women were included with a median follow-up of 5.2 years (interquartile range (IQR), 2.4-11.1 years) and maximum follow-up of 19 years. Younger women entered the cohort earlier and thus had greater follow-up (for example, median of 10.8 years (IQR, 3.8-17.2 years) for 46,484 women younger than 60 years with baseline mammogram before 2000). Median time between mammograms was 1.8 years (IQR, 1.0-2.0 years) and the median number of mammograms per woman was 3 (IQR, 2-6), with 32,010 women (24.2%) having a baseline mammogram only. The number of mammograms was similar across different ages at baseline and throughout the follow-up (appendix A.II). In total, 2704 women (2.0%) were diagnosed with invasive breast cancer during the follow-up.

3.3.2 Baseline characteristics

At baseline, median age was 50 years (IQR, 44-58 years), median BMI was 26.8 kg/m^2 (IQR, 23.2-31.1 kg/m^2) and the majority of women had dense breasts (75,846 (57.3%) of women had heterogeneous or extremely dense breasts). In the univariate model of variables at baseline, most statistical information was in age (LR-$\chi^2(1)$=309.1), followed by density (LR-$\chi^2(1)$=91.8) then BMI (LR-$\chi^2(1)$=5.5) (all $p<0.05$) (Table 3.1).

3.3.3 Model building of the linear mixed model

Hereafter, the model building strategy for the linear mixed model is described. Initially, the model included fixed effects for an intercept, age (per 5yr; continuous) and most recent BMI (per kg/m^2; continuous). Age was centred at 40yr and modelled per 5yr to aid interpretation of the age coefficient, and BMI was centred at 25kg/m^2. Random effects were also included for the intercept and age effect to account for each woman's deviation from the population mean density and population mean age effect. Model fit improved considerably with an unstructured covariance matrix allowing for correlation between the random effects (ΔLR-$\chi^2(1)$=2473, $p=2x10^{-308}$). Age was modelled as a quartic polynomial because this was determined to be the best fit after likelihood ratio tests of goodness of fit and visual assessment of plotted graphs (appendix A.III). For example, an improvement in model fit was seen until the 7th power of age (all $p<0.05$). However, when assessing plots of predicted density against age, polynomial terms past the 4th power appeared to be overfitting to the data for ages above 73yr (density is expected to decrease or plateau but it increased for these older women). For BMI, including a quadratic age term had little improvement on model fit (ΔLR-$\chi^2(1)$=4, $p=0.06$); but a cubic term improved the fit considerably (ΔLR-$\chi^2(1)$=655, $p=1x10^{-144}$). This improvement continued until the 10th power of BMI (all $p<0.05$). However, the plots for predicted density against BMI were somewhat unchanged when quadratic or cubic terms were added and there was evidence of

overfitting with BMI to the 4th power or more due to the peaking curve for the relationship between BMI and density (appendix A.IV). Thus, no polynomial BMI terms were included in the model. An interaction between the linear age term and BMI further improved model fit ($\Delta LR\text{-}\chi^2(1)=619$, $p=1x10^{-136}$) and was therefore included. In order to minimise overfitting and to create a parsimonious model, only fixed effects were tested for polynomials and only the linear age fixed effect was tested for an interaction with BMI. Table 3.2 shows the results for the final linear mixed model.

3.3.4 Model building of the proportional-hazards Cox models

In the proportional-hazards Cox model (model 3), the addition of a quadratic longitudinal density term improved model fit (relative to the model including linear longitudinal density only: $\Delta LR\text{-}\chi^2(1)=15.0$, $p<0.001$), but a cubic term did not improve model fit further ($\Delta LR\text{-}\chi^2(1)=1.1$, $p=0.3$). This non-linear relationship could also be seen in a plot of hazard ratios against longitudinal density. Figure 3.2 shows that the proportional-hazards model including a quadratic term for continuous longitudinal density mirrored the curved relationship seen between risk and longitudinal density when it was categorised into 6 arbitrary groups. There was little difference in risk between the highest 2 groups of longitudinal density; further showing the need for a quadratic term to capture the attenuating rate of increasing risk for the higher longitudinal density values. A non-linear relationship was also tested between breast cancer risk and age at baseline, BMI at baseline or most recent BMI, but model fit did not improve with inclusion of polynomial terms for these variables (likelihood ratio tests had $p>0.05$).

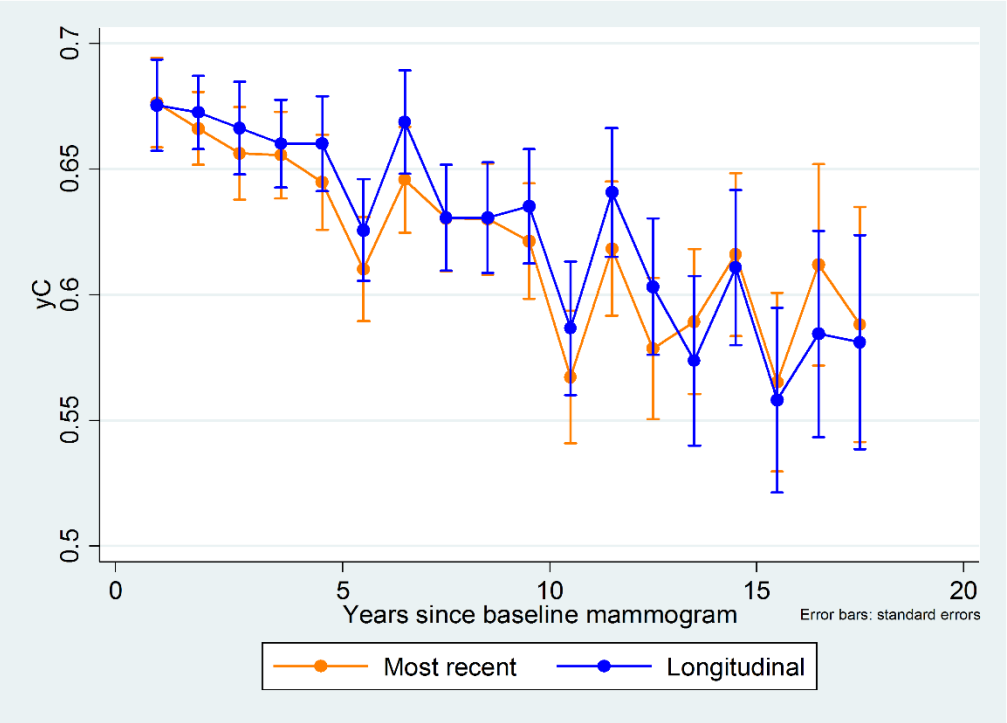
Table 3.3: Multivariable hazard ratios and statistical information on model fit from proportional-hazards Cox models using different breast density measures

Model	BMI	Density	No. of women at baseline (%)	Multivariable Hazard Ratio (95% CI)	df (model)	LR-χ^2 (model)	AIC (model)	df (density)	ΔLR-χ^2 (density)	ΔAIC (density)
1	Baseline	Baseline BI-RADS			5	607.4	58,361.0	3	296.2	290.2
		Fatty	10,387 (7.8)	0.59 (0.48-0.72)						
		Scattered	46,206 (34.9)	1 [Reference]						
		Heterogeneous	57,158 (43.2)	1.76 (1.61-1.93)						
		Extremely dense	18,688 (14.1)	2.31 (2.04-2.63)						
2	Most recent	Most recent BI-RADS			5	624.6	58,343.8	3	307.7	301.7
		Fatty	10,387 (7.8)	0.49 (0.40-0.60)						
		Scattered	46,206 (34.9)	1 [Reference]						
		Heterogeneous	57,158 (43.2)	1.71 (1.56-1.86)						
		Extremely dense	18,688 (14.1)	2.11 (1.84-2.42)						
3	Most recent	Longitudinal density (continuous)			4	696.5	58,269.9	2	379.6	375.6
		Linear (per unit)	132,439 (100)	5.53 (3.34-9.14)						
		Quadratic (per unit^2)	132,439 (100)	0.83 (0.76-0.92)						
4	Most recent	Longitudinal density (4 category)			5	629.3	58,339.2	3	312.3	306.3
		'Fatty'	10,383 (7.8)	0.59 (0.51-0.68)						
		'Scattered'	46,208 (34.9)	1 [Reference]						
		'Heterogeneous'	57,160 (43.2)	1.67 (1.52-1.83)						
		'Extremely dense'	18,688 (14.1)	2.15 (1.89-2.46)						

Table 3.3 continued

5	Most recent	Longitudinal density (8 category)			9	692.1	58,284.4	7	375.1	361.1
		'Fatty' I	5,191 (3.9)	0.48 (0.36-0.63)						
		'Fatty' II	5,192 (3.9)	0.83 (0.69-1.00)						
		'Scattered' I	23,103 (17.4)	1 [Reference]						
		'Scattered' II	23,105 (17.5)	1.43 (1.24-1.64)						
		'Heterogeneous' I	28,569 (21.6)	1.82 (1.59-2.07)						
		'Heterogeneous' II	28,591 (21.6)	2.44 (2.11-2.81)						
		'Extremely dense' I	9,341 (7.1)	2.66 (2.23-3.17)						
		'Extremely dense' II	9,347 (7.1)	3.03 (2.48-3.70)						

Hazard Ratios from Proportional-hazards Cox models for baseline BI-RADS density (model 1), most recent BI-RADS density (model 2), continuous longitudinal density (model 3), 4 category longitudinal density (cut-points chosen so that 4 category longitudinal density and BI-RADS density have the same distribution of women at baseline (frequency matched)) and 8 category longitudinal density (cut-points chosen so that the distribution in each 4 category longitudinal density is halved); longitudinal density: predicted density for each woman from linear mixed model; all models adjusted for age at baseline; baseline density additionally adjusted for baseline body mass index (BMI); most recent density and longitudinal density (continuous, 4 category and 8 category) additionally adjusted for most recent BMI; age, BMI and continuous longitudinal density fit as continuous variables; continuous longitudinal density fit with a quadratic term; baseline, most recent, 4 category longitudinal and 8 category longitudinal density fit as factor variables; 95% confidence intervals (CIs) from Wald tests; ΔLR-χ^2 represents the difference in likelihood ratio statistics (LR-χ^2) between a model fit to age and BMI and a model additionally incorporating the density term(s); ΔAIC represents the difference in Akaike Information Criterion (AIC) between a model fit to age and BMI and a model additionally incorporating the density term(s).



Time (yr)		0-5	5-10	10-15	15+
No. at start of period		132,439	68,307	37,302	19,563
No. of events		1,159	864	487	194
ΔLR-χ^2 (density)*	**Most recent**	187.2	98.7	31.1	13.9
	Longitudinal	221.7	133.6	43.6	8.3

Figure 3.3: Yearly mean concordance index (yC) through time for most recent density and longitudinal density.

**ΔLR-χ^2 represents the difference in likelihood ratio statistics (LR-χ^2) between a model fit to age at baseline and most recent body mass index (BMI) and a model additionally incorporating the density term(s).*

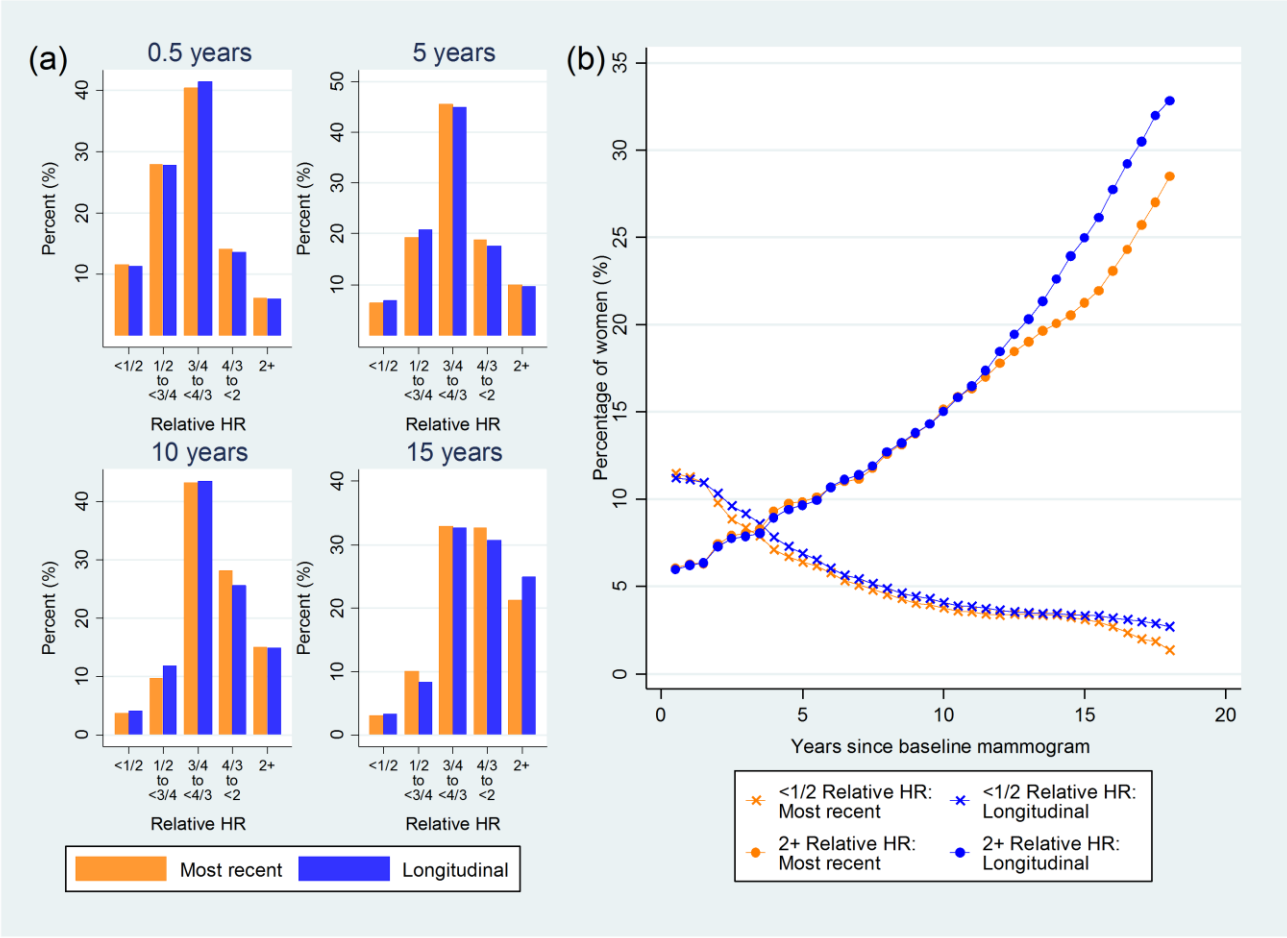


Figure 3.4: Comparison of observed relative risk distributions for most recent density and longitudinal density.

(a) Histograms showing the distribution of relative hazard ratios (HRs) for each model at 0.5yr, 5yr, 10yr and 15yr (HRs relative to the average HR at 0.5yr for each model); (b) Graph showing the percentage of women in the lowest (<1/2 relative HR) and highest (2+ relative HR) risk groups at each 6 month period.

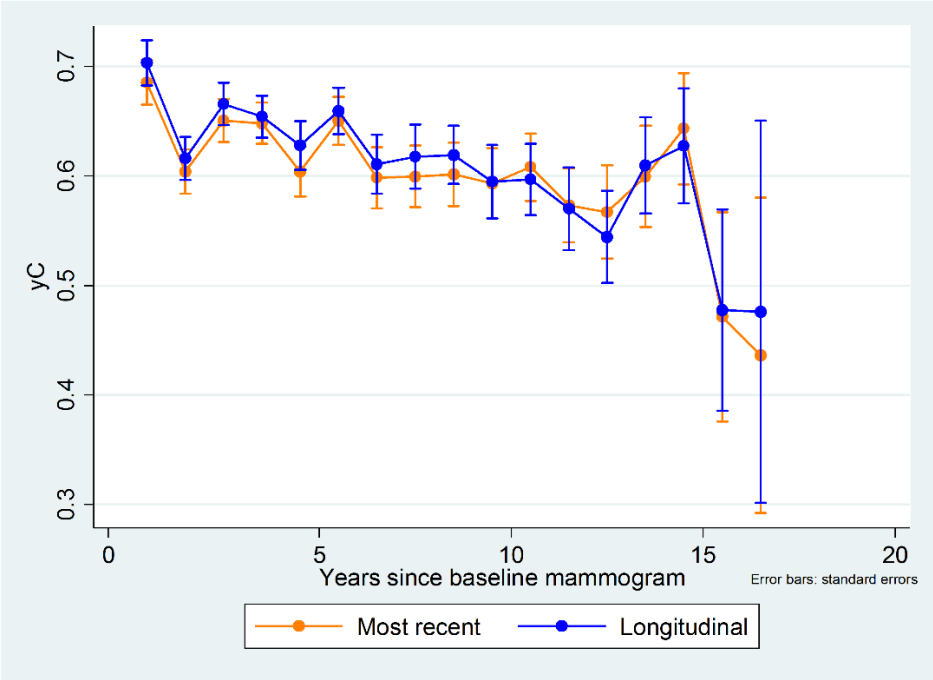
Table 3.4: Multivariable hazard ratios and statistical information on model fit from proportional-hazards Cox models using different breast density measures (subgroup of women with at least 3 mammograms (n=76,313), starting follow-up at third mammogram)

Model	BMI	Density	No. of women at baseline (%)	Multivariable Hazard Ratio (95% CI)	df (model)	LR-χ^2 (model)	AIC (model)	df (density)	ΔLR-χ^2 (density)	ΔAIC (density)
1	Baseline	Baseline BI-RADS			5	397.8	45,677.0	3	178.3	172.3
		Fatty	4,901 (6.4)	0.68 (0.54-0.84)						
		Scattered	26,151 (34.3)	1 [Reference]						
		Heterogeneous	36,351 (47.6)	1.65 (1.49-1.83)						
		Extremely dense	8,910 (11.7)	2.11 (1.83-2.43)						
2	Most recent	Most recent BI-RADS			5	411.6	45,663.2	3	186.1	180.1
		Fatty	4,901 (6.4)	0.53 (0.42-0.66)						
		Scattered	26,151 (34.3)	1 [Reference]						
		Heterogeneous	36,351 (47.6)	1.58 (1.43-1.74)						
		Extremely dense	8,910 (11.7)	1.85 (1.58-2.17)						
3	Most recent	Longitudinal density (continuous)			4	477.9	45,594.9	2	252.5	248.5
		Linear (per unit)	76,313 (100)	5.35 (3.10-9.24)						
		Quadratic (per unit^2)	76,313 (100)	0.82 (0.74-0.91)						
4	Most recent	Longitudinal density (4 category)			5	434.4	45,640.4	3	208.9	202.9
		'Fatty'	4,899 (6.4)	0.51 (0.41-0.64)						
		'Scattered'	26,152 (34.3)	1 [Reference]						
		'Heterogeneous'	36,348 (47.6)	1.63 (1.48-1.81)						
		'Extremely dense'	8,914 (11.7)	2.09 (1.78-2.46)						

Table 3.4 continued

5	Most recent	Longitudinal density (8 category)			9	465.2	45,617.6	7	239.8	225.8
		'Fatty' I	2,449 (3.2)	0.49 (0.35-0.69)						
		'Fatty' II	2,450 (3.2)	0.65 (0.48-0.88)						
		'Scattered' I	13,080 (17.1)	1 [Reference]						
		'Scattered' II	13,072 (17.1)	1.28 (1.10-1.50)						
		'Heterogeneous' I	18,178 (23.8)	1.69 (1.47-1.95)						
		'Heterogeneous' II	18,170 (23.8)	2.21 (1.90-2.58)						
		'Extremely dense' I	4,457 (5.8)	2.46 (1.99-3.05)						
		'Extremely dense' II	4,457 (5.8)	2.57 (2.03-3.25)						

Hazard Ratios from Proportional-hazards Cox models for baseline BI-RADS density (model 1), most recent BI-RADS density (model 2), continuous longitudinal density (model 3), 4 category longitudinal density (cut-points chosen so that 4 category longitudinal density and BI-RADS density have the same distribution of women at baseline (frequency matched)) and 8 category longitudinal density (cut-points chosen so that the distribution in each 4 category longitudinal density is halved); longitudinal density: predicted density for each woman from linear mixed model; all models adjusted for age at baseline; baseline density additionally adjusted for baseline body mass index (BMI); most recent density and longitudinal density (continuous, 4 category and 8 category) additionally adjusted for most recent BMI; age, BMI and continuous longitudinal density fit as continuous variables; continuous longitudinal density fit with a quadratic term; baseline, most recent, 4 category longitudinal and 8 category long itudinal density fit as factor variables; 95% confidence intervals (CIs) from Wald tests; ΔLR-χ^2 represents the difference in likelihood ratio statistics (LR-χ^2) between a model fit to age and BMI and a model additionally incorporating the density term(s); ΔAIC represents the difference in Akaike Information Criterion (AIC) between a model fit to age and BMI and a model additionally incorporating the density term(s).



Time (yr)		0-5	5-10	10-15	15+
No. at start of period		76,313	41,326	21,731	3,743
No. of events		1,308	556	281	24
ΔLR-χ^2 (density)*	Most recent	119.1	57.4	15.4	6.4
	Longitudinal	168.5	78.3	13.8	3.2

Figure 3.5: Yearly mean concordance index (yC) through time for most recent density and longitudinal density (subgroup of women with at least 3 mammograms (n=76,313), starting follow-up at third mammogram).

**ΔLR-χ^2 represents the difference in likelihood ratio statistics (LR-χ^2) between a model fit to age at baseline and most recent body mass index (BMI) and a model additionally incorporating the density term(s).*

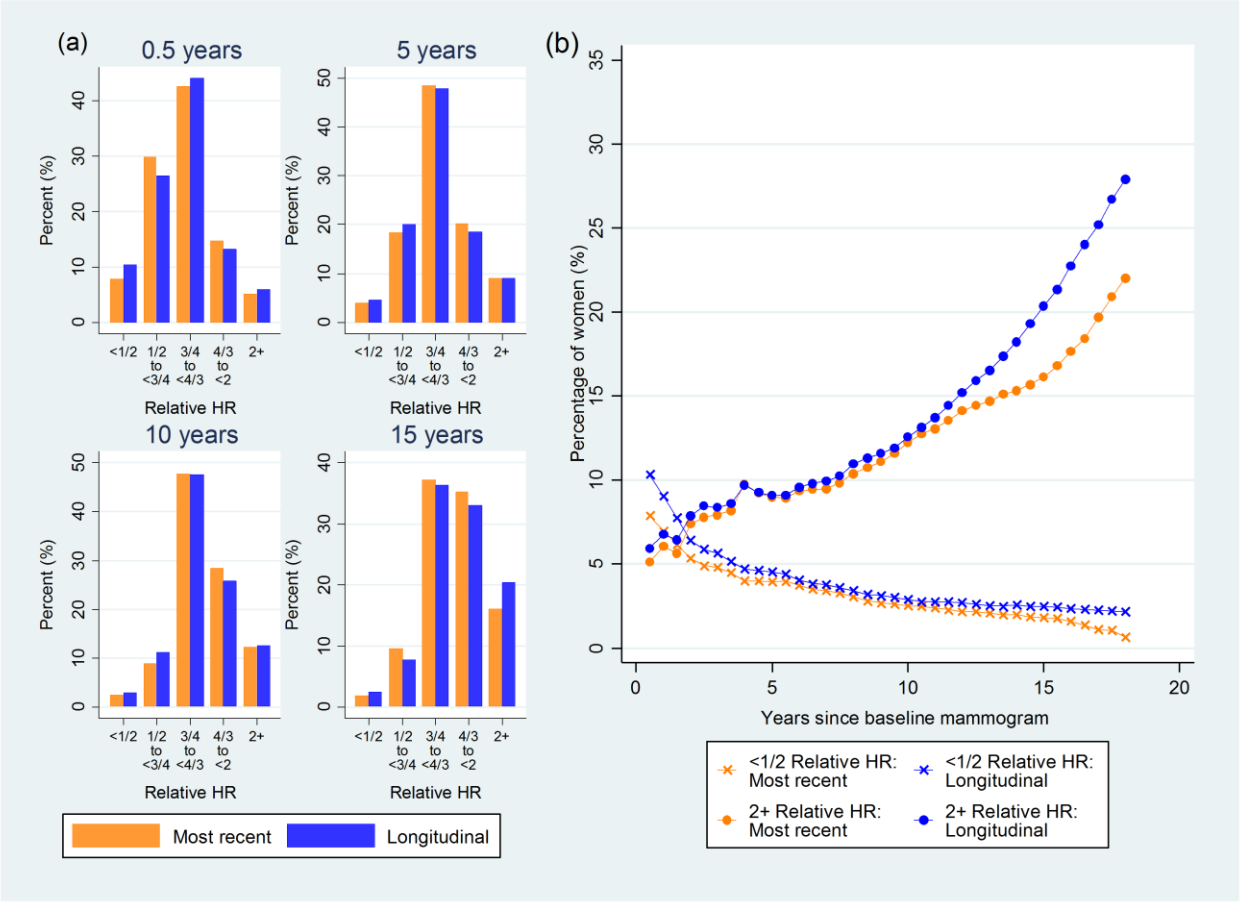


Figure 3.6: Comparison of observed relative risk distributions for most recent density and longitudinal density (subgroup of women with at least 3 mammograms (n=76,313), starting follow-up at third mammogram).

(a) Histograms showing the distribution of relative hazard ratios (HRs) for each model at 0.5yr, 5yr, 10yr and 15yr (HRs relative to the average HR at 0.5yr for each model); (b) Graph showing the percentage of women in the lowest (<1/2 relative HR) and highest (2+ relative HR) risk groups at each 6 month period.

3.3.5 Measures of predictive ability

As expected, baseline, most recent and longitudinal density were strongly associated with risk (Table 3.3). Longitudinal density added 28% more statistical information to a model with age and BMI than baseline density (ΔLR-$\chi^2(2)$=379.6 vs. ΔLR-$\chi^2(3)$=296.2, respectively) and 23% more statistical information than most recent density (ΔLR-$\chi^2(2)$=379.6 vs. ΔLR-$\chi^2(3)$=307.7, respectively). Similar results were observed with AIC statistics. Moreover, this improvement was attained with fewer degrees of freedom. Longitudinal density was also fit as a categorical variable to compare the model performance at different levels of granularity (frequency matched with baseline BI-RADS density). Eight-category longitudinal density had better model fit than four-category longitudinal density (ΔLR-$\chi^2(7)$=375.1 vs. ΔLR-$\chi^2(3)$=312.3), which improved further (and with fewer degrees of freedom) when fitting a continuous variable (ΔLR-$\chi^2(2)$=379.6). The gradient of risk for longitudinal density also increased with finer granularity. With eight-category longitudinal density, the densest breasts had a six-fold greater risk than the fattiest breasts ('Extremely dense' II HR=3.03 (95% CI, 2.48 to 3.70) vs. 'Fatty' I HR=0.48 (95% CI, 0.36 to 0.63)), but only a four-fold increased risk was seen with baseline density (Extremely dense HR=2.31 (95% CI, 2.04 to 2.63) vs. Fatty HR=0.59 (95% CI, 0.48 to 0.72)) and most recent density (Extremely dense HR=2.11 (95% CI, 1.84 to 2.42) vs. Fatty HR=0.49 (95% CI, 0.40 to 0.60)).

3.3.6 Measures of discriminatory accuracy

Figure 3.3 compares the yearly mean concordance index measures for most recent density (model 2) and longitudinal density (model 3). Baseline density was not included at this stage because it was the worst performing density measure in terms of predictive ability. In the first 13 years of follow-up, longitudinal density had better discriminatory accuracy than most recent density, as reflected in the likelihood ratio statistics for the first 15 years of follow-up. Of note, the concordance index measures differed at baseline between most recent density and longitudinal density because longitudinal density is adjusted for age and BMI and therefore takes into account their population effects. The overall mean concordance index was 0.634 (95% CI, 0.623 to 0.645) for most recent density and 0.642 (95% CI, 0.631 to 0.652) for longitudinal density.

3.3.7 Assessment of risk stratification

In Figure 3.4, the relative distributions of risk using most recent and longitudinal density were similar, with a comparable proportion of women categorised as highest and lowest risk using either model 2 or 3. Therefore, the ability for risk stratification was somewhat similar when using longitudinal density or most recent density. When stratified by age at baseline,

longitudinal density categorised a greater proportion of women as high- and low-risk than most recent density in women 60 years or older (appendix A.V-A.VII).

3.3.8 Secondary analyses

In subgroup analyses, longitudinal density continued to provide greater statistical information than baseline density and most recent density regardless of age at baseline, before or after 2007 (proxy for film or digital mammography), before or after 2003 (proxy for the 3rd or 4th BI-RADS density lexicon), or menopausal status. Longitudinal density also provided greater statistical information than baseline density and most recent density when assessed in the subgroup of women with long follow-up, including those likely to have been transitioning from premenopausal to postmenopausal status (appendix A.X-A.XV).

Tests for the addition of density trajectories to the longitudinal model (LR-$\chi^2(5)$=700.5 and AIC=58,267.9) showed only a small improvement on model 3 in terms of predictive ability (ΔLR-$\chi^2(1)$=4.0, p=0.046, ΔAIC=2.0). This resulted in an additional 1% statistical information output than using longitudinal density alone. Tests for an interaction between longitudinal density and age or BMI were not significant (HR for the interaction with age at baseline=1.00 (95% CI, 0.99 to 1.01), p=0.5; HR for the interaction with BMI=1.02 (95% CI, 1.00 to 1.03), p=0.08). Exploratory tests for an interaction between baseline density and age at baseline or BMI at baseline were not significant (p>0.05). Similarly, exploratory tests for an interaction between most recent density and age at baseline or most recent BMI were not significant (p>0.05).

When analyses were restricted to women with at least three mammograms (n=76,313 (58% of the cohort), 2,169 invasive breast cancers), statistical information output increased to 42% more than baseline density (compared with a model including age and BMI only: ΔLR-$\chi^2(2)$=252.5 vs. ΔLR-$\chi^2(3)$=178.3, respectively) and 36% more than most recent density (compared with a model including age and BMI only: ΔLR-$\chi^2(2)$=252.5 vs. ΔLR-$\chi^2(3)$=186.1, respectively), with similar results for AIC statistics (Table 3.4). Longitudinal density had better model fit as an eight-category variable than a four-category variable (ΔLR-$\chi^2(7)$=239.8 vs. ΔLR-$\chi^2(3)$=208.9), and an even better fit as a continuous variable (ΔLR-$\chi^2(2)$=252.5). Risk gradient between the densest and fattiest breasts was five-fold with eight-category longitudinal density ('Extremely dense' II HR=2.57 (95% CI, 2.03 to 3.25) vs. 'Fatty' I HR=0.49 (95% CI, 0.35 to 0.69)), but only three-fold with baseline density (Extremely dense HR=2.11 (95% CI, 1.83 to 2.43) vs. Fatty HR=0.68 (95% CI, 0.54 to 0.84)) and most recent density (Extremely dense HR=1.85 (95% CI, 1.58 to 2.17) vs. Fatty HR=0.68 (95% CI, 0.54 to 0.84)). The overall mean concordance index measures for most recent and longitudinal density were 0.623 (95% CI,

0.611 to 0.635) and 0.633 (95% CI, 0.621 to 0.644), respectively. The graph of yearly mean concordance index measures in the subgroup of women with at least three mammograms is shown in Figure 3.5. This, however, is not comparable to Figure 3.3 since follow-up starts at the third mammogram so all women are older at baseline and the age range of the cohort decreases; therefore discriminatory accuracy reduces for both most recent and longitudinal density. In Figure 3.6, the relative distributions of risk using longitudinal density and most recent density in the subgroup of women with at least three mammograms were again similar. When stratified by age at baseline, longitudinal density categorised a greater proportion of women as high-risk for women 40-50 years at baseline, a greater proportion as low-risk for women 50-60 years at baseline, and a greater proportion as low- and high-risk for women 60 years at baseline (appendix A.XVI-A.XVIII).

3.3.9 Sensitivity analyses

In sensitivity analyses, longitudinal density provided 29% and 30% more information than baseline density and most recent density, respectively, when mammograms with a missing corresponding BMI were removed (n=129,748 (98% of the cohort), 2,668 invasive breast cancers) (appendix A.VIII). Furthermore, longitudinal density provided 31% and 23% more information than baseline density and most recent density, respectively, when screen-detected mammograms were removed (n=132,439 (100% of the cohort)) (appendix A.IX).

3.4 Discussion

This cohort study found that using a woman's longitudinal history of breast density may improve risk prediction beyond using her baseline density or most recent density. Longitudinal density had the greatest predictive ability of the density measures, providing approximately a quarter more statistical information than baseline or most recent density. Women in the highest category of longitudinal density had a six-fold greater risk of developing breast cancer than women in the lowest category; but only a four-fold greater risk was seen with BI-RADS density at baseline or most recent mammogram. The benefit of longitudinal density for breast cancer risk estimation was not limited by age, menopausal status, image type or BI-RADS density classification lexicon. Discriminatory ability was also greatest with longitudinal density, whereby a small proportion more women were correctly classified as having breast cancer when using longitudinal density than when using the single measure for most recent density.

These results support previous findings that suggest an improvement in predictive ability of breast cancer risk estimation when using breast density values from more than one time point (262). In 2015, Kerlikowske et al. assessed BI-RADS density in a screening cohort of over

700,000 women from the Breast Cancer Surveillance Consortium, where a two-measure density score was developed combining each woman's first and last BI-RADS density taken on average 1.8 years apart (262). They found a slight improvement to the Breast Cancer Surveillance Consortium 5-year risk model when using the two-measure BI-RADS score compared with the one-measure BI-RADS score, whereby AUC for the two-measure score was 0.005 units higher than that for the one-measure score. Several other studies have made use of two serial mammograms (254-261, 323), however these studies aimed to assess the association between change in density and breast cancer risk (no specific intervention) as opposed to assessing the predictive ability of using both mammograms compared with just one. Results from most of these studies suggest that density change between two serial mammograms is associated with change in breast cancer risk (254, 257-261), indicating the benefit of considering more than one time point for breast cancer risk estimation. However, the Breast Cancer Surveillance Consortium study is the only other known study to have evaluated predictive ability when using more than one density measure (262).

The results suggest that the main advantage of longitudinal density is in its ability to act as a shrinkage estimator, making use of multiple data points to reduce measurement error (324). This is apparent because the predictive ability of longitudinal density improved substantially when it was modelled as a finer-grained variable which allowed density to be measured to a greater level of precision. Additionally, only a small benefit was seen when including random slopes to the model which represented each woman's density trajectory over time; suggesting a high level of density tracking (which has been seen before (325, 326)). The ability of longitudinal density to act as a shrinkage estimator also makes it a potentially useful tool for other aspects of density assessment including use as an outlier detection technique whereby observed values that deviate significantly from a predicted value could be flagged-up for investigation.

In this study, predictive and discriminatory ability improved with longitudinal density, however the capacity for risk stratification was somewhat similar for longitudinal density and most recent density. This may have been driven by the adjustment for age and BMI only. The ability to separate out extreme risk groups with longitudinal density might improve with the inclusion of other risk factors in the proportional-hazards Cox model alongside age and BMI, and remains a point for further investigation. There was better separation of high- and low-risk women with longitudinal density than most recent density for older ages (particularly 60 years and over). This is because the range of density values for most recent density in older women was small (all women were likely to have had fatty breasts), whereas longitudinal density accounted for previously high density values, hence giving it a greater range and ability to stratify risk. Of note, the proportion of women classified as high-risk increased and the proportion of women classified as low-risk decreased throughout follow-up when using either most recent or

longitudinal density. This is due to aging of the cohort. Risk scores were determined by age, which increased throughout the follow-up and hence increased the risk scores. Therefore, through time, women moved from a lower 10-year risk into a higher 10-year risk. Another notable point is the decreasing concordance index values for both most recent and longitudinal density throughout the follow-up. Again, this is due to aging of the cohort and narrowing of the age range through time. For example, at baseline the women at-risk were aged 40-73 years, but after 10 years the women at-risk would have been 50-74 years (the maximum is 74 years because of censoring at 75 years of age). Since the concordance index was based on a risk score that was indicated by age, it lost discriminatory ability through the follow-up regardless of the density measure. It is also noted that the improvement in discriminatory ability for longitudinal density compared with most recent density potentially attenuated towards the end of follow-up; which was possibly driven by fewer events occurring in this latter stage of follow-up which limited statistical power.

The major strength of this study is the ability to model a woman's entire history of density; including an unlimited number of mammograms arbitrarily spaced through time. This makes longitudinal density a particularly useful tool for clinical practice where women can have a number of mammograms taken at any point in time. Furthermore, predicting a woman's longitudinal density at each time point using only her current and previous densities would allow for the measure to be continually updated at each screening visit. Longitudinal density is not limited to any one density measurement technique, and it could just as easily be developed using semi-automated or fully-automated area-based or volumetric techniques. Predictive accuracy of breast cancer risk models that estimate personal breast cancer risk scores, such as the Tyrer-Cuzick, Gail or BCSC model, may also improve with the inclusion of longitudinal density. Using longitudinal density to assess risk of breast cancer may also prevent fluctuations in classifying women into different risk categories. For example, a women who has always had a high BI-RADS category 4 density that decreases to a BI-RADS category 3 could drop into a lower risk category that excludes her from supplemental screening or eligibility for chemoprevention. However, assessing her breast cancer risk using longitudinal density would take into account all of her previous measures and hence be more conservative with decreasing her risk.

A limitation of the study is that BI-RADS density categories were modelled as quantitative integer values to crudely approximate a linear association between density and the age and BMI predictors in the linear mixed model. It was considered best to first investigate longitudinal density using this simple linear model to identify if there is indeed an added benefit in assessing a woman's history of density, and from this, a more complex model could be developed. It may be that other models, for example a multinomial or ordinal logit model, better fit the data and

perhaps outperform the linear model (327). Furthermore, the linear mixed model used to develop longitudinal density was adjusted for age and BMI only. Including additional confounders of density such as HRT use, benign breast disease or reproductive factors might improve model fit and also the approximation of longitudinal density. Finally, the distribution of breast cancer risk was somewhat similar when using most recent or longitudinal density. Again, including additional breast cancer risk factors could potentially improve this risk stratification.

There are many ideas for future work on longitudinal density. These include evaluating the benefit of longitudinal density in different cohorts of women such as younger or older women outside of the routine screening age, or women at increased risk due to a family history of breast cancer. Assessment using other density measures including volumetric or semi/fully-automated methods would also be useful. Improving the prediction of longitudinal density is another area of future work. It is not yet known whether all previous density values are needed to predict longitudinal density, or whether the value of historical density measures reduces in time. It may be useful to up-weight more recent density measures or perhaps apply 'forgetting factors' to the linear mixed model to down-weight older measures (328). Moreover, previous research suggests a possible benefit of assessing the extent of density fluctuation through time (329), which has the potential to further improve the prediction of longitudinal density. Moreover, as mentioned earlier, a multinomial or ordinal logit model may improve model fit for the longitudinal density measure (327). The value of longitudinal density in assessing response to treatment also requires assessment in a future study. It may be that an observed decrease in density greater than that predicted from individual density trajectories is indicative of a response to treatment. Finally, the assessment of breast cancer risk with longitudinal density could potentially be improved by using different approaches for modelling risk. These include combining additional risk factors into the proportional-hazards Cox model, incorporating a longitudinal BMI measure (predicted using a similar linear mixed model approach), or using a joint longitudinal-survival model to maximise the likelihood of random effects that are common to both the mixed model and the proportional-hazards model simultaneously (327, 330, 331).

3.5 Conclusion

In this study, longitudinal density was shown to have greater predictive ability, better discriminatory accuracy and a higher risk gradient between the extreme density categories than a single measure of baseline or most recent density. Including information on a woman's history of mammographic density has the potential to improve the accuracy of breast cancer risk estimation and its implementation in breast cancer prevention strategies should be considered.

Chapter 4: Mammographic density, endocrine therapy and breast cancer risk: a prognostic and predictive biomarker review

4.1 Background

4.1.1 Description of the intervention

Selective oestrogen receptor modulators (SERMs) and aromatase inhibitors (AIs) are two types of endocrine drug used as therapy for ER+ breast cancers. SERMs prevent breast cancer (198, 200), and in the adjuvant setting, they reduce the chance that breast cancer will reoccur when it has been diagnosed at an early stage (332, 333). SERMs work by competing with oestrogen molecules for oestrogen receptor binding sites, hence reducing the amount of oestrogen uptake in breast tumours. SERMs are therefore effective in ER+ breast cancers only.

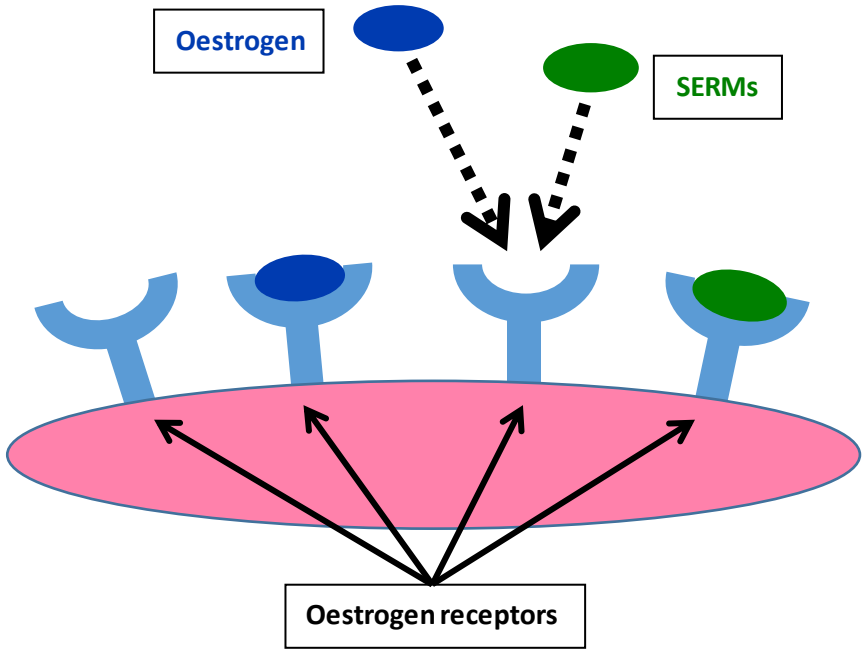


Figure 4.1: Mechanism for selective oestrogen receptor modulators (SERMs) competing with oestrogen for binding sites.

AIs are suitable for postmenopausal women only, and they are associated with greater average reductions in the risk of breast cancer (208, 334), and recurrence than SERMs (335). Like SERMs, AIs reduce oestrogen levels, but they instead work by inhibiting oestrogen synthesis in peripheral tissue by preventing the aromatase enzyme from converting adrenal androgens (androstenedione and testosterone) into oestrogens (estrone and estradiol). This process, known as aromatisation, is the main source of oestrogen after the menopause.

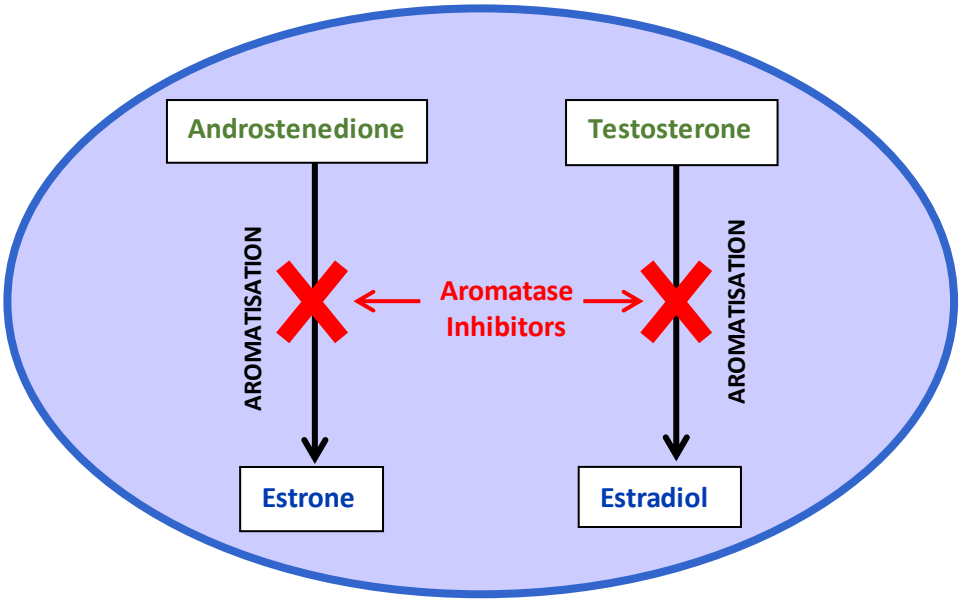


Figure 4.2: Mechanism for aromatase inhibitors (AIs) blocking oestrogen production in peripheral tissue (subcutaneous adipose tissue, liver, muscle, or brain).

4.1.2 How the biomarker might be related to treatment response

The biomarker for this review is mammographic density. As described in Chapter 1, exogenous hormones can change a woman's mammographic density. For instance, HRT increases density as well as breast cancer risk (188, 189). However, after cessation of HRT, mammographic density has been shown to decrease in as little as four weeks (195), and within a couple of years, breast cancer risk is likely to return to the same level as a non-HRT user (197). Treatment with certain SERMs can also decrease density more so than would be expected with age (203-206, 336), but the evidence for AIs is less clear (267-271, 337). There is evidence to suggest that increased risk from combination HRT is mediated by density (142, 338, 339), and that change in density may be an appropriate biomarker for response to SERMs used for both prevention (19) and treatment (263-266). A working hypothesis is that density reduction in women receiving endocrine therapy for treatment or prevention might indicate who is responding to the drug, making it a reliable surrogate outcome and biomarker for treatment efficacy. The underlying biological mechanism is still unclear, but one theory is that decreases in density reflect the body's ability to effectively metabolise the drug (340).

4.1.3 Why it was important to do this review

This review firstly aimed to assess the evidence that endocrine therapy-induced change in mammographic density is a prognostic biomarker (341). A prognostic biomarker is a measure

associated with a clinical outcome of interest in a defined group of patients. This is standard in the adjuvant setting when the group of patients has a health condition such as breast cancer. Several prognostic factors in the adjuvant setting include tumour size, grade and lymph node involvement, as well as biomarkers such as Ki67 and genetic scores such as OncotypeDX (342, 343). The terminology for prognostic biomarkers associated with breast cancer in healthy women in the preventive setting is less frequently used. Prognostic biomarkers in healthy women are more commonly called risk factors. These include age, a family history of breast cancer, BMI and reproductive factors such as age at first full term birth and number of children (223).

The second aim of this review was to assess the evidence that endocrine therapy-induced change in mammographic density is a predictive biomarker, such that it is a measure that is different in the presence of treatment and is therefore associated with response to treatment (344). Some, but not all, prognostic biomarkers are predictive biomarkers. Two examples for women with breast cancer are human epidermal growth factor receptor (HER-2) and ER status. HER-2 was first identified as a prognostic factor for breast cancer, and was subsequently recognised as a predictive biomarker whereby treatment with trastuzumab was shown to be effective for women with HER-2 breast cancer. ER status is a prognostic biomarker and a predictive biomarker for SERM and AI treatments, whereby these treatments improve clinical outcomes in ER+ patients only.

There are currently no systematic reviews that focus on the evidence that mammographic density reduction in women receiving endocrine therapy is a prognostic or predictive biomarker. However, some other reviews on the topic have been published, including a study by Shawky et al. (345). This reported seven studies of density change as a prognostic factor for women receiving a SERM or AI in the adjuvant setting, but there were no data from a randomised trial or otherwise to evaluate change in mammographic density after initiation of adjuvant treatment as a predictive biomarker. For prevention, only one study (a case-control study from within a randomised control trial) was identified that evaluated density change as a prognostic and predictive biomarker. Another recent review by Kanbayti et al. assessed the relationship between mammographic density reduction following breast cancer treatment and patient outcomes, although this was not specific to women receiving endocrine therapy (346). This review reported nine studies of density reduction as a prognostic factor for women receiving breast cancer treatment, but, again, there were no data to evaluate change in mammographic density as a predictive biomarker in the adjuvant setting.

This review should help to guide clinical decisions about whether to continue treatment or switch to another treatment regime, understand the aetiology of breast cancer development,

improve the design of trials in terms of implementing a surrogate marker, and improve personalised risk assessment (347). Findings are likely to be important to: clinicians and their patients undergoing or considering endocrine therapy by helping to predict response to treatment beyond the current 'wait and see' approach; regulators and ethics boards considering trials of products that use mammographic density reduction as a surrogate endpoint; and those with an interest in mechanisms by which endocrine therapy improves clinical outcomes. Additionally, as discussed in the study by Mullooly et al., had the randomised trials of SERMS and AIs included density change as a potential prognostic or predictive biomarker, then different conclusions might have been drawn regarding their effectiveness (348). For instance, in the ATAC trial, the AI anastrozole was shown to be more effective than the SERM tamoxifen in reducing risk of postmenopausal breast cancer recurrence (349). However, it is possible that women who had density reductions with tamoxifen might have had greater treatment benefits than those on anastrozole. Another possibility is that women who see density increases following a short-term decrease might in fact show resistance to the treatment (350, 351), but this is still unknown and requires investigation.

4.2 Objectives

The objective of the review was to synthesise available evidence testing whether mammographic density reduction in the preventive or adjuvant setting is (i) a prognostic biomarker and (ii) a predictive biomarker. Both prognostic and predictive biomarker reviews considered prevention and treatment populations separately, and within these, SERMs and AIs were considered separately.

4.3 Methods

This review was written according to PRISMA (352) and Remark (353, 354) guidelines.

The aim was to conduct a literature-based analysis to identify relevant studies that could then be used in a subsequent individual-level analysis. This individual-level meta-analysis is not included in the thesis, but is instead proposed as a future review that should help to account for heterogeneity between the studies in terms of the participants, length of follow-up, mammographic density measures, cut-points and overall study design (347, 355).

The methods described in this chapter outline the proposed procedure for conducting the systematic review. A version of this review plan is published as a Cochrane review protocol (356). At the review stage, it was decided that the studies were too heterogeneous to be able to combine into a meaningful meta-analysis. However, the full methodology is still presented

because it contains important information regarding other elements of the review and it explains the approach taken for conducting the literature-based analysis.

4.3.1 Criteria for considering studies for this review

4.3.1.1 Types of study designs

Randomised and non-randomised observational studies (prospective and retrospective cohort and case-control studies) were included for both the prognostic and predictive reviews. Studies based on exploratory biomarkers whereby density was one of several biomarkers were included.

4.3.1.2 Types of participants

Studies were included if they had subsets that met the following participant criteria, but only the relevant subset data was to be extracted for the meta-analysis.

For both the prognostic and predictive biomarker reviews, all adult women aged 18 years or older, with or without breast cancer (denoted respectively as treatment and prevention) were to be included based on the following criteria.

- Treatment: women with early stage hormone receptor (oestrogen (ER) or progesterone (PgR))-positive breast cancer. This was defined to be women who had been diagnosed with histologically proven operable invasive hormone receptor-positive breast cancer or DCIS, and who were candidates to receive adjuvant endocrine therapy. There was to be no clinical evidence of metastatic disease to minimise the risk of a recurrence or contralateral breast cancer being a misclassified metastasis. Women were to be considered ineligible if their breast density measurements were not made on the contralateral breast because there was a risk that tumours may have been misclassified as dense tissue. For this same reason, women were to be considered ineligible if they had bilateral breast cancer.

- Prevention: women who had not previously been diagnosed with invasive breast cancer or DCIS. Women of all levels of increased risk due to genetic factors (including BRCA1/2 gene mutations or a family history of the disease, or both) or otherwise assessed by an absolute or relative risk prediction model were to be included. If women had breast implants or if they had undergone risk-reducing mastectomies, they were to be excluded. This was considered because these factors affect the ability to produce accurate density estimates.

Women were to be at-risk for at least the length of time between baseline and follow-up mammogram. Women could be included if they changed treatment or discontinued treatment throughout their follow-up, but they were to be excluded if they changed treatment between their mammograms as this may have affected the change in density. However, women could be included if they discontinued treatment between mammograms. Women were to be excluded if they received another SERM or AI before treatment because these effects may have continued into the second period of treatment.

For AI comparisons, women had to be postmenopausal at the start of treatment; for SERM comparisons, they were allowed to be pre- or postmenopausal. The definition of postmenopausal women included women who had undergone a bilateral oophorectomy, or women who were aged more than 60 years, or women who were aged 40 to 59 years with an intact uterus and who were amenorrhoeic for at least 12 months. Women were to be excluded if they were rendered temporarily postmenopausal through medical interventions (e.g. gonadotropin-releasing hormone (GnRH) analogues).

4.3.1.3 Types of interventions

4.3.1.3.1 Interventions

Studies were included if they had subsets that met the following intervention criteria, but only the relevant subset data was to be extracted for the meta-analysis. Studies including women receiving doses lower or higher than those outlined below were included, but for the meta-analysis, these women were to be included in a secondary dose-response analysis only. Studies involving a mixture of women receiving SERMs and AIs were included, but for the meta-analysis, these studies were to be included in the main analysis if the results could be separated by treatment; otherwise they were to be included in a secondary analysis only.

For both the prognostic and predictive biomarker reviews, women were to be included if they received SERMs at the following minimum doses (357): Tamoxifen, 20 mg daily; Raloxifene, 60mg daily; Lasofoxifene, 0.25mg daily; Arzoxifene, 20mg daily; Droloxifene, 40 mg daily; Bazedoxifene, 20 mg daily; and Fulvestrant, 250 mg monthly. Women were to be included if they received AIs at the following minimum doses: Anastrozole, 1 mg daily; Letrozole, 2.5 mg daily; and Exemestane, 25 mg daily. All treatments were to be orally-consumed, except Fulvestrant (intramuscular). Treatment was to be received for at least the length of time between baseline and follow-up mammogram (i.e. intended for at least 1 year).

4.3.1.3.2 Co-interventions

Studies were included if they had subsets that met the following co-intervention criteria, but only the relevant subset data was to be extracted for the meta-analysis. The same types of co-interventions were allowed for both the prognostic and predictive biomarker reviews.

For treatment, women were to be considered ineligible if they had not completed primary loco-regional treatment (surgery or radiotherapy, or both) and systemic treatment (chemotherapy or targeted therapy, either neoadjuvant or adjuvant) with curative intent. Women were to be considered ineligible if there was a gap of more than eight weeks between different treatment interventions, for example, between surgery and the start of radiotherapy, or if endocrine treatment was started more than 28 days before surgery.

If women used HRT either during the study or up to 2 years before baseline, they could be included, but this was to be noted in the 'Risk of bias' assessment where relevant. Other co-interventions were permitted, including exercise and diet advice, but these were also to be noted in the 'Risk of bias' assessment where relevant.

4.3.1.3.3 Comparators

The main difference between the prognostic and predictive biomarker review was the comparator.

- Prognostic biomarker review: The comparison was within each intervention group (SERM or AI), whereby assessment was on the association between density change and outcome in women receiving the treatment.

- Predictive biomarker review: The comparison was within each study, whereby assessment was on the association between density change and outcome in the intervention group compared with a control group. The within-study comparator group was defined as a corresponding randomised placebo group, or a non-randomised control group of women not receiving endocrine therapy.

4.3.1.4 Biomarker

The same definition of biomarker was used for both the prognostic and predictive reviews. A measure of mammographic density was required at baseline (start of endocrine therapy or study entry in those from the control group) and follow-up mammogram.

Studies were included if they had subsets that met the following biomarker criteria, but only the relevant subset data was to be extracted for the meta-analysis.

For treatment, baseline mammograms could be taken before or after diagnosis, but they were to be no more than 2 years before the initial breast cancer diagnosis so that they represented the breast at the time of diagnosis as closely as possible. For treatment and prevention, baseline mammograms had to be taken before the start of treatment (or study entry) so that they reflected the breast phenotype before the effects of endocrine treatment. A follow-up mammogram had to be performed 90 days to 3 years after the start of endocrine treatment (or study entry), with the density closest to 1 year from the start of endocrine therapy (or study entry) selected if there was a choice.

Range and average timings were recorded for the following (if they were available): time between baseline mammogram and diagnosis, time between diagnosis and start of endocrine therapy (or study entry), and time between start of endocrine therapy (or study entry) and follow-up mammogram.

Density methods had to have been shown in more than one study (outside of the review studies) to have a relationship with breast cancer risk. Acceptable density methods included (but were not limited to) the following percentage methods:

- Visual assessment by expert in 5% bands (%).
- Visual assessment by expert in 20% bands (Boyd categories).
- Visual assessment by expert as continuous percentage (%).
- Semi-automated thresholding such as using 'Cumulus' software (23) by expert (or trained) reader (%).
- Fully-automated percentage (based on area of density) (%).
- Fully-automated volumetric percentage (e.g. Volpara, (44)) (%).

Acceptable absolute density methods included (but were not limited to) the following:

- Semi-automated thresholding such as using 'Cumulus' software (23) by expert (or trained) reader (cm^2).
- Fully-automated absolute density (based on area of density) (cm^2).
- Fully-automated volumetric absolute density (e.g. Volpara, (44)) (cm^3).

Acceptable categorical density measures included (but were not limited to) the following:

- BI-RADS density (18).
- Wolfe grade (358).
- Tabar grade (15).

Information on the reliability of density measures was also used to qualitatively assess the 'Risk of bias' due to measurement of the biomarker. Such information included:

- The correlation between repeated measures from repeat mammograms.
- Whether different readers of density were used and whether the same reader assessed mammograms from the same woman.
- Intra-class correlation coefficients and Bland-Altman limits of agreement (359) to assess intra- and inter-reader reliability.
- Whether the reader was blinded to case status.
- Whether the reader was blinded to treatment allocation.
- Whether randomisation was per mammogram (mammograms read independently) or per woman (mammograms for each woman read with the knowledge of her other mammograms).
- Whether the order of per woman mammograms was sequential or random and assessed one at a time or simultaneously.

If different definitions or measures of mammographic density were used between the time-points used to assess density change, these women or studies were excluded.

4.3.1.5 Types of outcome measures

The same outcome measures were used for both the prognostic and predictive reviews.

Primary outcomes

Potential benefits from treatment:

- Treatment: breast cancer mortality (time to death caused by breast cancer).
- Prevention: incidence of invasive breast cancer and DCIS.

Potential harms from treatment:

- Treatment and prevention: rate of all serious adverse events. These included serious side effects noted for Tamoxifen (cataracts, pulmonary embolism or deep vein thrombosis and endometrial cancer) and Anastrozole (osteoporosis and bone fractures).

Secondary outcomes

Potential benefits from treatment:

- Treatment: recurrence.
- Treatment: incidence of a secondary primary breast cancer (e.g. in the contralateral breast).
- Treatment: any recurrence or any death (disease-free survival).
- Treatment: distant metastases.
- Treatment: death from all causes (all-cause mortality).
- Treatment: recurrence of invasive cancer only.
- Treatment: recurrence of DCIS cancer only.
- Prevention: incidence of invasive cancer only.
- Prevention: incidence of DCIS cancer only.

Potential harms from treatment:

- Treatment and prevention: troublesome but not serious side effects observed for SERMs and AIs, including vasomotor symptoms and joint or muscle pain.

'Summary of findings' table for assessing the quality of the evidence

A 'Summary of findings' table was produced for each of the prognostic and predictive biomarker reviews, following the approach outlined by GRADE (360) and using GRADEpro GDT software (361).

4.3.2 Search methods for identification of studies

4.3.2.1 Electronic searches

The following databases were searched:

- The Cochrane Breast Cancer Group's (CBCG's) Specialised Register. Details of the search strategies used by the Group for the identification of studies and the procedure used to code references are outlined on the Group's website (362). Trials were extracted and considered

for inclusion if they included the key words "Tamoxifen, Raloxifene, Lasofoxifene, Arzoxifene, Droloxifene, Bazedoxifene, Fulvestrant, Anastrozole, Letrozole, Exemestane, selective estrogen receptor modulator, aromatase inhibitor".

- CENTRAL (The Cochrane Library, latest issue). See appendix B.I.
- MEDLINE (via OvidSP) from 1996 to present. See appendix B.II.
- Embase (via OvidSP) from 1996 to present. See appendix B.III.
- The World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) search portal (363) for all prospectively registered and on-going trials. See appendix B.IV.
- ClinicalTrials.gov (364). See appendix B.V.

4.3.2.2 Searching other resources

Bibliographic searching: Further studies were sort out from reference lists of identified relevant studies. A copy of the full article was to be obtained for each reference reporting a potentially eligible study.

4.3.3 Data collection and analysis

4.3.3.1 Selection of studies

All retrieved titles and abstracts were independently reviewed by the author of this thesis (Emma Atakpa) and Dr Brentnall to assess eligibility against the inclusion criteria. Any disagreements at this stage were discussed and resolved. Full-text copies of all potentially eligible studies were then obtained and reviewed by the thesis author (Emma Atakpa) and Dr Brentnall. Any disagreements at this stage were resolved by Dr Mangesh Thorat (the clinical expert on the systematic review). There was only one disagreement which regarded the inclusion of grey literature such as conference abstracts. It was determined that (although not explicitly stated in the protocol) only studies published in a peer-reviewed journal would be included and abstract-only records would be excluded as implied by the quotes: "We will obtain a copy of the *full article* for each reference reporting a potentially eligible trial" and "We will only include studies *published* in English" (356). Duplicate studies were recorded as one reference (for example, the same study but multiple papers with slightly different aims or follow-up). In this situation, the reference considered to be the most recent or up-to-date (largest number of participants, longest follow-up time, or correction to previous analysis) was included as the primary reference. Only studies published in English were included. The selection process was recorded in a PRISMA flow diagram (352) in the Review Manager 5 software (365). The whole process was recorded using the Covidence system (366).

4.3.3.2 Data extraction and management

Data extraction was independently completed by the thesis author (Emma Atakpa) and Dr Brentnall using custom forms (appendix B.VI). Again, any disagreement at this stage was to be resolved by Dr Thorat (although this was not required). The following information was collected (if available):

- Study design: type of study. For example, a nested case-control study from a randomised trial, or a non-randomised cohort study, or a case-control study. If there was matching, then what matching was by and to what level (e.g. age to plus/minus 2 years). Control group (women without treatment): yes/no. Whether a prognostic or predictive study, or both. For prognostic factor studies, what phase (following (355, 367)).
- Participants: demographic information, including the number of participants, age, BMI, ethnicity, education. Summary statistics such as mean, interquartile range (or standard deviation) and range for age, BMI and absolute or relative baseline risk, or both, from a risk model (e.g. Gail model (222), Tyrer-Cuzick (223), BCSC (239)). Total number (percentage) postmenopausal, perimenopausal or premenopausal. For the predictive review, the previous variables were split by treatment or control group.
- Biomarker: whether mammograms were from film (digitised for density or not) or full field digital mammography. Manufacturer of digital mammogram machine. Whether any pre-processing was carried out for quality control of mammographic density. Density measure(s), and the range and average time between baseline mammogram and diagnosis, between diagnosis and start of endocrine therapy (or study entry), and between start of endocrine therapy (or study entry) and the follow-up mammogram.
- Setting: country, whether in a high-risk clinic, a treatment clinic, time period, urban/rural.
- Co-interventions: HRT use, chemotherapy use (treatment), targeted therapy use (treatment), radiotherapy use (treatment), neoadjuvant endocrine therapy use (treatment).
- Follow-up time period: minimum, mean, median, interquartile range, standard deviation, maximum follow-up.
- Sources of funding and stated conflicts of interest: descriptive text copied from sections in each paper.

4.3.3.3 Assessment of risk of bias in included studies

For the prognostic biomarker review, a modified version of the QUIPS tool (368) was used to assess the risk of bias affecting the included studies (369) using six domains:

- Study participation.
- Attrition.
- Measurement of density.
- Measurement of the outcomes.
- Confounding.
- Statistical analysis.

For the predictive biomarker review, the QUIPS tool was augmented with the ROBINS-I tool (370) to assess the risk of bias in estimation of an interaction between mammographic density change and treatment.

'Risk of bias' was independently conducted by the thesis author (Emma Atakpa) and Dr Brentnall, with disagreements to be resolved by Dr Thorat (although this was not required). For both prognostic and predictive biomarker reviews, the included studies were considered together and an individual 'Risk of bias' table was produced for each study (appendix B.VII). In the results, a narrative is presented identifying the risk of bias in the six domains across studies. Studies that had substantial potential for bias were to be excluded in the meta-analysis for a sensitivity analysis of the results.

4.3.3.4 Measures of biomarker response

4.3.3.4.1 Effect measure

Studies were included in the quantitative synthesis if they had subsets that reported the following effect measures, but only the relevant subset data was to be extracted for the meta-analysis.

In both the prognostic and predictive biomarker reviews, the primary measure was to be the mean effect of treatment-induced density change over a five-year follow-up period. Other time periods could be included, but if they were split into different periods (e.g. 0 to 5 years, 5 to 10 years) then periods outside of the initial five years were to be treated as a secondary analysis only in the meta-analysis. Results of the meta-analysis were to be presented as subgroups of similar cut-points and biomarkers using continuous measures, and reported as ratios whereby less than 1.0 was to favour a risk reduction associated with a decrease in mammographic density and greater than 1.0 was to indicate a risk increase.

- Prognostic biomarker review: The primary measure was to be a hazard ratio (cohort study with time to event) or an odds ratio (case-control study) for the effect of density change.

Odds ratios were to be treated as an equivalent measure of the hazard ratio, unless the rate of breast cancer outcome was high. In this case, the odds ratio estimates were to be included in the meta-analysis as a secondary analysis only.

- Predictive biomarker review: The primary measure was to be the interaction between treatment and density change, expressed as a relative hazard (cohort study) or odds ratio (case-control study).

4.3.3.4.2 Adjustment

- Prognostic biomarker review: The primary effect estimate was to be adjusted. Estimates of the effect of prognostic factors tend to be more relevant when they are adjusted for potential confounders than when they are unadjusted (347). However, unadjusted estimates were to be included if adjusted estimates were not available because it was not expected that change in density would be associated with the baseline value of most other prognostic factors. Nonetheless, it is noted that changes in BMI may have affected changes in density because BMI is negatively associated with breast density, so one would ideally adjust for this in any analysis of density change as a prognostic biomarker.

- Predictive biomarker review: The primary effect estimate was to be adjusted. There are currently no established predictive biomarkers for either prevention or treatment in the groups of included women, so the adjustments are less clear than for the prognostic density change biomarker.

4.3.3.4.3 Dealing with missing data

Where data were missing, contact was to be made with the study authors in an attempt to obtain the data.

4.3.3.4.4 Assessment of heterogeneity

Heterogeneity in the meta-analysis was to be measured using the estimated variance in a random-effects model (Tau^2), and publication bias was to be measured using a funnel plot and Egger's test (371).

4.3.3.4.5 Subgroup analysis and investigation of heterogeneity

When sufficient studies existed, the following a priori subgroup analyses were to be conducted in the meta-analysis to explore reasons for heterogeneity within the predefined homogeneous groups described above.

Between-studies:

- Drug within SERM (Tamoxifen, Raloxifene, Lasofoxifene, Arzoxifene, Droloxifene, Bazedoxifene, Fulvestrant) and AI grouping (Anastrozole, Letrozole, Exemestane).
- Type of study: case-control, observational cohort, randomised trial (nested case-control).
- Type of cancer at baseline (treatment): (percentage DCIS).
- Severity of cancer at baseline (treatment): stage (percentage regional spread).
- Co-interventions (treatment): chemotherapy/targeted therapy.
- Hormone therapy use during therapy (yes/no, percentage if it was available), or in previous two years (yes/no, percentage if it was available).
- Time between start of therapy (or study entry) and follow-up mammogram (mean and range).
- Menopausal status (percentage premenopausal).
- Age (mean).
- BMI (mean).
- Digital or film mammography (percentage digital).
- Distribution of density at baseline (some studies may have excluded women with low density).

Within-study estimates of effect:

- Type of cancer at baseline (treatment): DCIS vs. invasive.
- Severity of cancer at baseline (treatment): stage (percentage regional spread).
- Co-interventions (treatment): chemotherapy/targeted therapy.
- Hormone therapy use: no HRT prior to endocrine therapy, some HRT 2 years or more than 2 years prior to endocrine therapy, some HRT less than 2 years prior to endocrine therapy, some HRT during endocrine therapy.
- Menopausal status (pre-, peri- or postmenopausal).
- Age group (< 50 years or ≥50 years) as a proxy for menopausal status.
- BMI (<25, 25 to <30, 30 to <35, ≥35 kg/m^2).
- Baseline density.

4.3.3.4.6 Data synthesis

Heterogeneity between studies was expected because this is common in reviews of prognostic biomarkers (347). Therefore, it was decided that a meta-analysis would only be conducted for studies within predefined groups that were believed to be homogenous enough in advance to be meaningful for data synthesis. Namely, those with the same class of drug, same outcome, same density measure, same effect measure (same cut-point or continuous variable assessment). Where more than one study was available, estimates were to be combined using an inverse-variance weighting (fixed-effect estimation); and if there was substantial variability then results were to be presented but it was to be stated that the overall effect estimate has very limited interpretation. Additionally, subgroups (4.3.3.4.5) were to be investigated to help to explain the heterogeneity.

4.4 Results

4.4.1 Description of studies

4.4.1.1 Results of the search

The database search identified 1180 records (see PRISMA follow diagram: Figure 4.3), and after deduplication, there were 888 records. Of these, 801 records were deemed ineligible according to their title and abstract and 87 records were selected for full-text review. Seventy-nine of the 87 full-text articles were excluded, and eight eligible studies that fulfilled the inclusion criteria were included in the qualitative synthesis, with six of these contributing to the quantitative synthesis. A bibliography search of the reference lists in the eight included studies was also conducted. Titles were reviewed and if they were considered potentially eligible (and not already included in the 888 records), their abstracts were reviewed. Nine potential records were identified but these were considered ineligible after abstract review, hence no additional studies were found through the bibliographic search.

4.4.1.2 Included studies

There was a large amount of variation across the eight included studies (see 'Characteristics of included studies' table: Table 4.1). All were observational studies, with four case-controls (19, 266, 372, 373) and four retrospective cohorts (263-265, 374). Of the four case-control studies, three studies used a matched design (266, 372, 373). Two studies were sub-studies from randomised controlled trials (one nested case-control (19) and one cohort study (374)). The studies ranged in size from 349 (266) to 1066 (263) women. Six studies included women from

Western populations (the UK (19), Finland (19), USA (266, 372), Canada (372), Sweden (265, 373) and the Netherlands (374)) and two studies were in women from South Korea (263, 264). Follow-up ranged from 5 years (263) to 14 years (265), and only one study (19) was in the preventive setting, with the rest being in the adjuvant treatment setting for women with breast cancer. Half of the studies assessed tamoxifen treatment only (19, 263, 265, 266), two studies assessed tamoxifen and an AI (264, 374), and two studies were not specific to a particular endocrine therapy, whereby only a subset of women were on endocrine therapy during their adjuvant treatment (372, 373). Two studies included a placebo (19) or control group (265), although the latter study did not compare across the interventions and could only be used in the prognostic review. Therefore, only one study was assessed in the predictive biomarker review (19). There was a mixture of premenopausal and postmenopausal women, and in the treatment setting, the two South Korean studies included women with DCIS or invasive breast cancer, whereas the other studies included women with invasive disease only.

The two sub-studies from clinical trials used visually-assessed density, with one assessing density to the nearest 5% (19) and the other assessing density in 20% Boyd categories (374). Two studies used a machine learning-based density assessment trained on Cumulus images (265, 373), three studies used Cumulus percent density (264, 266, 372) and one study used BI-RADS density (263). Density change cut-points varied greatly, with some adopting a 5% (264), 10% (19, 372, 373) or top tertile (determined by the distribution of controls) (266) absolute percent density reduction cut-point, one using a 20% relative dense area reduction cut-point (265), and another using reduction in BI-RADS category (263). One study did not report their definition of density change (374).

Two studies had recurrence (recurrence-free survival) as their endpoint (263, 264), two studies had mortality as their endpoint (265, 266), two studies looked at incidence of contralateral breast cancer (372, 373), one study assessed incidence of contralateral breast cancer and recurrence as its endpoint (374), whilst the final study assessed risk of developing invasive or DCIS breast cancer in a sample of at-risk women (19).

A detailed description of the included studies can be found in the data capture forms (appendix B.VI).

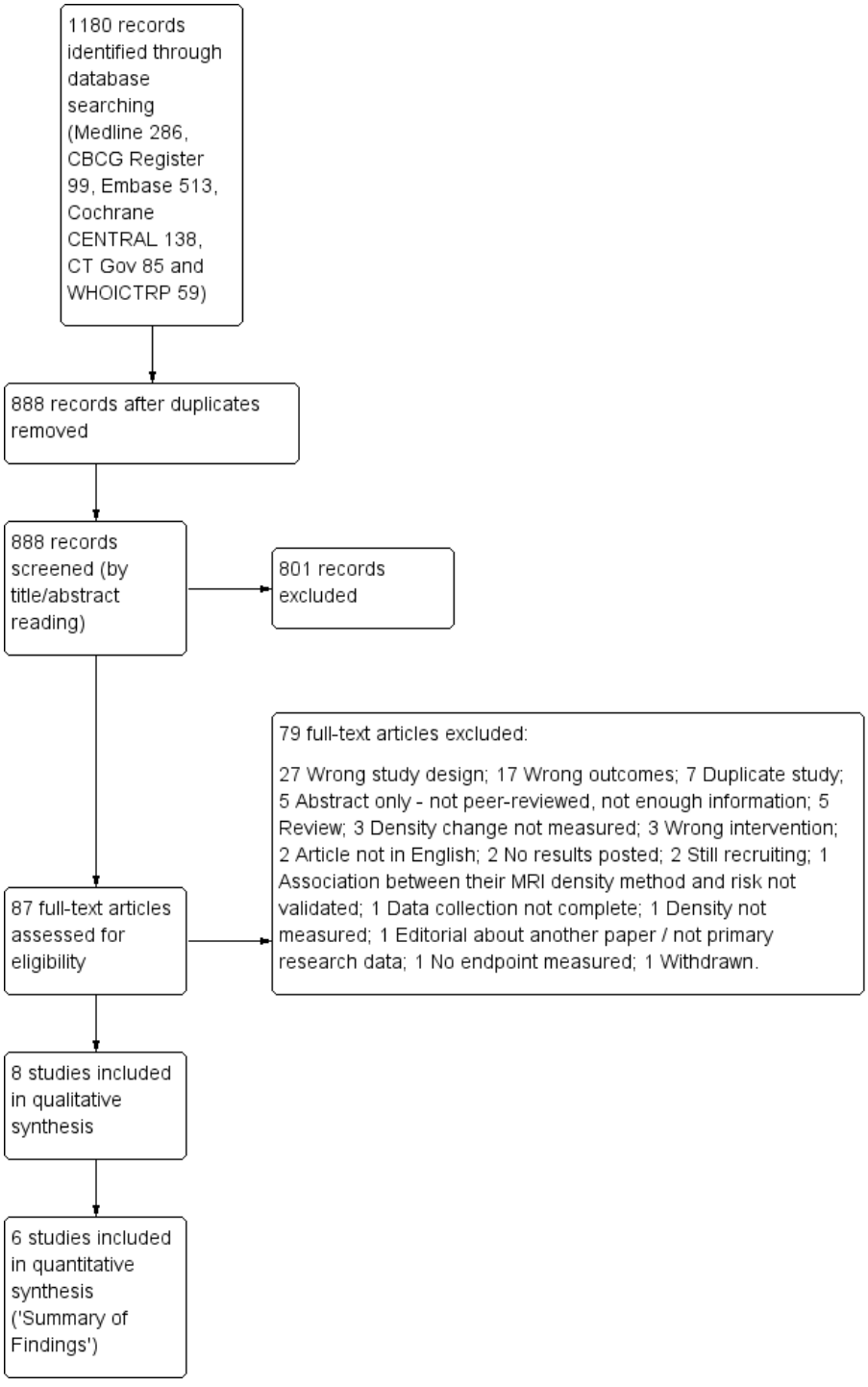


Figure 4.3: PRISMA flow diagram for the review.

Table 4.1: Characteristics of included studies

Cuzick 2011

Methods	Nested case-control within a multi-centre international randomised controlled trial (IBIS-I). Recruitment April 1992-March 2001, diagnosis before 1 October 2007. Prognostic and predictive biomarker. Prevention setting.	
Participants	123 cases from the UK and Finland, 942 controls from the UK. Age 35-70 years at recruitment to IBIS-I trial. Premenopausal and postmenopausal women. Approximately twice the population risk of developing breast cancer.	
Interventions	Tamoxifen 20mg daily (n=507), placebo daily (n=558), 5 years of treatment. Visually-assessed percent density. Density reduction 10% or more vs. no change at 12-18 month follow-up mammogram in tamoxifen arm (prognostic biomarker). Density reduction 10% or more vs. less than 10% at 12-18 month follow-up mammogram in tamoxifen arm compared with placebo arm (interaction, predictive biomarker).	
Outcomes	Incidence of invasive breast cancer and DCIS.	
Notes		
Item	**Authors' judgement**	**Support for judgement**
Study participation	Low risk	Source population described, demographic factors (and tumour characteristics for cases) similar in included and not included women from the IBIS-I main trial.
Study attrition	Low risk	Only 44 women withdrew who were not included in this sub-study (referenced (200)).
Prognostic factor measurement	Moderate risk	Valid and reliable density measure, although density change measure was determined by the data (the cut-point was chosen as "the minimum change that could be reproducibly detected"), the reader was not blinded to case-control status.

Outcome measurement	Low risk	From trial database.
Study confounding	Low risk	Adequate adjustment.
Statistical analysis and reporting	Low risk	Adequate analysis, although no interaction reported in the paper (worked out from raw data provided by the study authors).

Kim 2012

Methods	Retrospective cohort, non-randomised. Initial ER+ breast cancer diagnosis October 2003-December 2006. Follow-up median 69 months. Prognostic biomarker. Treatment setting.	
Participants	Seoul National University Hospital, South Korea. 1065 women. Age 24-77 years. No information on menopausal status but likely includes premenopausal and postmenopausal women. 12% DCIS, 88% invasive at first diagnosis.	
Interventions	Tamoxifen 5 years (n=657), tamoxifen 2-3 years + AIs (total 5 years) (n=41), tamoxifen 5 years + AIs (unknown total time) (n=192), AIs 5 years (n=175), at least 2 years of treatment. Cumulus percent density. Density reduction 5% or more vs. less than 5% at 8-20 month follow-up mammogram in tamoxifen. Density reduction 5% or more vs. less than 5% at 8-20 month follow-up mammogram in AIs.	
Outcomes	Recurrence.	
Notes		
Item	**Authors' judgement**	**Support for judgement**
Study participation	Moderate risk	No information on source population or key characteristics in women included vs. not included from source population.
Study attrition	Moderate risk	No information on participant drop-out, loss to follow-up or reasons for censoring.
Prognostic factor	Moderate	No information on reader experience, blinding to

measurement	risk	treatment or time between baseline and follow-up mammogram.
Outcome measurement	Moderate risk	No information on start of follow-up or reasons for censoring.
Study confounding	High risk	No information on when confounding factors were measured, unclear adjustments (if any) for analysis separated by treatment.
Statistical analysis and reporting	High risk	Unclear when follow-up started or reasons for censoring, unclear adjustments, unclear if subgroup analyses include women on tamoxifen only, AIs only or women who switched treatment.

Knight 2018

Methods	Case-control (matched on follow-up time, geographical area, birth year, diagnosis year and ethnicity), non-randomised. Initial breast cancer diagnosis 1990-2008, recruitment 2009-2012. Follow-up mean 8 years. Treatment setting.	
Participants	WECARE study (USA and Canada). 224 cases and 243 controls with mammograms at both time points. Mean age 46 years at mammogram before or at first diagnosis. Premenopausal and postmenopausal women. All invasive at first diagnosis.	
Interventions	Mainly tamoxifen, but specific treatments not reported. Cumulus percent density. Density reduction 10% or more vs. less than 10% at 6 month-4 year follow-up mammogram.	
Outcomes	Incidence of a secondary primary breast cancer (e.g. in the contralateral breast).	
Notes	Cannot include as a prognostic or predictive biomarker because the analysis adjusted for tamoxifen use.	
Item	**Authors' judgement**	**Support for judgement**
Study participation	Moderate risk	No information on source population but comparisons conducted on key characteristics between women included vs. not included from source population.

Study attrition	High risk	Potential for survival bias whereby women included were more likely to have survived at the time of interview than the wider cohort. Unsure about density or outcome in women who died before the study or who did not have available mammograms. "All women had to be alive at the time of contact for interview" and "Women in whom we could not obtain a mammogram in an appropriate time window (see below) were more likely to have an earlier year of first breast cancer diagnosis (65% diagnosed in 1990–1996 vs. 40% in 1990–1996)".
Prognostic factor measurement	Moderate risk	No information on blinding to treatment, or why 435/467 women were used in the analysis (could have been digital mammograms (instead of film like the rest of the study sample) or poor quality mammograms).
Outcome measurement	Low risk	From population registry.
Study confounding	Moderate risk	435/467 women used in the analysis (could have been missing data on adjusting factors), risk factors were obtained retrospectively by a telephone survey (potential for recall bias).
Statistical analysis and reporting	High risk	Analysis adjusted for treatment so unable to extract the effect as a prognostic or predictive biomarker, the study may have included other endocrine therapies besides tamoxifen.

Ko 2013

Methods	Retrospective cohort, non-randomised. Initial ER+ breast cancer diagnosis January 2003-December 2008. Follow-up mean 59 months. Prognostic biomarker. Treatment setting.
Participants	National Cancer Centre, Goyang, South Korea. 1066 women. Age 25-78 years. Unclear information on menopausal status but likely includes premenopausal and postmenopausal women.

	13% DCIS, 87% invasive at first diagnosis.	
Interventions	Tamoxifen (all women), at least 2 years of treatment. BI-RADS density. Reduction of at least 1 category vs. no reduction of at least 1 category at 10-34 month follow-up mammogram.	
Outcomes	Recurrence.	
Notes		
Item	**Authors' judgement**	**Support for judgement**
Study participation	Moderate risk	No information on source population or key characteristics in women included vs. not included from source population.
Study attrition	Moderate risk	No information on participant drop-out, loss to follow-up or reasons for censoring.
Prognostic factor measurement	Moderate risk	No information on whether restricted to contralateral breast or time between baseline and follow-up mammogram, no test of intra-reader reproducibility.
Outcome measurement	Moderate risk	No information on start of follow-up or reasons for censoring.
Study confounding	Moderate risk	No information on when confounding factors were measured, no adjustment for chemotherapy although it was associated with mammographic density reduction.
Statistical analysis and reporting	High risk	Unclear when follow-up started or reasons for censoring, title says 'premenopausal' women but likely includes postmenopausal women too since age range 25-78 years, no adjustment for confounding factors such as chemotherapy.

Li 2013

Methods	Retrospective cohort, non-randomised. Initial breast cancer diagnosis 1993-1995, follow-up until 31 December 2008. Follow-up median 14 years. Prognostic biomarker. Treatment setting.

Participants	Sweden. 974 women. Median age 62-63 years at diagnosis. Postmenopausal women. All invasive at first diagnosis.	
Interventions	Tamoxifen 20mg daily (n=231), tamoxifen 40mg daily (n=123), tamoxifen 20+40mg daily (n=108), tamoxifen 'other' dose daily (n=12), median 60 months of treatment. Fully-automated area-based method measuring absolute dense area. Relative dense area reduction more than 20% vs. stable dense area (≤9% increase to ≤10% reduction) at 6-36 month follow-up mammogram.	
Outcomes	Breast cancer mortality (time to death caused by breast cancer).	
Notes		
Item	**Authors' judgement**	**Support for judgement**
Study participation	Moderate risk	No information on source population or key characteristics in women included vs. not included from source population.
Study attrition	Low risk	Follow-up information from population registry, participant drop-out and loss to follow-up as a result of emigration likely to be small.
Prognostic factor measurement	Moderate risk	Cut-points chosen "a priori" but without justification.
Outcome measurement	Low risk	From population registry, clear definitions of start of follow-up and reasons for censoring.
Study confounding	Low risk	Adequate adjustment.
Statistical analysis and reporting	Low risk	Adequate analysis.

Nyante 2015

Methods	Case-control (matched on age at diagnosis, diagnosis year and disease stage), non-randomised. Initial ER+ breast cancer diagnosis 1990-2008, recruitment 1 January 1991-31 December 2010 (end of follow-up).

	Prognostic biomarker. Treatment setting.	
Participants	Kaiser Permanente Northwest, USA. 97 cases and 252 controls. Age 32-87 years at first diagnosis. No information on menopausal status but likely includes premenopausal and postmenopausal women. All invasive at first diagnosis.	
Interventions	Tamoxifen (all women), at least 1 tamoxifen prescription started within 1 year of diagnosis. Cumulus percent density. Density reduction more than 8.7% vs. less than 0.5% at 3-26 month follow-up mammogram.	
Outcomes	Breast cancer mortality (time to death caused by breast cancer).	
Notes		
Item	**Authors' judgement**	**Support for judgement**
Study participation	Moderate risk	No information on source population or key characteristics in women included vs. not included from source population.
Study attrition	Low risk	Follow-up information from population registry, reasons for censoring discussed.
Prognostic factor measurement	Low risk	Valid and reliable density and density change measures (based on tertiles).
Outcome measurement	Low risk	From population registry, clear definitions of start of follow-up and reasons for censoring.
Study confounding	Low risk	Adequate adjustment.
Statistical analysis and reporting	Low risk	Adequate analysis.

Sandberg 2013

Methods	Case-control (matched on age and calendar period of first breast cancer diagnosis, adjuvant therapy and follow-up time), non-randomised. Initial breast cancer diagnosis 1976-2005. Follow-up mean 8 years.

	Prognostic biomarker. Treatment setting.
Participants	Sweden. 211 cases and 211 controls. Age at first diagnosis: ≤45 years (n cases =37, n controls=37), 45-55 years (n cases=68, n controls=68), 55-65 years (n cases=56, n controls=56) and ≥65 years (n cases =50, n controls=50). Premenopausal and postmenopausal women. All invasive at first diagnosis.
Interventions	Endocrine therapy (n cases=87, n controls=87), but specific treatments not reported. Fully-automated area-based method measuring percentage density. Density reduction 10% or more vs. stable density (<10% increase to <10% reduction) at 1-5 year follow-up mammogram.
Outcomes	Incidence of a secondary primary breast cancer (e.g. in the contralateral breast).
Notes	

Item	**Authors' judgement**	**Support for judgement**
Study participation	Moderate risk	No information on source population but comparisons conducted on key characteristics between women included vs. not included from source population.
Study attrition	Moderate risk	Follow-up information from population registry, but no information on reasons for censoring or loss to follow-up.
Prognostic factor measurement	Moderate risk	Large variability in time between baseline and follow-up mammograms (follow-up mammogram 1-5 years after first breast cancer diagnosis), 66 women excluded if baseline percent density <10% or >90% because they could not undergo some of the defined density changes but these numbers were unknown for the subgroup of women on endocrine therapy.
Outcome measurement	Low risk	From population registry.
Study confounding	Moderate risk	Adjusted for age through matching, but other adjustments are unclear in the subgroup of women on endocrine therapy.

Statistical analysis and reporting	Moderate risk	Appropriate analysis for the study's primary objective, but the analysis in the subgroup of women on endocrine therapy was a secondary objective. Numbers unknown and unclear adjustments for the subgroup of women on endocrine therapy, cannot separate out endocrine therapies.

van Nes 2015

Methods	Retrospective sub-cohort within a multi-centre randomised controlled trial (TEAM). Start of TEAM trial enrolment in 2001, but unknown time period of sub-cohort study. Follow-up median 6 years. Prognostic biomarker. Treatment setting.	
Participants	The Netherlands. 378 women. Age 45-91 years at baseline. Postmenopausal women. All invasive at first diagnosis.	
Interventions	Exemestane 25mg daily for 5 years (n=197), tamoxifen 20mg daily for 2-3 years followed by 3-2 years of exemestane (totalling 5 years) (n=181). Visually-assessed percent density in 20% bands (Boyd categories). Unclear comparison: "change in breast density".	
Outcomes	Recurrence and incidence of a secondary primary breast cancer (e.g. in the contralateral breast) combined (loco-regional recurrence, distance recurrence or contralateral breast cancer).	
Notes	Cannot include as a prognostic biomarker because there were no results to extract.	
Item	**Authors' judgement**	**Support for judgement**
Study participation	Low risk	Source population not described but referenced (375), comparisons conducted between women included vs. not included from source population.
Study attrition	Moderate	No information on participant drop-out, loss to

	risk	follow-up or reasons for censoring.
Prognostic factor measurement	Low risk	Valid and reliable density and density change measures, although no information on which follow-up mammograms were used for the density change measure (therefore no information on time between baseline and follow-up mammogram).
Outcome measurement	Low risk	From trial database, clear definition of start of follow-up but unclear reasons for censoring (per protocol analysis so women were censored when they stopped treatment but no information on other reasons).
Study confounding	Moderate risk	Unclear adjustments (if any).
Statistical analysis and reporting	High risk	Insufficient presentation of data, adjustments and results for density change, both treatment arms combined so unable to separate out endocrine therapies.

4.4.1.3 Excluded studies

Seventy-nine studies were excluded after reading their full-text articles. Reasons for exclusion included wrong outcome (mammographic density change was modelled as the outcome), wrong study design (study was not designed to address the review question) and reviews (discussion regarding density change as a potential biomarker for endocrine therapy, but without any novel data). Abstracts from conference presentations were excluded because these had not undergone peer-review to be published as full-texts. For further details, see PRISMA flow diagram (Figure 4.3) and 'Characteristics of excluded studies' table (Table 4.2) which highlights some potentially relevant records that were excluded and the reasons for exclusions.

Table 4.2: Characteristics of excluded studies

AllianceforClinicalTrialsinOncology 2006 (376)

Reason for exclusion	No results posted

AllianceforClinicalTrialsinOncology 2007 (377)

Reason for exclusion	Wrong outcomes

Andersson 2017 (331)

Reason for exclusion	Density change not measured (not defined as a measure between a

	baseline and a follow-up mammogram)

Atkinson 1999 (205)

Reason for exclusion	Wrong outcomes

Becker 2009 (378)

Reason for exclusion	Review

Boyd 2001 (78)

Reason for exclusion	Wrong study design

Boyd 2011 (379)

Reason for exclusion	Wrong study design

CaseComprehensiveCancer 2007 (380)

Reason for exclusion	Still recruiting

Chlebowski 2003 (381)

Reason for exclusion	Wrong study design

Cosmacini 1993 (382)

Reason for exclusion	Abstract only - not peer-reviewed, not enough information

Cuzick 2012 (383)

Reason for exclusion	Wrong study design

Decensi 2004 (384)

Reason for exclusion	Abstract only - not peer-reviewed, not enough information

Decensi 2009 (385)

Reason for exclusion	Density change not measured (not defined as a measure between a baseline and a follow-up mammogram because repeated measures ANOVA was used, additionally digital density was calibrated by adjusting for different variables at different time points so the density measure was not the same at baseline and at follow-up mammogram)

Ekpo 2016 (296)

Reason for exclusion	Wrong outcomes

Engmann 2017 (271)

Reason for exclusion	Wrong outcomes

Fabian 2006 (386)

Reason for exclusion	Wrong study design

Fabian 2007 (387)

Reason for exclusion	Wrong study design

Fabian 2016 (388)

Reason for exclusion	Review

Ghosh 2010 (389)

Reason for exclusion	Wrong study design

Kim 2014 (390)

Reason for exclusion	Association between MRI density method and breast cancer risk not validated

Kmietowicz 2013 (391)

Reason for exclusion	Editorial about another paper / not primary research data

Macis 2011 (392)

Reason for exclusion	No results posted

Martin 2009 (339)

Reason for exclusion	Wrong study design

Martin 2016 (393)

Reason for exclusion	Abstract only - not peer-reviewed, not enough information

Mullooly 2016 (348)

Reason for exclusion	Review

NCICClinicalTrialsGroup 2000 (394)

Reason for exclusion	Wrong study design

NCICClinicalTrialsGroup 2001 (395)

Reason for exclusion	Wrong study design

NorthwesternUniversity 2003 (396)

Reason for exclusion	Wrong outcomes

Ozhand 2013 (397)

Reason for exclusion	Abstract only - not peer-reviewed, not enough information

Redfern 2016a (398)

Reason for exclusion	Wrong study design

Redfern 2016b (399)

Reason for exclusion	Abstract only - not peer-reviewed, not enough information

SeoulNationalUniversityHospital 2013 (400)

Reason for exclusion	Still recruiting

SeoulNationalUniversityHospital 2018 (401)

Reason for exclusion	Data collection not complete

Shawky 2017 (345)

Reason for exclusion	Review

UniversityofCalifornia 2013 (402)

Reason for exclusion	Withdrawn

UniversityofVirginia 2004 (403)

Reason for exclusion	Wrong study design

Ursin 1996 (404)

Reason for exclusion	Wrong outcomes

Vachon 2013a (123)

Reason for exclusion	Wrong study design

Vachon 2013b (267)

Reason for exclusion	Wrong outcomes

Whitman 2000 (405)

Reason for exclusion	Review

4.4.2 Risk of bias in included studies

Study participation

The risk of bias for study participation was moderate for most studies because the source populations were not adequately described. The two sub-studies within clinical trials (19, 374) were the only studies to give an indication of key characteristics of the source population, either in the text or through another referenced study. Four studies (19, 372-374) conducted an analysis to test the difference in key characteristics between women in the source population and women included in the study sample. The two sub-studies within clinical trials included information on both of these criteria and were rated low risk.

Study attrition

The risk of bias for study attrition was mixed across studies. Four studies were rated moderate risk because of a lack of information on participant drop-out, loss to follow-up or reasons for censoring. One study (372) required women to still be alive at the time of interview for the study, therefore it lacked information on women who had died before recruitment. Women who had a later breast cancer diagnosis were also more likely to be in the study, leading to a potential survival bias. This study was therefore given a high risk of bias for study attrition.

Prognostic factor measurement

The risk of bias for the measurement of mammographic density was moderate or low. Studies were deemed to be at a higher risk of bias if they did not provide information on the experience of the reader(s), the time between baseline and follow-up mammogram, blinding to treatment (if the sample of women could be on more than one drug) or case-control status, reliability of measurements, number of women who had density measured, or reasons for missing density data. Two studies (19, 265) were given a moderate rating for risk of bias for the prognostic factor measurement because the density reduction cut-point was determined by the data (19) or an 'a priori' definition that was not justified (31).

Outcome measurement

The risk of bias for the outcome measurement was the domain considered to be at the lowest risk of bias. The two studies (263, 264) rated moderate risk conducted a survival analysis but did not define a start of follow-up or reasons for censoring, and there was no indication of women lost to follow-up. It is essential to determine these factors when conducting a survival analysis. The other studies to not report reasons for censoring or loss to follow-up were deemed to be low risk because they were linked to population registries or clinical trials.

Study confounding

The risk of bias for study confounding was mixed across studies. Two studies did not give clear definitions for age (263, 264), adjustments were unclear in three studies (264, 373, 374) and one study was potentially affected by recall bias since women recalled their confounding risk factors over telephone (372). One study (263) did not adjust for the confounding factor, chemotherapy, even though this was shown to have an effect on density reduction. Chemotherapy can induce menopause in premenopausal women, causing an oestrogen deprivation and reduction in density as well as recurrence. Therefore, adjustments should be made for chemotherapy in adjuvant studies where necessary.

Statistical analysis and reporting

The risk of bias for statistical analysis and reporting was considered to be the highest risk domain. The statistical methods for survival analysis were not adequately defined in two studies (263, 264) and the subgroup analysis of interest in one study (373) was not described in detail. The study by Knight et al. was not designed to look at women on endocrine therapy specifically and hence the analysis was adjusted for tamoxifen use, making the results ineligible for the prognostic or predictive biomarker review (372). Additionally, the results for van Nes et al. were not reported, so this study was given a high risk of bias for reporting (374).

The risk of bias in each domain and by each study is outlined in Figure 4.4. A detailed description of the risk of bias judgements can be found in appendix B.VII.

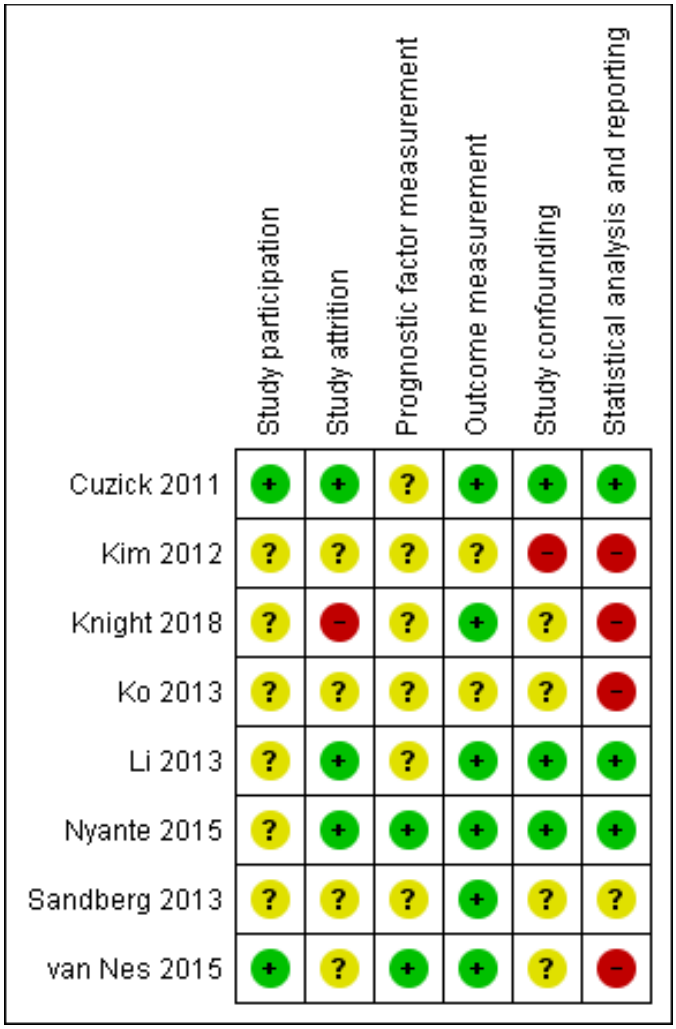


Figure 4.4: Risk of bias plot (red: high risk, yellow: moderate risk, green: low risk).

4.4.3 Effects of interventions

The different classes of drugs, outcomes, mammographic density measures and effect measures (for instance, the cut-points used) of the included studies were deemed too heterogeneous to be able to conduct a meaningful meta-analysis. Instead, the results of each study were reported in the 'Summary of findings' tables (Table 4.3 and Table 4.4). The study by Knight et al. (372) could not be included in the prognostic or predictive biomarker review due to the adjustment for endocrine therapy, and the study by van Nes et al. (374) was not included in the prognostic biomarker review because of a lack of reported results.

In the prognostic biomarker review (prevention), Cuzick et al. reported a 68% reduction in breast cancer risk with prophylactic tamoxifen for women who had a 12-18 month visually-assessed percent density reduction ≥10% compared with no change (OR=0.32 (95% CI, 0.14 to 0.72)) (19). For the prognostic biomarker review in the treatment setting, Kim et al. reported HRs of 0.66 (95% CI, 0.40 to 1.09) and 0.14 (95% CI, 0.02 to 1.11) for risk of recurrence with an 8-20 month Cumulus-assessed percent density reduction ≥5% compared with <5% whilst on tamoxifen or AIs, respectively (264). Similarly, in the prognostic biomarker review (treatment), Ko et al. reported a 65% reduction in risk of recurrence for women with a 10-34 month tamoxifen-induced reduction in BI-RADS density compared with no reduction (HR=0.35 (95% CI, 0.17 to 0.68)) (263). For mortality (treatment) in the prognostic biomarker review, Li et al. reported a 50% reduction in risk of breast cancer death with 6-36 month tamoxifen-induced relative reduction in dense area (machine learning area-based method) >20% compared with little change (≤9% increase to ≤10% reduction) (HR=0.50 (95% CI, 0.27 to 0.93)) ((265)). In another prognostic biomarker review study (treatment), Nyante et al. reported a 56% decreased risk of breast cancer death with a 3-26 month tamoxifen-induced reduction in Cumulus-assessed percent density of >8.7% compared with <0.5% (OR=0.44 (95% CI, 0.22 to 0.88)) (266). The final prognostic biomarker study (treatment) reported an OR of 0.52 (95% CI, 0.18 to 1.51) for risk of contralateral breast cancer with a 1-5 year reduction in percent density (machine learning area-based method) of ≥10% compared with little change (<10% reduction to <10% increase) whilst on endocrine therapy (373). In the predictive biomarker review (prevention), the OR of risk of breast cancer for an interaction between prophylactic tamoxifen and 12-18 month visually-assessed percent density reduction (≥10% or <10%) was not reported, but it was calculated as 0.53 (95% CI (0.21 to 1.32)) from raw data provided by the study authors (19).

Table 4.3: Endocrine therapy-induced mammographic density reduction vs. No endocrine therapy-induced mammographic density reduction as a prognostic biomarker

Endocrine therapy-induced mammographic density reduction vs. No endocrine therapy-induced mammographic density reduction as a prognostic biomarker			
Patient or population: Women on endocrine therapy (SERMs or AIs) **Setting:** Prevention or Treatment **Intervention:** Endocrine therapy-induced mammographic density reduction **Comparison:** No endocrine therapy-induced mammographic density reduction			
Outcomes	**Impact**	**№ of participants (studies)**	**Certainty of the evidence (GRADE)**
Incidence of invasive breast cancer and DCIS: Prevention, Tamoxifen	One study reported an OR of 0.32 (95% CI, 0.14 to 0.72) for 12-18 month visually-assessed percent density reduction ≥10% compared with no density change.	51 cases 456 controls (1 observational study)	⨁⨁⊖⊖ LOW
Recurrence: Treatment, Tamoxifen	One study reported an HR of 0.66 (95% CI, 0.40 to 1.09) for 8-20 month Cumulus-assessed percent density reduction ≥5% compared with <5%. Another study reported an HR of 0.35 (95% CI, 0.17 to 0.68) for 10-34 month reduction in BI-RADS density compared with no reduction (or increase).	1956 (2 observational studies)	⨁⊖⊖⊖ VERY LOW [1 2]
Recurrence: Treatment, AIs	One study reported an HR of 0.14 (95% CI, 0.02 to 1.11) for 8-20 month Cumulus-assessed percent density reduction ≥5% compared with <5%.	175 (1 observational study)	⨁⊖⊖⊖ VERY LOW [3 4]

Table 4.3 continued

Breast cancer mortality: Treatment, Tamoxifen	One study reported an OR=0.44 (95% CI, 0.22 to 0.88) for 3-26 month Cumulus-assessed percent density reduction >8.7% compared with <0.5%. Another study reported an HR of 0.50 (95% CI, 0.27 to 0.93) for 6-36 month relative reduction in dense area (machine learning area-based method) >20% compared with stable density (≤9% increase to ≤10% reduction).	1st study: 97 cases 252 controls; 2nd study: 26 events/217 exposed, 49 events/257 unexposed (2 observational studies)	⨁⨁⊖⊖ LOW
Incidence of a secondary primary breast cancer (e.g. in the contralateral breast): Treatment, unknown endocrine therapy	One study reported an OR of 0.52 (95% CI, 0.18 to 1.51) for 1-5 year reduction in percent density (machine learning area-based method) ≥10% compared with stable density (<10% increase to <10% reduction).	87 cases 87 controls (1 observational study)	⨁⊖⊖⊖ VERY LOW [5 6]

CI: Confidence interval; **HR:** Hazard ratio; **OR:** Odds ratio

GRADE Working Group grades of evidence

High certainty: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1] *Little information about study participation and attrition, no information on start of follow-up or reasons for censoring, unclear if adjustments made in one study and no adjustment for important confounding factors such as chemotherapy in another study, there may have been switching of endocrine therapy between baseline and follow-up mammogram in one study, density assessment may have been made on the ipsilateral breast in one study without an assessment of reliability of the density measure.*

[2] *Confidence interval includes the null effect for one study, no more than 147 events (one study does not give the number of events in the tamoxifen subgroup, but there are 80 events in the study overall + 67 events in the other study).*

[3] *Little information about study participation and attrition, no information on start of follow-up or reasons for censoring, unclear if adjustments made, there may have been switching of endocrine therapy between baseline and follow-up mammogram.*

[4] *Confidence interval includes the null effect, no more than 80 events (study does not give the number of events in the AIs subgroup, but there are 80 events in the study overall).*

[5] *The analysis in the subgroup of women on endocrine therapy was only a secondary objective of the study, therefore the data for the women included in the subgroup analysis are not reported and the individual endocrine therapies cannot be separated.*

[6] *Confidence interval includes the null effect, no more than 87 events (study does not give the number of events in the endocrine therapy subgroup analysis, but there are 87 cases on endocrine therapy in the study overall), large range of 1-5 years between diagnosis and follow-up mammogram might capture an effect on density change other than endocrine therapy, for example, weight change (adjustment for change in adiposity between mammograms was not considered).*

Table 4.4: Effect of mammographic density reduction in endocrine therapy group vs. Effect of mammographic density reduction in control group as a predictive biomarker

Effect of mammographic density reduction in endocrine therapy group vs. Effect of mammographic density reduction in control group as a predictive biomarker			
Patient or population: Women on endocrine therapy (SERMs or AIs) or a control group **Setting:** Prevention or Treatment **Intervention:** Effect of mammographic density reduction in endocrine therapy group **Comparison:** Effect of mammographic density reduction in control group			
Outcomes	**Impact**	**№ of participants (studies)**	**Certainty of the evidence (GRADE)**
Incidence of invasive breast cancer and DCIS: Prevention, Tamoxifen	One study reported an OR of 0.53 (95% CI, 0.21 to 1.32) for an interaction between prophylactic tamoxifen and 12-18 month visually-assessed percent density reduction (≥10% or <10%).	123 cases 942 controls (1 observational study)	⨁⨁⨀⨀ LOW [1] [2]
CI: Confidence interval; **OR:** Odds ratio			
GRADE Working Group grades of evidence **High certainty:** We are very confident that the true effect lies close to that of the estimate of the effect **Moderate certainty:** We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different **Low certainty:** Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect **Very low certainty:** We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect			

[1] *This is the only study found to investigate whether density change can be used as a predictive biomarker for women on endocrine therapy.*

[2] *Confidence interval includes the null effect, 123 cases provides small power to detect an interaction effect.*

4.5 Discussion

4.5.1 Summary of main results

This review demonstrated a potential use of mammographic density as a prognostic or predictive biomarker for endocrine therapy, but the evidence was limited by some quality issues in the included studies.

All studies included in the quantitative synthesis were observational and the number of studies testing each outcome was considerably small, with no more than two studies contributing to an outcome. Most of the studies focused on the mammographic density biomarker whilst on tamoxifen, and only one study provided results for the density change biomarker for treatment with AIs (264). One study included women on 'endocrine therapy' but did not stipulate a particular drug (373). Only one study contributed to the predictive biomarker review (19), whereas six studies were included in the prognostic biomarker review (19, 263-266, 373). Similarly, only one study was in the preventive setting (19), whilst five studies were in the adjuvant setting (263-266, 373).

There was a great deal of variation between studies. Density measurements included visual percentage score, BI-RADS categorical assessment, Cumulus semi-automated percentage score and machine-learned fully-automated percentage and absolute scores. Multiple cut-points for density reduction were tested, including 5%, 10% or tertile cut-points (absolute percentage density reduction), a 20% cut-point (relative dense area reduction), and BI-RADS categories. The characteristics of participants also varied. Studies were conducted in European, North American and Asian populations of premenopausal and postmenopausal women, with different disease status at first diagnosis in treatment studies.

As seen in the 'Summary of findings' tables, the GRADE certainty of evidence for outcomes was either low or very low, which was mainly driven by all studies starting at a low level of certainty due to their observational design. Additionally, some studies were downgraded because of high risk of bias or indirectness whereby studies were not designed to address the objectives of this review.

4.5.2 Overall completeness and applicability of evidence

The evidence in this review is reasonably limited and more studies are required to improve the certainty of evidence for mammographic density to be used as a biomarker for endocrine therapy. The findings of this review suggest that a mammographic density biomarker is

currently applicable to further research, but is not yet reliable enough to be applied to clinical practice. More research is needed on the use of mammographic density as a biomarker in response to other SERMs beyond tamoxifen as well as AIs.

4.5.3 Quality of the evidence

There was a large amount of heterogeneity across studies and the amount of evidence was low, with only six studies included in the quantitative synthesis (19, 263-266, 373). There was very low certainty of evidence for the recurrence and contralateral breast cancer outcomes which was driven by high risk of bias and indirectness, respectively. Additionally, the recurrence outcome on AI treatment and contralateral breast cancer incidence outcome (with a mixture of endocrine therapies) had evidence from only one study each (264, 373). There was a low certainty of evidence for the 'Incidence of invasive breast cancer and DCIS' outcome and the 'Breast cancer mortality' outcome. The influence of bias was relatively low for these outcomes, but this was counterbalanced by the small number of contributing studies. Breast cancer incidence was assessed in only one study (prognostic and predictive) (19) and breast cancer specific mortality was assessed in two studies (265, 266); perhaps making the latter the outcome with the highest certainty of evidence. For the predictive biomarker review, the addition of the ROBINS-I tool in assessment of risk of bias up-weighted the quality of evidence of the observational study from low to high because the placebo comparison group gave more support for the effect being treatment-induced. Nonetheless, this was downgraded again because there was only one contributing study that reported an imprecise effect with a confidence interval covering the null effect (19).

4.5.4 Potential biases in the review process

A potential bias of the review process is the exclusion of grey literature which included five conference abstracts (382, 384, 393, 397, 399). These were regarded as ineligible because they had not been peer-reviewed and the abstracts were limited in the amount of information they could report, such as full statistical methodology, justification for chosen techniques, and number of women with density reductions by outcome. These studies may be eligible for inclusion in a subsequent updated version of this review once they are reported as peer-reviewed full texts. Similarly, there were two studies without reported results (376, 392), two studies still recruiting (380, 400) and one study still collecting data (401), which may also be eligible for inclusion in an updated review once results are published as peer-reviewed full texts. One study used a non-validated MRI density method (390), which may be eligible in an updated review if more than one study (outside of the review studies) can show an association between the density method and breast cancer risk.

4.5.5 Agreements and disagreements with other studies or reviews

A systematic review by Shawky et al. entitled "Mammographic density: a potential monitoring biomarker for adjuvant and preventative breast cancer endocrine therapies" (345) reported findings similar to this review, although an assessment of quality was not included in their paper. Shawky et al.'s review included a study that was excluded from this review because it was a conference abstract (399). The study by Knight et al. that was included in this review was published after Shawky et al.'s paper and was hence not included in their list of studies. A recent systematic review by Kanbayti et al. (346) reported similar findings to this review, although their review assessed the relationship between mammographic density reduction and patient outcomes in women receiving any type of breast cancer treatment, therefore the focus was not specific to endocrine therapy. Furthermore, the review included two studies that were excluded at the full text screening stage of this review (331, 390). This is due to differences in the inclusion criteria: Kanbayti's study did not specify that the density measurement was to have been validated as a breast cancer risk factor in more than one study (outside of the review studies), and the density biomarker did not have to be derived from a baseline and a follow-up mammogram as was stipulated in this review. Moreover, Kanbayti et al. did not include the study by Cuzick et al. (19) because their population of interest was breast cancer patients treated in the adjuvant setting only.

4.6 Conclusions

4.6.1 Implications for practice

If mammographic density is determined to be a prognostic or predictive biomarker for endocrine therapy, it may be useful as a tool for measuring a woman's response to treatment. Currently, women have to rely on a 'wait-and-see' approach to assess if a course of endocrine therapy is working for them or not, but a density biomarker would provide an early indication of response. Knowing this information would allow women to make more informed decisions about whether to stay on an initial course of endocrine therapy or to change to another form of treatment. This would be particularly useful for those experiencing side effects of treatment, who may wish to balance the benefits and harms of continuing on the drug, which provides a more personalised approach for adjuvant or chemo-preventive care.

4.6.2 Implications for research

This review highlighted the need for more studies on density as a biomarker for endocrine therapy, to provide sufficient evidence before implementation of the biomarker into clinical practice. Some important points for consideration in future studies also arose as a result of the review. These are discussed in detail below:

- As described in Shawky et al. (345), it is important to establish the best density change predictor (qualitative vs. quantitative, area vs. volume, percent density vs. absolute density etc.), the best density change measurement method (manual vs. semi-automated vs. fully-automated etc.), the best density change cut-point, and whether absolute or relative density change is the better predictor of breast cancer risk and mortality.
- Compliance may have a confounding effect on the relationship between endocrine therapy-induced density change and breast cancer outcome. A lower compliance means less treatment is administered, hence a lower potential for density to reduce as well as a greater increase in risk and mortality. However, a lower compliance may be due to side effects of the treatment, which have been linked to a lower risk of recurrence (406, 407). Therefore, low compliance (as a result of side effects on endocrine therapy) might be a marker of treatment efficacy, and efforts should be made to test and control for the confounding effect of compliance in future studies.
- Another factor to consider is the negative confounding effect on results caused by masking. Women who experience a decrease in density whilst on treatment have a lower masking effect compared with women who do not have a reduction in density (provided all women have a similar starting density), so it is more likely that a cancer will be found on the follow-up mammograms of women who experience a treatment-induced density reduction. This would therefore cause an attenuation of the true effect on risk, recurrence and contralateral disease, suggesting that the association of treatment-induced density reduction on these breast cancer outcomes may in fact be stronger than what is typically found. It is therefore essential to allow for a long enough follow-up to avoid masking bias and to ensure that cancers occurring after measuring density change are adequately recorded. Nonetheless, follow-up should not be so long as to capture a possible increase in density after treatment cessation (350, 351) which could further influence the association between treatment-induced density change and breast cancer outcome.
- Efforts should be made to ensure there has been no previous treatment with SERMs or AIs before women enter the study (for example, for DCIS or prevention). This is important because residual effects of these previous treatments might influence density change and outcome during the study. For instance, the prolonged benefit of tamoxifen (200) may confound results if the number of women with residual effects is imbalanced between the

new treatment groups (predictive biomarker). Additionally, women previously treated with an endocrine drug may have already experienced a density decrease which cannot decrease further. However, they may simultaneously experience the benefit of treatment, thereby attenuating the relationship between density reduction and breast cancer outcome in the study.

- For studies of density change as a predictive biomarker, a suitable control group must be defined. Control groups defined as 'no tamoxifen' (as in Li et al. (265)) may have been treated with another SERM or AI, therefore changing the comparison group definition. Furthermore, women may have not received endocrine therapy because they had ER- breast cancers, which would be problematic if comparing groups of women with different starting prognoses.
- Finally, studies assessing density change and breast cancer outcome need to have clear definitions of start of follow-up to avoid a potential immortal time bias. This occurs when patients are defined as having a longer follow-up time than they actual received because they were not truly at-risk for all of the follow-up time. For instance, women have to be alive in order to undergo follow-up mammography, so a breast cancer death event cannot occur before a follow-up mammogram by definition of the study design. Therefore, women are only at-risk of death from follow-up mammogram. Only one study (265) accounted for this by starting follow-up at the follow-up mammogram. Incorrect follow-up times can be problematic if the 'immortal time' differs between women exposed to and not exposed to the predictor. Additionally, studies should ideally only include women who are still at-risk between mammograms so that the prognostic biomarker occurs before the event. Therefore (by definition) follow-up has to start at follow-up mammogram. If a study stipulates that women have to be on treatment for a certain amount of time (for example, 1 year), then women are only at-risk from the latter of 1 year from start of treatment or follow-up mammogram, and follow-up should start from that point in time. These timing issues are mainly relevant for cohort studies assessing time-to-event, but they can also be applicable to case-control studies that match on follow-up time. For example, if follow-up is started when treatment begins and the time between start of treatment and follow-up mammogram differs between cases and controls, then they may not be matched correctly on follow-up time.

The points outlined above indicate the implications for research derived from this review, and they are important considerations when designing future studies of mammographic density, endocrine therapy and breast cancer outcome.

Chapter 5: Anastrozole and mammographic density reduction in women at increased risk of breast cancer

5.1 Introduction

As mentioned in Chapter 4, SERMs and AIs are effective endocrine therapies for preventing and treating ER+ breast cancers. A particular selective oestrogen receptor modulator, tamoxifen, is a well-established drug used for preventing recurrence and reducing mortality in women with early stage ER+ breast cancer (408). As early as 1985, Cuzick and Baum also showed the benefit of tamoxifen in preventing new contralateral breast cancers (409), and in 2007, the IBIS-I trial showed that prophylactic tamoxifen reduced the risk of ER+ breast cancer in women at an increased risk of breast cancer by 30-40% (410). Further still, the benefit of chemoprevention with tamoxifen can be seen at least 10 years after an initial course of treatment (200).

However, AIs, including anastrozole, have been shown to be more effective in reducing the recurrence of early-stage ER+ disease than tamoxifen (349); and in 2014, analysis from IBIS-II showed that anastrozole reduced the risk of ER+ breast cancer in high-risk postmenopausal women by 60% (208). AIs have a low toxicity profile and tend to be well tolerated with fewer side effects than tamoxifen (411), making them a promising endocrine therapy for routine chemoprevention in postmenopausal women at increased risk of breast cancer.

It has been shown in numerous studies that tamoxifen reduces density in the preventive and adjuvant setting (203-206, 336), and most importantly, a reduction in density may be a marker of concurrent reduction in risk. The IBIS-I trial showed that high-risk women who experienced ≥10% density reduction after 12-18 months of prophylactic tamoxifen had approximately 63% lower risk of developing breast cancer compared with women on placebo, whilst women who experienced <10% density reduction on tamoxifen had a similar risk to women on placebo (19) (Figure 5.1). However, studies looking into the relationship between preventive AIs and density have so far shown modest or insignificant results (268-270), which is perhaps due to their small sample size; and larger studies looking into this association are in the adjuvant setting only (267, 271, 337). It is still unknown whether, like tamoxifen, preventive anastrozole treatment reduces density, and whether this is more than the natural decline occurring with age.

This study aims to compare mammographic density changes between the placebo and anastrozole arms of the IBIS-II Prevention trial to determine whether women on anastrozole experience different density changes (between baseline and first follow-up mammogram, and baseline and final follow-up mammogram) to women on placebo. Then, in Chapter 6, the effect of this treatment-induced density change will be assessed on breast cancer risk.

a) b)

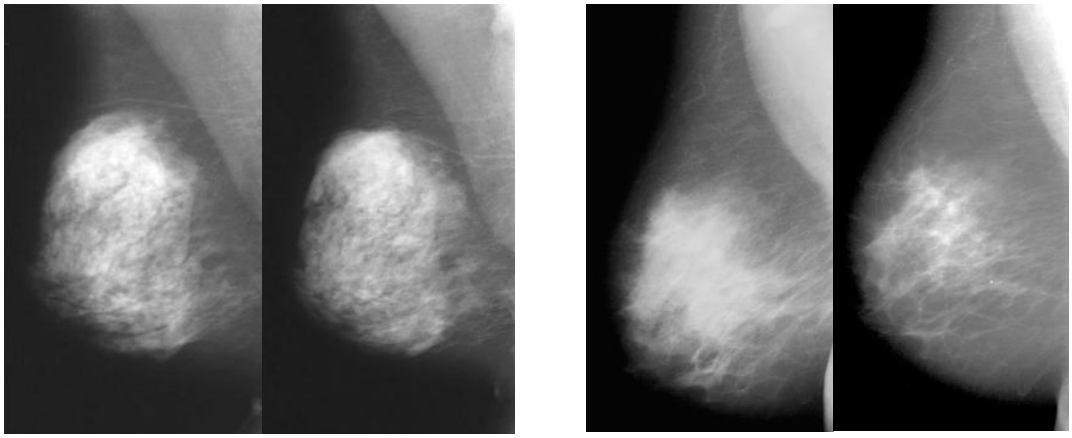


Figure 5.1: Visually assessed density change between baseline and follow-up mammogram (12-18 months after randomisation) in two women from the IBIS-I trial treated with tamoxifen.

a) No density change for one woman (left: baseline, right: 12-18 month follow-up mammogram), b) 15% reduction in density for another woman (left: baseline, right: 12-18 month follow-up mammogram). A density change similar to b) was associated with a concurrent approximately 63% reduction in breast cancer risk relative to women on placebo, whilst no density change, such as a), had a breast cancer risk similar to women on placebo.

A detailed description of the primary hypothesis and secondary hypotheses is outlined below:

Primary hypothesis

- H_0: There is no difference in age-adjusted change in density from baseline to first follow-up mammogram between patients in the anastrozole arm and patients in the placebo arm (continuous effect is primary).
- H_1: Age-adjusted change in density from baseline to first follow-up mammogram is different between patients in the anastrozole arm and patients in the placebo arm (continuous effect is primary).

Secondary hypothesis I

- H_0: There is no difference in change in density from baseline to first follow-up mammogram between patients in the anastrozole arm and patients in the placebo arm, after adjustment for age at randomisation, body mass index at randomisation, hormone replacement therapy use up to 12 months before randomisation, age at menopause, image type, and time between baseline and first follow-up mammogram.
- H_1: Change in density from baseline to first follow-up mammogram is different between patients in the anastrozole arm and patients in the placebo arm, after adjustment for age at randomisation, body mass index at randomisation, hormone replacement therapy use up to

12 months before randomisation, age at menopause, image type, and time between baseline and first follow-up mammogram.

Secondary hypothesis II

- H_0: There is no difference in age-adjusted change in density from baseline to final follow-up mammogram between patients in the anastrozole arm and patients in the placebo arm.
- H_1: Age-adjusted change in density from baseline to final follow-up mammogram is different between patients in the anastrozole arm and patients in the placebo arm.

Secondary hypothesis III

- H_0: There is no difference in change in density from baseline to final follow-up mammogram between patients in the anastrozole arm and patients in the placebo arm, after adjustment for age at randomisation, body mass index at randomisation, hormone replacement therapy use up to 12 months before randomisation, age at menopause, image type, and time between baseline and final follow-up mammogram.
- H_1: Change in density from baseline to final follow-up mammogram is different between patients in the anastrozole arm and patients in the placebo arm, after adjustment for age at randomisation, body mass index at randomisation, hormone replacement therapy use up to 12 months before randomisation, age at menopause, image type, and time between baseline and final follow-up mammogram.

Secondary hypothesis IV

- H_0: There is no difference in age-adjusted anastrozole-induced change in density from baseline to first follow-up mammogram between subgroups of covariates (age at randomisation, body mass index at randomisation, age at menarche, age at menopause, Tyrer-Cuzick 10-year risk, baseline density, age at first birth, oral contraception use, hormone replacement therapy use up to 12 months before randomisation, smoking status, history of atypical hyperplasia or LCIS, image type, and time between baseline and first follow-up mammogram).
- H_1: Age-adjusted anastrozole-induced change in density from baseline to first follow-up mammogram is different between subgroups of covariates (age at randomisation, body mass index at randomisation, age at menarche, age at menopause, Tyrer-Cuzick 10-year risk, baseline density, age at first birth, oral contraception use, hormone replacement therapy use up to 12 months before randomisation, smoking status, history of atypical hyperplasia or LCIS, image type, and time between baseline and first follow-up mammogram).

Secondary hypothesis V

- H_0: There is no difference in age-adjusted anastrozole-induced change in density from baseline to final follow-up mammogram between subgroups of covariates (age at randomisation, body mass index at randomisation, age at menarche, age at menopause, Tyrer-Cuzick 10-year risk, baseline density, age at first birth, oral contraception use, hormone replacement therapy use up to 12 months before randomisation, smoking status, history of atypical hyperplasia or LCIS, image type, and time between baseline and final follow-up mammogram).
- H_1: Age-adjusted anastrozole-induced change in density from baseline to final follow-up mammogram is different between subgroups of covariates (age at randomisation, body mass index at randomisation, age at menarche, age at menopause, Tyrer-Cuzick 10-year risk, baseline density, age at first birth, oral contraception use, hormone replacement therapy use up to 12 months before randomisation, smoking status, history of atypical hyperplasia or LCIS, image type, and time between baseline and final follow-up mammogram).

5.2 Methods

5.2.1 Study design

This study uses mammograms collected as part of a case-control study from the IBIS-II trial; a double-blind multicentre randomised placebo-controlled trial of 3864 postmenopausal women aged 40-70yr at an increased risk of breast cancer, aimed at determining whether chemoprevention of breast cancer with anastrozole is beneficial in this population of women. In brief, increased risk was determined by family history, previous benign disease with proliferation, nulliparity, LCIS, atypical hyperplasia (ductal or lobular), DCIS, and mammographic density ≥50% without use of HRT in the previous 3 months. Women were breast cancer-free at randomisation, and had not had a cancer in the previous 5 years (except non-melanoma skin cancer or in situ cancer of the cervix). Women were ineligible if they had taken SERMs for more than 6 months previously (unless they were taken as part of the IBIS-I trial and treatment had been completed at least 5 years prior to study entry), and women were not allowed to take a concurrent SERM or HRT whilst enrolled on the trial. Women were also excluded due to: premenopausal status, a prophylactic mastectomy, evidence of severe osteoporosis, concomitant disease, life expectancy <10 years, psychiatric or physical co-morbidities that could affect their ability to take part in the trial, and treatment with non-approved drugs up to 3 months before randomisation. Postmenopausal status was defined as meeting at least one of the following criteria: bilateral oophorectomy, aged over 60yr, aged ≤60yr with a uterus and amenorrhoea for at least 12 months, or aged ≤60yr without a uterus and FSH >30 IU/L. Eligible women were recruited between 2nd Feb 2003 and 31st Jan 2012 from 18

countries and were randomly assigned to a treatment arm on a 1:1 basis by central computer allocation. Women either received 1mg of orally-consumed anastrozole or a placebo alternative, to be taken every day for 5 years (n=1914 anastrozole and 1937 placebo). The primary endpoint was histologically confirmed breast cancer (invasive or non-invasive including ductal carcinoma in situ). A baseline questionnaire was completed by participants at recruitment to enable collection of information on confounding factors for risk of breast cancer. A mammogram was taken at baseline (after enrolment but before randomisation) and screening mammograms were taken at intervals as decided by the local co-ordinating centres, but at least every 2 years. These mammograms were subsequently sent to the trial co-ordinating centre.

For this nested case-control study, a cohort at the start of follow-up was defined, from which cases and controls were chosen. This study was designed such that the defined cohort was all randomised women who had participated in the IBIS-II main trial and had at least one available baseline MLO mammogram and at least one available first follow-up MLO mammogram collected as of May 2017 (n=1,274: 43 cases and 1,231 controls). Baseline mammograms were defined to be ≥0 months to <12.5 months prior to date of randomisation, first follow-up mammograms were ≥8.5 months to <38.5 months after date of randomisation, and final follow-up mammograms (if available) were ≥47.5 months to <60.5 months after date of randomisation. These time frames were chosen to mirror those used in IBIS-I (19) and in accordance with standard operating procedures for the local IBIS-II co-ordinating centres. Only MLO views were included to emulate the IBIS-I study (19) and because MLO views were predominantly collected during the trial. Follow-up began at each woman's first follow-up mammogram (not including the actual time of first follow-up mammogram) and ended at the earliest of: date of diagnosis (cases) or May 2017 if disease-free at this time (controls). This was done so that breast cancer events happened only after the density change predictor had occurred. Starting follow-up at first follow-up instead of treatment initiation prevented an immortal time bias (4.6.2) whereby the section of follow-up that was not truly 'at-risk' was excluded. Later in the chapter, cases are excluded if their event occurred before or at first follow-up mammogram, so (by definition) all women were breast cancer-free at first follow-up mammogram and the clock should start from this point. This bias is most relevant when conducting time-to-event analyses and is therefore not applicable in this study, but for completeness and to plan for potential further analyses, this was considered.

Cases were defined as women who developed breast cancer (invasive or non-invasive including DCIS) anytime throughout follow-up until (and including) May 2017 and controls were a random sample of the defined cohort who were breast cancer-free as of May 2017.

5.2.2 Power calculation

Density change for postmenopausal women in the IBIS-I nested study (19) was weighted based on age at randomisation of the IBIS-II cohort, to estimate the expected density change in a sample of women with the same age structure as IBIS-II. This was done separately for controls on placebo, controls on tamoxifen, cases on placebo and cases on tamoxifen (appendix C.I-C.IV). Different effect sizes (1/2 and 3/4) for tamoxifen were also tested by taking a weighting of the placebo and tamoxifen density change distributions corresponding to the proposed effect size (appendix C.V). This was done to allow anastrozole to have a weaker effect on density change than tamoxifen. For example, 1.11% of placebo controls and 0.23% of anastrozole controls were predicted to have a 30% reduction. If anastrozole was 3/4 as effective as tamoxifen at reducing density compared with placebo: 1.11% - (3/4)*(1.11%-0.23%) = 0.45% of anastrozole controls were predicted to have a 30% reduction. So, the proportions in each treatment arm became more similar as anastrozole became less effective. Then, an overall density change distribution for each treatment arm was formulated by weighting the density change distributions according to the distribution of cases and controls by treatment arm in IBIS-II (appendix C.VI-C.VII). The empirical cumulative distribution of density change in each treatment arm was then modelled, and two uniformly distributed random numbers were generated between (0,1) to find the inverse of the empirical cumulative distribution functions and simulate a density change value for placebo and anastrozole expected under the alternative hypothesis. This was repeated a number of times corresponding to the sample size in each treatment arm: sample sizes between 400 and 600 women per treatment arm were tested. Next, a linear regression model tested density change on treatment arm using the simulated data, and the simulation was recorded as a 'pass' if the (t-test) p-value for the treatment effect was <0.05 (based on a test of superiority). Simulations were repeated 10,000 times and the percentage of passes was counted to give the power. This was also done for dichotomised density change (≥10% reduction and <10% reduction), using a logistic regression model with treatment arm as the predictor. Results for different sample sizes can be found in appendix C.VIII.

Power was calculated to be 76%-94% for an anastrozole effect size 3/4 to 1 times that of tamoxifen, with 600 women per arm, to show a difference in density change (≥10% reduction and <10% reduction) from baseline to first follow-up mammogram between the two treatment arms at the 5% type-I error level (appendix C.IX). A proportion of women were also added to the required sample size to account for women with baseline density <10% that would later be excluded (based on the number of postmenopausal women with baseline density <10% in IBIS-I) (appendix C.X). In total, 1473 controls and 44 cases were required. A sample size larger than this was impracticable given the resources and number of mammograms received.

5.2.3 Data collection

An active effort was made to collect baseline and first follow-up mammograms for all IBIS-II participants. This was an attempt to include as many cases as possible in order to maximise statistical power. All national and international centres were emailed to request mammograms that had not yet been received by the co-ordinating centre. Response was greatest from the UK centres, ANZ centres (Australia and New Zealand), Belgium, Denmark, Italy and Finland, although the availability of case mammograms was somewhat disappointing. Many case mammograms were missing (lost, archived or destroyed in accordance with the local archiving policy) or had not been recorded at a participating site because the patient had moved address.

The multicentre aspect of the trial also caused unavoidable variation in the types of mammograms sent to the trial co-ordinating centre. Mammograms were taken using different machines and different image sizes. Mammograms could come in hardcopy form (viewed on a light box) or softcopy form (viewed on a computer screen). Film mammograms received by the trial co-ordinating centre mainly came in the form of original hardcopy or digitised softcopy images, but a handful (either hardcopy or softcopy) had gone through a number of iterations of conversion. Digital mammograms were mainly received in the form of original DICOM images (softcopy), however, as with film mammograms, some had been through various conversions before being sent to the trial co-ordinating centre. This created variability in image quality (for instance, different scanners at different sites). There was also a change of scanner at the trial co-ordinating centre throughout the trial, meaning that film mammograms received early on in the trial were scanned using a Vidar digitiser and saved as TIFF files, whilst films received later in the trial were digitised using an Array 2905 digitiser and saved as DICOM files.

5.2.4 Updating the Standard Operating Procedures

A number of issues were discovered in the standard operating procedure (SOP) outlining instructions for the IBIS-II members of staff to process, store and batch mammograms. Consequently, the SOP was updated by the author of this thesis (Emma Atakpa) and Dr Brentnall to rectify these issues. The main issues and implemented solutions are described below:

- Mammograms had a ‘levels’ conversion applied to them to try to standardise images. The ‘levels’ conversion was initially intended for film mammograms, however, its application to FFDM images was potentially detrimental to image quality since FFDM undergo their own optimisation when they are processed from raw to ‘for-presentation’. Additionally, on

inspection of the mammograms, it was apparent that this conversion had not been applied to all of the mammograms.

- Mammograms no longer underwent the 'levels' conversion. It was left to the radiologist (Dr Linda Metaxa) to edit contrast and brightness settings on a DICOM viewer to an optimum level for visual assessment.

- All mammograms were opened in Photoshop and had a black box placed over any patient identifiable information (PII) including name, date of birth and address. The image header was also anonymised if any PII was contained within it. All images were then saved as DICOM images. Whilst this was reasonable for film mammograms, this lost header information in FFDM images, which contained potentially useful imaging factors.
- FFDM were no longer opened in Photoshop, but were instead copied directly from the file directory and headers were anonymised automatically using anonymisation software provided by Volpara (no PII was contained on the FFDM images themselves). Raw images were also saved (although they were not included in this study, but were saved for potential future studies). This ensured that the original high quality DICOM image was intact and it also increased efficiency by reducing manual processing workload.

- Some images were of a very low resolution because they had been saved as JPEG images instead of TIFF or DICOM images. This appeared to be the case after digital mammograms had been opened in various DICOM viewers provided by the local sites that did not provide an option to save the image as a DICOM file. JPEG files compress images and remove some of the image information to make them more portable (412), and hence they can change the quality of the original mammogram.
- FFDM images were no longer opened in DICOM viewers, but were instead copied directly from the file directory, and hence they were saved in their original DICOM format. Film mammograms were to continue being saved in DICOM format after anonymisation.

It was decided that mammograms should undergo two stages of quality control (one by the thesis author (Emma Atakpa) at the batching stage and another by the radiologist, Dr Metaxa, at the density scoring stage) to reduce the remaining variation caused by differing mammogram types and quality.

After implementation of the updated SOP, IBIS-II bio-specimen staff were trained by the thesis author (Emma Atakpa) to request, process and store mammograms using this new process. With the IBIS-II staff and the thesis author (Emma Atakpa) requesting outstanding images and working through a backlog of retrieved deliveries, the number of women with at least one available MLO baseline and at least one available MLO first follow-up mammogram reached

1,274 as of May 2017. These 1,274 women made up the defined cohort which consisted of 43 cases and 1,231 controls.

5.2.5 Exclusions

Contralateral mammograms were kept for cases and mammograms from a randomly selected breast side (chosen using a random number generator) were kept for breast cancer-free controls to ensure that no cancers were present on the mammograms which could be misinterpreted as dense tissue. Women without information on diagnostic breast side or with bilateral breast cancer were excluded because it could not be guaranteed that either breast was breast cancer-free at the time of follow-up mammogram.

Duplicate mammograms were deleted so that each woman had only one mammogram per time point. Judgement of the best quality image was made by the thesis author (Emma Atakpa).

To ensure that density change from baseline to first follow-up mammogram could be used as a predictor of breast cancer risk, breast cancer events had to occur after the first follow-up mammogram. Therefore, women with an event before or at first follow-up mammogram were excluded. If cases had an incomplete diagnosis date that could not be reasonably rounded to the 1st day of the month, or if no diagnosis date had been entered, they were excluded. Final follow-up mammograms at or after the event were also removed to ensure that all women were breast cancer-free at all mammograms.

5.2.6 Exclusions – quality control (1)

The method for anonymisation at the processing stage was subject to human error (placing a black box around PII on the image and deleting sensitive information in DICOM headers), and a number of FFDM images produced an error after Volpara-provided anonymisation software. Therefore, further checks were conducted to ensure sufficient anonymisation before batching to send to the radiologist. Dr Brentnall ran Python code on all images to remove sensitive PII that was still contained in the DICOM headers. The mammograms were checked over by the thesis author (Emma Atakpa) to find any PII still contained on the image itself. If sensitive PII remained, these were anonymised following the method outlined in the updated SOP (open mammogram in Photoshop, place a black box over the PII and save as a DICOM file).

The mammograms were pseudo-anonymised using a random ID and suffix a, b or c for a woman's baseline, first and final follow-up mammograms, respectively. These new identifiers

were then burned onto the bottom centre of the mammograms using Python code written by Dr Brentnall to aid Dr Metaxa's analysis of density.

Image quality was then assessed by the thesis author (Emma Atakpa) and any mammograms that were judged to be below an expected standard for density assessment were removed. Reasons for exclusion were:

- Low resolution (JPEG images).
- Mammograms were too dark or light and could not be adequately seen after adjustment of brightness or contrast.
- Mammograms were too grainy.
- Incorrect mammographic view or breast side.
- Digital breast tomosynthesis slice – not the same imaging technique as the other mammograms i.e. 2D mammography.
- Writing pre-burned onto the mammogram (by study centres and hospitals) that covered the breast and affected density assessment.
- Wire localisation (this is an indicator of an abnormality possibly requiring a surgical biopsy) that affected the density assessment.
- Breast implant – breast tissue removed for insertion.
- Staples in the breast – indication of previous surgery which may have affected breast tissue composition.

After exclusions, there were 973 women (35 cases and 938 controls) with one MLO baseline and one MLO first follow-up mammogram.

5.2.7 Batching

All images were sent to Dr Metaxa in one batch using the IBIS-II secure file transfer protocol. Dr Metaxa viewed the mammograms using the DICOM viewer 'Sante DICOM Editor' (413) on a workstation used to view mammograms for detection. After initial testing, it was noted that some DICOM files produced an error on opening. Therefore, all mammograms were converted to TIFF format to allow easier reading of the mammograms. A Python code was used by Dr Brentnall to convert DICOM images to TIFF format, and for those files that produced an error, the thesis author (Emma Atakpa) used the 'Sante DICOM Editor' program (413) to convert the images manually. The TIFF and DICOM files produced the same image (no processing when converting the files).

5.2.8 Mammographic density scoring

Density was measured using visual assessment by an experienced radiologist (Dr Metaxa), who also made a judgement on whether she believed the original mammogram format to be film or digital.

Before undertaking the readings, an assessment was made on inter-reader reliability of VAS between Dr Metaxa and Dr Ruth Warren (expert radiologist who read the mammograms for IBIS-I). A test set of 100 IBIS-I mammograms had been previously scored by Dr Warren and were compared with Dr Metaxa's scores (Figure 5.2 and Figure 5.3). Correlation between the two readers was very good (Pearson correlation coefficient = 0.99, p<0.001, Spearman correlation coefficient = 0.98, p<0.001).

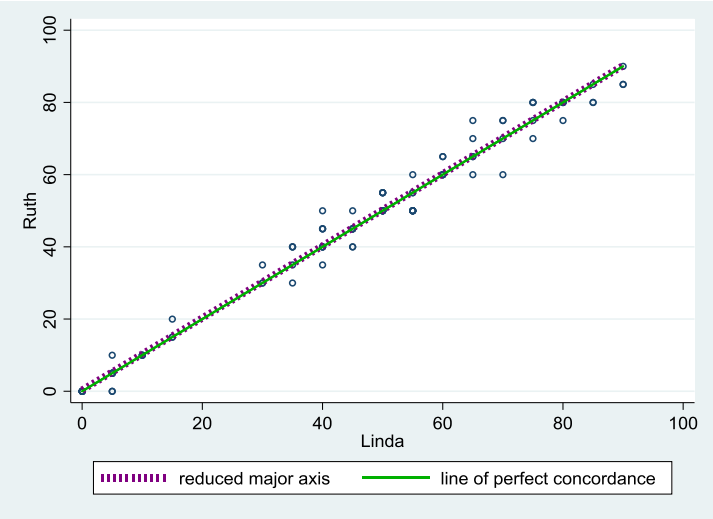


Figure 5.2: Correlation between Dr Ruth Warren's scores and Dr Linda Metaxa's scores.

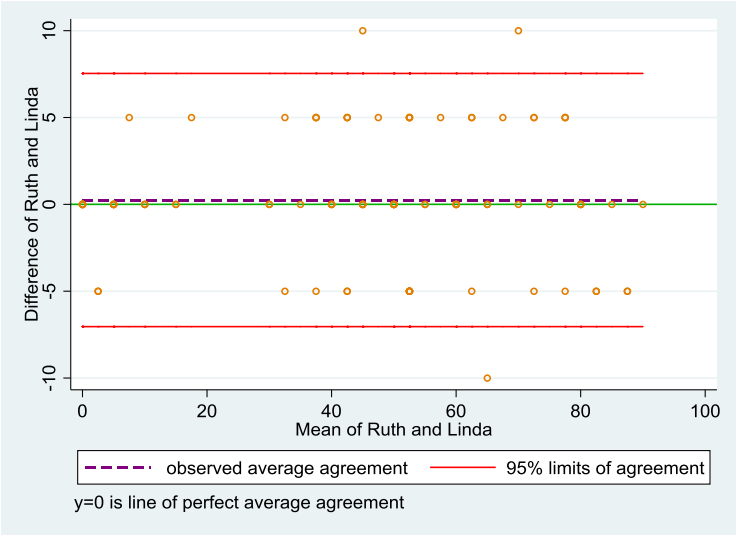


Figure 5.3: Bland Altman plots of concordance (359) between Dr Ruth Warren's scores and Dr Linda Metaxa's scores.

Each mammogram was scored in 5% increments (on a 21 point scale from 0 to 100%), following the same method as in IBIS-I (19, 203). Mammograms were read on a per-woman basis and each mammogram was read sequentially, in date order, in comparison with the previous mammograms for that woman, i.e. baseline first, followed by first follow-up mammogram (compared with baseline mammogram) and finally, final follow-up mammogram (compared with both baseline and first follow-up mammograms). Dr Metaxa was blinded to treatment group, case status and other information including risk factors.

5.2.9 Quality control (2)

After assessment by Dr Metaxa, mammograms were removed for the following reasons:

- Image quality was too low to reliably assess density.
- Incorrect breast side or mammographic view (for example, medio-lateral (ML) view instead of MLO view).
- Scarring indicating surgery (for example, a vacuum biopsy for a benign condition) which affected density assessment.
- Breast partially cut off as it was too big for the compression plate which affected the ability to assess density.
- Mammograms per-woman were of a different type (film or digital) – if baseline and first follow-up mammograms differed, the woman was excluded; if final follow-up mammogram differed from baseline and first follow-up mammogram, this mammogram was removed. These were excluded because digital mammograms tend to be darker than film mammograms and may therefore look less dense than if the breast had been imaged onto a film.
- Digital mammograms per-woman were processed using different machines (e.g. Fuji, Philips, Siemens) – if baseline and first follow-up mammograms differed (and this was deemed to affect the true measure of density change), the woman was excluded; if final follow-up mammogram differed from baseline and first follow-up mammogram, this mammogram was removed.
- Mammograms per-woman had substantially different radiographer techniques (e.g. different orientation of the breast positioning) which affected the spread of tissue and hence affected the true measure of density change – if baseline and first follow-up mammograms differed, the woman was excluded; if final follow-up mammogram differed from baseline and first follow-up mammogram, this mammogram was removed.

The resulting number of women with a baseline and a first follow-up mammogram was 842 (31 cases and 811 controls).

5.2.10 Baseline mammographic density at least 10%

First and final density change were defined as the difference between baseline density and first or final follow-up mammogram density, respectively. This was measured continuously and dichotomised into <10% or ≥10% absolute reduction (and <5% or ≥5% absolute reduction). To be able to lose 10% density, women had to start with at least 10% density at baseline. Therefore, women with a baseline density <10% were excluded.

The resulting number of women was 576 (19 cases and 557 controls) who had one available MLO baseline mammogram and one available MLO first follow-up mammogram.

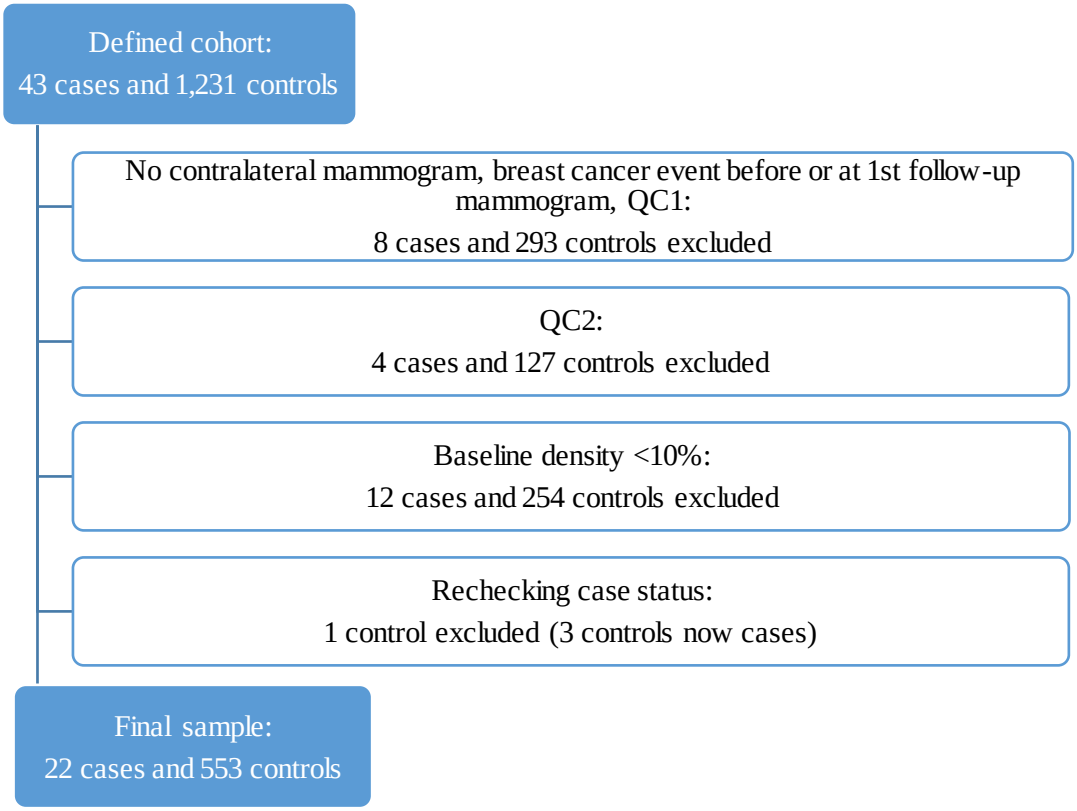


Figure 5.4: Flow diagram of study sample; QC=quality control.

Case status was updated (October 2018), to reassign women who were initially recorded as controls to now be cases. New cases whose ipsilateral breast had been assessed were excluded; as were new cases with diagnosis before or at first follow-up mammogram (i.e. there was a delay in reporting their diagnosis to the trials unit after mammograms had been sent to the radiologist). Final follow-up mammograms at or after diagnosis were removed for these new cases.

The new definition for cases was therefore described as women who had been diagnosed with breast cancer as of October 2018, and breast cancer-free controls were defined as women who had not been diagnosed with breast cancer as of October 2018.

The final number of women with one available MLO baseline mammogram and one available MLO first follow-up mammogram was 575 (22 cases and 553 controls). These women contributed mammograms from 46 participating centres across 6 countries: the UK, Italy, Finland, Denmark, Ireland and ANZ (Australia and New Zealand were grouped into one co-ordinating centre).

5.2.11 Statistical methods

A statistical analysis plan was developed for the study (appendix C.XXIII). All statistical analysis was conducted using Stata (316), and tests were two-sided with a significance level of 5%. The study used an intention-to-treat analysis (so the study design was based on the initial treatment intent, not the treatment that was eventually administered); therefore time on treatment was not included in adjustments. Only the trial statistician (Dr Ivana Sestak) was un-blinded to treatment allocation, therefore a set of Stata code was sent to Dr Sestak to run on the un-blinded data. A proforma was developed and sent to the IBIS-II Trial Steering Committee, who approved the study (appendix C.XXIV).

5.2.11.1 Baseline characteristics

Baseline covariates (collected via questionnaires as part of the IBIS-II trial) were: age at randomisation (years), body mass index (BMI) at randomisation (kg/m^2), age at menarche (years), age at menopause (years), Tyrer-Cuzick 10-year risk (%; version 7 excluding breast density), baseline density (%), age at first birth (nulliparous/>27/21-27/≤20), oral contraception use (never/previously/currently), hormone replacement therapy use up to 12 months before randomisation (no/yes; categorised in line with the IBIS-II main study (208); since HRT was not allowed during the trial, its use up to 12m before randomisation was considered to be a confounding factor which could have increased baseline density), smoking status (never/former/current), history of atypical hyperplasia or LCIS (no/yes) and image type (film/digital). Baseline covariates were summarised overall and by treatment arm using frequency tables. Frequency counts and percentages were provided for categorical data and means (standard deviation, SD) and medians (interquartile range, IQR) were provided for continuous data. Two-sample t-tests (for mean difference) and Wilcoxon rank sum tests (for median difference) were applied between treatment arms for continuous data and Pearson chi-

squared tests (Fisher's exact tests if cell size <5) tested independence between treatment arms in categorical data.

5.2.11.2 Effect of covariates on baseline mammographic density

An exploratory analysis assessed the effect of baseline covariates on baseline density using multivariable linear and logistic (baseline density dichotomised into <50% or ≥50%) regression models; adjusted for age at randomisation only and again for all covariates (except baseline density). Non-parametric empirical bootstrap 95% confidence intervals were used in linear regression models with continuous density outcome (2.2.4.3).

5.2.11.3 Change in mammographic density

There was a left skew on the density change data because density was assessed in comparison with previous mammograms and was more likely to decrease or stay the same than to increase (appendix C.XI-C.XXII), giving a small variance in per-woman readings (414). Therefore, density change did not follow a normal distribution, hence non-parametric methods (medians (interquartile range), Wilcoxon rank sum tests, and empirical bootstrap 95% confidence intervals in linear regression models (2.2.4.3)) were used as well as parametric methods (means (standard deviations) and two-sided t-tests). Furthermore, density change was assessed as a dichotomous variable (<10% absolute reduction or ≥10% absolute reduction, and <5% absolute reduction or ≥5% absolute reduction). The 5% cut-point was chosen because density was measured in 5% increments and the 10% cut-point was chosen because it was the minimum change that could be reproducibly detected in IBIS-I (19).

5.2.11.4 Change in mammographic density – Boyd categories

A cross tabulation was used to show the number of women in each Boyd category (0%, 1-10%, 11-25%, 26-50%, 51-75%, 76-100%) at baseline and first and final follow-up, overall and by treatment.

5.2.11.5 Change in mammographic density – unadjusted tests

Two-sample t-tests (for mean difference) and Wilcoxon rank sum tests (for median difference) assessed whether there was a difference in first and final follow-up density change between treatment arms, and Pearson chi-squared tests (Fisher's exact tests if cell size <5) tested whether there was independence in dichotomous first and final follow-up density change between treatment arms.

5.2.11.6 Change in mammographic density – adjusted regression models

The primary analysis used linear (continuous density change) regression models to examine the association between treatment arm and change in density from baseline mammogram to first follow-up mammogram, adjusted for age at randomisation (years). This was also assessed using logistic regression models for dichotomous density change. The secondary analysis (I) repeated the primary analysis with adjustment for age at randomisation (years), body mass index at randomisation (kg/m^2), hormone replacement therapy use up to 12 months before randomisation (no/yes), age at menopause (years), image type (film/digital) and time between baseline and follow-up mammogram (years). As an exploratory analysis, models were also adjusted for baseline density (%) and age at randomisation (years) only, and baseline density (%) and all other adjusting factors. The secondary analysis (II) repeated the primary analysis and the secondary analysis (III) repeated the secondary analysis (I) (including exploratory adjustments) but for final follow-up density change, in a subgroup of women who had an available final mammogram density score.

The adjusting covariates for regression models were chosen based on literature which suggests that they have a confounding effect on density change, including those shown to be significant in the IBIS-I trial (203). Age at randomisation was retained in all regression models, regardless of significance, because age is a strong confounder of density and density change. To aid interpretation, continuous adjusting variables were centred about their median in regression models.

5.2.11.7 Change in mammographic density – subgroup analyses

The secondary analyses (IV) and (V) used two-sided t-tests and Wilcoxon rank-sum tests (2 subgroups) or ANOVA F-tests and non-parametric Cuzick trend tests (>2 ordered subgroups) to assess heterogeneity, namely whether the effect of anastrozole on first and final density change varied between different covariate subgroups. Univariate logistic regression models were also used to assess the odds of a high density reduction (≥10% or ≥5% absolute reduction) in one subgroup relative to another subgroup, in anastrozole treated patients only. The covariates assessed in this subgroup analysis were the same as those at baseline, plus time between baseline and follow-up mammograms (years). Continuous variables were separated into subgroups by their median value.

5.2.11.8 Change in mammographic density – sensitivity analyses

As a sensitivity analysis, all analyses of density change at first follow-up mammogram were repeated in the subgroup of women with an available final mammogram.

5.3 Results

Table 5.1: Baseline characteristics overall and by treatment arm.

Variable	n	Overall		n	Placebo		n	Anastrozole	
		Mean (SD)	Median (IQR)		Mean (SD)	Median (IQR)		Mean (SD)	Median (IQR)
Age at randomisation (yr)	566	58.9 (5.5)	59 (55-63)	271	58.9 (5.7)	59 (55-63)	295	58.9 (5.4)	59 (54-63)
P*				0.91, 0.77					
Body Mass Index (kg/m^2)	562	27.1 (4.6)	26.5 (23.8-29.7)	268	27.0 (4.8)	26.5 (23.8-29.5)	294	27.1 (4.4)	26.5 (23.8-29.9)
P*				0.77, 0.57					
Age at menarche (yr)	562	12.9 (1.6)	13 (12-14)	269	12.9 (1.6)	13 (12-14)	293	12.8 (1.7)	13 (12-14)
P*				0.84, 0.69					
Age at menopause (yr)	563	48.3 (6.2)	50 (46-52)	268	48.4 (6.0)	50 (46-52)	294	48.4 (5.7)	50 (45-52)
P*				0.97, 0.91					
Tyrer-Cuzick 10-year risk (%)	568	8.6 (4.2)	7.8 (6.1-10.1)	273	8.7 (3.8)	7.9 (6.3-10.6)	295	8.6 (4.5)	7.6 (5.9-9.7)
P*				0.72, 0.15					
Baseline density (%)	575	43.5 (24.8)	45 (20-65)	276	44.3 (24.9)	45 (20-65)	299	42.7 (24.8)	40 (20-65)
P*				0.45, 0.45					
		n	**%**		**n**	**%**		**n**	**%**
Age at first birth (yr)									
Nulliparous	566	98	17.3	271	54	19.9	295	44	14.9
>27		121	21.4		57	21.0		64	21.7
21-27		245	43.3		110	40.6		135	45.8
≤20		102	18.0		50	18.5		52	17.6
P**				0.39					
Oral contraception use									
Never	565	120	21.4	270	65	24.1	295	55	18.6
Previously		441	78.1		205	75.9		236	80.0
Currently		4	0.7		0	0.0		4	1.4
P**				0.04#					
HRT use < 12 months before randomisation									
No	566	534	94.4	271	253	93.4	295	281	95.3
Yes		32	5.7		18	6.6		14	4.8
P**				0.33					
Smoking status									
Never	565	317	56.1	271	156	57.6	294	161	54.8
Former		73	12.9		32	11.8		41	14.0
Current		175	31.0		83	30.6		92	31.3
P**				0.70					
History of Atypical Hyperplasia or LCIS									
No	566	519	91.7	271	249	91.9	295	270	91.5
Yes		47	8.3		22	8.1		25	8.5
P**				0.88					
Image type									
Film	575	190	33.0	276	83	30.1	299	107	35.8
Digital		385	67.0		193	69.9		192	64.2
P**				0.15					

**P-value from two-sample t-test (for means) and Wilcoxon rank sum test (for medians), respectively, for continuous variables by treatment arm; **P-value from Pearson chi-squared test (# Fisher's exact test if cell size <5) for variable categories by treatment arm; interquartile range (IQR), standard deviation (SD).*

5.3.1 Baseline characteristics

Baseline characteristics were well balanced between treatment arms for all women (Table 5.1). Women who had used or were currently using oral contraception were slightly more likely to be in the anastrozole group, but the difference in numbers was small. Nine women transferred to a different participating centre where their study number changed. Therefore, some covariates, such as age, could not be matched to the study number and were missing for these women.

5.3.2 Timings of mammograms

The median (IQR) and range of time between baseline mammogram and randomisation was 0.2 years (0.1 years-0.4 years) and 0.0 years-1.0 years (i.e. 2.4 months (1.1 months-5.2 months) and 0 months-12.2 months). The median (IQR) and range of time between randomisation and first follow-up mammogram was 2.0 years (1.4 years-2.2 years) and 0.7 years-3.2 years, and the median (IQR) and range of time between randomisation and final follow-up mammogram was 4.4 years (4.1 years-4.8 years) and 4.0 years-5.0 years. The median (IQR) and range of time between baseline mammogram and first follow-up mammogram was 2.1 years (1.6 years-2.8 years) and 0.9 years-3.9 years. The median (IQR) and range of time between baseline mammogram and final follow-up mammogram was 4.6 years (4.4 years-5.0 years) and 4.0 years-5.9 years.

5.3.3 Effect of covariates on baseline mammographic density

Results from the age-adjusted and multivariable linear and logistic regression models for breast density at baseline are summarized in Table 5.2. Older age at randomisation was associated with reduced breast density (-0.45% (95% CI, -0.82 to -0.09) per year increase in age), but less so in multivariable models (-0.26% (95% CI, -0.64 to 0.11) per year increase in age). Higher body mass index was associated with a reduction in breast density in both age-adjusted (-1.69% (95% CI, -2.13 to -1.26) per kg/m^2 increase) and multivariable (-1.73% (95% CI, -2.17 to -1.31) per kg/m^2 increase) models. Age at first birth was another predictor of breast density in both age-adjusted and multivariable models. Compared with nulliparous women (n=98), women who had their first full-term birth below the age of 20 years (n=102) had approximately 15% lower absolute breast density. The other statistically significant variable was image type, which showed lower density for digital images. This is expected since digital mammograms appear darker and thus less dense than film mammograms.

Table 5.2: Association between baseline covariates and baseline breast density (continuous (%) and dichotomised into <50% or ≥50%) in age-adjusted and multivariable linear and logistic regression models.

Variable	n+	Age-adjusted linear regression		Multivariable linear regression		Age-adjusted logistic regression		Multivariable logistic regression	
		β-coefficient (95% CI)#	P-value##	β-coefficient (95% CI)#	P-value##	OR (95% CI)###	P-value###	OR (95% CI)###	P-value###
Age at randomisation (yr)*	566/553	-0.45 (-0.82,-0.09)	0.01	-0.26 (-0.64,0.11)	0.18	0.97 (0.94,1.00)	0.04	0.98 (0.94,1.02)	0.27
Body Mass Index (kg/m^2)*	562/553	-1.69 (-2.13,-1.26)	<0.01	-1.73 (-2.17,-1.31)	<0.01	0.87 (0.83,0.91)	<0.01	0.86 (0.82,0.90)	<0.01
Age at menarche (yr)*	562/553	0.55 (-0.73,1.84)	0.41	0.18 (-0.99,1.39)	0.77	1.04 (0.93,1.15)	0.50	1.02 (0.90,1.14)	0.79
Age at menopause (yr)*	562/553	0.11 (-0.23,0.46)	0.51	0.03 (-0.29,0.36)	0.86	1.00 (0.97,1.03)	0.99	0.99 (0.96,1.03)	0.61
Tyrer-Cuzick 10-year risk (%)*	564/553	0.45 (-0.05,0.94)	0.08	0.47 (-0.22,1.19)	0.19	1.03 (0.99,1.08)	0.11	1.02 (0.96,1.10)	0.48
Age at first birth (yr)**									
Nulliparous	566/553	Ref	-	Ref	-	Ref	-	Ref	-
>27		-7.67 (-14.50,-0.78)	0.03	-9.82 (-16.61,-2.79)	0.01	0.74 (0.43,1.27)	0.28	0.64 (0.35,1.15)	0.13
21-27		-8.97 (-15.00,-2.91)	<0.01	-9.92 (-16.04,-3.48)	<0.01	0.57 (0.35,0.91)	0.02	0.52 (0.30,0.89)	0.02
≤20		-15.72 (-22.48,-8.68)	<0.01	-15.40 (-22.30,-8.09)	<0.01	0.33 (0.19,0.60)	<0.01	0.31 (0.16,0.61)	<0.01
Oral contraception use**									
Never	565/553	Ref	-	Ref	-	Ref	-	Ref	-
Previously		-1.51 (-6.47,3.49)	0.55	1.39 (-3.30,6.18)	0.57	0.83 (0.54,1.27)	0.39	1.04 (0.66,1.66)	0.86
Currently		-11.27 (-32.00,36.28)	0.50	-2.95 (-28.48,42.58)	0.88	0.47 (0.05,4.65)	0.52	1.15 (0.10,13.3)	0.91

Table 5.2 continued

HRT use up to 12 months before randomisation**									
No	566/553	Ref	-	Ref	-	Ref	-	Ref	-
Yes		3.83 (-5.38,13.18)	0.42	3.23 (-5.80,12.83)	0.50	1.16 (0.56,2.39)	0.69	1.16 (0.51,2.62)	0.72
Smoking status**									
Never	565/553	Ref	-	Ref	-	Ref	-	Ref	-
Former		4.01 (-2.83,10.86)	0.25	1.74 (-4.29,7.72)	0.57	1.19 (0.71,2.00)	0.51	0.94 (0.53,1.68)	0.84
Current		2.55 (-1.98,7.06)	0.27	3.72 (-0.92,8.05)	0.10	1.08 (0.74,1.57)	0.69	1.17 (0.77,1.78)	0.47
History of Atypical Hyperplasia or LCIS**									
No	566/553	Ref	-	Ref	-	Ref	-	Ref	-
Yes		4.83 (-3.19,12.89)	0.24	2.46 (-7.41,12.31)	0.63	1.60 (0.88,2.92)	0.13	1.66 (0.63,4.35)	0.30
Image type**									
Film	566/553	Ref	-	Ref	-	Ref	-	Ref	-
Digital		-4.94 (-9.15,-0.67)	0.02	-7.59 (-11.61,-3.24)	<0.01	0.72 (0.51,1.03)	0.07	0.57 (0.38,0.84)	0.01

*All covariates adjusted for age at randomisation (yr) in age-adjusted models (except for age at randomisation (yr)), all covariates included in multivariable models; continuous variables centred about their median (see Table 5.1 overall column); * β-coefficient represents effect on baseline density per unit increase in covariate, odds ratio (OR) represents odds of having ≥50% baseline density per unit increase in covariate; ** β-coefficient represents difference in baseline density from reference category, OR represents odds of having ≥50% baseline density relative to the reference category; #empirical bootstrap 95% CI; ##P-value from z-test with known sample mean and standard deviation (the population is to the sample as the sample is to the bootstrap sample); ###95% CI and P-value from a Wald test; [+]number in age-adjusted model/number in multivariable model.*

Table 5.3: Cross tabulation of number of women in each Boyd category at entry to the study with category at first and final follow-up.

Boyd category at entry	Number of women												
	Boyd category at first follow-up						Total	Boyd category at first follow-up (women with a final mammogram)						Total
	0%	1-10%	11-25%	26-50%	51-75%	76-100%		0%	1-10%	11-25%	26-50%	51-75%	76-100%	
0%	-	-	-	-	-	-	0(0/0)	-	-	-	-	-	-	0(0/0)
1-10%	-	49(22/27)	1(1/0)	-	-	-	50(23/27)	-	14(9/5)	-	-	-	-	14(9/5)
11-25%	-	13(7/6)	147(65/82)	-	-	-	160(72/88)	-	3(2/1)	37(18/19)	-	-	-	40(20/20)
26-50%	-	-	5(2/3)	134(69/65)	1(0/1)	-	140(71/69)	-	-	2(1/1)	35(20/15)	-	-	37(21/16)
51-75%	-	-	-	6(1/5)	153(78/75)	-	159(79/80)	-	-	-	3(0/3)	41(20/21)		44(20/24)
76-100%	-	-	-	-	8(4/4)	58(27/31)	66(31/35)	-	-	-	-	2(2/0)	14(7/7)	16(9/7)
Total	0(0/0)	62(29/33)	153(68/85)	140(70/70)	162(82/80)	58(27/31)	575(276/299)	0(0/0)	17(11/6)	39(19/20)	38(20/18)	43(22/21)	14(7/7)	151(79/72)

Boyd category at entry	Number of women						
	Boyd category at final follow-up						Total
	0%	1-10%	11-25%	26-50%	51-75%	76-100%	
0%	-	-	-	-	-	-	0(0/0)
1-10%	-	14(9/5)	-	-	-	-	14(9/5)
11-25%	-	4(3/1)	36(17/19)	-	-	-	40(20/20)
26-50%	-	-	2(1/1)	35(20/15)	-	-	37(21/16)
51-75%	-	-	-	3(1/2)	41(19/22)	-	44(20/24)
76-100%	-	-	-	-	2(2/0)	14(7/7)	16(9/7)
Total	0(0/0)	18(12/6)	38(18/20)	38(21/17)	43(21/22)	14(7/7)	151(79/72)

The first number in each cell is the total number of subjects; numbers in parentheses are the placebo and anastrozole groups, respectively; '-' indicates no entries.

Table 5.4: Continuous change in density (%) by treatment arm.

	n	Mean (95% CI)	Standard deviation	P-value*	Median	IQR	P-value**
First follow-up							
Placebo	276	-0.82 (-1.12, -0.51)	2.59	0.28	0	(0,0)	0.13
Anastrozole	299	-1.05 (-1.36, -0.75)	2.65		0	(0,0)	
First follow-up (women with a final mammogram)							
Placebo	79	-1.08 (-1.71, -0.44)	2.85	0.94	0	(0,0)	0.58
Anastrozole	72	-1.04 (-1.56, -0.52)	2.21		0	(0,0)	
Final follow-up							
Placebo	79	-2.15 (-3.01, -1.30)	3.81	0.47	0	(-5,0)	0.43
Anastrozole	72	-1.74 (-2.48, -0.99)	3.16		0	(-5,0)	

**P-value from two-sided t-test (for means); **P-value from Wilcoxon rank sum test (for medians); interquartile range (IQR), 95% confidence interval (95% CI).*

Table 5.5: Dichotomised change in density (%) by treatment arm.

	n (% of follow-up)		P-value	n (% of follow-up)		P-value
	<10% reduction	≥10% reduction		<5% reduction	≥5% reduction	
First follow-up						
Placebo	268 (97.1%)	8 (2.9%)	0.49	236 (85.5%)	40 (14.5%)	0.14
Anastrozole	293 (98.0%)	6 (2.0%)		242 (80.9%)	57 (19.1%)	
First follow-up (women with a final mammogram)						
Placebo	75 (94.9%)	4 (5.1%)	0.37#	67 (84.8%)	12 (15.2%)	0.49
Anastrozole	71 (98.6%)	1 (1.4%)		58 (80.6%)	14 (19.4%)	
Final follow-up						
Placebo	75 (94.9%)	4 (5.1%)	1.00#	50 (63.3%)	29 (36.7%)	0.43
Anastrozole	69 (95.8%)	3 (4.2%)		50 (69.4%)	22 (30.6%)	

P-value from Pearson chi-squared test (#Fisher's exact test if cell size <5).

5.3.4 Change in mammographic density – Boyd categories

Change in density was summarised in terms of the number of women in each treatment group by Boyd scale at baseline and at first and final follow-up mammogram (Table 5.3). Movement between Boyd categories was minimal, and women who had a decrease in density moved by no more than one category below their baseline category. A similar percentage of women on placebo and anastrozole moved to a lower Boyd category at first follow-up mammogram (14/276=5% and 18/299=6%, respectively), with slightly more women moving down a category on placebo than anastrozole at final follow-up mammogram (7/79=9% and 4/72=6%, respectively). Increases in breast density were rare, and the two women (1 anastrozole control and 1 placebo control) who had increased density at first follow-up mammogram moved up by no more than one category. On inspection of the mammograms for these two women, both appeared to be caused by increases in dense tissue as opposed to weight-loss (decrease in breast fat), but the reason for the increase is unclear. These two women were not on HRT throughout the trial and they did not go on to develop breast cancer.

5.3.5 Change in mammographic density – first follow-up mammogram

At baseline, the mean breast density was 44.3% (95% CI, 41.4% to 47.3%) for the placebo group and 42.7% (95% CI, 39.9% to 45.6%) for the anastrozole group (p=0.45 from two-sample t-test and Wilcoxon rank sum test). By the first follow-up mammogram, breast density had fallen to an average 43.5% (95% CI, 40.5% to 46.4%) in the placebo group, with a change from baseline of -0.82% (95% CI, -1.12% to -0.51%) (Table 5.4). By the first follow-up mammogram, breast density had fallen to an average 41.7% (95% CI, 38.9% to 44.5%) in the anastrozole group, with a change from baseline of -1.05% (95% CI, -1.36% to -0.75%). The difference in density change at first follow-up mammogram between treatment arms (anastrozole minus placebo) was not significant (mean 0.24%, 95% CI, -0.19% to 0.67%, p=0.28 from two-sample t-test and p=0.13 from Wilcoxon rank sum test).

5.3.6 Change in mammographic density – final follow-up mammogram

By the final follow-up mammogram, breast density had fallen to an average 41.5% (95% CI, 35.8% to 47.3%) in the placebo group, with a change from baseline of -2.15% (95% CI, -3.01% to -1.30%). For anastrozole, breast density had fallen to an average of 42.5% (95% CI, 36.7% to 48.3%), with a change from baseline of -1.74% (95% CI, -2.48% to -0.99%). The difference in density change at final follow-up mammogram between treatment arms (anastrozole minus placebo) was not significant (mean -0.42%, 95% CI, -1.55% to 0.72%, p=0.47 from two-sample t-test and p=0.43 from Wilcoxon rank sum test) Table 5.4.

At first glance, anastrozole appeared to decrease more than placebo at first follow-up mammogram (-1.05% vs. -0.82%), but less than placebo at final follow-up mammogram (-1.74% vs. -2.15%), suggesting a possible slowing down of anastrozole-induced rate of change in the latter stages of follow-up. However, when assessing the subgroup of women with an available final follow-up mammogram, anastrozole-treated women had less of a density reduction than those on placebo at both first and final follow-up mammogram (-1.04% vs. -1.08% and -1.74% vs. -2.15%, respectively). Therefore, the effects seen at final follow-up mammogram may have been specific to this subgroup of women only.

Results were similar for dichotomised density, where tests for differences in density change between treatment arms were non-significant (Table 5.5). The number of women losing at least 10% density at first and final follow-up mammograms was small in both treatment arms, but numbers were larger for at least 5% density reduction. Although not significant, there was some suggestion that more women lost at least 5% density on anastrozole than placebo at first follow-up mammogram (19.1% vs. 14.5%), but this was not the case at final follow-up mammogram. Overall, breast density fell over the course of the study for both anastrozole and placebo, but changes were not significantly different between treatment arms.

Table 5.6: Association between treatment arm and change in density (continuous (%), <10%/≥10% reduction and <5%/≥5% reduction) in adjusted linear and logistic regression models.

Treatment	Continuous[#]							
First follow-up								
	Age-adjusted (n=566)		**Adjusted[1] (n=559)**		**Adjusted[2] (n=566)**		**Adjusted[3] (n=559)**	
	β-coefficient (95% CI)	**P-value**	**β-coefficient (95% CI)**	**P-value**	**β-coefficient (95% CI)**	**P-value**	**β-coefficient (95% CI)**	**P-value**
Placebo	Ref	-	Ref	-	Ref	-	Ref	-
Anastrozole	-0.17 (-0.59,0.26)	0.43	-0.27 (-0.69,0.13)	0.19	-0.18 (-0.60,0.26)	0.41	-0.29 (-0.70,0.13)	0.18
First follow-up (women with a final mammogram)								
	Age-adjusted (n=150)		**Adjusted[1] (n=148)**		**Adjusted[2] (n=150)**		**Adjusted[3] (n=148)**	
	β-coefficient (95% CI)	**P-value**	**β-coefficient (95% CI)**	**P-value**	**β-coefficient (95% CI)**	**P-value**	**β-coefficient (95% CI)**	**P-value**
Placebo	Ref	-	Ref	-	Ref	-	Ref	-
Anastrozole	0.01 (-0.77,0.81)	0.98	-0.07 (-0.87,0.72)	0.86	0.01 (-0.78,0.81)	0.98	-0.08 (-0.87,0.73)	0.84
Final follow-up								
	Age-adjusted (n=150)		**Adjusted[1] (n=148)**		**Adjusted[2] (n=150)**		**Adjusted[3] (n=148)**	
	β-coefficient (95% CI)	**P-value**	**β-coefficient (95% CI)**	**P-value**	**β-coefficient (95% CI)**	**P-value**	**β-coefficient (95% CI)**	**P-value**
Placebo	Ref	-	Ref	-	Ref	-	Ref	-
Anastrozole	0.39 (-0.71,1.46)	0.48	0.27 (-0.81,1.30)	0.61	0.39 (-0.70,1.46)	0.48	0.26 (-0.84,1.31)	0.63

Treatment	≥10% reduction*							
First follow-up								
	Age-adjusted (n=566)		**Adjusted[1] (n=559)**		**Adjusted[2] (n=566)**		**Adjusted[3] (n=559)**	
	OR (95% CI)	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**
Placebo	Ref	-	Ref	-	Ref	-	Ref	-
Anastrozole	0.57 (0.18,1.77)	0.33	0.82 (0.25,2.73)	0.75	0.59 (0.19,1.83)	0.36	0.84 (0.25,2.79)	0.77
First follow-up (women with a final mammogram)								
	Age-adjusted (n=150)		**Adjusted[1] (n=148)**		**Adjusted[2] (n=150)**		**Adjusted[3] (n=148)**	
	OR (95% CI)	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**
Placebo	Ref	-	Ref	-	Ref	-	Ref	-
Anastrozole	0.27 (0.03,2.44)	0.24	0.37 (0.04,3.83)	0.40	0.26 (0.03,2.43)	0.24	0.36 (0.03,3.77)	0.40
Final follow-up								
	Age-adjusted (n=150)		**Adjusted[1] (n=148)**		**Adjusted[2] (n=150)**		**Adjusted[3] (n=148)**	
	OR (95% CI)	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**
Placebo	Ref	-	Ref	-	Ref	-	Ref	-
Anastrozole	0.83 (0.18,3.91)	0.82	1.06 (0.20,5.62)	0.95	0.82 (0.17,3.92)	0.81	1.04 (0.19,5.61)	0.96

Table 5.6 continued

Treatment	≥5% reduction**							
First follow-up								
	Age-adjusted (n=566)		**Adjusted[1] (n=559)**		**Adjusted[2] (n=566)**		**Adjusted[3] (n=559)**	
	OR (95% CI)	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**
Placebo	Ref	-	Ref	-	Ref	-	Ref	-
Anastrozole	1.33 (0.85,2.09)	0.21	1.51 (0.95,2.40)	0.08	1.33 (0.85,2.09)	0.21	1.51 (0.95,2.41)	0.08
First follow-up (women with a final mammogram)								
	Age-adjusted (n=150)		**Adjusted[1] (n=148)**		**Adjusted[2] (n=150)**		**Adjusted[3] (n=148)**	
	OR (95% CI)	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**
Placebo	Ref	-	Ref	-	Ref	-	Ref	-
Anastrozole	1.40 (0.58,3.38)	0.45	1.76 (0.68,4.51)	0.24	1.40 (0.58,3.37)	0.45	1.75 (0.68,4.49)	0.25
Final follow-up								
	Age-adjusted (n=150)		**Adjusted[1] (n=148)**		**Adjusted[2] (n=150)**		**Adjusted[3] (n=148)**	
	OR (95% CI)	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**
Placebo	Ref	-	Ref	-	Ref	-	Ref	-
Anastrozole	0.75 (0.38,1.51)	0.42	0.87 (0.41,1.82)	0.71	0.75 (0.37,1.51)	0.42	0.87 (0.41,1.82)	0.71

*Treatment adjusted for age at randomisation (yr) in age-adjusted models; [1]treatment adjusted for age at randomisation (yr), body mass index at randomisation (kg/m^2), hormone replacement therapy use up to 12 months before randomisation (no/yes), age at menopause (yr), image type (film/digital) and time between baseline and follow-up mammogram (yr); [2]treatment adjusted for age at randomisation (yr) and baseline density (%); [3]treatment adjusted for age at randomisation (yr), body mass index at randomisation (kg/m^2), hormone replacement therapy use up to 12 months before randomisation (no/yes), age at menopause (yr), image type (film/digital), time between baseline and follow-up mammogram (yr) and baseline density (%); # β-coefficient represents difference in density change from placebo, empirical bootstrap 95% CI, P-value from z-test with known sample mean and standard deviation (the population is to the sample as the sample is to the bootstrap sample); * odds ratio (OR) represents odds of ≥10% density reduction relative to placebo, 95% confidence interval (95% CI) and P-value from a Wald test; ** OR represents odds of ≥5% density reduction relative to placebo, 95% CI and P-value from a Wald test.*

Table 5.7: Change in density (continuous, %) and odds ratios of relative risk of high density reduction (≥10% and ≥5%) by subgroups of covariates in the anastrozole arm only.

Variable	First follow-up							First follow-up (women with a final mammogram)							Final follow-up						
	n	Continuous		≥10% reduction[#]		≥5% reduction[##]		n	Continuous		≥10% reduction[#]		≥5% reduction[##]		n	Continuous		≥10% reduction[#]		≥5% reduction[##]	
		Mean (SD)	Median (IQR)	OR (95% CI)	P	OR (95% CI)	P		Mean (SD)	Median (IQR)	OR (95% CI)	P	OR (95% CI)	P		Mean (SD)	Median (IQR)	OR (95% CI)	P	OR (95% CI)	P
Age at randomisation (yr)*																					
<59 yr	145	-1.07 (2.84)	0 (0,0)	Ref	-	Ref	-	33	-1.21 (2.18)	0 (0,0)	Ref	-	Ref	-	33	-2.12 (3.54)	0 (-5,0)	Ref	-	Ref	-
≥59 yr	150	-0.93 (2.20)	0 (0,0)	1.46 (0.24,8.86)	0.68	0.77 (0.43,1.38)	0.38	39	-0.90 (2.26)	0 (0,0)	n/a	n/a	0.57 (0.17,1.85)	0.35	39	-1.41 (2.80)	0 (-5,0)	0.41 (0.04,4.71)	0.47	0.79 (0.29,2.15)	0.64
P**		0.65	0.56						0.55	0.38						0.35	0.49				
Body Mass Index (kg/m^2)*																					
<26.5 kg/m^2	146	-1.03 (2.19)	0 (0,0)	Ref	-	Ref	-	38	-0.92 (1.96)	0 (0,0)	Ref	-	Ref	-	38	-1.71 (3.14)	0 (-5,0)	Ref	-	Ref	-
≥26.5 kg/m^2	148	-0.98 (2.84)	0 (0,0)	1.49 (0.25,9.05)	0.67	0.94 (0.52,1.69)	0.84	34	-1.18 (2.48)	0 (0,0)	n/a	n/a	1.15 (0.36,3.69)	0.82	34	-1.76 (3.23)	0 (-5,0)	2.31 (0.20,26.71)	0.50	1.17 (0.43,3.20)	0.75
P**		0.87	0.69						0.63	0.77						0.94	0.86				
Age at menarche (yr)*																					
<13 yr	126	-1.19 (2.93)	0 (0,0)	Ref	-	Ref	-	33	-0.91 (1.96)	0 (0,0)	Ref	-	Ref	-	33	-1.36 (2.87)	0 (-5,0)	Ref	-	Ref	-
≥13 yr	167	-0.87 (2.19)	0 (0,0)	1.13 (0.19,6.89)	0.89	0.68 (0.37,1.22)	0.19	38	-1.18 (2.45)	0 (0,0)	n/a	n/a	1.20 (0.37,3.90)	0.76	38	-2.11 (3.42)	0 (-5,0)	1.78 (0.15,20.54)	0.65	1.39 (0.50,3.84)	0.53
P**		0.28	0.22						0.61	0.73						0.33	0.41				
Age at menopause (yr)*																					
<50 yr	138	-1.05 (2.86)	0 (0,0)	Ref	-	Ref	-	39	-1.28 (2.49)	0 (0,0)	Ref	-	Ref	-	39	-1.92 (3.74)	0 (-5,0)	Ref	-	Ref	-
≥50 yr	156	-0.96 (2.21)	0 (0,0)	1.33 (0.22,8.10)	0.76	0.82 (0.46,1.48)	0.51	32	-0.78 (1.84)	0 (0,0)	n/a	n/a	0.62 (0.18,2.07)	0.44	32	-1.56 (2.35)	0 (-5,0)	n/a	n/a	1.02 (0.37,2.81)	0.97
P**		0.76	0.72						0.35	0.41						0.64	0.99				

Table 5.7 continued

Tyrer-Cuzick 10-year risk (%)*																				
<7.8%	154	-0.94 (2.61)	0 (0,0)	Ref	-	Ref	-	43	-0.93 (2.25)	0 (0,0)	Ref	-	Ref	-	43	-1.86 (3.62)	0 (-5,0)	Ref	-	Ref	-
≥7.8%	141	-1.13 (2.70)	0 (0,0)	1.09 (0.22,5.51)	0.91	1.07 (0.59,1.92)	0.83	28	-1.07 (2.09)	0 (0,0)	n/a	n/a	1.40 (0.42,4.72)	0.58	28	-1.43 (2.30)	0 (-5,0)	n/a	n/a	0.92 (0.32,2.63)	0.88
P**		0.53	0.66						0.79	0.62						0.58	0.86				
Baseline density (%)*																				
<45%	151	-0.86 (1.98)	0 (0,0)	Ref	-	Ref	-	35	-0.86 (1.91)	0 (0,0)	Ref	-	Ref	-	35	-1.57 (2.36)	0 (-5,0)	Ref	-	Ref	-
≥45%	148	-1.25 (3.19)	0 (0,0)	n/a	n/a	1.17 (0.66,2.08)	0.60	37	-1.22 (2.47)	0 (0,0)	n/a	n/a	1.33 (0.41,4.33)	0.63	37	-1.89 (3.79)	0 (-5,0)	n/a	n/a	0.92 (0.34,2.52)	0.88
P**		0.21	0.51						0.50	0.60						0.67	0.93				
Age at first birth (yr)																				
Nulliparous	44	-0.91 (2.23)	0 (0,0)	Ref	-	Ref	-	9	-1.11 (3.33)	0 (0,0)	Ref	-	Ref	-	9	-1.11 (3.33)	0 (0,0)	Ref	-	Ref	-
>27	64	-0.86 (2.10)	0 (0,0)	n/a	n/a	1.22 (0.44,3.39)	0.70	18	-1.39 (2.30)	0 (-5,0)	n/a	n/a	3.08 (0.30,31.33)	0.34	18	-2.22 (2.56)	0 (-5,0)	n/a	n/a	6.40 (0.66,62.40)	0.11
21-27	135	-1.19 (2.81)	0 (0,0)	1.31 (0.14,12.07)	0.81	1.26 (0.51,3.15)	0.62	35	-0.86 (1.91)	0 (0,0)	n/a	n/a	1.66 (0.17,15.82)	0.66	35	-1.43 (3.55)	0 (0,0)	0.48 (0.04,6.04)	0.57	2.37 (0.26,21.90)	0.45
≤20	52	-0.77 (2.50)	0 (0,0)	n/a	n/a	1.26 (0.44,3.64)	0.67	10	-1.00 (2.11)	0 (0,0)	n/a	n/a	2.00 (0.15,26.73)	0.60	10	-2.50 (2.64)	-2.5 (-5,0)	n/a	n/a	8.00 (0.71,90.00)	0.09
P**		0.71	0.76						0.88	0.93						0.65	0.52				
Oral contraception use																				
Never	55	-1.27 (2.59)	0 (0,0)	Ref	-	Ref	-	10	-1.00 (2.11)	0 (0,0)	Ref	-	Ref	-	10	0.00 (2.36)	0 (0,0)	Ref	-	Ref	-
Previously	236	-0.95 (2.54)	0 (0,0)	0.34 (0.06,2.09)	0.25	0.80 (0.39,1.64)	0.54	62	-1.05 (2.24)	0 (0,0)	-0.05 (-1.56,1.46)	0.95	n/a	n/a	62	-2.02 (3.20)	0 (-5,0)	n/a	n/a	4.61 (0.55,38.86)	0.16
Currently	4	0.00 (0.00)	0 (0,0)	n/a	n/a	n/a	n/a	0	-	-	-	-	-	-	0	-	-	-	-	-	-
P**		0.51	0.30						0.95	0.98						0.06	0.06				

Table 5.7 continued

HRT use up to 12 months before randomisation																				
No	281	-0.98 (2.54)	0 (0,0)	Ref	-	Ref	-	67	-0.97 (2.17)	0 (0,0)	Ref	-	Ref	-	67	-1.64 (3.18)	0 (-5,0)	Ref	-	Ref	-
Yes	14	-1.43 (2.34)	0 (-5,0)	n/a	n/a	1.80 (0.54,5.98)	0.34	5	-2.00 (2.74)	0 (-5,0)	n/a	n/a	3.06 (0.46,20.33)	0.25	5	-3.00 (2.74)	0 (-5,0)	n/a	n/a	3.79 (0.59,24.50)	0.16
P**		0.52	0.34						0.32	0.25						0.36	0.18				
Smoking status																				
Never	161	-1.02 (2.80)	0 (0,0)	Ref	-	Ref	-	34	-1.18 (2.48)	0 (0,0)	Ref	-	Ref	-	34	-1.47 (2.89)	0 (-5,0)	Ref	-	Ref	-
Former	41	-0.61 (1.66)	0 (0,0)	n/a	n/a	0.58 (0.21,1.61)	0.30	14	-1.07 (2.13)	0 (0,0)	n/a	n/a	1.05 (0.23,4.83)	0.95	14	-2.14 (3.23)	0 (-5,0)	2.54 (0.15,43.67)	0.52	1.33 (0.36,4.99)	0.67
Current	92	-1.14 (2.36)	0 (0,0)	1.17 (0.19,7.14)	0.87	1.09 (0.58,2.07)	0.79	24	-0.83 (1.90)	0 (0,0)	n/a	n/a	0.77 (0.20,3.00)	0.71	24	-1.88 (3.55)	0 (-5,0)	1.43 (0.09,24.13)	0.80	0.99 (0.31,3.12)	0.98
P**		0.53	0.73						0.85	0.69						0.78	0.82				
History of Atypical Hyperplasia or LCIS																				
No	270	-0.94 (2.50)	0 (0,0)	Ref	-	Ref	-	66	-0.98 (2.19)	0 (0,0)	Ref	-	Ref	-	66	-1.74 (3.22)	0 (-5,0)	Ref	-	Ref	-
Yes	25	-1.60 (2.78)	0 (-5,0)	2.77 (0.30,25.79)	0.37	1.80 (0.71,4.55)	0.22	6	-1.67 (2.58)	0 (-5,0)	n/a	n/a	2.25 (0.37,13.73)	0.38	6	-1.67 (2.58)	0 (-5,0)	n/a	n/a	1.15 (0.19,6.80)	0.88
P**		0.22	0.18						0.47	0.39						0.96	0.90				
Image type																				
Film	107	-0.89 (2.82)	0 (0,0)	Ref	-	Ref	-	19	-1.05 (2.09)	0 (0,0)	Ref	-	Ref	-	19	-1.58 (4.10)	0 (-5,0)	Ref	-	Ref	-
Digital	192	-1.15 (2.56)	0 (0,0)	2.83 (0.33,24.58)	0.35	1.26 (0.68,2.33)	0.46	53	-1.04 (2.27)	0 (0,0)	n/a	n/a	0.87 (0.24,3.20)	0.84	53	-1.79 (2.79)	0 (-5,0)	0.71 (0.06,8.26)	0.78	1.32 (0.41,4.27)	0.64
P**		0.42	0.40						0.98	0.87						0.80	0.51				

Table 5.7 continued

Time between baseline and first follow-up mammogram (yr)*																					
<2.1 years	151	-1.03 (2.73)	0 (0,0)	Ref	-	Ref	-	29	-0.86 (2.34)	0 (0,0)	Ref	-	Ref	-	-	-	-	-	-	-	-
≥2.1 years	148	-1.08 (2.58)	0 (0,0)	0.50 (0.09,2.79)	0.43	0.98 (0.55,1.75)	0.95	43	-1.16 (2.14)	0 (0,0)	n/a	n/a	1.89 (0.53,6.75)	0.33	-	-	-	-	-	-	-
P**		0.86	0.90						0.58	0.37						-	-				
Time between baseline and final follow-up mammogram (yr)*																					
<4.6 years	-	-	-	-	-	-	-		-	-	-	-	-	-		40	-2.13 (3.38)	Ref	-	Ref	-
≥4.6 years	-	-	-	-	-	-	-		-	-	-	-	-	-		32	-1.25 (2.84)	0.61 (0.05,7.08)	0.70	0.62 (0.22,1.74)	0.36
P**		-	-						-	-						0.25	0.28				

**Continuous variables dichotomised by their median (Table 5.1: median time between baseline and first follow-up mammogram=2.1 years, median time between baseline and final follow-up mammogram=4.6 years); ** P-value from two-sample t-test (corresponding to mean column) or Wilcoxon rank sum test (corresponding to median column) for covariates with 2 subgroups, P-value from ANOVA F-test (corresponding to mean column) or Cuzick's trend test (corresponding to median column) for covariates with >2 ordered subgroups; # odds ratio (OR) represents odds of ≥10% density reduction relative to the reference category, adjusted for age at randomisation (yr) (except for age at randomisation (yr)), 95% confidence interval (95% CI) and P-value from a Wald test; ## OR represents odds of ≥5% density reduction relative to the reference category, adjusted for age at randomisation (yr) (except for age at randomisation (yr)), 95% CI and P-value from a Wald test; n/a represents no results since subgroups perfectly predicted dichotomous density change; interquartile range (IQR); standard deviation (SD).*

5.3.7 Adjusted change in mammographic density – first follow-up mammogram (primary analysis)

Results from the linear and logistic regression models examining the adjusted associations between treatment arm and change in breast density are given in Table 5.6. The more negative the coefficient for continuous density change, the greater the effect of anastrozole on decreasing breast density from baseline mammogram than placebo. The primary analysis at first follow-up mammogram found that anastrozole had a mean -0.17% (95% CI, -0.59% to 0.26%, p=0.43) decrease in density compared with placebo after adjustment for age at baseline. This changed to a mean -0.27% (95% CI, -0.69% to 0.13%, p=0.19) decrease in the fully adjusted linear model. Accounting for baseline density only slightly strengthened the effect in both adjusted models. Baseline density was only marginally (non-significantly) associated with density change (mean density change per 10% increase in baseline density: -0.06% (95% CI, -0.14 to 0.03), p=0.21). If anything, it was suggestive that women on anastrozole were less likely to see a reduction in density of at least 10% at first follow-up mammogram than women on placebo (OR for anastrozole relative to placebo=0.57, 95% CI, 0.18 to 1.77, p=0.33 (age-adjusted), OR for anastrozole relative to placebo=0.82, 95% CI, 0.25 to 2.73, p=0.75 (fully-adjusted)), although the number of women experiencing at least 10% density reduction was small. Secondary analyses also assessed density change when dichotomised by a 5% reduction. In the age-adjusted model, there was some suggestion that women on anastrozole were more likely to see a density reduction of at least 5% at first follow-up mammogram than placebo (OR for anastrozole relative to placebo=1.33, 95% CI, 0.85 to 2.09, p=0.21). This changed to an OR of 1.51 (95% CI, 0.95 to 2.40, p=0.08) for anastrozole relative to placebo when adjusted for other factors in the fully-adjusted model.

The study was underpowered to find a difference in density change at first follow-up mammogram between anastrozole and placebo. With 575 women (22 cases, 553 controls), the power to detect a difference in density change (continuous) from baseline to first follow-up mammogram between the two treatment arms at the 5% type-I error level was only 8%. The power to detect a difference in density change (≥10% reduction and <10% reduction) from baseline to first follow-up mammogram between the two treatment arms at the 5% type-I error level was 11% (6.2.1). The power to detect a difference in density change (≥5% reduction and <5% reduction) from baseline to first follow-up mammogram between the two treatment arms at the 5% type-I error level was 32% (6.2.1).

5.3.8 Adjusted change in mammographic density – final follow-up mammogram

At final follow-up mammogram, anastrozole had a mean 0.39% (95% CI, -0.71% to 1.46%, p=0.48) increase in density compared with placebo after adjustment for age at baseline, which changed to 0.27% (95% CI, -0.81% to 1.30%, p=0.61) in the fully adjusted linear model (Table 5.6). Accounting for baseline density, again, had little effect. There was no clear difference in the odds of at least a 10% density reduction at final follow-up mammogram between placebo and anastrozole (OR for anastrozole relative to placebo=0.83, 95% CI, 0.18 to 3.91, p=0.82 (age-adjusted), OR for anastrozole relative to placebo=1.06, 95% CI, 0.20 to 5.62, p=0.95 (fully-adjusted)). When dichotomised by a 5% reduction and adjusted for age at baseline, women on anastrozole had an odds ratio of experiencing a density reduction of at least 5% at final follow-up mammogram (relative to placebo) of 0.75 (95% CI, 0.38 to 1.51, p=0.42). When adjusted for other covariates, the odds ratio relative to placebo changed to 0.87 (95% CI, 0.41 to 1.82, p=0.71).

5.3.9 Missing covariate data

As a sensitivity analysis, the age-adjusted regression models (n=566) were run in the subgroup of women with non-missing data for all adjusting variables (i.e. the subgroup included in fully-adjusted multivariable regression models, n=559), to test whether adjusted results were robust to missing data. There was only a small amount of missing data for adjusting covariates and the results of these sensitivity models were similar to those in the main analysis (results not reported); hence the analysis was robust to missing data.

5.3.10 Subgroup analysis

Table 5.7 shows the effect of anastrozole on density change by different subgroups of covariates in the anastrozole arm only. Continuous covariates were dichotomised by their medians in all women since there were no differences by treatment arm (Table 5.1). There was no discernible difference in the effect of anastrozole on density change compared to placebo in these subgroups.

5.3.11 Potential impact of compliance

To test the impact of compliance, Kaplan–Meier curves and log rank tests were conducted in cases (censored 3 months before cancer diagnosis) and controls, to assess the difference in time to stopping treatment between: anastrozole cases with ≥5% vs. <5% reduction in density (no cases had ≥10% reduction in density), placebo cases with ≥5% vs. <5% reduction in density (no cases had ≥10% reduction in density), anastrozole controls with ≥5% vs. <5% and ≥10% vs.

<10% reduction in density, and placebo controls with ≥5% vs. <5% and ≥10% vs. <10% reduction in density. Better compliance may have been associated with a decrease in density since more treatment would have been administered. Figure 5.5 shows the results for anastrozole cases (log-rank p=0.15), Figure 5.6 shows the results for placebo cases (log-rank p=0.53), Figure 5.7 (log-rank p=0.13) and Figure 5.8 (log-rank p= 0.43) show the results for anastrozole controls with 10% and 5% cut-points, respectively, and Figure 5.9 (log-rank p= 0.75) and Figure 5.10 (log-rank p= 0.48) show the results for placebo controls with 10% and 5% cut-points, respectively. There did not appear to be a difference in compliance between the two treatment arms (by case-control status).

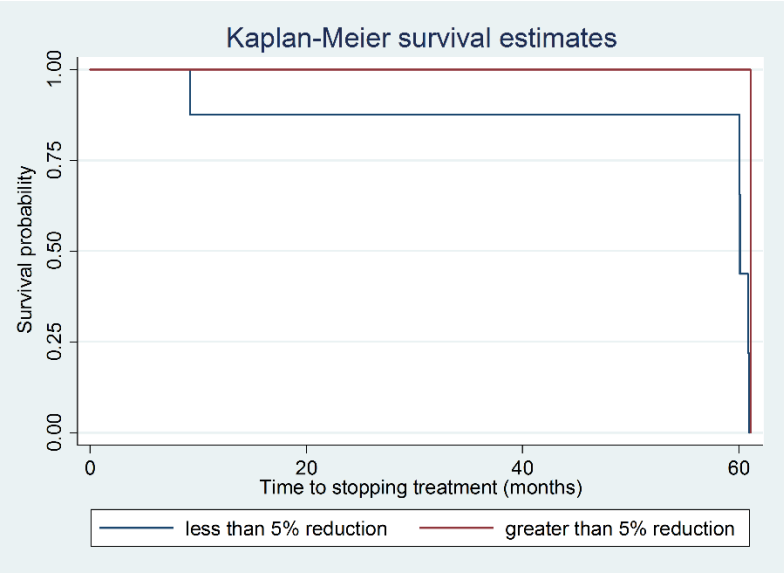


Figure 5.5: Kaplan-Meier graph for anastrozole cases (5% cut-point).

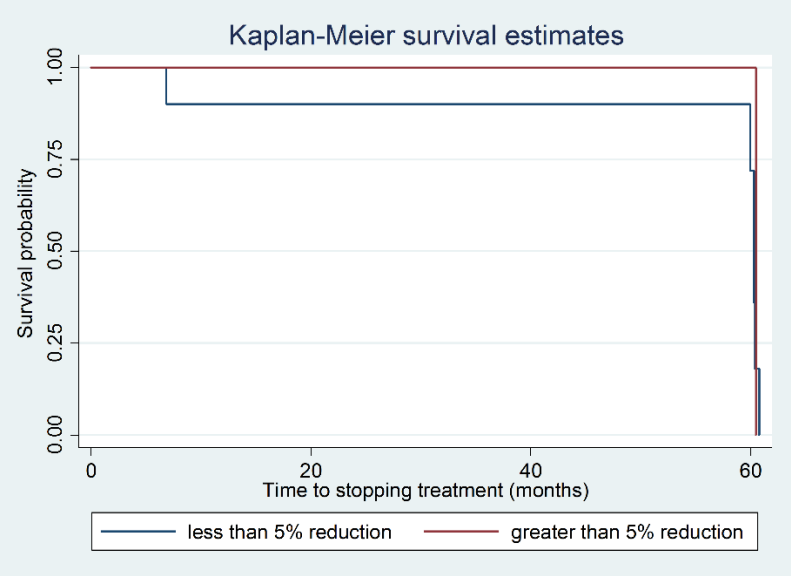


Figure 5.6: Kaplan-Meier graph for placebo cases (5% cut-point).

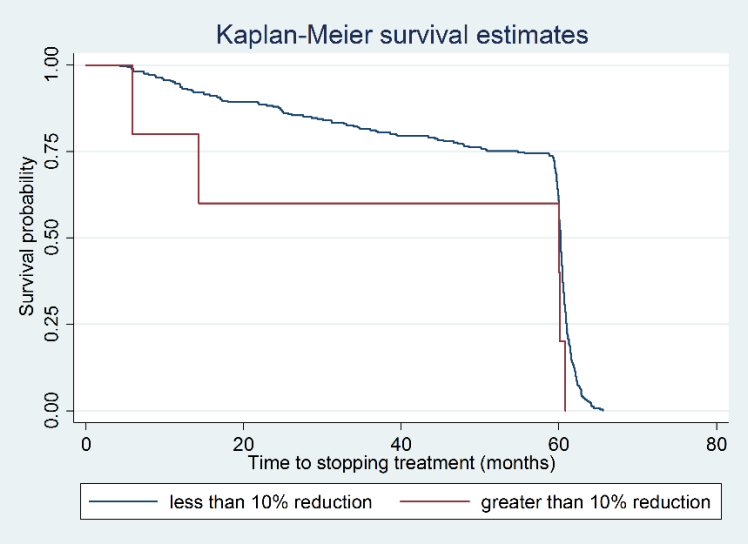


Figure 5.7: Kaplan-Meier graph for anastrozole controls (10% cut-point).

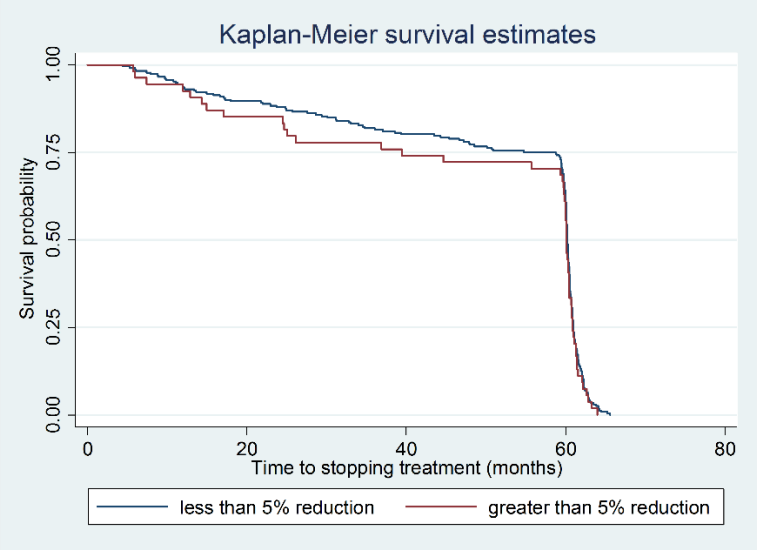


Figure 5.8: Kaplan-Meier graph for anastrozole controls (5% cut-point).

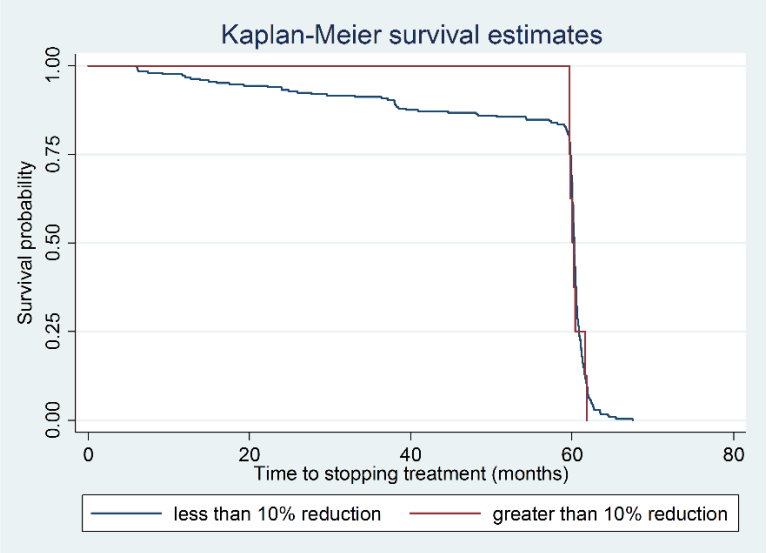


Figure 5.9: Kaplan-Meier graph for placebo controls (10% cut-point).

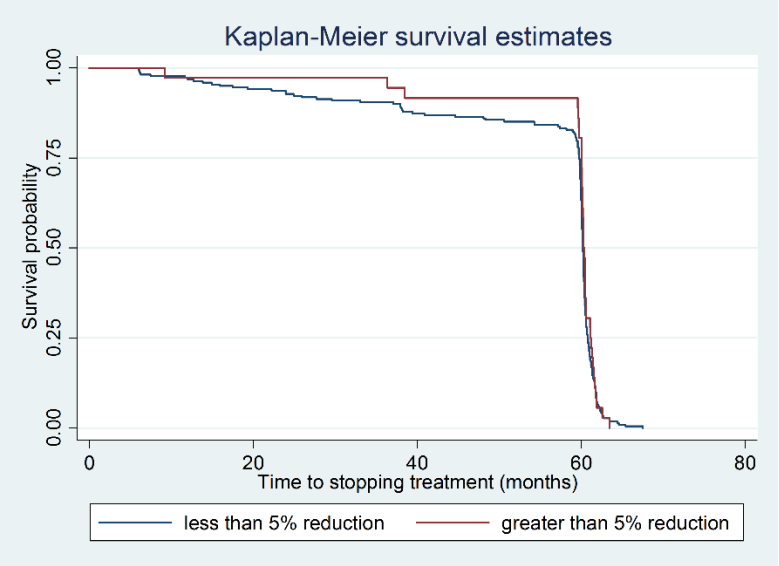


Figure 5.10: Kaplan-Meier graph for placebo controls (5% cut-point).

5.3.12 Missing mammograms

Since the sampling frame for this study is based on the availability of mammograms at the time of study design, the nested sample may not have truly represented the source population (IBIS-II main cohort). To test this, comparisons were made between cases that were included in this study and those from the main IBIS-II study who were not included, and between controls that were included in this study and those from the main IBIS-II study who were not included (Table 5.8). Included and non-included controls were similar in terms of age, Tyrer-Cuzick 10-year risk and HRT use, whilst included and non-included cases were similar in terms of age, BMI and HRT use. However, there was a significant difference in BMI between included and non-included controls, with the latter being somewhat heavier (mean non-included=28.4kg/m^2; mean included=27.0kg/m^2, $p<0.01$). When this was separated by country, the difference in BMI appeared to be driven by the UK centres. There was also a marginally significant difference in Tyrer-Cuzick 10-year risk between included and non-included cases, with the latter having a lower risk (mean non-included=9.2%; mean included=11.5%, $p=0.06$). There was also some indication that non-included controls had a slightly lower risk than included controls for the UK centres (mean non-included=8.2%; mean included=8.6%, $p=0.05$). The difference in HRT use between included and non-included controls in Ireland was driven by only 1 woman, and there did not appear to be a difference in any other country.

Table 5.8: Comparison of main confounding variables between women included in this study and women in the main IBIS-II study who were not included, overall and by country of the included women

Country of included controls	Controls													
	N		Age at randomisation (yr)			Body Mass Index (kg/m^2)			Tyrer-Cuzick 10-year risk (%)			HRT use up to 12 months before randomisation		
	Included	Not included	Statistic	Included	Not included	Statistic	Included	Not included	Statistic	Included	Not included	Statistic	Included	Not included
All	553	3165	Mean (SD)	58.9 (5.6)	58.7 (5.8)	Mean (SD)	27.0 (4.6)	28.4 (5.8)	Mean (SD)	8.5 (4.1)	8.3 (4.2)	no/yes	513/31	2870/237
			P*	0.38		P*	<0.01		P*	0.41		P**	0.11	
Australia and New Zealand	24	782	Mean (SD)	59.8 (4.7)	59.8 (5.7)	Mean (SD)	28.3 (5.0)	29.0 (5.7)	Mean (SD)	7.6 (5.0)	8.4 (3.4)	no/yes	19/2	678/73
			P*	0.99		P*	0.58		P*	0.29		P**	1.00#	
Denmark	5	47	Mean (SD)	54.4 (4.2)	55.7 (6.2)	Mean (SD)	26.3 (4.3)	25.3 (4.2)	Mean (SD)	10.0 (6.1)	8.5 (4.6)	no/yes	5/0	42/5
			P*	0.64		P*	0.62		P*	0.52		P**	1.00#	
Finland	24	100	Mean (SD)	61.9 (5.8)	60.3 (5.0)	Mean (SD)	28.1 (4.1)	27.9 (5.5)	Mean (SD)	7.0 (2.8)	7.7 (3.0)	no/yes	23/1	89/11
			P*	0.18		P*	0.88		P*	0.32		P**	0.46#	
Ireland	1	63	Mean (SD)	52 (-)	56.8 (5.6)	Mean (SD)	27 (-)	29.3 (5.7)	Mean (SD)	5.6 (-)	9.3 (4.2)	no/yes	0/1	60/1
			P*	-		P*	-		P*	-		P**	0.03#	
Italy	58	135	Mean (SD)	58.6 (6.1)	58.7 (5.5)	Mean (SD)	24.9 (4.0)	25.9 (4.7)	Mean (SD)	8.4 (4.1)	8.7 (5.7)	no/yes	55/3	129/5
			P*	0.93		P*	0.15		P*	0.66		P**	0.70#	
UK	441	1663	Mean (SD)	58.8 (5.5)	58.7 (5.7)	Mean (SD)	27.2 (4.6)	28.6 (5.9)	Mean (SD)	8.6 (4.0)	8.2 (3.9)	no/yes	411/24	1535/110
			P*	0.68		P*	<0.01		P*	0.05		P**	0.38	

Table 5.8 continued

Country of included cases	Cases													
	N		Age at randomisation (yr)			Body Mass Index (kg/m^2)			Tyrer-Cuzick 10-year risk (%)			HRT use up to 12 months before randomisation		
	Included	Not included	Statistic	Included	Not included	Statistic	Included	Not included	Statistic	Included	Not included	Statistic	Included	Not included
All	22	225	Mean (SD)	57.8 (4.5)	58.9 (5.6)	Mean (SD)	28.6 (5.1)	29.5 (6.7)	Mean (SD)	11.5 (5.5)	9.2 (5.4)	no/yes	21/1	206/18
			P*	0.36		P*	0.50		P*	0.06		P**	1.00#	
Australia and New Zealand	1	43	Mean (SD)	61.0 (-)	59.8 (5.7)	Mean (SD)	31.8 (-)	29.6 (6.5)	Mean (SD)	15.4 (-)	8.4 (3.1)	no/yes	1/0	38/5
			P*	-		P*	-		P*	-		P**	1.00#	
Denmark	1	3	Mean (SD)	59.0 (-)	54.3 (3.1)	Mean (SD)	26.8 (-)	28.5 (7.3)	Mean (SD)	12.7 (-)	13.9 (9.2)	no/yes	1/0	3/0
			P*	-		P*	-		P*	-		P**	-	
Finland	1	3	Mean (SD)	58.0 (-)	57.7 (7.4)	Mean (SD)	25.5 (-)	31.9 (1.9)	Mean (SD)	13.5 (-)	7.8 (4.5)	no/yes	1/0	3/0
			P*	-		P*	-		P*	-		P**	-	
Italy	1	17	Mean (SD)	49.0 (-)	60.7 (5.3)	Mean (SD)	22.9 (-)	27.8 (4.0)	Mean (SD)	20.1 (-)	11.3 (8.6)	no/yes	1/0	17/0
			P*	-		P*	-		P*	-		P**	-	
UK	18	132	Mean (SD)	58.0 (4.4)	58.9 (5.3)	Mean (SD)	29.0 (5.4)	30.2 (7.4)	Mean (SD)	10.7 (5.5)	9.4 (5.2)	no/yes	17/1	121/10
			P*	0.52		P*	0.49		P*	0.32		P**	1.00#	

**P-value from two-sided t-test; **P-value from Pearson chi-squared test (#Fisher's exact test if cell size <5); '-' indicates that value could not be calculated due to small numbers; standard deviation (SD).*

5.4 Discussion

In this study, breast density was shown to decrease with anastrozole in the first 2 years of therapy, but this effect was not significantly different from the density decreases seen in women on placebo. It might be that the early anastrozole-induced density reduction seen in this study was only attributable to aging; or perhaps it was a true but small effect, however there was limited power to detect it. Only 8% power was obtained to detect a continuous difference in density change between treatment arms, even with a reasonable number of women (n=575). This further suggests that the effect size of anastrozole-induced density reduction is very small.

The results for this study are consistent with previous findings that show modest reductions in density with use of aromatase inhibitors (compared with a control group). In 2007, Vachon et al. (2007) conducted a study of 9-15 month change in Cumulus percent density with letrozole in women with early-onset breast cancer who had previously undergone 5 years of tamoxifen treatment (270). They found no difference in density reduction relative to placebo (mean percent density reduction of 0.8% in letrozole and 0.6% in placebo (p=0.76)). In the preventive setting, Ciglar et al. (2010) showed similar results in their analysis of healthy postmenopausal women with or without a history of breast cancer, but with a baseline density greater than 25% from the NCIC CTG MAP.1 trial. After 12 months, Cumulus percent density was similar between women treated with letrozole or placebo (mean percent density reduction of 1.74% on letrozole and 0.24% on placebo (p=0.67)) (268). A similar trial by Ciglar et al. in 2011 (the NCIC CTG MAP.2 trial) reported a mean Cumulus percent density increase of 0.56% on exemestane and 0.58% on placebo (p=0.91) after 12 months of treatment (269). In both studies by Ciglar et al., density change was similar in both treatment arms even after 24 months of treatment. However, this potentially weak effect is not always seen. An aromatase inhibitor-induced density reduction was reported in a small cohort study by Mousa et al. who assessed density in 40 women on either HRT alone or HRT plus AIs (415). However, it should be noted that the chosen method of density assessment (integrated pixel intensity) is not a common or verified density measurement technique. It could be that this method of assessment is measuring a mammographic feature other than density (for example, a textural feature) so these results would require validation with an established density measurement technique.

One reason why these studies may not have seen an effect of AIs on density reduction is their lack of statistical power due to their small sample sizes. Density change after approximately 12 months of treatment was assessed in only 68 women in Vachon et al. (2007), 49 women in Ciglar et al. (2010) and 65 women in Ciglar et al. (2011). A large amount of starting density is required to see a substantial absolute density reduction in postmenopausal women. Whilst studies such as Ciglar et al. (2010) included women with at least 25% starting density and still

found little effect, women were also eligible to enter the trial if they had taken tamoxifen up to 3-months prior to recruitment, which has been shown to have anti-estrogenic effects for more than 10 years after treatment cessation (200). This may also extend to a prolonged effect on density, which would be particularly relevant if more placebo subjects were previously on tamoxifen, hence attenuating the effect of the AI. Most other studies looking into the effect of aromatase inhibitors on density do not have a placebo or 'no treatment' control group to assess its effect beyond that of aging (296).

A recent study by Engmann et al. did find a difference in density change between women on AIs and women not on endocrine treatment. They showed that 403 breast cancer cases on AIs experienced, on average, 0.3% greater 2-3 year reduction in Volpara percent density and 0.6% greater 2-3 year reduction in Quantra percent density than 1,618 breast cancer-free controls. These results were also statistically significant (Volpara p=0.02, Quantra p=0.03) (271). This is one of only a few studies to assess the effect of aromatase inhibitors on density in a considerably large sample of women. Another study by Vachon et al. in 2013, investigated the difference in Cumulus percent density 10 month change between 369 early-stage postmenopausal breast cancer cases on adjuvant anastrozole and 369 matched breast cancer-free controls. Unlike Engmann et al., they found modest and non-significant results, with the cases experiencing a median density reduction that was only 0.1% lower than their matched controls (p=0.51) (267). In a more recent study in 355 postmenopausal breast cancer patients, Eriksson et al. again showed little association between AIs and density change (OR for a density reduction greater than 15% in women treated with an AI relative to those not treated with any endocrine therapy=0.91 (95% CI, 0.65 to 1.26)) (337).

One explanation for these dissimilar findings could be the difference in measurement techniques. Vachon et al. measured 2-dimensional density using Cumulus and Eriksson measured 2-dimensional density using the 'STRATUS' tool which aligns mammograms to reduce measurement error between sequential mammograms (416); whereas Engmann used 3-dimensional Volpara and Quantra. It may be that small changes in density with use of AIs are best measured using volumetric methods.

Several strengths of this study are listed below:

- This is the largest known study to date to assess the effect of an aromatase inhibitor on density in the preventive setting.
- Inclusion of a placebo control group enabled a comparison between reductions in density whilst on treatment and reductions that would occur naturally with age.
- Including a first and final follow-up mammogram allowed for assessment of density changes throughout the course of anastrozole treatment.

- The study is nested within a double-blind placebo-controlled randomised trial which is subject to minimal bias.
- The multicentre aspect of the study enabled recruitment of women from different countries and ethnicities, increasing generalisability of results.
- Exclusion of women with bilateral breast cancer, implants or preventive mastectomies, and assessment of the contralateral breast in cases ensured density estimates were not affected by these confounding factors. Since women were not allowed to take HRT during the trial, potential confounding from HRT was also reduced.
- There was no switching of treatment throughout the trial, ensuring that the recorded effects were as a result of anastrozole treatment only.
- Questionnaire data was collected for all women, allowing for adjustment of age and other density change confounding variables such as BMI, reproductive factors, HRT use and a history of benign disease.
- There was a high inter-reader correlation between Dr Metaxa and the radiologist for IBIS-I (Dr Ruth Warren).
- Dr Metaxa was blinded to treatment, case status and other risk factors, reducing bias from these factors in the measurement of density and density change.

Several limitations of this study are listed below:

- There was limited power to detect a change in density between treatment arms. However, the sample size in this study was similar to or larger than other studies assessing density change on AIs (as outlined in the discussion).
- The multicentre aspect of the study resulted in different imaging modalities, imaging technologies and scanning techniques. This may have introduced variability in the density measurements, therefore making the signal of an anastrozole effect on density change harder to detect through added noise.
- The measurement of covariates (such as BMI) would have been more reliable if measured using accurate measuring devices instead of questionnaires as used in this study.
- Use of a volumetric measurement method such as Volpara or Quantra may have been better at detecting small changes in density than the 2-dimensional visual assessment method used in this study. However, since the study included both digital and film mammograms, not all images could be measured using these methods which require raw images from digital mammography.

5.5 Conclusion

Findings of this study indicate that visually-assessed percent breast density may be marginally reduced by prophylactic anastrozole treatment, but that this reduction is likely to be minimal. However, use of different density measurement techniques that measure density change to a finer grain of detail (for instance, volumetric methods) may be useful for assessing the effect of anastrozole on density change, hence further examination is required on this topic. Nonetheless, these results suggest that the risk reduction from anastrozole observed in IBIS-II is unlikely to be fully mediated through density, as was suggested with tamoxifen in IBIS-I.

Chapter 6: Anastrozole-induced reduction in mammographic density and breast cancer risk reduction: a case-control study

6.1 Introduction

As described in Chapters 4 and 5, the SERM, tamoxifen, decreases risk (410) as well as density (203-206, 336). This reduction in density might be a marker of concurrent reduction in breast cancer risk, making endocrine therapy-induced density reduction a potential biomarker for risk reduction. It is not yet known whether anastrozole-induced density reduction can similarly be used as a biomarker for risk reduction, whereby an early anastrozole-induced density reduction of at least 10% would be associated with a lower risk of breast cancer compared with <10% anastrozole-induced density reduction. There is some suggestion of this biomarker effect with adjuvant aromatase inhibitors (264), but no study has so far examined the effect with preventive anastrozole therapy. Validation of aromatase inhibitor-induced density reduction would be useful as a biomarker for risk reduction in postmenopausal women whose risk may be lowered more than that with tamoxifen (349), and who may experience fewer adverse effects and better tolerance of symptoms than if treated with tamoxifen (348, 411). This study aims to assess early anastrozole-induced change in density as a biomarker for breast cancer risk reduction in patients from the IBIS-II Prevention trial. A detailed description of the primary and secondary hypotheses is described below:

Primary hypothesis (Prognostic biomarker)

- H_0: There is no difference in age-adjusted risk of breast cancer between anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram and anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram
- H_1: Age-adjusted risk of breast cancer in anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram is different to age-adjusted risk of breast cancer in anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram

Secondary hypothesis I (Prognostic biomarker)

- H_0: There is no difference in age-adjusted risk of breast cancer between anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram and anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram

- H_1: Age-adjusted risk of breast cancer in anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram is different to age-adjusted risk of breast cancer in anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram

Secondary hypothesis II (Prognostic biomarker)

- H_0: There is no difference in risk of breast cancer between anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram and anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.
- H_1: Risk of breast cancer in anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram is different to risk of breast cancer in anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.

Secondary hypothesis III (Prognostic biomarker)

- H_0: There is no difference in risk of breast cancer between anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram and anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.
- H_1: Risk of breast cancer in anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram is different to risk of breast cancer in anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.

Secondary hypothesis IV (Predictive biomarker I)

- H_0: (1) There is no difference in age-adjusted risk of breast cancer between anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram and placebo-treated patients, (2) there is no difference in age-adjusted risk of breast cancer between anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram and placebo-treated patients.
- H_1: (1) Age-adjusted risk of breast cancer in anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram is different to age-adjusted risk of breast cancer in placebo-treated patients, (2) age-adjusted risk of breast cancer in

anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram is different to age-adjusted risk of breast cancer in placebo-treated patients.

Secondary hypothesis V (Predictive biomarker I)

- H_0: (1) There is no difference in age-adjusted risk of breast cancer between anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram and placebo-treated patients, (2) there is no difference in age-adjusted risk of breast cancer between anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram and placebo-treated patients
- H_1: (1) Age-adjusted risk of breast cancer in anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram is different to age-adjusted risk of breast cancer in placebo-treated patients, (2) age-adjusted risk of breast cancer in anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram is different to age-adjusted risk of breast cancer in placebo-treated patients.

Secondary hypothesis VI (Predictive biomarker I)

- H_0: (1) There is no difference in risk of breast cancer between anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram and placebo-treated patients and (2) there is no difference in risk of breast cancer between anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram and placebo-treated patients, both after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.
- H_1: (1) Risk of breast cancer in anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram is different to risk of breast cancer in placebo-treated patients and (2) risk of breast cancer in anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram is different to risk of breast cancer in placebo-treated patients, both after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.

Secondary hypothesis VII (Predictive biomarker I)

- H_0: (1) There is no difference in risk of breast cancer between anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram and placebo-treated patients and (2) there is no difference in risk of breast cancer between anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram and placebo-treated patients, both after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.
- H_1: (1) Risk of breast cancer in anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram is different to risk of breast cancer in

placebo-treated patients and (2) risk of breast cancer in anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram is different to risk of breast cancer in placebo-treated patients, both after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.

Secondary hypothesis VIII (Predictive biomarker II)

- H_0: There is no difference in the age-adjusted effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) between anastrozole-treated patients and placebo-treated patients (interaction between treatment and density change).
- H_1: The age-adjusted effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) in anastrozole-treated patients is different to the age-adjusted effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) in placebo-treated patients (interaction between treatment and density change).

Secondary hypothesis IX (Predictive biomarker II)

- H_0: There is no difference in the age-adjusted effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) between anastrozole-treated patients and placebo-treated patients (interaction between treatment and density change).
- H_1: The age-adjusted effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) in anastrozole-treated patients is different to the age-adjusted effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) in placebo-treated patients (interaction between treatment and density change).

Secondary hypothesis X (Predictive biomarker II)

- H_0: There is no difference in the effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) between anastrozole-treated patients and placebo-treated patients, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk (interaction between treatment and density change).
- H_1: The effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) in anastrozole-treated patients is different to the effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) in placebo-treated patients, after adjustment for age at randomisation,

body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk (interaction between treatment and density change).

Secondary hypothesis XI (Predictive biomarker II)

- H_0: There is no difference in the effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) between anastrozole-treated patients and placebo-treated patients, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk (interaction between treatment and density change).
- H_1: The effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) in anastrozole-treated patients is different to the effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) in placebo-treated patients, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk (interaction between treatment and density change).

Secondary hypothesis XII (Prognostic and Predictive biomarker I)

- H_0: There is no difference in age-adjusted breast cancer risk between subgroups of covariates (tumour ER status, age at randomisation, body mass index at randomisation, baseline density, history of atypical hyperplasia or LCIS, hormone replacement therapy use up to 12 months before randomisation, Tyrer-Cuzick 10-year risk, image type, and time between baseline mammogram and first follow-up mammogram) in women who experience an anastrozole-induced ≥10% reduction in density from baseline to first follow-up mammogram.
- H_1: Age-adjusted breast cancer risk in women who experience an anastrozole-induced ≥10% reduction in density from baseline to first follow-up mammogram is different between subgroups of covariates (tumour ER status, age at randomisation, body mass index at randomisation, baseline density, history of atypical hyperplasia or LCIS, hormone replacement therapy use up to 12 months before randomisation, Tyrer-Cuzick 10-year risk, image type, and time between baseline mammogram and first follow-up mammogram).

Secondary hypothesis XIII (Prognostic and Predictive biomarker I)

- H_0: There is no difference in age-adjusted breast cancer risk between subgroups of covariates (tumour ER status, age at randomisation, body mass index at randomisation, baseline density, history of atypical hyperplasia or LCIS, hormone replacement therapy use up to 12 months before randomisation, Tyrer-Cuzick 10-year risk, image type, and time between baseline mammogram and first follow-up mammogram) in women who experience

an anastrozole-induced ≥5% reduction in density from baseline to first follow-up mammogram.

- H_1: Age-adjusted breast cancer risk in women who experience an anastrozole-induced ≥5% reduction in density from baseline to first follow-up mammogram is different between subgroups of covariates (tumour ER status, age at randomisation, body mass index at randomisation, baseline density, history of atypical hyperplasia or LCIS, hormone replacement therapy use up to 12 months before randomisation, Tyrer-Cuzick 10-year risk, image type, and time between baseline mammogram and first follow-up mammogram).

6.2 Methods

The 'Study design', 'Data collection', 'Updating the Standard Operating Procedures', 'Exclusions', 'Exclusions - quality control (1)', 'Batching', 'Mammographic density scoring', 'Quality control (2)' and 'Baseline mammographic density at least 10%' sections are described in Chapter 5 (5.2).

6.2.1 Power calculation

The estimated distribution of IBIS-II cases and controls with <10% and ≥10% density reduction per treatment arm was calculated by weighting the distributions that were observed in IBIS-I (19) with hazard ratios relative to placebo from IBIS-I (HR=0.7) (200) and IBIS-II (HR=0.5) (208). The numbers in the table below relate to the distribution of the 123 cases in IBIS-I, whereby 57 cases on placebo lost less than 10% density, 15 cases on placebo lost at least 10% density, 36 cases on tamoxifen lost less than 10% density and 15 cases on tamoxifen lost at least 10% density. Since fewer women on anastrozole in IBIS-II developed breast cancer than women on tamoxifen in IBIS-I, some cases were removed from the anastrozole row. The parameters x_1 and x_2 were the number of cases needed to add to the placebo row (from the anastrozole row) to ensure that the 2x2 table equalled 123 (total number of cases in IBIS-I density study). The expected case distribution was:

	<10% reduction	≥10% reduction
Placebo	$57 + x_1$	$15 + x_2$
Anastrozole	$36\left(\frac{0.5}{0.7}\right)$	$15\left(\frac{0.5}{0.7}\right)$

There were therefore $36\left(\frac{0.2}{0.7}\right) + 15\left(\frac{0.2}{0.7}\right)$ fewer cases on anastrozole who needed to go into the placebo group. Thus, $x_1 = \frac{57}{(57+15)} \frac{(36+15)\ 0.2}{0.7}$ and $x_2 = \frac{15}{(57+15)} \frac{(36+15)\ 0.2}{0.7}$.

The expected case distribution for anastrozole was therefore estimated to be:

	<10% reduction	≥10% reduction	
Placebo	68	18	
Anastrozole	26	11	$\Sigma = 123$

For 3 controls per 1 case, the number of controls needed was 3x123=369. In IBIS-I, there were 361 controls on placebo who lost less than 10% density, 125 controls on placebo who lost at least 10% density, 239 controls on tamoxifen who lost less than 10% density and 217 controls on tamoxifen lost at least 10% density, totalling 942 controls. Reweighting this distribution to have a total of 369 gave an expected control distribution of:

	<10% reduction	≥10% reduction	
Placebo	$369\left(\frac{361}{942}\right) = 141$	$369\left(\frac{125}{942}\right) = 49$	
Anastrozole	$369\left(\frac{239}{942}\right) = 94$	$369\left(\frac{217}{942}\right) = 85$	$\Sigma = 369$

Therefore, the proportion of expected breast cancer events was:

	<10% reduction	≥10% reduction
Placebo	$\frac{68}{68+141} = 0.33$	$\frac{18}{18+49} = 0.27$
Anastrozole	$\frac{26}{26+94} = 0.22$	$\frac{11}{11+85} = 0.11$

These distributions were then weighted using chosen multipliers to obtain sample sizes for 50, 100, 150 and 200 cases, with 3 controls per 1 case. A difference of proportions power calculation was then estimated (superiority test) (417).

The power for different sample sizes (3 controls per 1 case) was therefore:

Sample size	200 (50 cases)	400 (100 cases)	600 (150 cases)	800 (200 cases)
Power	0.295	0.516	0.688	0.808

As a result, there was 81% power with 247 cases and 1013 controls to show a difference in risk between anastrozole-treated patients experiencing ≥10% density change and anastrozole-treated patients experiencing <10% density change from baseline to first follow-up mammogram at the 5% type-I error level. This number also accounted for exclusions with baseline density <10%

based on the number of postmenopausal women with baseline density <10% in IBIS-I. A sample size larger than this was impracticable given the resources, number of mammograms received and number of breast cancer cases in the entire IBIS-II trial (n=approximately 200 at the time of study design). Assuming anastrozole to be 1/2 or 3/4 the effect size of tamoxifen would have made the power smaller. There was therefore not enough power to complete the primary objective. However, Kim et al.'s suggestion of increased risk of recurrence in ER+ breast cancer cases on AIs who lost <5% density after 8-20 months of treatment relative to similarly treated women who lost ≥5% density (HR=7.11, 95% CI, 0.90 to 56.37, p=0.06) (264) indicated that this study possibly had adequate power. The number of recurrences on AIs was not reported in Kim et al. but it was estimated to be 13 from other numbers reported in the paper. Assuming 32% of the 35 cases in this study that were sent to Dr Metaxa were on anastrozole (40 anastrozole cases/125 cases in Cuzick et al. (208)) it was estimated that there would be approximately 11 anastrozole cases in this study. Therefore, there was potentially enough power to detect an effect if density change was dichotomised into <5% and ≥5% reduction (secondary objectives).

6.2.2 Statistical methods

As in Chapter 5, a statistical analysis plan was developed for the study (appendix C.XXV). All statistical analysis was conducted using Stata (316), and tests were two-sided with a significance level of 5%. Time on treatment was not included in adjustments (intention-to-treat analysis). A set of Stata code was sent to Dr Sestak to run on un-blinded data and a proforma was developed and sent to the IBIS-II Trial Steering committee, who approved the study (appendix C.XXVI).

6.2.2.1 Baseline characteristics

Baseline covariates were summarised using frequency tables (as described in 5.2.11.1) by treatment arm and case status.

6.2.2.2 Change in mammographic density

Density change was assessed as a dichotomous variable (<10% absolute reduction or ≥10% absolute reduction, and <5% absolute reduction or ≥5% absolute reduction), where cut-points were chosen to emulate previous studies reporting that dichotomous density change is a useful biomarker for breast cancer risk (Chapter 4).

6.2.2.3 Change in mammographic density – Boyd categories

A cross tabulation of Boyd categories was used as in Chapter 5, but with the primary interest in case-control status and density change between baseline and first follow-up mammogram only.

6.2.2.4 Change in mammographic density – unadjusted tests

Two-sample t-tests, Wilcoxon rank sum tests and Pearson chi-squared tests (Fisher's exact tests if cell size <5) were used as in 5.2.11.5, but with the primary interest in case-control status and density change between baseline and first follow-up mammogram only.

6.2.2.5 Change in mammographic density – adjusted regression models

Age at randomisation was retained in all regression models, regardless of significance, because age is a strong confounder of density change and breast cancer risk. Continuous adjusting variables were centred about their median in regression models. The adjusting covariates for regression models were chosen based on literature which suggests that they have a confounding effect on breast cancer risk and density change, including those shown to be significant in the IBIS-I trial (19).

The primary analysis used logistic regression models to examine the association between risk of breast cancer and change in density from baseline mammogram to first follow-up mammogram (dichotomised into <10% absolute reduction and ≥10% absolute reduction, reference category: <10% absolute reduction), adjusted for age at randomisation (years), in anastrozole-treated patients only. The secondary analysis (I) repeated the primary analysis for density change dichotomised into <5% absolute reduction and ≥5% absolute reduction (reference category: <5% absolute reduction). The secondary analysis (II) repeated the primary analysis with adjustment for age at randomisation (years), body mass index at randomisation (kg/m^2), baseline density (%) and Tyrer-Cuzick 10-year risk (%, version 7 excluding breast density). The secondary analysis (III) repeated the secondary analysis (II) for density change dichotomised into <5% absolute reduction and ≥5% absolute reduction (reference category: <5% absolute reduction).

The secondary analysis (IV) used logistic regression models to examine the association between risk of breast cancer and a factor variable for change in density and treatment (categories: <10% anastrozole-induced absolute reduction in density from baseline mammogram to first follow-up mammogram, ≥10% anastrozole-induced absolute reduction in density from baseline mammogram to first follow-up mammogram, and placebo, reference category: placebo) in all women, adjusted for age at randomisation (years). The secondary analysis (V) repeated the

secondary analysis (IV) for density change dichotomised into <5% absolute reduction and ≥5% absolute reduction. The secondary analysis (VI) repeated the secondary analysis (IV) with adjustment for age at randomisation (years), body mass index at randomisation (kg/m^2), baseline density (%) and Tyrer-Cuzick 10-year risk (%, version 7 excluding breast density). The secondary analysis (VII) repeated the secondary analysis (IV) for density change dichotomised into <5% absolute reduction and ≥5% absolute reduction, with adjustment for age at randomisation (years), body mass index at randomisation (kg/m^2), baseline density (%) and Tyrer-Cuzick 10-year risk (%, version 7 excluding breast density).

The secondary analysis (VIII) used logistic regression models to examine the association between risk of breast cancer and change in density from baseline mammogram to first follow-up mammogram (dichotomised into <10% absolute reduction and ≥10% absolute reduction, reference category: <10% absolute reduction), treatment arm (placebo or anastrozole, reference category: placebo), and an interaction between density change and treatment, in all women, adjusted for age at randomisation (years). The secondary analysis (IX) repeated the secondary analysis (VIII) for density change dichotomised into <5% absolute reduction and ≥5% absolute reduction (reference category: <5% absolute reduction). The secondary analysis (X) repeated the secondary analysis (VIII) with adjustment for age at randomisation (years), body mass index at randomisation (kg/m^2), baseline density (%) and Tyrer-Cuzick 10-year risk (%, version 7 excluding breast density). The secondary analysis (XI) repeated the secondary analysis (VIII) for density change dichotomised into <5% absolute reduction and ≥5% absolute reduction (reference category: <5% absolute reduction), with adjustment for age at randomisation (years), body mass index at randomisation (kg/m^2), baseline density (%) and Tyrer-Cuzick 10-year risk (%, version 7 excluding breast density).

6.2.2.6 Change in mammographic density – subgroup analyses

The secondary analysis (XII) used logistic regression to examine the association between risk of breast cancer and change in density from baseline mammogram to first follow-up mammogram (dichotomised into <10% absolute reduction and ≥10% absolute reduction, reference category: <10% absolute reduction), adjusted for age at randomisation (years), in anastrozole-treated patients only, in different covariate subgroups. Logistic regression models were also used to examine the association between risk of breast cancer and a factor variable for change in density and treatment (categories: <10% absolute anastrozole-induced reduction in density from baseline mammogram to first follow-up mammogram, ≥10% absolute anastrozole-induced reduction in density from baseline mammogram to first follow-up mammogram, and placebo, reference category: placebo) in all women, adjusted for age at randomisation (years), in different covariate subgroups. The secondary analysis (XIII) repeated the secondary analysis

(XII) for density change dichotomised into <5% absolute reduction and ≥5% absolute reduction. Covariate subgroups were: tumour ER status (negative/positive), age at randomisation (years), body mass index (BMI) at randomisation (kg/m^2), baseline density (%), history of atypical hyperplasia or LCIS (no/yes), hormone replacement therapy use up to 12 months before randomisation (no/yes), Tyrer-Cuzick 10-year risk (%, version 7 excluding breast density), image type (film/digital) and time between baseline mammogram and first follow-up mammogram (years). Continuous variables were separated into subgroups by their median value.

6.3 Results

6.3.1 Baseline characteristics

The distribution of baseline characteristics differed between cases and controls (Table 6.1). There was a significant difference between cases and controls in terms of Tyrer-Cuzick 10-year risk and history of atypical hyperplasia or LCIS, which was only apparent in the placebo arm and not the anastrozole arm.

Mean baseline density was 47% in cases and 43% in controls (mean difference cases minus controls=3.4%, 95% CI, -7.1% to 14.1%, two-sample t-test p=0.52); which was 47% in case subjects and 44% in control subjects of the placebo arm (mean difference cases minus controls=3.1%, 95% CI, -10.8% to 17.1%, two-sample t-test p=0.66), and 46% in case subjects and 43% in control subjects of the anastrozole arm (mean difference cases minus controls=3.5%, 95% CI, -13.1% to 20.0%, two-sample t-test p=0.68). The association between baseline density (continuous) and risk of developing breast cancer was not significant overall (OR=1.06 per 10% increase in baseline density (95% CI, 0.89 to 1.25), p=0.52). The suggested relative reduction in breast cancer risk associated with anastrozole in this study was 37% (OR relative to placebo=0.63 (95% CI, 0.26 to 1.49), p=0.29), which was consistent with the IBIS-II main study finding of 53% relative risk reduction (208).

Table 6.1: Baseline characteristics overall and by treatment, separated by case status

Variable	Overall						Placebo						Anastrozole					
	Cases			Controls			Cases			Controls			Cases			Controls		
	n	Mean (SD)	Median (IQR)	n	Mean (SD)	Median (IQR)	n	Mean (SD)	Median (IQR)	n	Mean (SD)	Median (IQR)	n	Mean (SD)	Median (IQR)	n	Mean (SD)	Median (IQR)
Age at randomisation (yr)	22	57.8 (4.5)	58 (54-61)	544	58.9 (5.6)	59 (55-63)	13	57.8 (4.6)	58 (56-61)	258	58.9 (5.7)	59 (55-63)	9	57.8 (4.5)	59 (54-60)	286	59.0 (5.4)	59 (54-63)
P*	0.33, 0.33						0.48, 0.40						0.52, 0.57					
Body Mass Index (kg/m^2)	22	28.6 (5.1)	29.3 (23.8-32.0)	540	27.0 (4.6)	26.4 (23.8-29.5)	13	28.1 (5.1)	27.8 (25.5-31.8)	255	26.9 (4.8)	26.4 (23.8-29.4)	9	29.1 (5.4)	30.1 (23.8-34.2)	285	27.1 (4.4)	26.4 (23.9-29.7)
P*	0.12, 0.11						0.38, 0.27						0.17, 0.23					
Age at menarche (yr)	22	12.8 (1.8)	13 (11-14)	540	12.9 (1.6)	13 (12-14)	13	12.4 (1.7)	12 (11-13)	256	12.9 (1.6)	13 (12-14)	9	13.3 (1.9)	13 (12-14)	284	12.8 (1.7)	13 (12-14)
P*	0.79, 0.63						0.244, 0.202						0.38, 0.48					
Age at menopause (yr)	22	49.2 (6.8)	50 (48-54)	541	48.3 (6.2)	50 (46-52)	13	49.8 (5.5)	50 (48-53)	255	48.3 (6.1)	50 (46-52)	9	48.3 (8.8)	50 (42-55)	285	48.4 (5.6)	50 (45-52)
P*	0.51, 0.31						0.39, 0.35						0.98, 0.67					
Tyrer-Cuzick 10-year risk (%)	22	11.5 (5.5)	10.2 (7.7-14.6)	546	8.5 (4.1)	7.7 (6.0-10.0)	13	12.1 (4.0)	12.7 (9.4-14.6)	260	8.5 (3.7)	7.8 (6.2-10.5)	9	10.8 (7.3)	7.7 (7.3-10.8)	286	8.5 (4.4)	7.5 (5.9-9.7)
P*	<0.01, <0.01						<0.01, <0.01						0.14, 0.36					
Baseline density (%)	22	46.8 (27.8)	50 (20-70)	553	43.4 (24.7)	45 (20-65)	13	47.3 (31.1)	30 (20-80)	263	44.2 (24.6)	45 (20-65)	9	46.1 (24.1)	55 (30-65)	290	42.6 (24.9)	40 (20-65)
P*	0.52, 0.58						0.66, 0.71						0.68, 0.70					

Table 6.1 continued

Variable	Overall						Placebo						Anastrozole					
	Cases			Controls			Cases			Controls			Cases			Controls		
	Total	n	%	Total	n	%	Total	n	%	Total	n	%	Total	n	%	Total	n	%
Age at first birth (yr)																		
Nulliparous	22	2	9.1	544	96	17.7	13	2	15.4	258	52	20.2	9	0	0.0	286	44	15.4
>27		7	31.8		114	21.0		4	30.8		53	20.5		3	33.3		61	21.3
21-27		11	50.0		234	43.0		6	46.2		104	40.3		5	55.6		130	45.5
≤20		2	9.1		100	18.4		1	7.7		49	19.0		1	11.1		51	17.8
P**	0.40#						0.70#						0.60#					
Oral contraception use																		
Never	22	4	18.2	543	116	21.4	13	3	23.1	257	62	24.1	9	1	11.1	286	54	18.9
Previously		18	81.8		423	77.9		10	76.9		195	75.9		8	88.9		228	79.7
Currently		0	0.0		4	0.7		0	0.0		0	0.0		0	0.0		4	1.4
P**	1.00#						1.00#						1.00#					
HRT use up to 12 months before randomisation																		
No	22	22	100.0	544	512	94.1	13	13	100.0	258	240	93.0	9	9	100.0	286	272	95.1
Yes		0	0.0		32	5.9		0	0.0		18	7.0		0	0.0		14	4.9
P**	0.63#						1.00#						1.00#					
Smoking status																		
Never	22	11	50.0	543	306	56.4	13	6	46.2	258	150	58.1	9	5	55.6	285	156	54.7
Former		2	9.1		71	13.1		2	15.4		30	11.6		0	0.0		41	14.4
Current		9	40.9		166	30.6		5	38.5		78	30.2		4	44.4		88	30.9
P**	0.64#						0.64#						0.51#					
History of Atypical Hyperplasia or LCIS																		
No	22	15	68.2	544	504	92.7	13	8	61.5	258	241	93.4	9	7	77.8	286	263	92.0
Yes		7	31.8		40	7.4		5	38.5		17	6.6		2	22.2		23	8.0
P**	<0.01						<0.01						0.17#					
Image type																		
Film	22	11	50.0	553	179	32.4	13	6	46.2	263	77	29.3	9	5	55.6	290	102	35.2
Digital		11	50.0		374	67.6		7	53.9		186	70.7		4	44.4		188	64.8
P**	0.09						0.20						0.29#					

**P-value from two-sample t-test (for means) and Wilcoxon rank sum test (for medians), respectively, for continuous variables by case status; **P-value from Pearson chi-squared test (#Fisher's exact test if cell size <5) for variable categories by case status; interquartile range (IQR), standard deviation (SD).*

Table 6.2: Cross tabulation of number of women in each Boyd category at entry to the study with category at first follow-up, by case status

	Number of women													
Boyd category at entry	**Boyd category at first follow-up: Cases**						**Total**	**Boyd category at first follow-up: Controls**						**Total**
	0%	**1-10%**	**11-25%**	**26-50%**	**51-75%**	**76-100%**		**0%**	**1-10%**	**11-25%**	**26-50%**	**51-75%**	**76-100%**	
0%	-	-	-	-	-	-	0(0/0)	-	-	-	-	-	-	0(0/0)
1-10%	-	2(0/2)	-	-	-	-	2(0/2)	-	47(22/25)	1(1/0)	-	-	-	48(23/25)
11-25%	-	1(1/0)	4(4/0)	-	-	-	5(5/0)	-	12(6/6)	143(61/82)	-	-	-	155(67/88)
26-50%	-	-	-	4(2/2)	-	-	4(2/2)	-	-	5(2/3)	130(67/63)	1(0/1)	-	136(69/67)
51-75%	-	-	-	-	7(2/5)	-	7(2/5)	-	-	-	6(1/5)	146(76/70)	-	152(77/75)
76-100%	-	-	-	-	-	4(4/0)	4(4/0)	-	-	-	-	8(4/4)	54(23/31)	62(27/35)
Total	0(0/0)	3(1/2)	4(4/0)	4(2/2)	7(2/5)	4(4/0)	22(13/9)	0(0/0)	59(28/31)	149(64/85)	136(68/68)	155(80/75)	54(23/31)	553(263/290)

The first number in each cell is the total number of subjects. Numbers in parentheses are the placebo and anastrozole groups, respectively; '-' indicates no entries.

Table 6.3: First follow-up change in density overall and by treatment, separated by case status

Change in breast density (%)	Overall		Placebo		Anastrozole		P-value*					
	Cases (n=22)	Controls (n=553)	Cases (n=13)	Controls (n=263)	Cases (n=9)	Controls (n=290)	Placebo vs. Anastrozole	Cases vs. Controls	Placebo Cases vs. Anastrozole Cases	Placebo Controls vs. Anastrozole Controls	Anastrozole Cases vs. Anastrozole Controls	Placebo Cases vs. Placebo Controls
Mean (SD)	-0.91 (1.97)	-0.94 (2.64)	-1.15 (2.19)	-0.80 (2.61)	-0.56 (1.67)	-1.07 (2.68)	0.28	0.96	0.50	0.23	0.57	0.63
Median (IQR)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0.13	0.83	0.49	0.09	0.56	0.37
<10% reduction: n (% of 2x2)	22 (3.8)	539 (93.7)	13 (4.7)	255 (92.4)	9 (3.0)	284 (95.0)						
≥10% reduction: n (% of 2x2)	0 (0.0)	14 (2.4)	0 (0.0)	8 (2.9)	0 (0.0)	6 (2.0)	0.49	1.00#	n/a	0.47	1.00#	1.00#
<5% reduction: n (% of 2x2)	18 (3.1)	460 (80.0)	10 (3.6)	226 (81.9)	8 (2.7)	234 (78.3)						
≥5% reduction: n (% of 2x2)	4 (0.7)	93 (16.2)	3 (1.1)	37 (13.4)	1 (0.3)	56 (18.7)	0.14	0.78#	0.62#	0.10	1.00#	0.41#
P trend**	0.96		0.63		0.56							

**P-value from two-sample t-test (corresponding to mean row) and Wilcoxon rank sum test (corresponding to median row) for continuous change in density, or Pearson chi-squared (#Fisher's exact test if cell size <5) for dichotomised density change (corresponding to ≥10% reduction row and ≥5% reduction row, as appropriate); **P-trend from a Wald test of change in density (continuous) from separate unadjusted logistic regression models of breast cancer risk on change in density (continuous), overall and in each treatment arm; n/a represents no results since no cases lost ≥10% density in placebo or anastrozole arm; interquartile range (IQR), standard deviation (SD).*

6.3.2 Change in mammographic density – Boyd categories

Change in density was shown in terms of the number of women in each treatment group by Boyd scale at baseline and first follow-up mammogram in cases and controls (Table 6.2). Movement between Boyd categories was minimal, and women who had a decrease in density moved by no more than one category below their baseline category. More controls moved to a lower Boyd category than cases, with only 1 (placebo) case moving downwards compared with 31 controls; however in terms of percentages, these numbers were similar between cases and controls (1/22=5% of cases, 31/553=6% of controls). Out of the controls moving to a lower Boyd category, a similar percentage were on anastrozole (18/290=6%) as placebo (13/263=5%).

6.3.3 Change in mammographic density – unadjusted tests

In Table 6.3, cases and controls lost similar amounts of density overall (controls: mean change (%)= -0.94, ≥10% reduction=2.4%, ≥5% reduction=16.2%; cases: mean change (%)= -0.91, ≥10% reduction=0.0%, ≥5% reduction=0.7%). This was similar in the anastrozole arm (controls: mean change (%)= -1.07, ≥5% reduction=18.7%; cases: mean change (%)= -0.56, ≥5% reduction=0.3%) and placebo arm (controls: mean change (%)= -0.80, ≥5% reduction=13.4%; cases: mean change (%)= -1.15, ≥5% reduction=1.1%). The proportions were smaller for the 10% density reduction cut-point (anastrozole controls: ≥10% reduction=2.0%; anastrozole cases: ≥10% reduction=0.0%; placebo controls: ≥10% reduction=2.9%; placebo cases: ≥10% reduction=0.0%).

6.3.4 Change in mammographic density – adjusted regression models

Table 6.4, Table 6.5 and Table 6.6 show the effect of density change in the anastrozole arm only (prognostic biomarker), and in all women as a predictive biomarker (compared with the placebo arm as a whole and as an interaction effect, respectively). There was no consistent association between continuous density reduction and breast cancer risk reduction. There were no case subjects who lost ≥10% density in the placebo or anastrozole arm, therefore density change dichotomised by 5% reduction was assessed. Density reduction of at least 5% on anastrozole (relative to <5% reduction) had an OR of 0.53 (95% CI, 0.06 to 4.31, p=0.55) in the age-adjusted model and 0.52 (95% CI, 0.06 to 4.26, p=0.54) in the fully-adjusted model (Table 6.4). Women in the anastrozole arm who experienced a 5% or greater reduction in breast density (relative to women in the placebo arm) had an OR of 0.36 (95% CI, 0.05 to 2.84, p=0.34) in the age-adjusted model and 0.34 (95% CI, 0.04 to 2.74, p=0.31) in the fully-adjusted model. Women who took anastrozole but experienced less than a 5% reduction in breast density (relative to women in the placebo arm) had an OR of 0.69 (95% CI, 0.28 to 1.69, p=0.42) in the

age-adjusted model and 0.68 (95% CI, 0.27 to 1.71, p=0.42) in the fully-adjusted model (Table 6.5). The interaction effect between density reduction (≥5% reduction) and treatment (anastrozole) had an OR of 0.32 (95% CI, 0.03 to 3.91, p=0.37) in the age-adjusted model and 0.23 (95% CI, 0.02 to 2.94, p=0.26) in the fully-adjusted model (Table 6.6). Only 3 women on placebo and 1 woman on anastrozole lost at least 5% density and went on to develop breast cancer. With 22 cases and 553 controls, the power to detect a difference in risk of breast cancer in anastrozole-treated patients who experienced a ≥5% reduction in density at first follow-up mammogram relative to anastrozole-treated patients who experienced a <5% reduction in density at first follow-up mammogram was only 7% at the 5% type-I error level.

Table 6.4: (Prognostic biomarker): Odds ratios for relative risk of breast cancer on first follow-up reduction in density in the anastrozole arm only, from adjusted logistic regression models

Variable	Age-adjusted (n=295)		Fully-adjusted# (n=292)	
	OR (95% CI)	P-value	OR (95% CI)	P-value
Density reduction				
Continuous reduction*[1]	0.92 (0.68,1.25)	0.59	0.92 (0.69,1.22)	0.55
<10% reduction[2]**	Ref	-	Ref	-
≥10% reduction[2]**	n/a	n/a	n/a	n/a
<5% reduction[3]**	Ref	-	Ref	-
≥5% reduction[3]**	0.53 (0.06,4.31)	0.55	0.52 (0.06,4.26)	0.54

*[1] Density reduction modelled as continuous reduction; [2] density reduction modelled as <10%/≥10% reduction; [3] density reduction modelled as <5%/≥5% reduction; density reduction adjusted for age at randomisation (yr) in age-adjusted models; #density reduction adjusted for age at randomisation (yr), body mass index at randomisation (kg/m^2), baseline density (%) and Tyrer-Cuzick 10-year risk (%); *odds ratio (OR) represents odds of breast cancer per unit decrease in continuous density; **OR represents odds of breast cancer relative to the reference category (<10% reduction or <5% reduction, as appropriate); 95% confidence intervals (95% CIs) and P-values from Wald tests; n/a represents no results since no cases lost ≥10% density in placebo or anastrozole arm.*

Table 6.5: (Predictive biomarker I): Odds ratios for relative risk of breast cancer on first follow-up reduction in density for the anastrozole arm (relative to the placebo arm) in all women, from adjusted logistic regression models

Variable	Age-adjusted (n=566)		Fully-adjusted# (n=560)	
	OR (95% CI)	P-value	OR (95% CI)	P-value
Density reduction				
Continuous reduction*[1]	1.01 (0.85,1.19)	0.92	1.01 (0.86,1.18)	0.93
<10% reduction[2]**	0.64 (0.27,1.52)	0.31	0.64 (0.26,1.55)	0.32
≥10% reduction[2]**	n/a	n/a	n/a	n/a
<5% reduction[3]**	0.69 (0.28,1.69)	0.42	0.68 (0.27,1.71)	0.42
≥5% reduction[3]**	0.36 (0.05,2.84)	0.34	0.34 (0.04,2.74)	0.31

*[1] Density reduction modelled as continuous reduction; [2] density reduction modelled as a factor variable with 3 categories: <10% reduction anastrozole, ≥10% reduction anastrozole and placebo; [3] density reduction modelled as a factor variable with 3 categories: <5% reduction anastrozole, ≥5% reduction anastrozole and placebo; density reduction adjusted for age at randomisation (yr) in age-adjusted models; # density reduction adjusted for age at randomisation (yr), body mass index at randomisation (kg/m^2), baseline density (%) and Tyrer-Cuzick 10-year risk (%); *odds ratio (OR) represents odds of breast cancer per unit decrease in continuous density; **OR represents odds of breast cancer relative to all women in the placebo arm; 95% confidence intervals (CIs) and P-values from Wald tests; n/a represents no results since no cases lost ≥10% density in placebo or anastrozole arm, therefore age-adjusted model ([2]) drops 5 anastrozole controls who lost ≥10% density and had non-missing age, hence n=561 and fully-adjusted model ([2]) drops 5 anastrozole controls who lost ≥10% density and had non-missing full covariates, hence n=555.*

Table 6.6: (Predictive biomarker II): Odds ratios for relative risk of breast cancer on first follow-up reduction in density, treatment, and an interaction between density reduction and treatment, in all women, from adjusted logistic regression models

Variable	Continuous[1]				<10%/≥10% reduction[2]				<5%/≥5% reduction[3]			
	Age-adjusted (n=566)		Fully-adjusted# (n=560)		Age-adjusted (n=566)		Fully-adjusted# (n=560)		Age-adjusted (n=566)		Fully-adjusted# (n=560)	
	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value
Density reduction												
Continuous reduction*[1]	1.03 (0.85,1.25)	0.73	1.08 (0.89,1.30)	0.44								
<10% reduction[2]**					Ref	-	Ref	-				
≥10% reduction[2]**					n/a	n/a	n/a	n/a				
<5% reduction[3]**									Ref	-	Ref	-
≥5% reduction[3]**									1.64 (0.42,6.34)	0.47	2.15 (0.54,8.58)	0.28
Treatment												
Placebo* **	Ref	-	Ref	-	Ref	-	Ref	-	Ref	-	Ref	-
Anastrozole* **	0.69 (0.28,1.72)	0.43	0.71 (0.28,1.80)	0.48	n/a	n/a	n/a	n/a	0.76 (0.29,1.95)	0.56	0.78 (0.30,2.05)	0.61
Interaction+	0.88 (0.62,1.28)	0.52	0.84 (0.59,1.19)	0.32	n/a	n/a	n/a	n/a	0.32 (0.03,3.91)	0.37	0.23 (0.02,2.94)	0.26

*[1] Density reduction modelled as continuous reduction; [2] density reduction modelled as <10%/≥10% reduction; [3] density reduction modelled as <5%/≥5% reduction; density reduction adjusted for age at randomisation (yr) in age-adjusted models; # density reduction adjusted for age at randomisation (yr), body mass index at randomisation (kg/m^2), baseline density (%) and Tyrer-Cuzick 10-year risk (%); *odds ratio (OR) represents odds of breast cancer per unit decrease in continuous density; **OR represents odds of breast cancer relative to the reference category (<10% reduction or <5% reduction, as appropriate); ***OR represents odds of breast cancer relative to placebo; + OR represents odds of an interaction effect between density reduction and treatment arm; 95% confidence intervals (CIs) and P-values from Wald tests; n/a represents no results since no cases lost ≥10% density in placebo or anastrozole arm.*

Table 6.7: Odds ratios for relative risk of breast cancer on anastrozole-induced first follow-up reduction in density (≥10% reduction and ≥5% reduction) by subgroups of covariates, from adjusted logistic regression models

Variable	Anastrozole, density reduction ≥10% (anastrozole arm only)		P	Anastrozole, density reduction ≥5% (anastrozole arm only)		P	Anastrozole, density reduction ≥10% (all women)		P	Anastrozole, density reduction ≥5% (all women)		P
	No. of anastrozole case subjects/total number in model	OR[#] (95% CI)		No. of anastrozole case subjects/total number in model	OR[##] (95% CI)		No. of anastrozole case subjects/total number in model	OR[+] (95% CI)		No. of anastrozole case subjects/total number in model	OR[++] (95% CI)	
Overall				1/295	0.53 (0.06,4.31)	0.55				1/566	0.36 (0.05,2.84)	0.34
Tumour ER status												
Negative				0/240	n/a	n/a				0/271	n/a	n/a
Positive				1/240	n/a	n/a				1/271	n/a	n/a
Age at randomisation (yr)*												
<59 yr				0/265	n/a	n/a				0/536	n/a	n/a
≥59 yr				1/265	1.73 (0.18,16.61)	0.64				1/536	1.13 (0.13,9.60)	0.91
Body Mass Index (kg/m^2)*												
<26.5 kg/m^2				0/267	n/a	n/a				0/537	n/a	n/a
≥26.5 kg/m^2				1/267	1.15 (0.14,9.56)	0.90				1/537	0.79 (0.10,6.27)	0.82
Baseline density (%)*												
<45%				0/269	n/a	n/a				0/540	n/a	n/a
≥45%				1/269	1.00 (0.12,8.31)	1.00				1/540	0.69 (0.09,5.49)	0.73
History of atypical hyperplasia or LCIS												
No				1/288	0.60 (0.07,4.95)	0.64				1/559	0.42 (0.05,3.27)	0.41
Yes				0/288	n/a	n/a				0/559	n/a	n/a

Table 6.7 continued

HRT use up to 12 months before randomisation								
No		1/291	0.57 (0.07, 4.66)	0.60		1/552	0.39 (0.05,3.08)	0.37
Yes		0/291	n/a	n/a		0/552	n/a	n/a
Tyrer-Cuzick 10-year risk (%)*								
<7.8%		1/267	1.07 (0.13,8.94)	0.95		1/537	0.74 (0.09,5.88)	0.77
≥7.8%		0/267	n/a	n/a		0/537	n/a	n/a
Image type								
Film		1/258	1.63 (0.19,13.87)	0.66		1/529	1.13 (0.14,9.17)	0.91
Digital		0/258	n/a	n/a		0/529	n/a	n/a
Time between baseline mammogram and first follow-up mammogram (yr)*								
<2.1 yr		1/268	1.07 (0.13,8.90)	0.95		1/539	0.74 (0.09,5.87)	0.77
≥2.1 yr		0/268	n/a	n/a		0/539	n/a	n/a

**Continuous variables dichotomised by their median (Table 5.1: overall column, median time between baseline and first follow-up mammogram=2.1 years); #odds ratio (OR) represents odds of breast cancer for women with ≥10% density reduction relative to women with<10% density reduction, in different subgroups, in the anastrozole arm only; ##OR represents odds of breast cancer for women with ≥5% density reduction relative to women with <5% density reduction, in different subgroups, in the anastrozole arm only; +OR represents odds of breast cancer for women with ≥10% density reduction relative to all women in the placebo arm, in different subgroups; ++OR represents odds of breast cancer for women with ≥5% density reduction relative to all women in the placebo arm, in different subgroups; density reduction adjusted for age at randomisation (yr); 95% confidence intervals (95% CIs) and P-values from Wald tests; no results for ≥10% reduction since no cases lost ≥10% density in placebo or anastrozole arm; n/a represents no results since subgroup numbers were small and perfectly predicted breast cancer.*

6.3.5 Subgroup analyses

Table 6.7 shows the estimated effect of anastrozole-induced density reduction as a prognostic biomarker (≥10% or ≥5% reduction relative to <10% or <5% reduction, respectively, in the anastrozole arm) and a predictive biomarker (≥10% or ≥5% anastrozole-induced density reduction relative to placebo in all women), in different subgroups of covariates. Continuous covariates were dichotomised by their medians in all women, because there were no differences by treatment arm (Table 5.1). There was too little data for the primary analysis, which was even smaller when conducting the subgroup analyses. Since there were no case subjects who lost ≥10% density in the placebo or anastrozole arm, only density change dichotomised by 5% reduction could be assessed. However, the number of women in each subgroup for the 5% cut-point was too small to obtain any useful results since there was only 1 anastrozole case who lost at least 5% density, thus the subgroup(s) that she did not belong to contained no anastrozole cases with ≥5% reduction and the odds ratios could not be calculated.

6.3.6 Impact of length of time between the baseline and first follow-up mammogram on breast cancer risk

The length of time between the baseline and first follow-up mammogram was not associated with breast cancer risk (OR per year=1.02 (95% CI, 0.96 to 1.07), p=0.56), and there was no statistically significant difference between cases and controls in relation to length of time between the baseline and first follow-up mammogram (mean difference cases minus controls = -1.0 months (95% CI, -4.3 months to 2.4 months), two-sample t-test p=0.56, Wilcoxon rank sum test p=0.52). In summary, length of time between baseline and first follow-up mammogram did not appear to be associated with risk of breast cancer.

6.3.7 Missing covariate data

As a sensitivity analysis, the age-adjusted regression models (n=295 and n=566) were run in the subgroup of women with non-missing data for all adjusting variables (i.e. the subgroup included in fully-adjusted multivariable regression models: n=292 and n=560, respectively), to test whether adjusted results were robust to missing data. There was only a small amount of missing data for adjusting covariates and the results of these sensitivity models were similar to those in the main analysis (results not reported); hence the analysis was robust to missing data.

6.4 Discussion

The sample size of this study was too small to determine whether anastrozole-induced density reduction is a prognostic and predictive biomarker of breast cancer risk reduction, therefore results could not be inferred. A larger study with more case subjects is needed to truly test this hypothesis. However, given there was little effect of prophylactic anastrozole on density change observed in Chapter 5, these data suggest that density change might not be an effective prognostic or predictive marker of breast cancer risk reduction with prophylactic anastrozole.

IBIS-I was the first trial to suggest that change in density might be a biomarker of the beneficial effect of endocrine therapy. Cuzick et al. found that women who had at least a 10% reduction in visually-assessed density in the first 12-18 months of prophylactic tamoxifen treatment had an approximate 63% reduction in breast cancer risk compared with women on placebo, whilst women who experienced <10% density reduction on tamoxifen had no difference in risk compared with women on placebo (19). It was suggested that women who experience the greatest density reductions after 12-18 months of tamoxifen may be responding to the drug and would perhaps benefit from continuing with their 5 year course of treatment (418). Women who see a more modest reduction or increase in density might not be responding to treatment and may benefit from alternative therapies such as exercise and dietary interventions to reduce weight, or chemoprevention with other SERMs or AIs.

Other studies have since tested Cuzick et al.'s results and provided evidence for density change to be used as a biomarker for tamoxifen treatment also in the adjuvant setting (Chapter 4). With respect to aromatase inhibitors, a study by Kim et al. found that the hazard ratio for risk of recurrence in AI-treated women who lost <5% density compared with women who lost ≥5% density after 8-20 months of treatment was 7.11 (95% CI, 0.90 to 56.37, p=0.06). However, as discussed in Chapter 4, there were some quality issues with this study, and one should bear this in mind when interpreting the results. It is therefore still unclear whether density reduction as a result of treatment with aromatase inhibitors can similarly be used as a biomarker for treatment efficacy.

As discussed in Chapter 4, low compliance (as a result of side effects of endocrine therapy) in the treatment arm might be a marker of treatment efficacy. A test of compliance between treatment arms by case-control status was completed in Chapter 5 (5.3.11). This suggested that the lower rate of breast cancers in women with larger density reductions was not due to better compliance because cases on either anastrozole or placebo had similar compliance regardless of density reduction. Nonetheless, the question still remains as to whether the joint association between compliance, side effects and density can be used as a biomarker for individual response

to treatment. In the Kaplan Meier graphs from Chapter 5 (5.3.11), anastrozole controls who lost at least 5% or 10% density were more likely to stop treatment earlier than women who lost less than 5% or 10% density, respectively (although log-rank tests were not significant). This effect was reversed in cases and placebo controls. This could potentially be a marker that treatment was working in these women, causing side effects and hence lower compliance, higher density reduction and breast cancer-free 'control' status. However, this hypothesis is purely speculative and requires validation and further assessment in another study.

There were several strengths of this study. Most of the strengths were outlined in Chapter 5, but some additional points relevant to this case-control study are listed below:

- This is the first known study to examine anastrozole-induced density reduction as a biomarker for breast cancer risk reduction in the preventive setting.
- The inclusion of a placebo arm enabled assessment of density reduction as a predictive biomarker.
- Exclusion of cases with a breast cancer diagnosis before or at first follow-up mammogram ensured that timings were appropriate and that the predictor occurred before the event.
- Measuring density change as both a continuous and dichotomous variable allowed for assessment of multiple density change predictors, which found that the 5% cut-point for dichotomised density change might be a better threshold than the 10% cut-point for smaller density reductions occurring with anastrozole.

There were several limitations of this study. Again, most of the limitations were outlined in Chapter 5, but some additional points relevant to this case-control study are listed below:

- The major limitation of the study is the small number of cases and hence limited power to detect an effect of treatment-induced density change on breast cancer risk. The main analysis included just one woman from the anastrozole case group who lost at least 5% density. Clearly, more women are required to see if this is an effect that may be representative of the population.
- Using an intention to treat analysis has the disadvantage that treatment administered does not necessarily mean treatment consumed. However, assessments of compliance showed no difference in adherence between cases or controls who did or did not lose density, therefore it is unlikely that compliance was a confounding factor.

6.5 Conclusion

The sample size of this study was too small to effectively test for an association between anastrozole-induced density reduction and breast cancer risk reduction, and therefore conclusions could not be drawn. Further assessment in a large sample of women with many breast cancer events is essential to investigate this hypothesis with enough statistical power. Nonetheless, given there was little effect of anastrozole on density change observed in Chapter 5, one would not expect density change to be an effective prognostic or predictive biomarker of breast cancer risk reduction with prophylactic anastrozole therapy.

Chapter 7: Conclusions

7.1 Conclusion of findings

This thesis investigated the association between breast cancer risk and changes in mammographic density. The central thesis hypothesis was that repeated measures of mammographic density would be valuable for personalised breast cancer prevention. This hypothesis was tested over five chapters; each assessing individual study aims centred on the evaluation of changes in mammographic density in the assessment of breast cancer risk. An introduction and description of study aims, methods, results, discussions and conclusions were presented separately for each chapter. In this final chapter, the combined thesis findings are summarised and ideas for future research on changes in mammographic density are discussed, along with the overall impact of the thesis findings.

Chapter 1 introduced mammographic density and gave an overview of the relevant literature relating to changes in mammographic density. The rationale for the thesis was also described as well as the study aims for each chapter.

Chapter 2 assessed whether changes in BMI were associated with changes in breast density during a one year dietary weight-loss intervention study. The aim was to evaluate whether mammographic density acts as a potential mediator for reduction in risk of postmenopausal breast cancer with premenopausal weight-loss. Overall, as women lost weight, their breast fat decreased but little change was seen in their dense tissue, leading to a higher percentage density. This negated the idea that density reduction may be a biomarker for risk reduction with weight-loss. It is likely that weight-loss-induced reductions in postmenopausal risk are driven by lower levels of adipose tissue, which reduce the amount of oestrogen production through aromatisation and hence breast cancer risk; but that this pathway is somewhat independent of fibroglandular dense tissue. There have been only a few studies to assess the effect of weight-loss on dense tissue in premenopausal women, with most investigating the effect of bariatric surgery and reporting mixed results (252, 253). However, a reduction in dense tissue in premenopausal women has been seen previously with a 2 year low-fat, high-carbohydrate diet (251), suggesting that certain weight-loss interventions that lower blood levels of estradiol and estrone may be required to see such a reduction in dense tissue (296). Density was measured using three methods (visual/Cumulus/volumetric 'Stepwedge'), which allowed for the assessment of percentage and absolute density as well as area-based and volumetric density. Area and volumetric measures gave similar results for the short-term association between BMI

and density: positive relationship with breast fat, inverse relationship with percent density and little relationship with dense tissue.

Chapter 3 aimed to assess the benefit of using a woman's longitudinal history of (BI-RADS) density to improve breast cancer risk estimation beyond using a single density measure. Longitudinal density was shown to have greater predictive ability and discriminatory accuracy than a single measure of most recent density. A quarter more statistical information was gained and a small proportion more women were correctly classified as having breast cancer when using longitudinal density instead of most recent density. Longitudinal density also predicted a six-fold increased risk of breast cancer for women in the highest vs. lowest longitudinal density categories, but only a four-fold increased risk was seen between baseline or most recent BI-RADS density categories 4 and 1. These findings supported the only other known study to evaluate predictive ability when using more than one density measure compared with a single measure (262). Kerlikowske et al.'s large cohort study reported an AUC for the two-measure BI-RADS predictor that was 0.005 units higher than the one-measure predictor, whereas this thesis found a slightly greater improvement when including an unlimited number of BI-RADS measures (concordance index 0.008 units higher than the most recent BI-RADS density measure). The benefit of longitudinal density was driven by shrinkage estimation and a reduction in measurement error. There was only a small amount of information gained when assessing individual density trajectories, supporting the idea that density has a high amount of tracking through time (325, 326). Longitudinal density was also shown to have several potential uses in the clinical setting. Predictive accuracy of breast cancer risk models may be improved with the addition of longitudinal density; which would be particularly useful for breast cancer prevention strategies that aim to stratify screening and standards of care based on risk. Additionally, longitudinal density could be easily applied to a screening environment where mammography examinations occur at arbitrary points in time. Since longitudinal density is predicted using only current and previous density values, its value can be continually updated at each woman's screening appointment. Moreover, risk assessed with longitudinal density would lead to more conservative changes between risk groups than most recent BI-RADS density, reducing the possibility of a woman receiving a radically different standard of care as a result of her moving into a different risk group.

Chapter 4 was a Cochrane systematic review investigating the association between endocrine therapy-induced density reduction and breast cancer risk and mortality. Density reduction was assessed as both a prognostic and a predictive biomarker, and within each of these, the preventive and treatment settings were considered separately, as were the effects of SERMs and AIs. A literature search identified 888 potential studies to include in the review. After assessment of the titles and abstracts of potential studies, 87 full texts were obtained for further

examination. Of these, 8 studies were identified as eligible for the review according to the inclusion criteria. One study (19) tested density reduction as a prognostic and predictive biomarker for prevention of breast cancer, whilst five studies tested density reduction as a prognostic biomarker for the treatment of breast cancer (263-266, 373). One study tested density reduction in a sample of women diagnosed with breast cancer, but the analysis adjusted for endocrine therapy; hence it could not be classified as either a prognostic or predictive biomarker (372). Another study did not report any results that could be extracted so it could not be included in the prognostic review (374). The different classes of drugs, outcomes, mammographic density measures and effect measures (for instance, the cut-points used) of the studies were deemed too heterogeneous to be able to conduct a meaningful meta-analysis. Instead, the results of each study were reported in 'Summary of findings' tables. In the prognostic biomarker review (prevention), one study reported a 68% reduction in breast cancer risk with prophylactic tamoxifen for women who had a 12-18 month visually-assessed percent density reduction ≥10% compared with no change (OR=0.32 (95% CI, 0.14 to 0.72)) (19). For the prognostic biomarker review in the treatment setting, one study indicated an HR of 0.66 (95% CI, 0.40 to 1.09) for risk of recurrence with an 8-20 month Cumulus-assessed percent density reduction ≥5% compared with <5% whilst on tamoxifen (264). For AIs, the HR was 0.14 (95% CI, 0.02 to 1.11) using the same biomarker (264). Another prognostic biomarker review study (treatment) reported a 65% reduction in risk of recurrence for women with a 10-34 month tamoxifen-induced reduction in BI-RADS density compared with no reduction (HR=0.35 (95% CI, 0.17 to 0.68)) (263). In terms of mortality, for the prognostic biomarker review (treatment), one study reported a 50% reduction in risk of breast cancer death with 6-36 month tamoxifen-induced relative reduction in dense area (machine-learned area-based method) >20% compared with little change (≤9% increase to ≤10% reduction) (HR=0.50 (95% CI, 0.27 to 0.93)) (265). Another prognostic biomarker review study (treatment) reported a 56% decreased risk of breast cancer death with a 3-26 month tamoxifen-induced reduction in Cumulus-assessed percent density of >8.7% compared with <0.5% (OR=0.44 (95% CI, 0.22 to 0.88)) (266). The final prognostic biomarker study (treatment) reported an OR of 0.52 (95% CI, 0.18 to 1.51) for risk of contralateral breast cancer with a 1-5 year reduction in percent density (machine-learned area-based method) of ≥10% compared with little change (<10% reduction to <10% increase) whilst on endocrine therapy (373). In the predictive biomarker review (prevention), an interaction between prophylactic tamoxifen and 12-18 month visually-assessed percent density reduction (≥10% or <10%) had an OR for risk of breast cancer of 0.53 (95% CI, 0.21 to 1.32) (19). Overall, there was some evidence to suggest that density reduction may be a prognostic and predictive biomarker for reduction in breast cancer risk, and prognostic biomarker for reduction in breast cancer risk of recurrence, mortality and contralateral breast cancer with tamoxifen. However, the level of evidence for this biomarker was limited by several study quality issues. The suggestion of density reduction as a prognostic biomarker for reduction in

risk of recurrence whilst on AIs was also limited by study quality issues and requires further investigation.

Chapter 5 aimed to assess the effect of prophylactic anastrozole on visually-assessed density reduction during the IBIS-II trial. After 2 years of anastrozole treatment, breast density decreased by an average of 1.05%, but this was not significantly different from the 0.82% mean density reduction on placebo. It is likely that prophylactic anastrozole treatment reduces breast density by only a small amount and this was not captured by the study due to limited power. However, this is the largest known study to assess the effect of a preventive AI on density, and the lack of a significant effect even with a considerable sample size (n=575) further suggests only a marginal effect of anastrozole on density. These results were in concordance with previous studies reporting a null or very minimal effect of AIs in both the preventive (268-270) and adjuvant (267, 337) settings. Whilst a study by Engmann et al. reported a significant reduction in volumetric density with adjuvant AIs, the effect was again small with only a 0.3% greater reduction in Volpara percent density and 0.6% greater reduction in Quantra percent density for breast cancer cases on 2-3 years of AIs compared with breast cancer-free controls on no treatment. To see such small changes in density with AI treatment, it may be necessary to use volumetric measures that account for overlapping dense tissue within the breast. Use of volumetric methods to test the effect of AIs on density requires further exploration.

Chapter 6 aimed to investigate visually-assessed density reduction with prophylactic anastrozole as a biomarker for concurrent breast cancer risk reduction using a case-control study from the IBIS-II trial. Unfortunately, the number of breast cancer cases was not large enough for the study to be adequately powered (n=22), and the number of cases who lost at least 10% or 5% density was minimal. Therefore, the sample size was too small to infer an effect and no definitive conclusions could be drawn.

Overall, changes in mammographic density were shown to be useful for the assessment of breast cancer risk. Repeated measures of density have great potential for use in personalised breast cancer prevention. Their use in improving the accuracy of breast cancer risk estimation and in indicating response to endocrine treatment could prove to be extremely valuable for breast cancer prevention strategies.

7.2 Future research: use of changes in mammographic density for breast cancer prevention

This thesis revealed two main avenues for further research: (a) further development and validation of the longitudinal density measure with the ultimate aim of incorporating it into established breast cancer risk models, (b) expanding evidence on mammographic density reduction as a biomarker for reduction in risk and mortality with endocrine therapy.

Several approaches can be used to further develop the longitudinal density measure. Firstly, it would be useful to test the benefit of applying different weightings to historical density values. For instance, more recent values may require up-weighting, whereas earlier mammograms may be less informative and require application of 'forgetting factors'. Secondly, including additional confounders of density in the linear mixed model or modelling the mixed effects in a multinomial or ordinal logit model may improve the prediction of the longitudinal density measure. Thirdly, the ability of longitudinal density to stratify breast cancer risk might be improved with the addition of other classical risk factors or a longitudinal measure of BMI in the proportional-hazards Cox model. Additionally, estimation of breast cancer risk with longitudinal density may be improved with a joint longitudinal-survival model that maximises likelihood in both models simultaneously (330, 331). It would then be important to validate the longitudinal density measure in other cohorts of women and using different density measures, as well as testing the benefit of this measure in established breast cancer risk models such as the Tyrer-Cuzick, Gail or BCSC model.

Establishing whether a reduction in density can be used as a biomarker for response to endocrine therapy is another priority for future research on changes in mammographic density. A meta-analysis of individual participant-level data from the best quality studies identified in the Cochrane systematic review would help to overcome some of the quality issues identified as well as account for heterogeneity between studies. Gathering further evidence to support this biomarker is essential if it is to be implemented into clinical practice. Additionally, in the Cochrane review, there was little evidence for other endocrine therapies besides tamoxifen. With AIs, it is important to first determine whether density reductions truly occur whilst on the drug. Only then can AI-induced density reduction be tested as a biomarker for breast cancer incidence or death. It is therefore important to continue to test the effect of AIs on density change, for instance using volumetric methods that may capture changing breast phenotypes that area-based methods have so far failed to do.

Another possible future study for consideration combines these two avenues of research. Whilst the addition of random slopes showed little benefit in estimating breast cancer risk in Chapter 3,

it should be noted that women in this cohort study were part of a population-based screening programme with a mixture of underlying baseline risks. Random slopes may in fact be more useful in high-risk populations of women undergoing active treatment to reduce their risk, or breast cancer patients on a course of adjuvant treatment. As seen in Chapter 4, multiple cut-points are currently used to assess endocrine therapy-induced density reduction and breast cancer outcomes. It may be more useful to assess within-women changes in density across the course of endocrine therapy to assess whether they differ from individual random slopes predicted from a linear mixed model using data from before treatment commencement. A density decline greater than a woman's predicted trajectory might signify a response to treatment, and it would be useful to investigate this hypothesis in a future study.

7.3 Impact of the thesis findings

The findings of this thesis have the potential to be useful for breast cancer prevention in several ways. Optimising the longitudinal density measure to better predict risk of breast cancer could help to identify women at low- and high-risk of developing breast cancer who would benefit from more tailored prevention regimes. For instance, risk-stratified screening would allow those at a higher risk of developing breast cancer to be screened more frequently, and those at a lower risk to be screened less frequently than the current practice of inviting all women for the same frequency of screening regardless of risk (134, 419). High-risk women may also benefit from supplemental imaging using modalities such as MRI or ultrasound (134), and having a greater amount of information on their risk of breast cancer could be useful for helping women to make more informed decisions about lifestyle choices that may be influencing their risk of developing the disease, such as diet, exercise and alcohol intake (420, 421). Alternatively, high-risk women may benefit from a course of treatment with a chemo-preventive drug to lower their risk (134, 421). Recommendations for chemoprevention are already outlined in NICE guidelines, whereby a 5-year course of tamoxifen or raloxifene is advised for women at high or moderate familial risk (422, 423), and in 2017, these guidelines were updated to include a 5-year course of anastrozole for postmenopausal women at high or moderate familial risk, without severe osteoporosis (422).

Chemoprevention highlights another potential use for the findings of this thesis. Chapter 4 suggested that endocrine therapy-induced density reduction is a promising biomarker for reduction in breast cancer risk and mortality. If this biomarker were to be implemented in practice, it would act as a quick and cost-effective tool for assessing response to treatment that would be less invasive than the alternative tissue and blood sample biomarkers. Additionally, density change is an early biomarker that can be measured approximately a year after the start of treatment. In theory, this biomarker should therefore result in fewer diagnoses of breast cancer if

non-responders are given the chance to try alternative risk-reducing treatments early on when there is time to prevent the development of a potential tumour. This would also be relevant in the adjuvant setting whereby changing the treatment of non-responders and intervening at an early stage would be an improvement on the current "wait-and-see" approach whereby a lack of treatment response is often only revealed at the point of breast cancer recurrence or death. This makes density change a particularly useful biomarker with the potential to save a number of lives.

To conclude, this thesis showed that changes in mammographic density were useful for the study of breast cancer risk and should be considered for personalised breast cancer prevention strategies. Future research into the area of changes in mammographic is a priority. It is important that health practitioners, policy makers and patients benefit from the findings of this thesis and make use of the great potential that changes in mammographic density have for breast cancer prevention.

References

1. Cancer Research UK. Breast Cancer Statistics [Internet] [cited 2019 25 July]. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2018;68(6):394-424.
3. Matthews TJ, Hamilton BE. Delayed childbearing: more women are having their first child later in life. NCHS data brief. 2009(21):1-8.
4. Kennedy DL, Baum C, Forbes MB. Noncontraceptive estrogens and progestins: use patterns over time. Obstet Gynecol. 1985;65(3):441-6.
5. Seidell JC. Prevalence and time trends of obesity in Europe. Journal of endocrinological investigation. 2002;25(10):816-22.
6. Althuis MD, Dozier JM, Anderson WF, Devesa SS, Brinton LA. Global trends in breast cancer incidence and mortality 1973-1997. Int J Epidemiol. 2005;34(2):405-12.
7. Torre LA, Siegel RL, Ward EM, Jemal A. Global Cancer Incidence and Mortality Rates and Trends-An Update. Cancer Epidemiol Biomarkers Prev. 2016;25(1):16-27.
8. Youlden DR, Cramb SM, Dunn NA, Muller JM, Pyke CM, Baade PD. The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality. Cancer Epidemiol. 2012;36(3):237-48.
9. Duffy SW, Tabar L, Olsen AH, Vitak B, Allgood PC, Chen TH, et al. Absolute numbers of lives saved and overdiagnosis in breast cancer screening, from a randomized trial and from the Breast Screening Programme in England. J Med Screen. 2010;17(1):25-30.
10. Tabar L, Vitak B, Chen TH, Yen AM, Cohen A, Tot T, et al. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. Radiology. 2011;260(3):658-63.
11. Peto R, Boreham J, Clarke M, Davies C, Beral V. UK and USA breast cancer deaths down 25% in year 2000 at ages 20-69 years. Lancet. 2000;355(9217):1822.
12. Wellings SR. A hypothesis of the origin of human breast cancer from the terminal ductal lobular unit. Pathology, research and practice. 1980;166(4):515-35.
13. Holland K, Sechopoulos I, Mann RM, den Heeten GJ, van Gils CH, Karssemeijer N. Influence of breast compression pressure on the performance of population-based mammography screening. Breast Cancer Research. 2017;19(126).
14. Wolfe JN. The prominent duct pattern as an indicator of cancer risk. Oncology. 1969;23(2):149-58.

15. Gram IT, Funkhouser E, Tabar L. The Tabar classification of mammographic parenchymal patterns. European Journal of Radiology. 1997;24:131-6.

16. Assi V, Warwick J, Cuzick J, Duffy SW. Clinical and epidemiological issues in mammographic density. Nat Rev Clin Oncol. 2011;9(1):33-40.

17. Spak DA, Plaxco JS, Santiago L, Dryden MJ, Dogan BE. BI-RADS((R)) fifth edition: A summary of changes. Diagn Interv Imaging. 2017;98(3):179-90.

18. D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA. ACR BI-RADS®, Atlas, Breast Imaging Reporting and Data System. Reston, VA: American College of Radiology; 2013.

19. Cuzick J, Warwick J, Pinney E, Duffy SW, Cawthorn S, Howell A, et al. Tamoxifen-induced reduction in mammographic density and breast cancer risk reduction: a nested case-control study. J Natl Cancer Inst. 2011;103(9):744-52.

20. Ciatto S, Houssami N, Apruzzese A, Bassetti E, Brancato B, Carozzi F, et al. Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories. The Breast. 2004;14(4):269-75.

21. Gao J, Warren R, Warren-Forward H, Forbes JF. Reproducibility of visual assessment on mammographic density. Breast Cancer Res Treat. 2008;108(1):121-7.

22. Boyd NF, Lockwood GA, Byng JW, Tritchler DL, Yaffe MJ. Mammographic densities and breast cancer risk. Cancer Epidemiology Biomarkers & Prevention. 1998;7(12):1133-44.

23. Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ. The quantitative analysis of mammographic densities. Phys Med Biol. 1994;39(10):1629-38.

24. Couwenberg AM, Verkooijen HM, Li J, Pijnappel RM, Charaghvandi KR, Hartman M, et al. Assessment of a fully automated, high-throughput mammographic density measurement tool for use with processed digital mammograms. Cancer Causes Control. 2014;25(8):1037-43.

25. Habel LA, Lipson JA, Achacoso N, Rothstein JH, Yaffe MJ, Liang RY, et al. Case-control study of mammographic density and breast cancer risk using processed digital mammograms. Breast Cancer Res. 2016;18(1):53.

26. Wolfe JN, Saftlas AF, Salane M. Mammographic parenchymal patterns and quantitative evaluation of mammographic densities: a case-control study. The American Journal of Roentgenology. 1987;148(6):1087–92.

27. 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. Journal of the National Cancer Institute. 1995;87(9):670-5.

28. 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(6):1587-98.

29. Nickson C, Arzhaeva Y, Aitken Z, Elgindy T, Buckley M, Li M, et al. AutoDensity: an automated method to measure mammographic breast density that predicts breast cancer risk and screening outcomes. Breast Cancer Research. 2013;15:R80.

30. Keller BM, Nathan DL, Wang Y, Zheng Y, Gee JC, Conant EF, et al. Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation. Med Phys. 2012;39(8):4903-17.

31. Li J, Szekely L, Eriksson L, Heddson B, Sundbom A, Czene K, et al. High-throughput mammographic-density measurement: a tool for risk prediction of breast cancer. Breast Cancer Res. 2012;14(4):R114.

32. Cancer Research UK. Breast screening [Internet] [cited 2017 03 February]. Available from: http://www.cancerresearchuk.org/about-cancer/breast-cancer/screening/breast-screening.

33. Highnam R, Pan X, Warren R, Jeffreys M, Davey Smith G, Brady M. Breast composition measurements using retrospective standard mammogram form (SMF). Phys Med Biol. 2006;51(11):2695-713.

34. Hologic. Understanding R2 Quantra 1.3 (Software Manual). In: Hologic Inc., editor. 2009.

35. Yaffe MJ. Measurement of mammographic density. Breast Cancer Res. 2008;10(3):209.

36. Highnam R, Jeffreys M, McCormack V, Warren RML, Smith DG, Brady M. Comparing measurements of breast density. Phys Med Biol. 2007;52(19):5881-95.

37. Alonzo-Proulx O, Packard N, Boone JM, Al-Mayah A, Brock KK, Shen SZ, et al. Validation of a method for measuring the volumetric breast density from digital mammograms. Phys Med Biol. 2010;55(11):3027-44.

38. Shepherd JA, Herve L, Landau J, Fan B, Kerlikowske K, Cummings SR. Novel use of Single X-Ray Absorptiometry for Measuring Breast Density. Technology in Cancer Research & Treatment. 2005;4(2).

39. Blot L, Zwiggelaar R. A volumetric approach to glandularity estimation in mammography: a feasibility study. Physics in Medicine and Biology. 2005;50(4):695-708.

40. van Engeland S, Snoeren PR, Huisman H, Boetes C, Karssemeijer N. Volumetric Breast Density Estimation From Full-Field Digital Mammograms. IEEE Transactions on medical imaging. 2006;25(3).

41. 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(6):1754-62.

42. Diffey J, Berks M, Hufton A, Chung C, Verow R, Morrison J, et al., editors. A stepwedge-based method for measuring breast density: observer variability and comparison with human reading. SPIE Medical Imaging; 2010; San Diego, CA, US: Medical Imaging 2010: Physics of Medical Imaging.

43. Hartman K, Highnam R, Warren R, Jackson V, editors. Volumetric Assessment of Breast Tissue Composition from FFDM Images. 9th International Workshop, IWDM 2008; 2008; Tucson, AZ, USA: Digital Mammography.

44. Highnam R, Brady M, Yaffe MK, N., Harvey J, editors. Robust breast composition measurement - VolparaTM. 10th International Workshop, IWDM; 2010; Girona, Catalonia, Spain: Digital Mammography.

45. Gilhuijs KGA, Gubern-Mérida A, Kallenberg M, Platel B, Mann RM, Martí R, et al. Volumetric Breast Density Estimation from Full-Field Digital Mammograms: A Validation Study. PLoS ONE. 2014;9(1).

46. Engelken F, Singh J, Fallenberg E, Bick U, Böttcher J, Miriam Renz D. Volumetric breast composition analysis: reproducibility of breast percent density and fibroglandular tissue volume measurements in serial mammograms. Nordic Society of Medical Radiology. 2014;55(1).

47. Ding J, Warren R, Warsi I, Day N, Thompson D, Brady M, et al. Evaluating the effectiveness of using standard mammogram form to predict breast cancer risk: case-control study. Cancer Epidemiol Biomarkers Prev. 2008;17(5):1074-81.

48. 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(2):418-28.

49. Brandt KR, Scott CG, Ma L, Mahmoudzadeh AP, Jensen MR, Whaley DH, et al. Comparison of Clinical and Automated Breast Density Measurements: Implications for Risk Prediction and Supplemental Screening. Radiology. 2016;279(3):710-9.

50. Astley SM, Harkness EF, Sergeant JC, Warwick J, Stavrinos P, Warren R, et al. A comparison of five methods of measuring mammographic density: a case-control study. Breast Cancer Res. 2018;20(1):10.

51. Jeffers AM, Sieh W, Lipson JA, Rothstein JH, McGuire V, Whittemore AS, et al. Breast Cancer Risk and Mammographic Density Assessed with Semiautomated and Fully Automated Methods and BI-RADS. Radiology. 2017;282(2):348-55.

52. Nielsen M, Karemore G, Loog M, Raundahl J, Karssemeijer N, Otten JD, et al. A novel and automatic mammographic texture resemblance marker is an independent risk factor for breast cancer. Cancer Epidemiol. 2011;35(4):381-7.

53. Torres-Mejía G, De Stavola B, Allen DS, Pérez-Gavilán 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(5):1052-9.

54. Häberle L, Wagner F, Fasching PA, Jud SM, Heusinger K, Loehberg CR, et al. Characterizing mammographic images by using generic texture features. Breast Cancer Res. 2012;14(2):R59.

55. Kontos D, Ikejimba LC, Bakic PR, Troxel AB, Conant EF, Maidment AD. Analysis of parenchymal texture with digital breast tomosynthesis: comparison with digital mammography and implications for cancer risk assessment. Radiology. 2011;261(1):80-91.

56. Manduca A, Carston MJ, Heine JJ, Scott CG, Pankratz VS, Brandt KR, et al. Texture features from mammographic images and risk of breast cancer. Cancer Epidemiol Biomarkers Prev. 2009;18(3):837-45.

57. He W, Juette A, Denton ER, Oliver A, Marti R, Zwiggelaar R. A Review on Automatic Mammographic Density and Parenchymal Segmentation. Int J Breast Cancer. 2015;2015:276217.

58. Oliver A, Freixenet J, Marti R, Pont J, Perez E, Denton ERE, et al. A Novel Breast Tissue Density Classification Methodology. IEEE Transactions on information technology in biomedicine 2008;12(1).

59. Heine JJ, Scott CG, Sellers TA, Brandt KR, Serie DJ, Wu FF, et al. A novel automated mammographic density measure and breast cancer risk. J Natl Cancer Inst. 2012;104(13):1028-37.

60. Kontos D, Bakic PR, Carton AK, Troxel AB, Conant EF, Maidment AD. Parenchymal texture analysis in digital breast tomosynthesis for breast cancer risk estimation: a preliminary study. Acad Radiol. 2009;16(3):283-98.

61. Wang C, Brentnall AR, Cuzick J, Harkness EF, Evans DG, Astley S. A novel and fully automated mammographic texture analysis for risk prediction: results from two case-control studies. Breast Cancer Res. 2017;19(1):114.

62. Wanders JOP, van Gils CH, Karssemeijer N, Holland K, Kallenberg M, Peeters PHM, et al. The combined effect of mammographic texture and density on breast cancer risk: a cohort study. Breast Cancer Res. 2018;20(1):36.

63. Vedantham S, Shi L, Karellas A, O'Connell AM. Dedicated breast CT: fibroglandular volume measurements in a diagnostic population. Med Phys. 2012;39(12):7317-28.

64. Moon WK, Lo CM, Goo JM, Bae MS, Chang JM, Huang CS, et al. Quantitative analysis for breast density estimation in low dose chest CT scans. J Med Syst. 2014;38(3):21.

65. Huang S, Boone J, Yang K, Packard NJ, McKenney SE, Prionas ND, et al. The characterization of breast anatomical metrics using dedicated breast CT. Medical Physics. 2011;38(4):2180–91.

66. Glide-Hurst CK, Duric N, Littrup P. Volumetric breast density evaluation from ultrasound tomography images. Medical Physics. 2008;35(9):3988-97.

67. Chen J, Huang C, Chien KC, Takada E, Moon WK, Wu JHK, et al. Breast density analysis for whole breast ultrasound images. Medical Physics. 2009;36(11):4933-43.

68. Thompson DJ, Leach MO, Kwan-Lim G, Gayther SA, Ramus SJ, Warsi I, et al. Assessing the usefulness of a novel MRI-based breast density estimation algorithm in a cohort

of women at high genetic risk of breast cancer: the UK MARIBS study. Breast Cancer Research. 2009;11:R80.

69. Tagliafico A, Tagliafico G, Houssami N. Differences in breast density assessment using mammography, tomosynthesis and MRI and their implications for practice. Br J Radiol. 2013;86(1032):20130528.

70. Maskarinec G, Morimoto Y, Daida Y, Laidevant A, Malkov S, Shepherd JA, et al. Comparison of breast density measured by dual energy X-ray absorptiometry with mammographic density among adult women in Hawaii. Cancer Epidemiol. 2011;35(2):188-93.

71. Niklason LT, Christian BT, Niklason LE, Kopans DB, Castleberry DE, Opsahl-Ong BH. Digital Tomosynthesis in Breast Imaging. Radiology. 1997;205:399-406.

72. Tromans C, Highnam R, Morrish O, Black R, Tucker L, Gilbert FJ, editors. Volumetric breast density estimation on conventional mammography versus digital breast tomosynthesis - Poster no. C-0363. European Congress of Radiology 2014; 2014; Vienna, Austria: Insights into Imaging - Springer Berlin Heidelberg.

73. Ekpo EU, Mello-Thoms C, Rickard M, Brennan PC, McEntee MF. Breast density (BD) assessment with digital breast tomosynthesis (DBT): Agreement between Quantra and 5th edition BI-RADS(R). Breast. 2016;30:185-90.

74. Hsieh J, Bakic PR, Samei E, Kontos D, Carton A, Maidment ADA, editors. Breast percent density estimation from 3D reconstructed digital breast tomosynthesis images. Proceedings of SPIE; 2008: Medical Imaging 2008: Physics of Medical Imaging.

75. Tagliafico A, Tagliafico G, Astengo D, Cavagnetto F, Rosasco R, Rescinito G, et al. Mammographic density estimation: one-to-one comparison of digital mammography and digital breast tomosynthesis using fully automated software. Eur Radiol. 2012;22(6):1265-70.

76. Tagliafico AS, Tagliafico G, Cavagnetto F, Calabrese M, Houssami N. Estimation of percentage breast tissue density: comparison between digital mammography (2D full field digital mammography) and digital breast tomosynthesis according to different BI-RADS categories. Br J Radiol. 2013;86(1031):20130255.

77. Ekpo EU, McEntee MF. Measurement of breast density with digital breast tomosynthesis - a systematic review. Br J Radiol. 2014;87(1043):20140460.

78. Boyd NF, Martin LJ, Stone J, Greenberg C, Minkin S, Yaffe MJ. Mammographic densities as a marker of human breast cancer risk and their use in chemoprevention. Current Oncology Reports. 2001;3:314–21.

79. 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(6):1159-69.

80. Warner E, Lockwood G, Tritchler D, Boyd NF. The risk of breast cancer associated with mammographic parenchymal patterns: a meta-analysis of the published literature to

examine the effect of method of classification. Cancer detection and prevention. 1992;16(1):67-72.

81. Gram IT, Bremnes Y, Ursin G, Maskarinec G, Bjurstam N, Lund E. Percentage density, Wolfe's and Tabar's mammographic patterns: agreement and association with risk factors for breast cancer. Breast Cancer Res. 2005;7(5):R854-61.

82. Eng A, Gallant Z, Shepherd J, McCormack V, Li J, Dowsett M, et al. Digital mammographic density and breast cancer risk: a case-control study of six alternative density assessment methods. Breast Cancer Research. 2014;16:439.

83. 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(7):1473-82.

84. Duffy SW, Morrish OWE, Allgood PC, Black R, Gillan MGC, Willsher P, et al. Mammographic density and breast cancer risk in breast screening assessment cases and women with a family history of breast cancer. Eur J Cancer. 2017;88:48-56.

85. Kerlikowske K, Scott CG, Mahmoudzadeh AP, Ma L, Winham S, Jensen MR, et al. Automated and Clinical Breast Imaging Reporting and Data System Density Measures Predict Risk for Screen-Detected and Interval Cancers: A Case-Control Study. Annals of internal medicine. 2018;168(11):757-65.

86. Kerlikowske K, Ma L, Scott CG, Mahmoudzadeh AP, Jensen MR, Sprague BL, et al. Combining quantitative and qualitative breast density measures to assess breast cancer risk. Breast Cancer Res. 2017;19(1):97.

87. Stone J, Ding J, Warren RM, Duffy SW, Hopper JL. Using mammographic density to predict breast cancer risk: dense area or percentage dense area. Breast Cancer Res. 2010;12(6):R97.

88. Rauh C, Hack CC, Häberle L, Hein A, Engel A, Schrauder MG, et al. Percent Mammographic Density and Dense Area as Risk Factors for Breast Cancer. Geburtshilfe Frauenheilkd. 2012;72(8):727-33.

89. Vachon CM, Brandt KR, Ghosh K, Scott CG, Maloney SD, Carston MJ, et al. Mammographic breast density as a general marker of breast cancer risk. Cancer Epidemiol Biomarkers Prev. 2007;16(1):43-9.

90. Pettersson A, Graff RE, Ursin G, dos Santos Silva I, McCormack V, Baglietto L, et al. Mammographic density phenotypes and risk of breast cancer: a meta-analysis. J Natl Cancer Inst. 2014;106(5).

91. Stone J, Ding J, Warren RM, Duffy SW. Predicting breast cancer risk using mammographic density measurements from both mammogram sides and views. Breast Cancer Res Treat. 2010;124(2):551-4.

92. Warren R, Thompson D, del Frate C, Cordell M, Highnam R, Tromans C, et al. A comparison of some anthropometric parameters between an Italian and a UK population: “proof

of principle" of a European project using MammoGrid. Clinical Radiology. 2007;62(11):1052–60.

93. Duffy SW, Nagtegaal ID, Astley SM, Gillan MG, McGee MA, Boggis CR, et al. Visually assessed breast density, breast cancer risk and the importance of the craniocaudal view. Breast Cancer Res. 2008;10(4):R64.

94. Harvey JA. Quantitative Assessment of Percent Breast Density: Analog versus Digital Acquisition. Technology in Cancer Research & Treatment. 2004;3(6):611-6.

95. 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(3):227-36.

96. Are You Dense Inc. Are You Dense? Exposing the Best-Kept Secret® [Internet] 2012 [cited 2017 02 February]. Available from: http://www.areyoudense.org/.

97. Price ER, Hargreaves J, Lipson JA, Sickles EA, Brenner RJ, Lindfors KK, et al. The California Breast Density Information Group: A Collaborative Response to the Issues of Breast Density, Breast Cancer Risk, and Breast Density Notification Legislation. Radiology. 2013;269(3).

98. DenseBreast-info.org. Legislation and Regulation [Internet]. Available from: https://densebreast-info.org/legislation.aspx.

99. Ekpo EU, Hogg P, Highnam R, McEntee MF. Breast composition: Measurement and clinical use. Radiography. 2015;21(4):324-33.

100. Irshad A, Leddy R, Ackerman S, Cluver A, Pavic D, Abid A, et al. Effects of Changes in BI-RADS Density Assessment Guidelines (Fourth Versus Fifth Edition) on Breast Density Assessment: Intra- and Interreader Agreements and Density Distribution. AJR Am J Roentgenol. 2016;207(6):1366-71.

101. Egan RL, Mosteller RC. Breast cancer mammography patterns. Cancer. 1977;40(5):2087-90.

102. van Gils CH, Otten JDM, Verbeek ALM, Hendriks JHCL. Mammographic breast density and risk of breast cancer: Masking bias or causality? European journal of epidemiology. 1998;14(4):315-20.

103. 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 Cancer Inst. 2000;92(13):1081-7.

104. 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. Journal of the National Cancer Institute. 1995;87(21):1622-9.

105. Rebolj M, Blyuss O, Chia KS, Duffy SW. Long-term excess risk of breast cancer after a single breast density measurement. Eur J Cancer. 2019;117:41-7.

106. Harvey JA, Bovbjerg VE. Quantitative Assessment of Mammographic Breast Density: Relationship with Breast Cancer Risk. Radiology. 2004;230:29–41.

107. Sala E, Warren R, McCann J, Duffy S, Day N, Luben R. Mammographic parenchymal patterns and mode of detection: implications for the breast screening programme. J Med Screen. 1998;5:207–12.

108. 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(1):33-8.

109. Pisano ED, Hendrick RE, Yaffe MJ, Baum JK, Acharyya S, Cormack JB, et al. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. Radiology. 2008;246(2).

110. Kerlikowske K, Hubbard RA, Miglioretti DL, Geller BM, Yankaskas BC, Lehman CD, et al. Comparative effectiveness of digital versus film-screen mammography in community practice in the United States: a cohort study. Ann Intern Med. 2011;155(8):493-502.

111. Kelly KM, Dean J, Comulada WS, Lee SJ. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. Eur Radiol. 2010;20(3):734-42.

112. Giuliano V, Giuliano C. Improved breast cancer detection in asymptomatic women using 3D-automated breast ultrasound in mammographically dense breasts. Clinical Imaging. 2013;37(3):480-6.

113. Kuhl CK, Schrading S, Leutner CC, Morakkabati-Spitz N, Wardelmann E, Fimmers R, et al. Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. J Clin Oncol. 2005;23(33):8469-76.

114. 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(13).

115. Rebolj M, Assi V, Brentnall A, Parmar D, Duffy SW. Addition of ultrasound to mammography in the case of dense breast tissue: systematic review and meta-analysis. British journal of cancer. 2018;118(12):1559-70.

116. Warner E, Messersmith H, Causer P, Eisen A, Shumak R, Plewes D. Systematic review: using magnetic resonance imaging to screen women at high risk for breast cancer. Ann Intern Med. 2008;148(9):671-9.

117. Freer PE. Mammographic Breast Density: Impact on Breast Cancer Risk and Implications for Screening. RadioGraphics. 2015;35:302–15.

118. Sprague BL, Stout NK, Schechter C, van Ravesteyn NT, Cevik M, Alagoz O, et al. Potential impact of legislation mandating breast density notification: benefits, harms, and cost effectiveness of supplemental ultrasound screening. Ann Intern Med. 2015;162(3):157–66.

119. Haas BM, Kalra V, Geisel J, Raghu M, Durand M, Philpotts LE. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. Radiology. 2013;269(3):694-700.

120. Rafferty EA, Niklason L, editors. FFDM versus FFDM with tomosynthesis for women with radiologically dense breasts: an enriched retrospective reader study. Radiological Society of North America annual meeting; 2011; Chicago, IL, USA.

121. Rafferty EA, Niklason L, Smith A, editors. Comparison of FFDM with breast tomosynthesis to FFDM alone: performance in fatty and dense breasts. Tomosynthesis Imaging Symposium; 2009; Duke University, North Carolina, USA.

122. Melnikow J, Fenton JJ, Whitlock EP, Miglioretti DL, Weyrich MS, Thompson JH, et al. Supplemental Screening for Breast Cancer in Women With Dense Breasts: A Systematic Review for the U.S. Preventive Service Task Force [Internet]. Evidence Synthesis No 126 Agency for Healthcare Research and Quality (US) Publication, Rockville, MD: Agency for Healthcare Research and Quality. 2016;No. 14–05201-EF-3.

123. Vachon CM, Ghosh K, Brandt KR. Mammographic Density: Potential as a Risk Factor and Surrogate Marker in the Clinical Setting. Current Breast Cancer Reports. 2013;5(3):183-93.

124. Skaane P, Bandos A, Eben EB, Jebsen 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(3):655-63.

125. Tice JA, Kerlikowske K. Supplemental Breast Cancer Screening: A Density Conundrum. J Gen Intern Med. 2017;32(6):593–4.

126. Richman I, Asch SM, Bendavid E, Bhattacharya J, Owens DK. Breast Density Notification Legislation and Breast Cancer Stage at Diagnosis: Early Evidence from the SEER Registry. J Gen Intern Med. 2017;32(6):603-9.

127. Emaus MJ, Bakker MF, Peeters PH, Loo CE, Mann RM, de Jong MD, et al. MR Imaging as an Additional Screening Modality for the Detection of Breast Cancer in Women Aged 50-75 Years with Extremely Dense Breasts: The DENSE Trial Study Design. Radiology. 2015;277(2):527-37.

128. 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(1):42-8.

129. Yaghjyan 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 Cancer Inst. 2011;103:1179–89.

130. Aiello EJ, Buist DSM, White E, Porter PL. Association between Mammographic Breast Density and Breast Cancer Tumor Characteristics. Cancer Epidemiol Biomarkers Prev. 2005;14(3):662-8.

131. Evans DG, Howell A. Can the breast screening appointment be used to provide risk assessment and prevention advice? Breast Cancer Res. 2015;17:84.

132. National Collaborating Centre for Cancer. 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. Update of clinical guideline 14 and 41. Clinical guideline 164. [Internet]. 2013 [cited 2017 7 May]. Available from: http://guidance.nice.org.uk/CG164.

133. Tice JA, Ollendorf DA, Lee JM, Pearson SD. The Comparative Clinical Effectiveness and Value of Supplemental Screening Tests Following Negative Mammography in Women with Dense Breast Tissue [Internet] 2013 [cited 2017 14 March]. Available from: http://www.ctaf.org/sites/default/files/assessments/ctaf-final-report-dense-breast-imaging-11.04.2013-b.pdf.

134. Evans DG, Brentnall AR, Harvie M, Dawe S, Sergeant JC, Stavrinos P, et al. Breast cancer risk in young women in the national breast screening programme: implications for applying NICE guidelines for additional screening and chemoprevention. Cancer Prev Res (Phila). 2014;7(10):993-1001.

135. Slanetz PJ, Freer PE, Birdwell RL. Breast-Density Legislation — Practical Considerations. New England Journal of Medicine. 2015;372(7):593-5.

136. Hooley RJ, Greenberg KL, Stackhouse RM, Geisel JL, Butler RS, Philpotts LE. Screening US in patients with mammographically dense breasts: initial experience with Connecticut Public Act 09-41. Radiology. 2012;265(59-69).

137. Hutson SW, Cowen PN, Bird CC. Morphometric studies of age related changes in normal human breast and their significance for evolution of mammary cancer. J Clin Pathol. 1985;38(3):281-7.

138. Vachon CM, Kuni CC, Anderson K, Anderson VE, Sellers TA. Association of mammographically defined percent breast density with epidemiologic risk factors for breast cancer (United States). Cancer Causes and Control. 2000;11:653-62.

139. Ursin G, Qureshi SA. Mammographic density - a useful biomarker for breast cancer risk in epidemiologic studies. Norsk Epidemiologi. 2009;19(1):59-68.

140. Burton A, Maskarinec G, Perez-Gomez B, Vachon C, Miao H, Lajous M, et al. Mammographic density and ageing: A collaborative pooled analysis of cross-sectional data from 22 countries worldwide. PLoS Med. 2017;14(6):e1002335.

141. Krishnan K, Baglietto L, Stone J, Simpson JA, Severi G, Evans CF, et al. Longitudinal Study of Mammographic Density Measures That Predict Breast Cancer Risk. Cancer Epidemiol Biomarkers Prev. 2017;26(4):651-60.

142. Boyd NF, Martin LJ, Li Q, Sun L, Chiarelli AM, Hislop G, et al. Mammographic density as a surrogate marker for the effects of hormone therapy on risk of breast cancer. Cancer Epidemiol Biomarkers Prev. 2006;15(5):961-6.

143. Pike MC, Krailo MD, Henderson BE, Casagrande JT, Hoel DG. 'Hormonal' risk factors, 'breast tissue age' and the age-incidence of breast cancer. Nature. 1983;303(5920):767-70.

144. Rosner B, Colditz GA. Nurses' health study: log-incidence mathematical model of breast cancer incidence. J Natl Cancer Inst. 1996;88(6):359-64.

145. Boyd NF, Stone J, Martin LJ, Jong R, Fishell E, Yaffe M, et al. The association of breast mitogens with mammographic densities. Br J Cancer. 2002;87(8):876-82.

146. Oza AM, Boyd NF. Mammographic parenchymal patterns: a marker of breast cancer risk. Epidemiol Rev. 1993;15(1):196-208.

147. Spicer DV, Ursin G, Parisky YR, Pearce JG, Shoupe D, Pike A, et al. Changes in mammographic densities induced by a hormonal contraceptive designed to reduce breast cancer risk. J Natl Cancer Inst. 1994;86(6):431-6.

148. Pollak MN. Endocrine effects of IGF-I on normal and transformed breast epithelial cells: potential relevance to strategies for breast cancer treatment and prevention. Breast Cancer Res Treat. 1998;47(3):209-17.

149. Pike MC, Spicer DV, Dahmoush L, Press MF. Estrogens, progestogens, normal breast cell proliferation, and breast cancer risk. Epidemiol Rev. 1993;15(1):17-35.

150. El-Bastawissi AY, White E, Mandelson MT, Taplin SH. Reproductive and hormonal factors associated with mammographic breast density by age (United States). Cancer Causes Control. 2000;11(10):955-63.

151. Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. Lancet Oncol. 2012;13(11):1141-51.

152. Gram IT, Funkhouser E, Tabar L. Reproductive and menstrual factors in relation to mammographic parenchymal patterns among perimenopausal women. Br J Cancer. 1995;71(3):647–50.

153. de Waard F. Hormonal factors in human carcinogenesis. J Cancer Res Clin Oncol. 1984;108(2):177-80.

154. Boyd N, Martin L, Stone J, Little L, Minkin S, Yaffe M. A longitudinal study of the effects of menopause on mammographic features. Cancer Epidemiology, Biomarkers & Prevention. 2002;11:1048–53.

155. Ghosh K, Hartmann LC, Reynolds C, Visscher DW, Brandt KR, Vierkant RA, et al. Association between mammographic density and age-related lobular involution of the breast. J Clin Oncol. 2010;28(13):2207-12.

156. Gram IT, Bremnes Y, Ursin G, Maskarinec G, Bjurstam N, Lund E. Percentage density, Wolfe's and Tabár's mammographic patterns: agreement and association with risk factors for breast cancer. Breast Cancer Research. 2000;7:R854.

157. Stone J, Warren RML, Pinney E, Warwick J, Cuzick J. Determinants of Percentage and Area Measures of Mammographic Density. American Journal of Epidemiology. 2009;170(12):1571-8.

158. Boyd NF, Lockwood GA, Byng JW, Little LE, Yaffe MJ, Tritchler DL. The relationship of anthropometric measures to radiological features of the breast in premenopausal women. British Journal of Cancer. 1998;78(9):1233-8.

159. Tseng M, Byrne C. Adiposity, adult weight gain and mammographic breast density in US Chinese women. Int J Cancer. 2011;128(2):418-25.

160. Reeves KW, Stone RA, Modugno F, Ness RB, Vogel VG, Weissfeld JL, et al. Longitudinal association of anthropometry with mammographic breast density in the Study of Women's Health Across the Nation. Int J Cancer. 2009;124(5):1169-77.

161. Hart V, Reeves KW, Sturgeon SR, Reich NG, Sievert LL, Kerlikowske K, et al. The effect of change in body mass index on volumetric measures of mammographic density. Cancer Epidemiol Biomarkers Prev. 2015;24(11):1724-30.

162. Dorgan JF, Klifa C, Shepherd JA, Egleston BL, Kwiterovich PO, Himes JH, et al. Height, adiposity and body fat distribution and breast density in young women. Breast Cancer Res. 2012;14(4):R107.

163. Jeffreys M, Warren R, Highnam R, Davey Smith G. Breast cancer risk factors and a novel measure of volumetric breast density: cross-sectional study. Br J Cancer. 2008;98(1):210-6.

164. Alimujiang A, Appleton C, Colditz GA, Toriola AT. Adiposity during early adulthood, changes in adiposity during adulthood, attained adiposity, and mammographic density among premenopausal women. Breast Cancer Res Treat. 2017;166(1):197-206.

165. Pereira A, Garmendia ML, Uauy R, Neira P, Lopez-Arana S, Malkov S, et al. Determinants of volumetric breast density in Chilean premenopausal women. Breast Cancer Res Treat. 2017;162(2):343-52.

166. Boyd NF, Martin LJ, Sun L, Guo H, Chiarelli A, Hislop G, et al. Body size, mammographic density, and breast cancer risk. Cancer Epidemiol Biomarkers Prev. 2006;15(11):2086-92.

167. Haars G, van Noord PAH, van Gils CH, Gobbee DE, Peeters PHM. Measurements of Breast Density: No Ratio for a Ratio. Cancer Epidemiol Biomarkers Prev. 2005;1411(11):2634–40.

168. Woolcott CG, Cook LS, Courneya KS, Boyd NF, Yaffe MJ, Terry T, et al. Associations of overall and abdominal adiposity with area and volumetric mammographic measures among postmenopausal women. Int J Cancer. 2011;129(2):440-8.

169. Harris HR, Tamimi RM, Willett WC, Hankinson SE, Michels KB. Body size across the life course, mammographic density, and risk of breast cancer. Am J Epidemiol. 2011;174(8):909-18.

170. Renehan AG, Tyson M, Egger M, Heller RF, M. Z. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. Lancet. 2008;371(9612):569-78.

171. Lam PB, Vacek PM, Geller BM, Muss HB. The association of increased weight, body mass index, and tissue density with the risk of breast carcinoma in Vermont. Cancer. 2000;89(2):369-75.

172. Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. BMJ. 2007;335(7630):1134.

173. Kyrgiou M, Kalliala I, Markozannes G, Gunter MJ, Paraskevaidis E, Gabra H, et al. Adiposity and cancer at major anatomical sites: umbrella review of the literature. BMJ. 2017:356:j477.

174. Sala E, Warren R, McCann J, Duffy S, Luben R, Day N. High-risk mammographic parenchymal patterns and anthropometric measures: a case-control study. British Journal of Cancer. 1999;81(7):1257–61.

175. Schoemaker MJ, Nichols HB, Wright LB, Brook MN, Jones ME, O'Brien KM, et al. Association of Body Mass Index and Age With Subsequent Breast Cancer Risk in Premenopausal Women. JAMA Oncol. 2018;4(11):e181771.

176. Bhaskaran K, Douglas I, Forbes H, dos Santos Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults. Lancet. 2014;384(9945):755-65.

177. Cancer Research UK. Anastrozole (Arimidex) [Internet] [cited 2017 04 March]. Available from: http://www.cancerresearchuk.org/about-cancer/breast-cancer/treatment/hormone-therapy/drugs/anastrozole-arimidex.

178. Bertrand KA, Eliassen AH, Hankinson SE, Rosner BA, Tamimi RM. Circulating Hormones and Mammographic Density in Premenopausal Women. Horm Cancer. 2018.

179. McCormack VA, Dowsett M, Folkerd E, Johnson N, Palles C, Coupland B, et al. Sex steroids, growth factors and mammographic density: a cross-sectional study of uk postmenopausal caucasian and afro-caribbean women. Breast Cancer Res. 2009;11(3):R38.

180. Tamimi RM, Hankinson SE, Colditz GA, Byrne C. Endogenous sex hormone levels and mammographic density among postmenopausal women. Cancer Epidemiology Biomarkers & Prevention. 2005;14(11):2641-7.

181. Tamimi RM, Byrne C, Colditz GA, Hankinson SE. Endogenous hormone levels, mammographic density, and subsequent risk of breast cancer in postmenopausal women. J Natl Cancer Inst. 2007;99(15):1178-87.

182. Vachon CM, Sasano H, Ghosh K, Brandt KR, Watson DA, Reynolds C, et al. Aromatase immunoreactivity is increased in mammographically dense regions of the breast. Breast Cancer Res Treat. 2011;125(1):243-52.

183. Diorio C, Pollak M, Byrne C, Masse B, Hebert-Croteau N, Yaffe M, et al. Insulin-like growth factor-I, IGF-binding protein-3, and mammographic breast density. Cancer Epidemiol Biomarkers Prev. 2005;14(5):1065-73.

184. Greendale GA, Huang MH, Ursin G, Ingles S, Stanczyk F, Crandall C, et al. Serum prolactin levels are positively associated with mammographic density in postmenopausal women. Breast Cancer Res Treat. 2007;105(3):337-46.

185. Gabrielson M, Ubhayasekera K, Ek B, Andersson Franko M, Eriksson M, Czene K, et al. Inclusion of Plasma Prolactin Levels in Current Risk Prediction Models of Premenopausal and Postmenopausal Breast Cancer. JNCI Cancer Spectrum. 2018;2(4).

186. Zhang X, Rice M, Tworoger SS, Rosner BA, AH E, Tamimi RM, et al. Addition of a polygenic risk score, mammographic density, and endogenous hormones to existing breast cancer risk prediction models: A nested case-control study. PLoS Med. 2018;15(9):e1002644.

187. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Lancet. 1997;350(9084):1047–59.

188. McTiernan A, Martin CF, Peck JD, Aragaki AK, Chlebowski RT, Pisano ED, et al. Estrogen-plus-progestin use and mammographic density in postmenopausal women: Women's Health Initiative randomized trial. J Natl Cancer Inst. 2005;97(18):1366-76.

189. Rutter CM, Mandelson MT, Laya MB, Taplin S. Changes in breast density associated with initiation, discontinuation, and continuing use of hormone replacement therapy. JAMA. 2001;285(2):171-6.

190. Warren R. Hormones and mammographic breast density. Maturitas. 2004;49 67-78.

191. Persson I, Thurfjell E, Holmberg L. Effect of estrogen and estrogen-progestin replacement regimens on mammographic breast parenchymal density. J Clin Oncol. 1997; 15:3201–7.

192. Vachon CM, Sellers TA, Vierkant RA, Wu FF, Brandt KR. Case-control study of increased mammographic breast density response to hormone replacement therapy. Cancer Epidemiology Biomarkers & Prevention. 2002;11(11):1382-8.

193. Greendale GA, Reboussin BA, Slone S, Wasilauskas C, Pike MC, Ursin G. Postmenopausal Hormone Therapy and Change in Mammographic Density. Journal of the National Cancer Institute. 2003;95(1):30-7.

194. Lundstrom E, Wilczek B, von Palffy Z, Soderqvist G, von Schoultz B. Mammographic breast density during hormone replacement therapy: Differences according to treatment. Am J Obstet Gynecol. 1999;181(2):348-52.

195. Harvey JA, Pinkerton JV, Herman CR. Short-term cessation of hormone replacement therapy and improvement of mammographic specificity. J Natl Cancer Inst. 1997;89:1623–5.

196. Colacurci N, Fornaro F, De Franciscis P, Mele D, Palermo M, del Vecchio W. Effects of a short-term suspension of hormone replacement therapy on mammographic density. Fertility and Sterility. 2001;76(3).

197. Beral V, Reeves G, Bull D, Green J. Breast cancer risk in relation to the interval between menopause and starting hormone therapy. J Natl Cancer Inst. 2011;103(4):296-305.
198. Cuzick J, Sestak I, Bonanni B, Costantino JP, Cummings S, DeCensi A, et al. Selective oestrogen receptor modulators in prevention of breast cancer: an updated meta-analysis of individual participant data. The Lancet. 2013;381(9880):1827-34.
199. Powles TJ, Ashley S, Tidy A, Smith IE, Dowsett M. Twenty-year follow-up of the Royal Marsden randomized, double-blinded tamoxifen breast cancer prevention trial. J Natl Cancer Inst. 2007;99(4):283-90.
200. Cuzick J, Sestak I, Cawthorn S, Hamed H, Holli K, Howell A, et al. Tamoxifen for prevention of breast cancer: extended long-term follow-up of the IBIS-I breast cancer prevention trial. The Lancet Oncology. 2015;16(1):67-75.
201. Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM, et al. Tamoxifen for Prevention of Breast Cancer: Report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. J Natl Cancer Inst. 1998;90(18):1371-88.
202. Cummings SR, Eckert S, Krueger KA, Grady D, Powles TJ, Cauley JA, et al. The effect of raloxifene on risk of breast cancer in postmenopausal women: results from the MORE randomized trial. Multiple Outcomes of Raloxifene Evaluation. JAMA. 1999;281(23):2189-97.
203. Cuzick J, Warwick J, Pinney E, Warren RML, Duffy SW. Tamoxifen and Breast Density in Women at Increased Risk of Breast Cancer. JNCI Journal of the National Cancer Institute. 2004;96(8):621-8.
204. Brisson J, Brisson B, Cote G, Maunsell E, Berube S, Robert J. Tamoxifen and mammographic breast densities. Cancer Epidemiology, Biomarkers & Prevention. 2000;9:911–5.
205. Atkinson C, Warren R, Bingham SA, Day NE. Mammographic patterns as a predictive biomarker of breast cancer risk: Effect of tamoxifen. Cancer Epidemiology, Biomarkers & Prevention. 1999;8:863–6.
206. Chow CK, Venzon D, Jones EC, Premkumar A, O'Shaughnessy J, Zujewski J. Effect of tamoxifen on mammographic density. Cancer Epidemiol Biomarkers Prev. 2000;9(9):917-21.
207. Cuzick J. Aromatase inhibitors for breast cancer prevention. J Clin Oncol. 2005;23(8):1636-43.
208. Cuzick J, Sestak I, Forbes JF, Dowsett M, Knox J, Cawthorn S, et al. Anastrozole for prevention of breast cancer in high-risk postmenopausal women (IBIS-II): an international, double-blind, randomised placebo-controlled trial. The Lancet. 2014;383(9922):1041-8.
209. Boyd NF, Dite GS, Stone J, Gunasekara A, English DR, McCredie MRE, et al. Heritability of mammographic density, a risk factor for breast cancer. N Engl J Med. 2002;347(12).

210. Huo Z, Giger ML, Olopade OI, Wolverton DE, Weber BL, Metz CE, et al. Computerized analysis of digitized mammograms of BRCA1 and BRCA2 gene mutation carriers. Radiology. 2002;225(2):519-26.

211. Mitchell G, Antoniou AC, Warren R, Peock S, Brown J, Davies R, et al. Mammographic Density and Breast Cancer Risk in BRCA1 and BRCA2 Mutation Carriers. Epidemiology and Prevention. 2006;66(3).

212. Vachon CM, Pankratz VS, Scott CG, Haeberle L, Ziv E, Jensen MR, et al. The contributions of breast density and common genetic variation to breast cancer risk. J Natl Cancer Inst. 2015;107(5).

213. Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 2007;447(7148):1087-93.

214. Odefrey F, Stone J, Gurrin LC, Byrnes GB, Apicella C, Dite GS, et al. Common genetic variants associated with breast cancer and mammographic density measures that predict disease. Cancer Res. 2010;70(4):1449-58.

215. Lindstrom S, Vachon CM, Li J, Varghese J, Thompson D, Warren R, et al. Common variants in ZNF365 are associated with both mammographic density and breast cancer risk. Nature Genetics. 2011;43:185–7.

216. Varghese JS, Thompson DJ, Michailidou K, Lindstrom S, Turnbull C, Brown J, et al. Mammographic breast density and breast cancer: evidence of a shared genetic basis. Cancer Res. 2012;72(6):1478-84.

217. Kelemen LE, Sellers TA, Vachon CM. Can genes for mammographic density inform cancer etiology? Nat Rev Cancer. 2008;8(10):812.

218. Dumas I, Diorio C. Polymorphisms in genes involved in the estrogen pathway and mammographic density. BMC Cancer. 2010;10:636.

219. Tamimi RM, Cox D, Kraft P, Colditz GA, Hankinson SE, Hunter DJ. Breast cancer susceptibility loci and mammographic density. Breast Cancer Res. 2008;10(4):R66.

220. Dumas I, Diorio C. Estrogen pathway polymorphisms and mammographic density. Anticancer Res. 2011;31(12):4369-86.

221. Vachon CM, Scott CG, Tamimi RM, Thompson DJ, Fasching PA, Stone J, et al. Joint association of mammographic density adjusted for age and body mass index and polygenic risk score with breast cancer risk. Breast cancer research. 2019;21(68).

222. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting Individualized Probabilities of Developing Breast Cancer for White Females Who Are Being Examined Annually. J Natl Cancer Inst. 1989;81(24):1879-86.

223. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. Stat Med. 2004;23(7):1111-30.

224. Tice JA, Bissell MCS, Miglioretti DL, Gard CC, Rauscher GH, Dabbous FM, et al. Validation of the breast cancer surveillance consortium model of breast cancer risk. Breast Cancer Res Treat. 2019;175(2):519-23.

225. Antoniou AC, Pharoah PPD, Smith S, Easton DF. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. British Journal of Cancer. 2004;91:1580–90.

226. Claus EB, Risch N, Thompson WD. The calculation of breast cancer risk for women with a first degree family history of ovarian cancer. Breast cancer research and treatment. 1993;28(2):115-20.

227. Ford D, Easton DF, Bishop DT, Narod SA, Goldgar DE. Risks of cancer in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. The Lancet. 1994;343(8899):692-5.

228. Costantino JP, Gail MH, Pee D, Anderson S, Redmond CK, Benichou J, et al. Validation studies for models projecting the risk of invasive and total breast cancer incidence. J Natl Cancer Inst. 1999;91:1541-8.

229. Amir E, Freedman OC, Seruga B, Evans DG. Assessing women at high risk of breast cancer: a review of risk assessment models. J Natl Cancer Inst. 2010;102:680–91.

230. Amir E, Evans DG, Shenton A, Lalloo F, Moran A, Boggis C, et al. Evaluation of breast cancer risk assessment packages in the family history evaluation and screening programme. J Med Genet. 2003;40:807–14.

231. Jacobi CE, de Bock GH, Siegerink B, van Asperen CJ. Differences and similarities in breast cancer risk assessment models in clinical practice: which model to choose? Breast Cancer Res Treat. 2009;115(2):381-90.

232. Cintolo-Gonzalez JA, Braun D, Blackford AL, Mazzola E, Acar A, Plichta JK, et al. Breast cancer risk models: a comprehensive overview of existing models, validation, and clinical applications. Breast Cancer Res Treat. 2017;164(2):263-84.

233. Terry MB, Liao Y, Whittemore AS, Leoce N, Buchsbaum R, Zeinomar N, et al. 10-year performance of four models of breast cancer risk: a validation study. The Lancet Oncology. 2019;20(4):504-17.

234. Quante AS, Whittemore AS, Shriver T, Strauch K, Terry MB. Breast cancer risk assessment across the risk continuum: genetic and nongenetic risk factors contributing to differential model performance. Breast Cancer Res. 2012;14(6):R144.

235. Brentnall AR, Harkness EF, Astley SM, Donnelly LS, Stavrinos P, Sampson S, et al. Mammographic density adds accuracy to both the Tyrer-Cuzick and Gail breast cancer risk models in a prospective UK screening cohort. Breast Cancer Research. 2015;17(147).

236. Santen RJ, Boyd NF, Chlebowski RT, Cummings S, Cuzick J, Dowsett M, et al. Critical assessment of new risk factors for breast cancer: considerations for development of an improved risk prediction model. Endocr Relat Cancer. 2007;14(2):169-87.

237. Tice JA, Cummings SR, Ziv E, Kerlikowske K. Mammographic Breast Density and the Gail Model for Breast Cancer Risk Prediction in a Screening Population. Breast Cancer Research and Treatment. 2005;94(2):115-22.

238. Darabi H, Czene K, Zhao W, Liu J, Hall P, Humphreys K. Breast cancer risk prediction and individualised screening based on common genetic variation and breast density measurement. Breast Cancer Research. 2012;14(R25).

239. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. Ann Intern Med. 2008;148(5):337-47.

240. Barlow WE, White E, Ballard-Barbash R, Vacek PM, Titus-Ernstoff L, Carney PA, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. J Natl Cancer Inst. 2006;98(17):1204-14.

241. Chen J, Pee D, Ayyagari R, Graubard B, Schairer C, Byrne C, et al. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. J Natl Cancer Inst. 2006;98(17):1215-26.

242. Warwick J, Birke H, Stone J, Warren RML, Pinney E, Brentnall AR, et al. Mammographic breast density refines Tyrer-Cuzick estimates of breast cancer risk in high-risk women: findings from the placebo arm of the International Breast Cancer Intervention Study I. Breast Cancer Research. 2014;16:451.

243. IBIS Breast Cancer Risk Evaluation Tool. [Software] Updated September 17, 2017 [cited 2018 23 Feb]. Available from: http://www.ems-trials.org/riskevaluator/.

244. Cuzick J, Brentnall A. Models for Assessment of Breast Cancer Risk. Breast Cancer, DI EUROPE. 2016:54-5.

245. Rice MS, Rosner BA, Tamimi RM. Percent mammographic density prediction: development of a model in the nurses' health studies. Cancer Causes Control. 2017;28(7):677-84.

246. Brentnall AR, Cohn WF, Knaus WA, Yaffe MJ, Cuzick J, Harvey JA. A Case-Control Study to Add Volumetric or Clinical Mammographic Density into the Tyrer-Cuzick Breast Cancer Risk Model. Journal of Breast Imaging. 2019;1(2):99–106.

247. Keum N, Greenwood DC, Lee DH, Kim R, Aune D, Ju W, et al. Adult weight gain and adiposity-related cancers: a dose-response meta-analysis of prospective observational studies. J Natl Cancer Inst. 2015;107(2).

248. Hardefeldt PJ, Penninkilampi R, Edirimanne S, Eslick GD. Physical Activity and Weight Loss Reduce the Risk of Breast Cancer: A Meta-analysis of 139 Prospective and Retrospective Studies. Clin Breast Cancer. 2018;18(4):e601-e12.

249. Winder AA, Kularatna M, MacCormick AD. Does Bariatric Surgery Affect the Incidence of Breast Cancer Development? A Systematic Review. Obes Surg. 2017;27(11):3014-20.

250. Harvie M, Howell A, Vierkant RA, Kumar N, Cerhan JR, Kelemen LE, et al. Association of Gain and Loss of Weight before and after Menopause with risk of postmenopausal breast cancer in the Iowa Women's Health Study. Cancer Epidemiol Biomarkers Prev. 2005;14(3):656–61.

251. Boyd NF, Greenberg C, Lockwood G, Little L, Martin L, Byng J, et al. Effects at Two Years of a Low-Fat, High-Carbohydrate Diet on Radiologic Features of the Breast: Results From a Randomized Trial. J Natl Cancer Inst. 1997;89(7):488-96.

252. Williams AD, So A, Synnestvedt M, Tewksbury CM, Kontos D, Hsiehm MK, et al. Mammographic breast density decreases after bariatric surgery. Breast Cancer Res Treat. 2017;165(3):565-72.

253. Vohra NA, Kachare SD, Vos P, Schroeder BF, Schuth O, Suttle D, et al. The Short-Term Effect of Weight Loss Surgery on Volumetric Breast Density and Fibroglandular Volume. Obes Surg. 2017;27(4):1013-23.

254. Salminen TM, Saarenmaa IE, Heikkilä MM, Hakama M. Risk of breast cancer and changes in mammographic parenchymal patterns over time. Acta Oncol. 1998;37(6):547-51.

255. Maskarinec G, Pagano I, Lurie G, Kolonel LN. A longitudinal investigation of mammographic density: the multiethnic cohort. Cancer Epidemiol Biomarkers Prev. 2006;15(4):732-9.

256. Vachon CM, Pankratz VS, Scott CG, Maloney SD, Ghosh K, Brandt KR, et al. Longitudinal trends in mammographic percent density and breast cancer risk. Cancer Epidemiol Biomarkers Prev. 2007;16(5):921-8.

257. van Gils CH, Hendriks JH, Holland R, Karssemeijer N, Otten JD, Straatman H, et al. Changes in mammographic breast density and concomitant changes in breast cancer risk. Eur J Cancer Prev. 1999;8(6):509-15.

258. Kerlikowske K, Ichikawa L, Miglioretti DL, Buist DS, Vacek PM, Smith-Bindman R, et al. Longitudinal measurement of clinical mammographic breast density to improve estimation of breast cancer risk. J Natl Cancer Inst. 2007;99(5):386-95.

259. Work ME, Reimers LL, Quante AS, Crew KD, Whiffen A, Terry MB. Changes in mammographic density over time in breast cancer cases and women at high risk for breast cancer. Int J Cancer. 2014;135(7):1740-4.

260. Lokate M, Stellato RK, Veldhuis WB, Peeters PH, van Gils CH. Age-related changes in mammographic density and breast cancer risk. Am J Epidemiol. 2013;178(1):101-9.

261. Roman M, Sala M, Bare M, Posso M, Vidal C, Louro J, et al. Changes in mammographic density over time and the risk of breast cancer: An observational cohort study. Breast. 2019;46:108-15.

262. Kerlikowske K, Gard CC, Sprague BL, Tice JA, Miglioretti DL, Breast Cancer Surveillance C. One versus Two Breast Density Measures to Predict 5- and 10-Year Breast Cancer Risk. Cancer Epidemiol Biomarkers Prev. 2015;24(6):889-97.

263. Ko KL, Shin IS, You JY, Jung SY, Ro J, Lee ES. Adjuvant tamoxifen-induced mammographic breast density reduction as a predictor for recurrence in estrogen receptor-positive premenopausal breast cancer patients. Breast Cancer Res Treat. 2013;142(3):559-67.
264. Kim J, Han W, Moon HG, Ahn S, Shin HC, You JM, et al. Breast density change as a predictive surrogate for response to adjuvant endocrine therapy in hormone receptor positive breast cancer. Breast Cancer Res. 2012;14(4):R102.
265. Li J, Humphreys K, Eriksson L, Edgren G, Czene K, Hall P. Mammographic density reduction is a prognostic marker of response to adjuvant tamoxifen therapy in postmenopausal patients with breast cancer. J Clin Oncol. 2013;31(18):2249-56.
266. Nyante SJ, Sherman ME, Pfeiffer RM, Berrington de Gonzalez A, Brinton LA, Aiello Bowles EJ, et al. Prognostic significance of mammographic density change after initiation of tamoxifen for ER-positive breast cancer. Journal of the National Cancer Institute. 2015;107(3).
267. Vachon CM, Suman VJ, Brandt KR, Kosel ML, Buzdar AU, Olson JE, et al. Mammographic breast density response to aromatase inhibition. Clin Cancer Res. 2013;19(8):2144-53.
268. Cigler T, Tu D, Yaffe MJ, Findlay B, Verma S, Johnston D, et al. A randomized, placebo-controlled trial (NCIC CTG MAP1) examining the effects of letrozole on mammographic breast density and other end organs in postmenopausal women. Breast Cancer Res Treat. 2010;120(2):427-35.
269. Cigler T, Richardson H, Yaffe MJ, Fabian CJ, Johnston D, Ingle JN, et al. A randomized, placebo-controlled trial (NCIC CTG MAP.2) examining the effects of exemestane on mammographic breast density, bone density, markers of bone metabolism and serum lipid levels in postmenopausal women. Breast Cancer Res Treat. 2011;126(2):453-61.
270. Vachon CM, Ingle JN, Suman VJ, Scott CG, Gottardt H, Olson JE, et al. Pilot study of the impact of letrozole vs. placebo on breast density in women completing 5 years of tamoxifen. Breast. 2007;16(2):204-10.
271. Engmann NJ, Scott CG, Jensen MR, Ma L, Brandt KR, Mahmoudzadeh AP, et al. Longitudinal Changes in Volumetric Breast Density with Tamoxifen and Aromatase Inhibitors. Cancer Epidemiol Biomarkers Prev. 2017;26(6):930-7.
272. Pettersson A, Hankinson SE, Willett WC, Lagiou P, Trichopoulos D, Tamimi RM. Nondense mammographic area and risk of breast cancer. Breast Cancer Res. 2011;13(5):R100.
273. van Kruijsdijk RC, van der Wall E, Visseren FL. Obesity and cancer: the role of dysfunctional adipose tissue. Cancer Epidemiol Biomarkers Prev. 2009;18(10):2569-78.
274. Travis RC, Key TJ. Oestrogen exposure and breast cancer risk. Breast Cancer Res. 2003;5:239-47.
275. Kim BK, Chang Y, Ahn J, Jung HS, Kim CW, Yun KE, et al. Metabolic syndrome, insulin resistance, and mammographic density in pre- and postmenopausal women. Breast Cancer Res Treat. 2015;153(2):425-34.

276. Amadou A, Ferrari P, Muwonge R, Moskal A, Biessy C, Romieu I, et al. Overweight, obesity and risk of premenopausal breast cancer according to ethnicity: a systematic review and dose-response meta-analysis. Obes Rev. 2013;14(8):665-78.

277. Harvie M, Cohen H, Mason C, Mercer T, Malik R, Adams J, et al. Adherence to a Diet and Exercise Weight Loss Intervention amongst Women at Increased Risk of Breast Cancer. The Open Obesity Journal. 2010;2:71-80.

278. Graffy HJ, Harvie MN, Warren RM, Boggis CR, Astley SM, Evans DG, et al. Abstract P4-13-09: The effect of weight change on breast adipose and dense tissue. Cancer Research. 2012;72(24 Supplement):P4-13-09-P4-13-09.

279. Patel HG, Astley SM, Hufton AP, Harvie M, Hagan K, Marchant TE, et al., editors. Automated Breast Tissue Measurement of Women at Increased Risk of Breast Cancer. Astley, SM, Brady, M, Rose, C, Zwiggelaar, R (eds) Digital Mammography IWDM Lecture Notes in Computer Science; 2006; Springer, Berlin, Heidelberg: vol 4046.

280. Diffey J, Hufton A, Astley S, editors. A new step-wedge for the volumetric measurement of mammographic density. Astley, SM, Brady, M, Rose, C Zwiggelaar, R (eds) Digital Mammography IWDM Lecture Notes in Computer Science; 2006; Springer, Berlin, Heidelberg: vol 4046.

281. Stevens J, Truesdale KP, McClain JE, Cai J. The definition of weight maintenance. Int J Obes (Lond). 2006;30(3):391-9.

282. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing [Software] 2015 [cited 2019 28 July]. Available from: https://www.R-project.org/.

283. Bland JM, Altman DG. Calculating correlation coefficients with repeated observations: Part 1 - Correlation within subjects. BMJ (Clinical research ed). 1995;310(6977):446.

284. Bland JM, Altman DG. Calculating correlation coefficients with repeated observations: Part 2 - Correlation between subjects. BMJ. 1995;310(6980):633.

285. Bakdash JZ, Marusich LR. Repeated Measures Correlation. Front Psychol. 2017;8:456.

286. Efron B, Tibshirani R. An Introduction to the Bootstrap. Boca Raton, FL: Chapman & Hall/CRC; 1993.

287. Crowder MJ, Hand DJ. Analysis of Repeated Measures (1st ed.), Monographs on Statistics and Applied Probability. London: Chapman & Hall; 1990.

288. Wilk MB, Gnanadesikan R. Probability plotting methods for the analysis of data. Biometrika. 1968;55(1):1-17.

289. Patterson HD, Thompson R. Recovery of inter-block information when block sizes are unequal. Biometrika. 1971;58(3):545–54.

290. O'Neill M. The mathematics of REML, A workshop conducted at Universitas Brawijaya, Malang, Indonesia [Internet]2013 [cited 2019 25 May]. Available from: www.stats.net.au.

291. Casella G, Berger RL. Statistical Inference (2nd ed.). CA, USA: Wadsworth Publishing, §8.2.1; 2002.

292. Wilks SS. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. Ann Math Statist. 1938;9(1):60-2.

293. Sung J, Song YM, Stone J, Lee K, Kim SY. Association of body size measurements and mammographic density in Korean women: the Healthy Twin study. Cancer Epidemiol Biomarkers Prev. 2010;19(6):1523-31.

294. Pollan M, Lope V, Miranda-Garcia J, Garcia M, Casanova F, Sanchez-Contador C, et al. Adult weight gain, fat distribution and mammographic density in Spanish pre- and post-menopausal women (DDM-Spain). Breast Cancer Res Treat. 2012;134(2):823-38.

295. Samimi G, Colditz GA, Baer HJ, Tamimi RM. Measures of energy balance and mammographic density in the Nurses' Health Study. Breast Cancer Res Treat. 2008;109(1):113-22.

296. Ekpo EU, Brennan PC, Mello-Thoms C, McEntee MF. Relationship Between Breast Density and Selective Estrogen-Receptor Modulators, Aromatase Inhibitors, Physical Activity, and Diet: A Systematic Review. Integrative Cancer Therapies. 2016;15(2):127–44.

297. Masala G, Assedi M, Sera F, Ermini I, Occhini D, Castaldo M, et al. Can dietary and physical activity modifications reduce breast density in postmenopausal women? The DAMA study, a randomized intervention trial in Italy. Cancer Epidemiol Biomarkers Prev. 2018.

298. Woolcott CG, Courneya KS, Boyd NF, Yaffe MJ, Terry T, McTiernan A, et al. Mammographic density change with 1 year of aerobic exercise among postmenopausal women: a randomized controlled trial. Cancer Epidemiol Biomarkers Prev. 2010;19(4):1112-21.

299. Magnusson C, Baron J, Persson I, Wolk A, Bergstrom R, Trichopoulos D, et al. Body size in different periods of life and breast cancer risk in post-menopausal women. Int J Cancer. 1998;76(29-34).

300. Eliassen AH, Colditz GA, Rosner B, Willett WC, Hankinson SE. Adult weight change and risk of postmenopausal breast cancer. JAMA. 2006;296(2):193-201.

301. Ahn J, Schatzkin A, Lacey Jr JV, Albanes D, Ballard-Barbash R, Adams KF, et al. Adiposity, adult weight change, and postmenopausal breast cancer risk. Arch Intern Med. 2007;167(19):2091-102.

302. Huang Z, Hankinson SE, Colditz GA, Stampfer MJ, Hunter DJ, Manson JE, et al. Dual effects of weight and weight gain on breast cancer risk. JAMA. 1997;278(17):1407-11.

303. Trentham-Dietz A, Newcomb PA, Egan KM, Titus-Ernstoff L, Baron JA, Storer BE, et al. Weight change and risk of postmenopausal breast cancer (United States). Cancer Causes Control. 2000;11(6):533-42.

304. Hugo HJ, Tourell MC, O'Gorman PM, Paige AE, Wellard RM, Lloyd T, et al. Looking beyond the mammogram to assess mammographic density: A narrative review. Biomedical Spectroscopy and Imaging. 2018;7(1-2):63-80.

305. Assi V, Massat NJ, Thomas S, MacKay J, Warwick J, Kataoka M, et al. A case-control study to assess the impact of mammographic density on breast cancer risk in women aged 40-49 at intermediate familial risk. Int J Cancer. 2015;136(10):2378-87.

306. Murphy S, Khaw KT, Sneyd MJ, Compston JE. Endogenous sex hormones and bone mineral density among community-based postmenopausal women. Postgrad Med J. 1992;68:908 -13.

307. Clemons M, Goss P. Estrogen and the Risk of Breast Cancer. N Engl J Med. 2001;344:276-85.

308. Kerlikowske K, Shepherd J, Creasman J, Tice JA, Ziv E, Cummings SR. Are breast density and bone mineral density independent risk factors for breast cancer? J Natl Cancer Inst. 2005;97(5):368-74.

309. Zain NM, Seriramulu VP, Chelliah KK. Bone Mineral Density and Breast Cancer Risk Factors among Premenopausal and Postmenopausal Women A Systematic Review. Asian Pac J Cancer Prev. 2016;17(7):3229-34.

310. Crandall CJ, Zheng Y, Karlamangla A, Sternfeld B, Habel LA, Oestreicher N, et al. The association between mammographic breast density and bone mineral density in the study of women's health across the nation. Ann Epidemiol. 2007;17(8):575-83.

311. Cuzick J. Bone versus breast density. Breast Cancer Res. 2006;8(3):104.

312. Diffey J, Hufton A, Beeston C, Smith J, Marchant T, Astley S, editors. Quantifying Breast Thickness for Density Measurement. International Workshop on Digital Mammography; 2008; Tucson, AZ, USA: Digital Mammography.

313. Shepherd J, Kerlikowske K. Do fatty breasts increase or decrease breast cancer risk? Breast Cancer Research. 2012;14(102).

314. Troy LM. The validity of recalled weight among younger women. Int J Obes Relat Metab Disord. 1995;19(8):570-2.

315. Brentnall AR, Cuzick J, Buist DSM, Bowles EJA. Long-term Accuracy of Breast Cancer Risk Assessment Combining Classic Risk Factors and Breast Density. JAMA Oncol. 2018:e180174.

316. StataCorp. Stata Statistical Software: Release 13 [Software]. College Station, TX: StataCorp LP. 2013.

317. Kelemen LE, Pankratz VS, Sellers TA, Brandt KR, Wang A, Janney C, et al. Age-specific trends in mammographic density: the Minnesota Breast Cancer Family Study. Am J Epidemiol. 2008;167(9):1027-36.

318. Huber PJ. The behavior of maximum likelihood estimates under nonstandard conditions. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics. Berkeley, Calif.: University of California Press; 1967.

319. White H. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. Econometrica. 1980;48(4):817-38.

320. White H. Maximum Likelihood Estimation of Misspecified Models. Econometrica. 1982;50(1):1-25.

321. Cox DR. Partial likelihood. biometrika. 1975;62(2):269-76.

322. Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. Parzen E, Tanabe K, Kitagawa G, editors. New York, NY: Springer; 1998.

323. Azam S, Eriksson M, Sjölander A, Hellgren R, Gabrielson M, Czene K, et al. Mammographic Density Change and Risk of Breast Cancer. JNCI: Journal of the National Cancer Institute. 2019;djz149.

324. Copas JB. Regression, Prediction and Shrinkage. Journal of the Royal Statistical Society, Series B. 1983;45(3):311–54.

325. McCormack VA, Perry NM, Vinnicombe SJ, dos Santos Silva I. Changes and tracking of mammographic density in relation to Pike's model of breast tissue aging: a UK longitudinal study. Int J Cancer. 2010;127(2):452-61.

326. Busana MC, De Stavola BL, Sovio U, Li J, Moss S, Humphreys K, et al. Assessing within-woman changes in mammographic density: a comparison of fully versus semi-automated area-based approaches. Cancer Causes Control. 2016;27(4):481-91.

327. Armero C, Forne C, Rue M, Forte A, Perpinan H, Gomez G, et al. Bayesian joint ordinal and survival modeling for breast cancer risk assessment. Stat Med. 2016;35(28):5267-82.

328. Van Vaerenbergh S, Santamaría I, Lázaro-Gredilla M, editors. Estimation of the forgetting factor in kernel recursive least squares. IEEE International Workshop on Machine Learning for Signal Processing; 2012; Santander, Spain.

329. Strand F, Humphreys K, Eriksson M, Li J, Andersson TML, Törnberg S, et al. Longitudinal fluctuation in mammographic percent density differentiates between interval and screen-detected breast cancer. International Journal of Cancer. 2016;140(1).

330. Crowther MJ, Andersson TM, Lambert PC, Abrams KR, Humphreys K. Joint modelling of longitudinal and survival data: incorporating delayed entry and an assessment of model misspecification. Stat Med. 2016;35(7):1193–209.

331. Andersson TML, Crowther MJ, Czene K, Hall P, Humphreys K. Mammographic density reduction as a prognostic marker for postmenopausal breast cancer; results using a joint longitudinal-survival modelling approach. Am J Epidemiol. 2017;186(9):1065-73.

332. Davies C, Godwin J, Gray R, Clarke M, Cutter D, Darby S, et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. Lancet. 2011;378(9793):771-84.

333. Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. Lancet. 1998;351(9114):1451-67.

334. Visvanathan K, Hurley P, Bantug E, Brown P, Col NF, Cuzick J, et al. Use of pharmacologic interventions for breast cancer risk reduction: American Society of Clinical Oncology clinical practice guideline. J Clin Oncol. 2013;31(23):2942-62.

335. Early Breast Cancer Trialists' Collaborative Group. Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. Lancet. 2015;386(10001):1341-52.

336. Nyante SJ, Sherman ME, Pfeiffer RM, Berrington de Gonzalez A, Brinton LA, Bowles EJ, et al. Longitudinal Change in Mammographic Density among ER-Positive Breast Cancer Patients Using Tamoxifen. Cancer Epidemiol Biomarkers Prev. 2016;25(1):212-6.

337. Eriksson L, He W, Eriksson M, Humphreys K, Bergh J, Hall P, et al. Adjuvant Therapy and Mammographic Density Changes in Women With Breast Cancer. JNCI Cancer Spectrum. 2018;2(4).

338. Byrne C, Ursin G, Martin CF, Peck JD, Cole EB, Zeng D, et al. Mammographic Density Change With Estrogen and Progestin Therapy and Breast Cancer Risk. J Natl Cancer Inst. 2017;109(9).

339. Martin LJ, Minkin S, Boyd NF. Hormone therapy, mammographic density, and breast cancer risk. Maturitas. 2009;64(1):20-6.

340. Jordan VC. New insights into the metabolism of tamoxifen and its role in the treatment and prevention of breast cancer. Steroids. 2007;72(13):829-42.

341. Altman DG. Systematic reviews of evaluations of prognostic variables. BMJ. 2001;323(7306):224-8.

342. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. J Clin Oncol. 2011;29(32):4273-8.

343. Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J Clin Oncol. 2007;25(33):5287-312.

344. Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. BMJ. 2013;346:e5793.

345. Shawky MS, Martin H, Hugo HJ, Lloyd T, Britt KL, Redfern A, et al. Mammographic density: a potential monitoring biomarker for adjuvant and preventative breast cancer endocrine therapies. Oncotarget. 2017;8(3):5578-91.

346. Kanbayti IH, Rae WID, McEntee MF, Ekpo EU. Are mammographic density phenotypes associated with breast cancer treatment response and clinical outcomes? A systematic review and meta-analysis. The Breast. 2019;47:62-76.

347. Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. PLoS Med. 2013;10(2):e1001380.

348. Mullooly M, Pfeiffer RM, Nyante SJ, Heckman-Stoddard BM, Perloff M, Jatoi I, et al. Mammographic Density as a Biosensor of Tamoxifen Effectiveness in Adjuvant Endocrine Treatment of Breast Cancer: Opportunities and Implications. Journal of Clinical Oncology. 2016;34(18):2093–7.

349. Cuzick J, Sestak I, Baum M, Buzdar A, Howell A, Dowsett M, et al. Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 10-year analysis of the ATAC trial. The Lancet Oncology. 2010;11(12):1135-41.

350. Kim WH, Cho N, Kim YS, Yi A. Mammographic density changes following discontinuation of tamoxifen in premenopausal women with oestrogen receptor-positive breast cancer. Eur Radiol. 2018;28(8):3176-84.

351. Howell A, Ashcroft L, Fallowfield L, Eccles DM, Eeles RA, Ward A, et al. RAZOR: a phase II open randomized trial of screening plus goserelin and raloxifene versus screening alone in pre-menopausal women at increased risk of breast cancer. Cancer Epidemiol Biomarkers Prev. 2018;27(1):58-66.

352. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009;339:b2700.

353. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. PLoS Med. 2012;9(5):e1001216.

354. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting Recommendations for Tumor Marker Prognostic Studies. Journal of Clinical Oncology. 2005;23(36):9067-72.

355. Riley RD, Sauerbrei W, Altman DG. Prognostic markers in cancer: the evolution of evidence from single studies to meta-analysis, and beyond. Br J Cancer. 2009;100(8):1219-29.

356. Atakpa EC, Thorat MA, Cuzick J, Brentnall AR. Mammographic density, endocrine therapy and breast cancer risk: A prognostic and predictive biomarker review. Cochrane Database of Systematic Reviews. 2018;Issue 8(CD013091).

357. Komm BS, Mirkin S. An overview of current and emerging SERMs. J Steroid Biochem Mol Biol. 2014;143:207-22.

358. Wolfe JN. Breast Patterns as an Index of Risk for Developing Breast Cancer. AJR Am J Roentgenol. 1976;126(6):1130-7.

359. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8(2):135-60.

360. Schunemann HJ, Oxman AD, Vist GE, Higgins JP, Deeks JJ, Glaziou P, et al. Chapter 12: Interpreting results and drawing conclusions. Cochrane Handbook for Systematic Reviews of Interventions, Version 510 (updated March 2011), Higgins JP and Green S (editors), handbookcochraneorg. 2011.

361. GRADEpro GDT: GRADEpro Guideline Development Tool [Software] 2015. McMaster University (developed by Evidence Prime, Inc.), Accessed 28 July 2019, gradepro.org.

362. Cochrane Breast Cancer Group. Available from: breastcancer.cochrane.org/specialised-register.

363. World Health Organization International Clinical Trials Registry Platform. Available from: apps.who.int/trialsearch.

364. ClinicalTrials.gov. Available from: ClinicalTrials.gov.

365. Review Manager 5 (RevMan 5) Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration. 2014.

366. Covidence systematic review software. Veritas Health Innovation [Software] 2018 Melbourne, Australia, www.covidence.org, Accessed 2019 28 July.

367. Altman DG, Lyman GH. Methodological challenges in the evaluation of prognostic factors in breast cancer. Breast Cancer Res Treat. 1998;52(1-3):289-303.

368. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. Annals of internal medicine. 2013;158(4):280-6.

369. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Annals of internal medicine. 2006;144(6):427-37.

370. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.

371. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629-34.

372. Knight JA, Blackmore KM, Fan J, Malone KE, John EM, Lynch CF, et al. The association of mammographic density with risk of contralateral breast cancer and change in density with treatment in the WECARE study. Breast Cancer Research. 2018;20(1):23.

373. Sandberg ME, Li J, Hall P, Hartman M, dos Santos Silva I, Humphreys K, et al. Change of mammographic density predicts the risk of contralateral breast cancer-a case-control study. Breast Cancer Res. 2013;15(4):R57.

374. van Nes JG, Beex LV, Seynaeve C, Putter H, Sramek A, Lardenoije S, et al. Minimal impact of adjuvant exemestane or tamoxifen treatment on mammographic breast density in postmenopausal breast cancer patients: a Dutch TEAM trial analysis. Acta Oncol. 2015;54(3):349-60.

375. van de Velde CJ, Rea D, Seynaeve C, Putter H, Hasenburg A, Vannetzel JM, et al. Adjuvant tamoxifen and exemestane in early breast cancer (TEAM): a randomised phase 3 trial. Lancet. 2011;377(9762):321-31.

376. Alliance for Clinical Trials in Oncology, National Cancer Institute, NCIC Clinical Trials Group. Breast Density, Hormone Levels, and Anticancer Drug Levels in Women With Invasive Breast Cancer Who Are Receiving Exemestane or Anastrozole. https://ClinicalTrials.gov/show/NCT00316836; 2006.

377. Alliance for Clinical Trials in Oncology, National Cancer Institute. Changes in Breast Density and Blood Hormone Levels in Postmenopausal Women Receiving Anastrozole or Exemestane for Breast Cancer. https://ClinicalTrials.gov/show/NCT00516698; 2007.

378. Becker S, Kaaks R. Exogenous and endogenous hormones, mammographic density and breast cancer risk: can mammographic density be considered an intermediate marker of risk? Recent Results Cancer Res. 2009;181:135-57.

379. Boyd NF. Tamoxifen, mammographic density, and breast cancer prevention. J Natl Cancer Inst. 2011;103(9):704-5.

380. Case Comprehensive Cancer Centre, National Cancer I. Changes in Breast Density and Breast Cancer Risk in Women With Breast Cancer and in Healthy Women. 2007.

381. Chlebowski RT, McTiernan A. Biological significance of interventions that change breast density. Journal of the National Cancer Institute. 2003;95(1):4-5.

382. Cosmacini P, Cassano E, Sacchini V, Coopmans de Yoldi G, Costa A. Chemoprevention of breast cancer by using tamoxifen: radiologic evaluation. Radiology. 1993;189(Suppl):245.

383. Cuzick J. Breast density predicts endocrine treatment outcome in the adjuvant setting. Breast Cancer Res. 2012;14(4):109.

384. Decensi A, Bonanni B, Guerrieri GA, Robertson C, Cazzaniga M, Mariette F, et al. A randomized 2x2 biomarker trial of low-dose tamoxifen and fenretinide in postmenopausal women at high-risk for breast cancer. Annual Meeting Proceedings of the American Society of Clinical Oncology. 2004:97.

385. Decensi A, Robertson C, Guerrieri-Gonzaga A, Serrano D, Cazzaniga M, Mora S, et al. Randomized double-blind 2 x 2 trial of low-dose tamoxifen and fenretinide for breast cancer prevention in high-risk premenopausal women. Journal of Clinical Oncology. 2009;27(23):3749-56.

386. Fabian CJ, Kimler BF. Mammographic density: Use in risk assessment and as a biomarker in prevention trials. Journal of Nutrition. 2006;136(10):2705S-8S.

387. Fabian CJ, Kimler BF. Use of biomarkers for breast cancer risk assessment and prevention. J Steroid Biochem Mol Biol. 2007;106(1-5):31-9.

388. Fabian CJ, Kimler BF. Incorporating biomarkers in studies of chemoprevention. Advances in Experimental Medicine and Biology. 2016;882:69-94.

389. Ghosh K, Vachon CM. Mammographic breast density, endocrine therapies, and breast cancer risk. Sexuality, Reproduction and Menopause. 2010;8(1):34-9.

390. Kim JY, Cho N, Jeyanth JX, Kim WH, Lee SH, Gweon HM, et al. Smaller reduction in 3D breast density associated with subsequent cancer recurrence in patients with breast cancer receiving adjuvant tamoxifen therapy. AJR American Journal of Roentgenology. 2014;202(4):912-21.

391. Kmietowicz Z. Breast density on mammography can help assess response to tamoxifen and predict survival, finds study. BMJ. 2013;346:f2645.

392. Macis D, Gandini S, Guerrieri-Gonzaga A, Harriet J, Serrano D, Cazzaniga M, et al. Associations between a polymorphic locus in the promoter region of the IGFBP-3 gene, circulating IGFs, mammographic breast density and survival in a 2x2 trial of low-dose tamoxifen and fenretinide in premenopausal women 6211. Cancer Research. 2011;71(8).

393. Martin HL, Yap F, Chung K, Stone J, Redfern A. Mammographic breast-density change as a predictor of outcome in hormone receptor positive breast cancer. Asia-Pacific Journal of Clinical Oncology. 2016;12 (Supplement 5):72.

394. NCIC Clinical Trials Group, Canadian Cancer Trials Group. Letrozole in Preventing Breast Cancer in Postmenopausal Women Who Are at Increased Risk for Breast Cancer Due to High Breast Density. https://ClinicalTrials.gov/show/NCT00238316; 2000.

395. NCIC Clinical Trials Group, Canadian Cancer Trials Group. Exemestane in Reducing Breast Density in Postmenopausal Women at Risk for Breast Cancer. https://ClinicalTrials.gov/show/NCT00066586; 2001.

396. Northwestern University, National Cancer Institute. Effect of Tamoxifen on Breast Density in Premenopausal Women With Breast Cancer or High Risk for Breast Cancer. https://ClinicalTrials.gov/show/NCT00086749; 2003.

397. Ozhand A, McKean-Cowdin R, Bernstein L, Ballard-Babash R, McTiernan A, Baumgartner KB. Short term reduction in mammographic density predicts survival in breast cancer. Cancer Research Conference: 104th Annual Meeting of the American Association for Cancer Research, AACR. 2013;73(8 SUPPL. 1).

398. Redfern A. Breast density change: Toward tailoring of adjuvant endocrine therapy. Asia-Pacific Journal of Clinical Oncology. 2016;12 (Supplement 5):100.

399. Redfern AD, Martin HL, Stone J, Davidson JA, Yap F, Chung K. Mammographic breast density as a predictor of hormone receptor positive breast cancer recurrence: A single centre longitudinal analysis. Cancer Research Conference: 38th Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Publication:. 2016;76(4 SUPPL. 1).

400. Seoul National University Hospital. Breast Density Change Predicting Response to Adjuvant Aromatase Inhibitor. https://ClinicalTrials.gov/show/NCT01765049; 2013.

401. Seoul National University Hospital. Density Lowering Effect of "OFS Add on to TMX"(DELFINO Trial). https://ClinicalTrials.gov/show/NCT03664895; 2018.

402. University of California Irvine, University of California San Francisco, Beckman Laser Institute University of California Irvine. Development of a Quantitative Tissue Optical Index of Breast Density For Prediction of Hormone Therapy Response. 2013.

403. University of Virginia, National Cancer Institute. Letrozole With or Without Zoledronate in Treating Healthy Postmenopausal Women With High Breast Density. https://ClinicalTrials.gov/show/NCT00114270; 2004.

404. Ursin G, Pike MC, Spicer DV, Porrath SA, Reitherman RW. Can mammographic densities predict effects of tamoxifen on the breast? Journal of the National Cancer Institute. 1996;88(2):128-9.

405. Whitman GJ, Stelling CB, Sunku K. Mammographic changes by endocrine therapy. Seminars in Breast Disease. 2000;3(2):70-80.

406. Fontein DB, Seynaeve C, Hadji P, Hille ET, van de Water W, Putter H, et al. Specific adverse events predict survival benefit in patients treated with tamoxifen or aromatase inhibitors: an international tamoxifen exemestane adjuvant multinational trial analysis. J Clin Oncol. 2013;31(18):2257-64.

407. Cuzick J, Sestak I, Cella D, Fallowfield L. Treatment-emergent endocrine symptoms and the risk of breast cancer recurrence: a retrospective analysis of the ATAC trial. The Lancet Oncology. 2008;9(12):1143-8.

408. Early Breast Cancer Trialists' Collaborative Group. Systematic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. Lancet. 1992;339:1-15, 71-85.

409. Cuzick J, Baum M. Tamoxifen and contralateral breast cancer. Lancet. 1985;2(8449):282.

410. Cuzick J, Forbes JF, Sestak I, Cawthorn S, Hamed H, Holli K, et al. Long-term results of tamoxifen prophylaxis for breast cancer - 96-month follow-up of the randomized IBIS-I trial. J Natl Cancer Inst. 2007;99(4):272-82.

411. The Arimidex Tamoxifen Alone or in Combination (ATAC) Trialists' Group. Comprehensive side-effect profile anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: long-term safety analysis of the ATAC trial. Lancet Oncol. 2006;7(8):633-43.

412. Varma DR. Managing DICOM images: Tips and tricks for the radiologist. The Indian journal of radiology & imaging. 2012;22(1):4-13.

413. Santesoft Ltd. Sante DICOM Editor [Software]. Athens, Greece2018.

414. Stone J, Gunasekara A, Martin LJ, Yaffe M, Minkin S, Boyd NF. The detection of change in mammographic density. Cancer Epidemiol Biomarkers Prev. 2003;12(7):625-30.

415. Mousa NA, Crystal P, Wolfman WL, Bedaiwy MA, Casper RF. Aromatase inhibitors and mammographic breast density in postmenopausal women receiving hormone therapy. Menopause. 2008;15(5):875-84.

416. Eriksson M, Li J, Leifland K, Czene K, Hall P. A comprehensive tool for measuring mammographic density changes over time. Breast Cancer Res Treat. 2018;169(2):371-9.

417. Agresti A. Categorical data analysis (2nd ed.). New York: John Wiley & Sons Ltd; 2002.

418. Mullooly M, Gierach GL. The Potential for Mammographic Breast Density Change as a Biosensor of Adjuvant Tamoxifen Therapy Adherence and Response. JNCI Cancer Spectr. 2018;2(4):pky072.

419. Cuzick J. Progress in preventive therapy for cancer: a reminiscence and personal viewpoint. Br J Cancer. 2018;118(9):1155-61.

420. Evans DG, Donnelly LS, Harkness EF, Astley SM, Stavrinos P, Dawe S, et al. Breast cancer risk feedback to women in the UK NHS breast screening population. Br J Cancer. 2016;114(9):1045-52.

421. Howell A, Astley S, Warwick J, Stavrinos P, Sahin S, Ingham S, et al. Prevention of breast cancer in the context of a national breast screening programme. J Intern Med. 2012;271(4):321-30.

422. National Institute for Health and Clinical Excellence (NICE) Clinical guideline [CG164]. Familial breast cancer: classification, care and managing breast cancer and related risks in people with a family history of breast cancer [Internet] 2013 [updated March 2017; cited 2017 22 April]. Available from: http://www.nice.org.uk/guidance/CG164/chapter/1-Recommendations.

423. Evans DG, Graham J, O'Connell S, Arnold S, Fitzsimmons D. Familial breast cancer: summary of updated NICE guidance. BMJ. 2013;346:f3829.

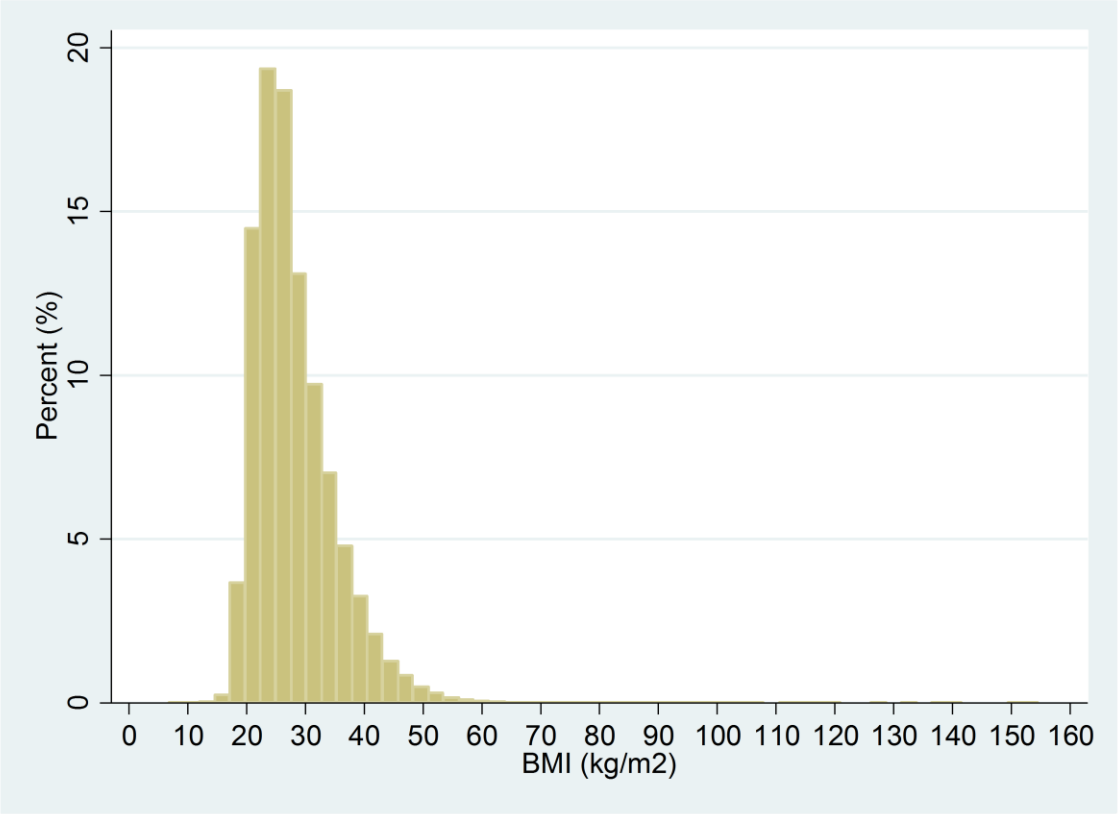
424. Beral V, Bull D, Doll R, Key T, Peto R, Reeves G, et al. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52 705 women with breast cancer and 108 411 women without breast cancer. The Lancet. 1997;350(9084):1047-59.

425. Cuzick J, Forbes J, Edwards R, Baum M, Cawthorn S, Coates A, et al. First results from the International Breast Cancer Intervention Study (IBIS-I): a randomised prevention trial. The Lancet. 2002;360(9336):817-24.

426. Prowell TM, Blackford AL, Byrne C, Khouri NF, Dowsett M, Folkerd E, et al. Changes in breast density and circulating estrogens in postmenopausal women receiving adjuvant anastrozole. Cancer Prev Res (Phila). 2011;4(12):1993-2001.

Appendix A: Supplementary material for Chapter 3

A.I Histogram of body mass index



Before imputation, before winsorising.

47,390 women had one or more mammograms with missing body mass index (BMI):

- 3,409 women had only baseline mammogram with missing BMI (only baseline imputed)
- 2,638 women had baseline mammogram with missing BMI (baseline imputed) and at least one follow-up mammogram with missing BMI (carried forward)
- 41,343 women had only follow-up mammograms with missing BMI (carried forward)

A.II Table for number and frequency of mammograms per woman (by age at baseline group)

Age at baseline (yr)	Measure	Median (IQR)			
		Time (yr)			
		0-5	5-10	10-15	15+
All	Number of mammograms (except baseline) per woman	1 (0-2)	2 (1-3)	2 (2-3)	2 (1-2)
	Frequency of mammograms (per yr) per woman	0.4 (0-0.6)	0.6 (0.4-0.8)	0.6 (0.4-0.8)	0.6 (0.4-0.9)
40-44	Number of mammograms (except baseline) per woman	1 (0-2)	2 (1-3)	2 (2-3)	2 (1-2)
	Frequency of mammograms (per yr) per woman	0.4 (0-0.6)	0.6 (0.4-0.8)	0.6 (0.4-0.8)	0.6 (0.5-0.9)
45-49	Number of mammograms (except baseline) per woman	1 (0-2)	2 (2-3)	3 (2-4)	2 (1-2)
	Frequency of mammograms (per yr) per woman	0.4 (0-0.6)	0.6 (0.4-0.8)	0.6 (0.4-0.8)	0.6 (0.4-0.9)
50-54	Number of mammograms (except baseline) per woman	2 (1-2)	2 (2-3)	2 (2-4)	2 (1-2)
	Frequency of mammograms (per yr) per woman	0.4 (0.2-0.6)	0.6 (0.4-0.8)	0.6 (0.4-0.8)	0.6 (0.4-0.9)
55-59	Number of mammograms (except baseline) per woman	2 (0-2)	2 (2-3)	2 (2-3)	1 (1-2)
	Frequency of mammograms (per yr) per woman	0.4 (0-0.6)	0.6 (0.4-0.8)	0.6 (0.4-0.8)	0.6 (0.4-0.9)
60-64	Number of mammograms (except baseline) per woman	2 (0-2)	3 (2-3)	1 (1-2)	-
	Frequency of mammograms (per yr) per woman	0.4 (0-0.6)	0.6 (0.4-0.8)	0.6 (0.4-0.9)	-
65+	Number of mammograms (except baseline) per woman	2 (1-2)	2 (1-2)	-	-
	Frequency of mammograms (per yr) per woman	0.4 (0.2-0.6)	0.6 (0.5-0.9)	-	-

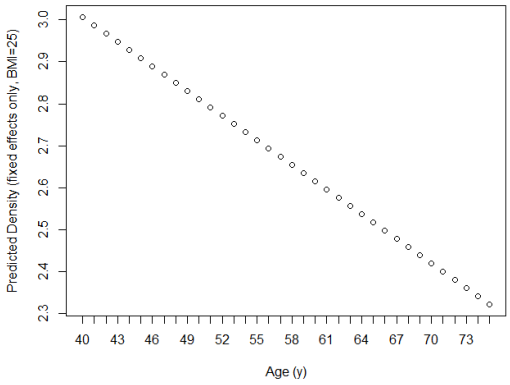
Frequency of mammograms is measured over the time at-risk for each woman in each time period.

A.III Plots of predicted density against age at baseline for a woman with body mass index 25kg/m²

Likelihood ratio tests for model compared with previous model:

Linear age

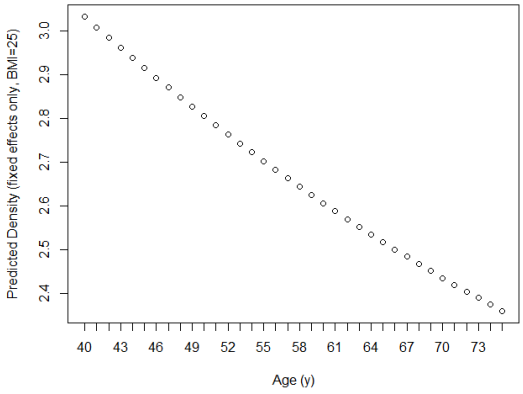
$$y_{ij} = \beta_0 + u_{0i} + \beta_1 BMI_{ij} + (\beta_2 + u_{1i}) age_{ij} + e_{ij}$$



Quadratic age

$$y_{ij} = \beta_0 + u_{0i} + \beta_1 BMI_{ij} + (\beta_2 + u_{1i}) age_{ij} + \beta_3 age_{ij}^2 + e_{ij}$$

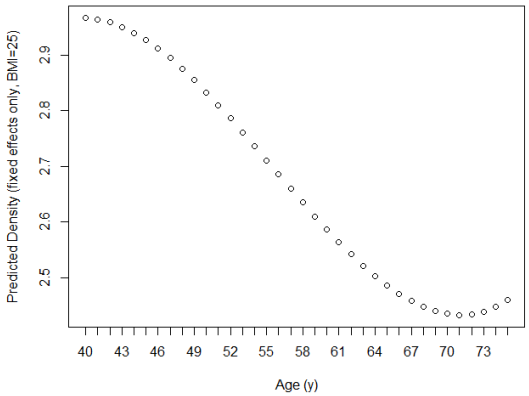
-> $\Delta LR\text{-}\chi^2(1)=140.9$, $p=1.7x10^{-32}$



Cubic age

$$y_{ij} = \beta_0 + u_{0i} + \beta_1 BMI_{ij} + (\beta_2 + u_{1i}) age_{ij} + \beta_3 age_{ij}^2 + \beta_4 age_{ij}^3 + e_{ij}$$

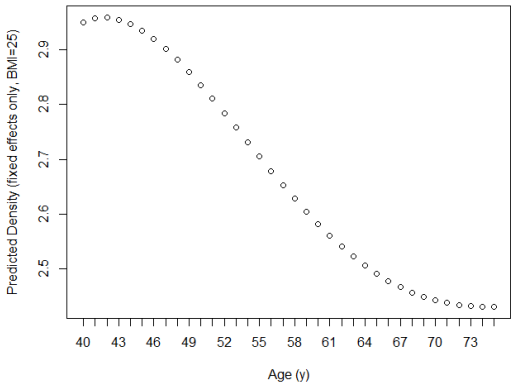
-> $\Delta LR\text{-}\chi^2(1)=1050.4$, $p=2.0x10^{-230}$



Quartic age

$$y_{ij} = \beta_0 + u_{0i} + \beta_1 BMI_{ij} + (\beta_2 + u_{1i}) age_{ij} + \beta_3 age_{ij}^2 + \beta_4 age_{ij}^3 + \beta_5 age_{ij}^4 + e_{ij}$$

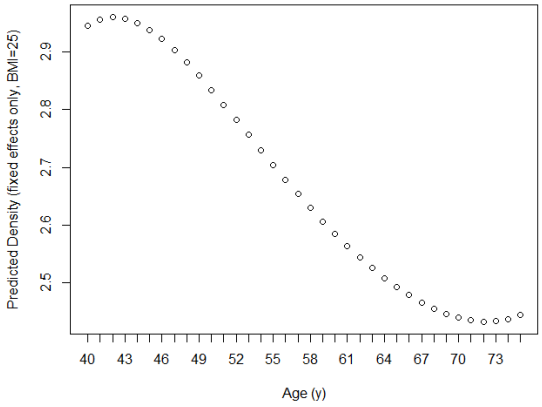
-> ΔLR-$\chi^2(1)$=77.1, p=1.6x10^{-18}



Age to the power of 5

$$y_{ij} = \beta_0 + u_{0i} + \beta_1 BMI_{ij} + (\beta_2 + u_{1i}) age_{ij} + \beta_3 age_{ij}^2 + \beta_4 age_{ij}^3 + \beta_5 age_{ij}^4 + \beta_6 age_{ij}^5 + e_{ij}$$

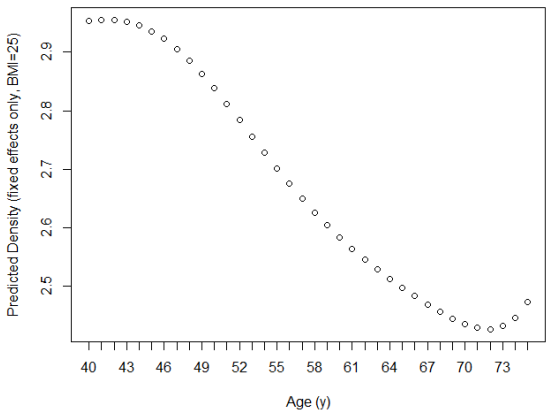
-> ΔLR-$\chi^2(1)$=9.6, p=0.002



Age to the power of 6

$$y_{ij} = \beta_0 + u_{0i} + \beta_1 BMI_{ij} + (\beta_2 + u_{1i}) age_{ij} + \beta_3 age_{ij}^2 + \beta_4 age_{ij}^3 + \beta_5 age_{ij}^4 + \beta_6 age_{ij}^5 + \beta_7 age_{ij}^6 + e_{ij}$$

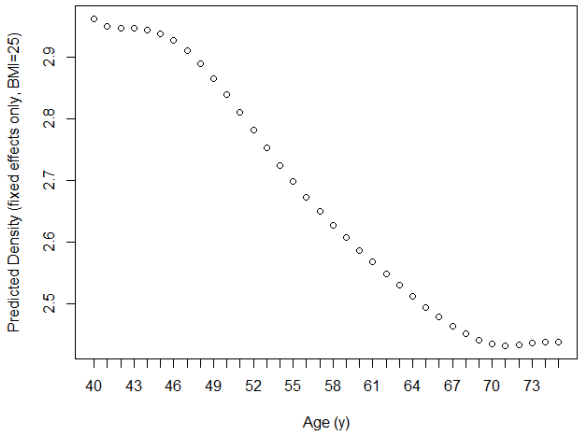
-> ΔLR-$\chi^2(1)$=33.3, p=8.1x10^{-09}



Age to the power of 7

$$y_{ij} = \beta_0 + u_{0i} + \beta_1 BMI_{ij} + (\beta_2 + u_{1i})age_{ij} + \beta_3 age_{ij}^2 + \beta_4 age_{ij}^3 + \beta_5 age_{ij}^4 + \beta_6 age_{ij}^5 + \beta_7 age_{ij}^6 + \beta_8 age_{ij}^7 + e_{ij}$$

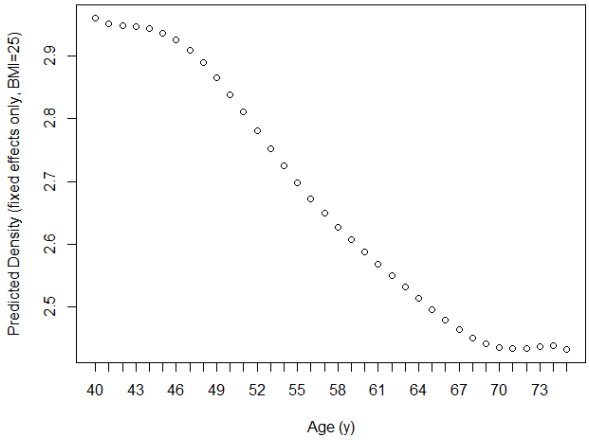
-> ΔLR-$\chi^2(1)$=31.9, p=1.6x10^{-08}



Age to the power of 8

$$y_{ij} = \beta_0 + u_{0i} + \beta_1 BMI_{ij} + (\beta_2 + u_{1i})age_{ij} + \beta_3 age_{ij}^2 + \beta_4 age_{ij}^3 + \beta_5 age_{ij}^4 + \beta_6 age_{ij}^5 + \beta_7 age_{ij}^6 + \beta_8 age_{ij}^7 + \beta_9 age_{ij}^8 + e_{ij}$$

-> ΔLR-$\chi^2(1)$=0.6, p=0.43

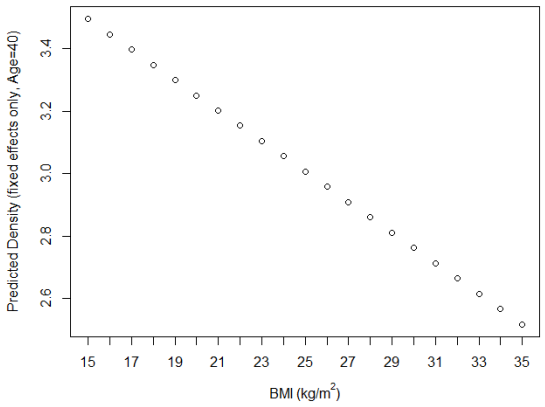


A.IV Plots of predicted density against body mass index for a woman 40 years old at baseline

Likelihood ratio tests for model compared with previous model:

Linear BMI

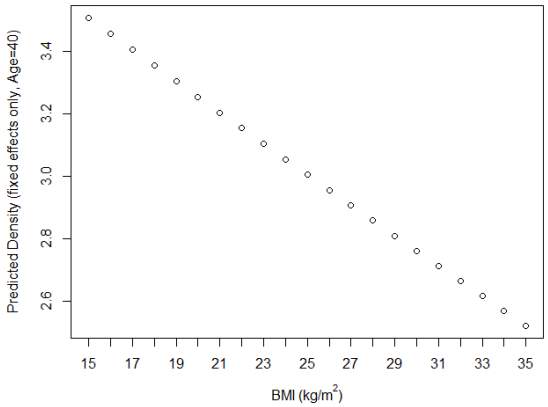
$$y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i})age_{ij} + \beta_2 BMI_{ij} + e_{ij}$$



Quadratic BMI

$$y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i})age_{ij} + \beta_2 BMI_{ij} + \beta_3 BMI_{ij}^2 + e_{ij}$$

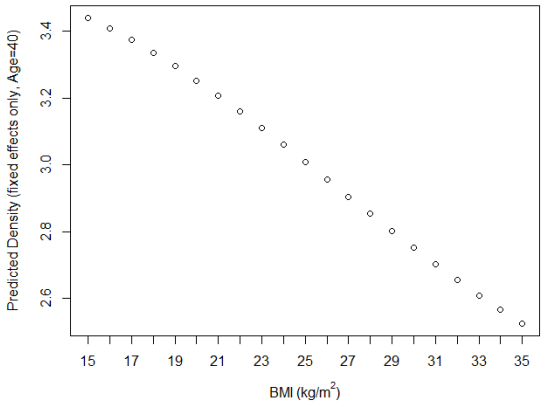
-> ΔLR-$\chi^2(1)$=3.6, p=0.06



Cubic BMI

$$y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i})age_{ij} + \beta_2 BMI_{ij} + \beta_3 BMI_{ij}^2 + \beta_4 BMI_{ij}^3 + e_{ij}$$

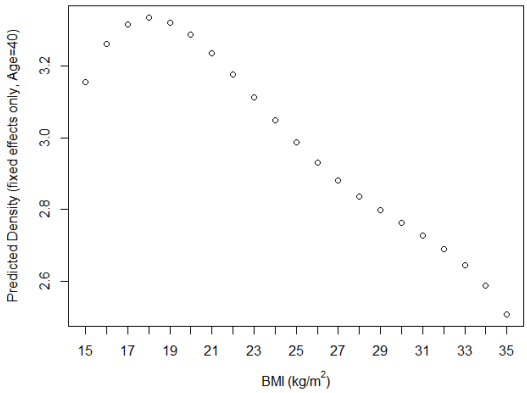
-> ΔLR-$\chi^2(1)$=43.1, p=5.1x10^{-11}



Quartic BMI

$$y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i})age_{ij} + \beta_2 BMI_{ij} + \beta_3 BMI_{ij}^2 + \beta_4 BMI_{ij}^3 + \beta_5 BMI_{ij}^4 + e_{ij}$$

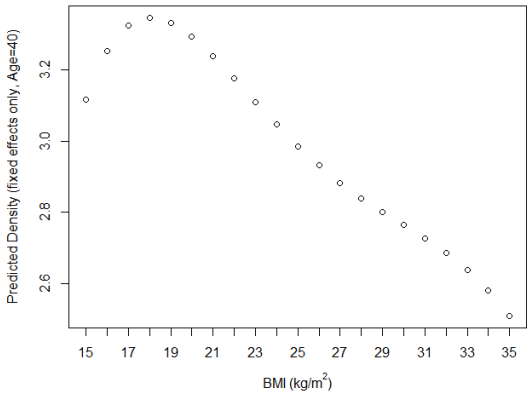
-> ΔLR-$\chi^2(1)$=655.5, p=1.4x10^{-144}



BMI to the power of 5

$$y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i})age_{ij} + \beta_2 BMI_{ij} + \beta_3 BMI_{ij}^2 + \beta_4 BMI_{ij}^3 + \beta_5 BMI_{ij}^4 + \beta_6 BMI_{ij}^5 + e_{ij}$$

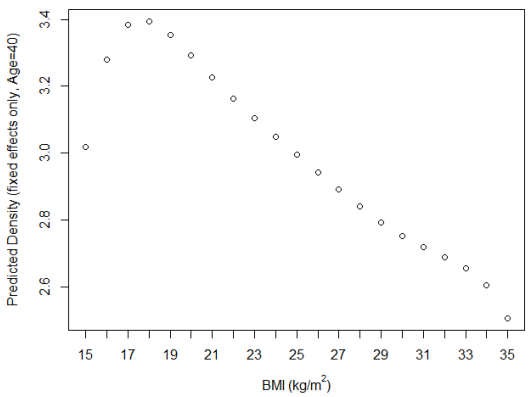
-> ΔLR-$\chi^2(1)$=15.7, p=7.3x10^{-05}



BMI to the power of 6

$$y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i})age_{ij} + \beta_2 BMI_{ij} + \beta_3 BMI_{ij}^2 + \beta_4 BMI_{ij}^3 + \beta_5 BMI_{ij}^4 + \beta_6 BMI_{ij}^5 + \beta_7 BMI_{ij}^6 + e_{ij}$$

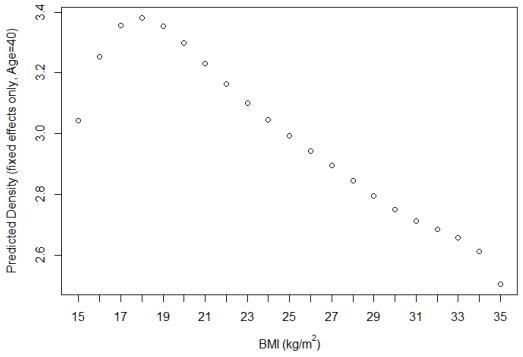
-> ΔLR-$\chi^2(1)$=171.4, p=3.7x10^{-39}



BMI to the power of 7

$$y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i}) age_{ij} + \beta_2 BMI_{ij} + \beta_3 BMI_{ij}^2 + \beta_4 BMI_{ij}^3 + \beta_5 BMI_{ij}^4 + \beta_6 BMI_{ij}^5 + \beta_7 BMI_{ij}^6 + \beta_8 BMI_{ij}^7 + e_{ij}$$

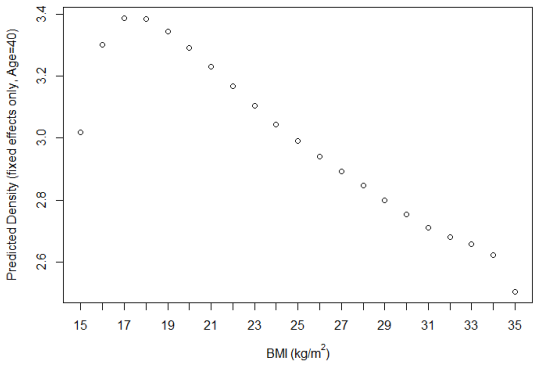
-> ΔLR-$\chi^2(1)=18.3$, $p=1.9 \times 10^{-05}$



BMI to the power of 8

$$y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i}) age_{ij} + \beta_2 BMI_{ij} + \beta_3 BMI_{ij}^2 + \beta_4 BMI_{ij}^3 + \beta_5 BMI_{ij}^4 + \beta_6 BMI_{ij}^5 + \beta_7 BMI_{ij}^6 + \beta_8 BMI_{ij}^7 + + \beta_9 BMI_{ij}^8 + e_{ij}$$

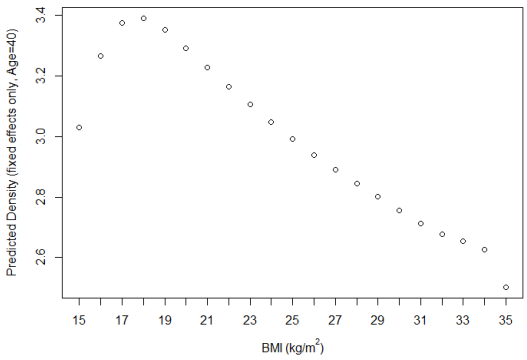
-> ΔLR-$\chi^2(1)=26.1$, $p= 3.2 \times 10^{-07}$



BMI to the power of 9

$$y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i})age_{ij} + \beta_2 BMI_{ij} + \beta_3 BMI_{ij}^2 + \beta_4 BMI_{ij}^3 + \beta_5 BMI_{ij}^4 + \beta_6 BMI_{ij}^5 + \beta_7 BMI_{ij}^6 + \beta_8 BMI_{ij}^7 + \beta_9 BMI_{ij}^8 + + \beta_{10} BMI_{ij}^9 + e_{ij}$$

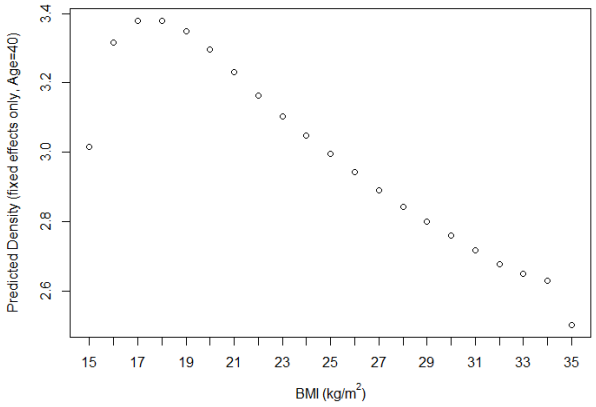
-> ΔLR-$\chi^2(1)=9.5$, p= 0.002



BMI to the power of 10

$$y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i})age_{ij} + \beta_2 BMI_{ij} + \beta_3 BMI_{ij}^2 + \beta_4 BMI_{ij}^3 + \beta_5 BMI_{ij}^4 + \beta_6 BMI_{ij}^5 + \beta_7 BMI_{ij}^6 + \beta_8 BMI_{ij}^7 + \beta_9 BMI_{ij}^8 + + \beta_{10} BMI_{ij}^9 + \beta_{11} BMI_{ij}^{10} + e_{ij}$$

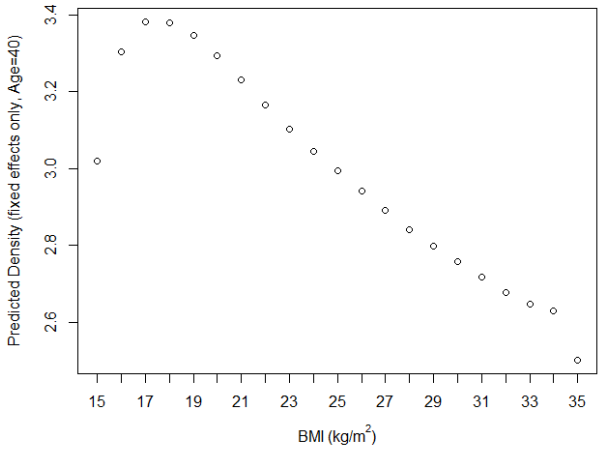
-> ΔLR-$\chi^2(1)=19.0$, p= 1.3×10^{-05}



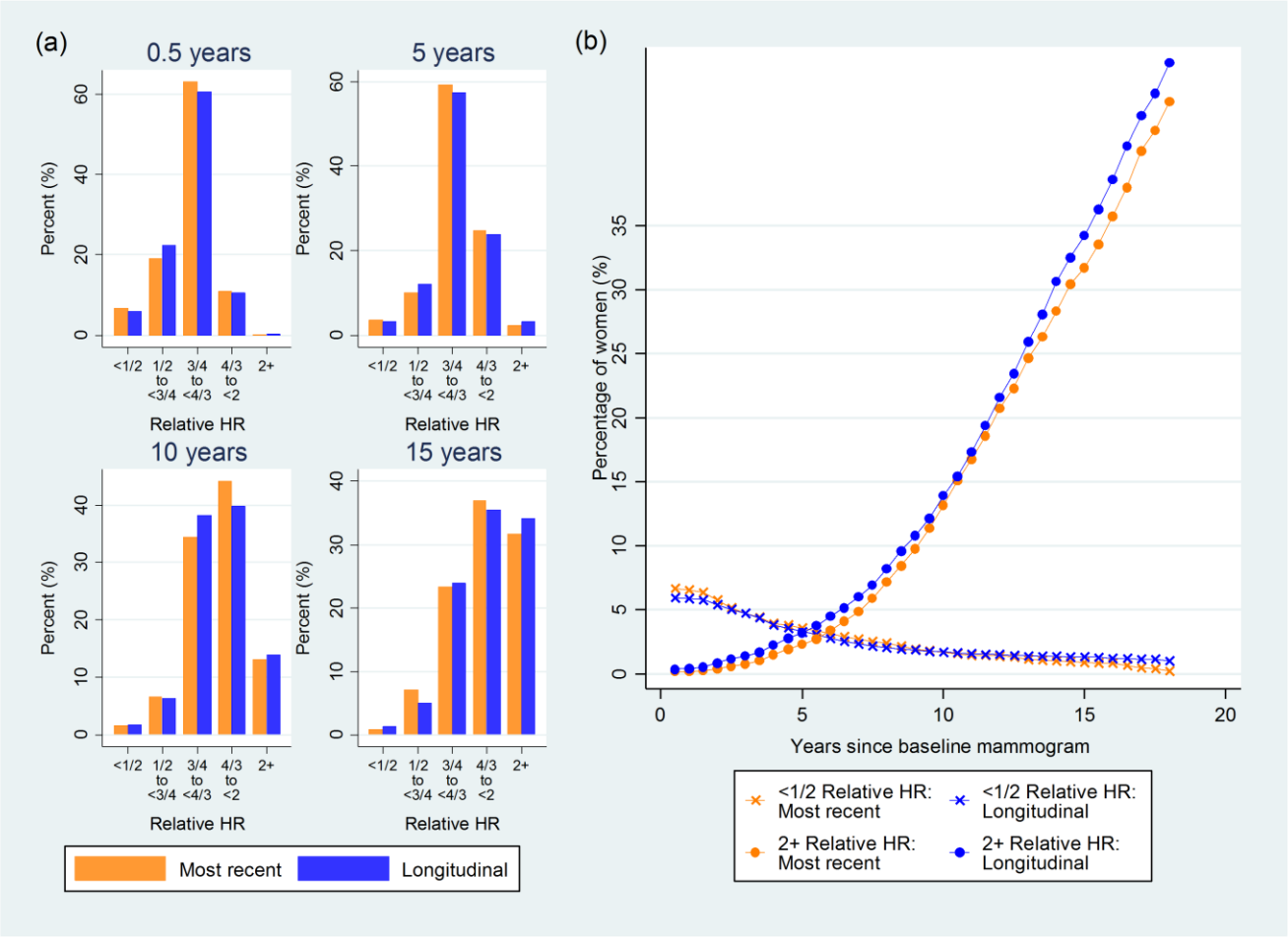
BMI to the power of 11

$$y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i})age_{ij} + \beta_2 BMI_{ij} + \beta_3 BMI_{ij}^2 + \beta_4 BMI_{ij}^3 + \beta_5 BMI_{ij}^4 + \beta_6 BMI_{ij}^5 + \beta_7 BMI_{ij}^6 + \beta_8 BMI_{ij}^7 + \beta_9 BMI_{ij}^8 + + \beta_{10} BMI_{ij}^9 + \beta_{11} BMI_{ij}^{10} + \beta_{12} BMI_{ij}^{11} + e_{ij}$$

-> $\Delta LR\text{-}\chi^2(1)=1.9$, p=0.17

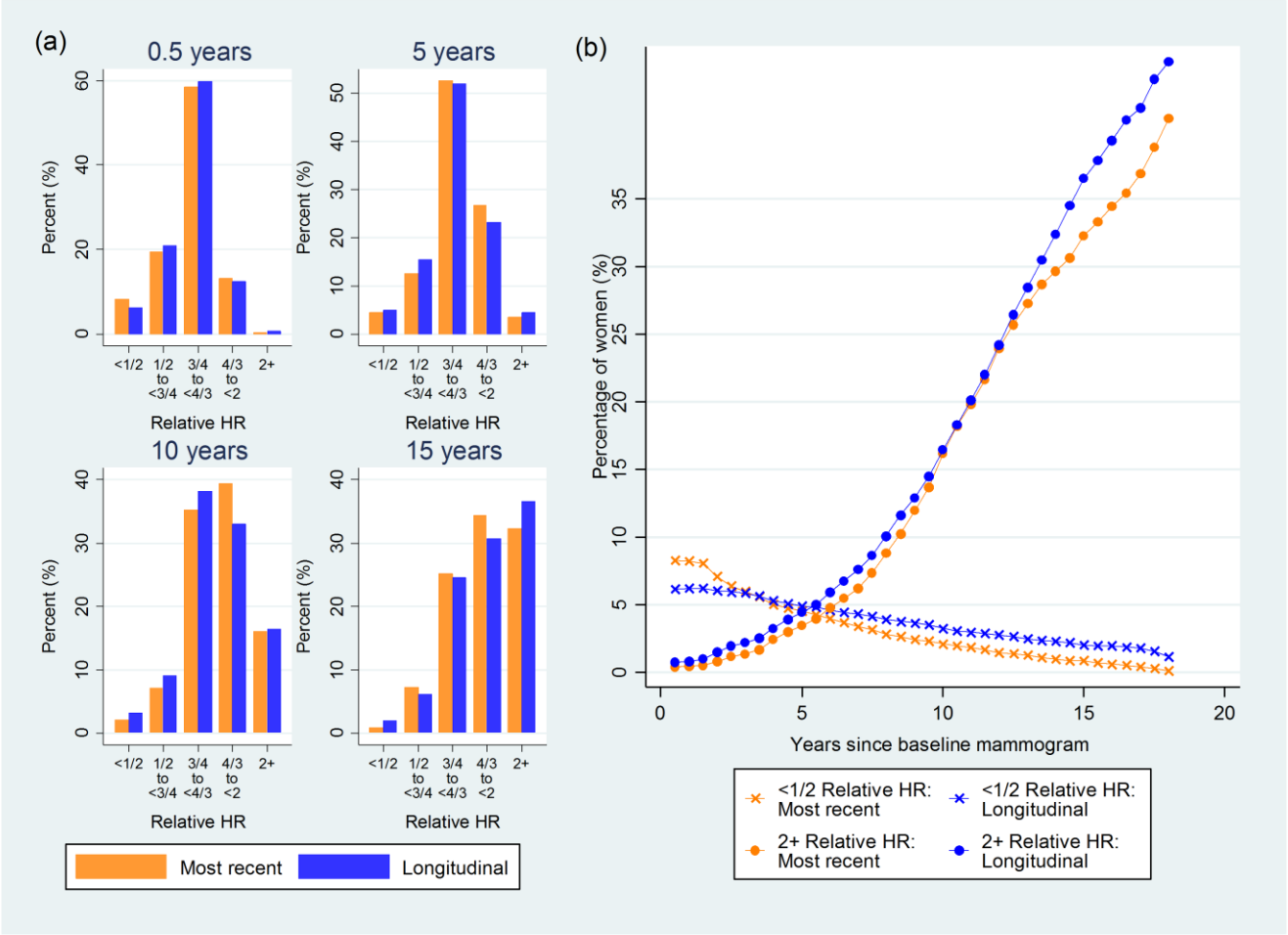


A.V Comparison of observed relative risk distributions for most recent density and longitudinal density (40-50yr)



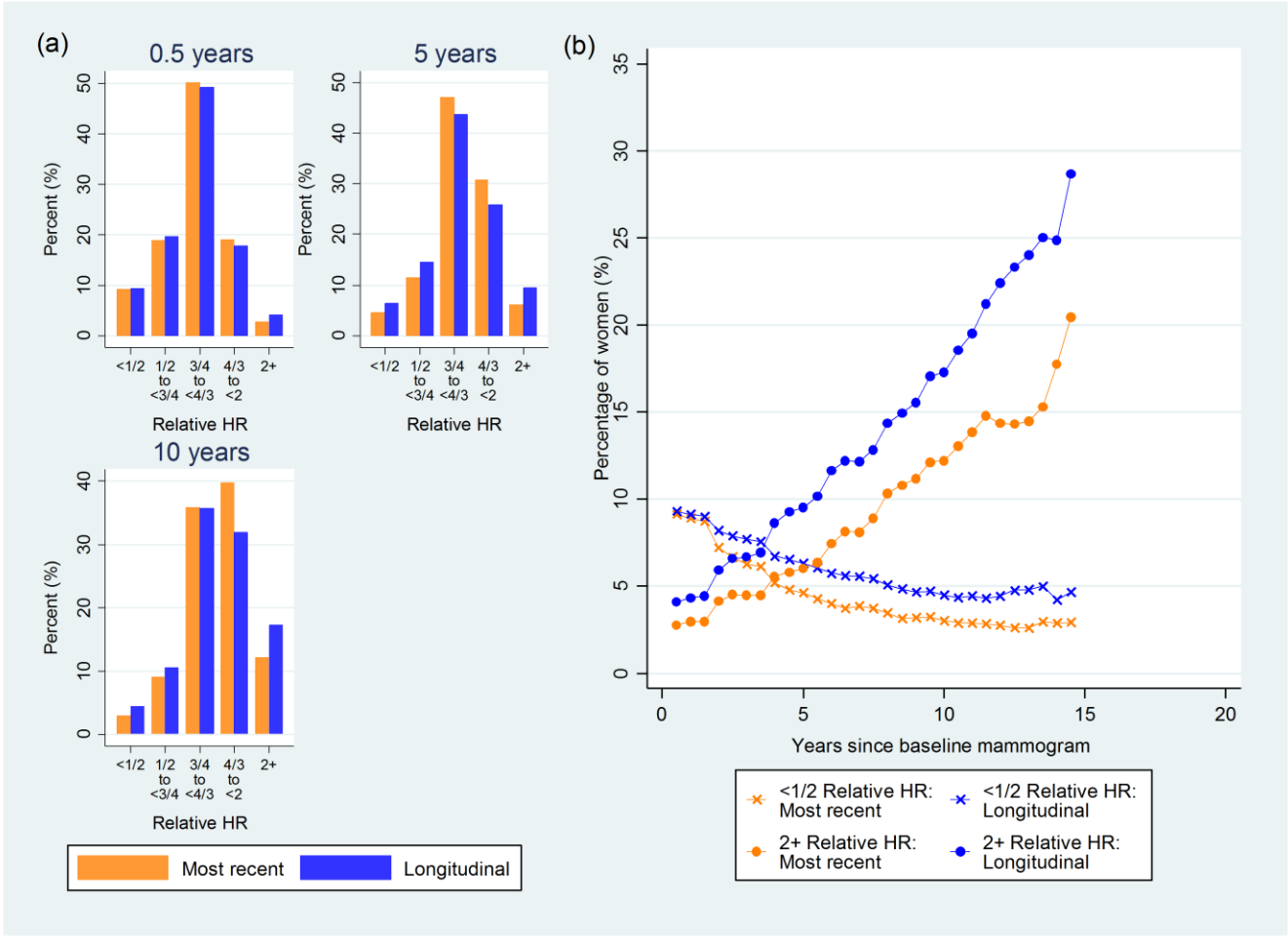
(a) Histograms showing the distribution of relative hazard ratios (HRs) for each model at 0.5yr, 5yr, 10yr and 15yr (HRs relative to the average HR at 0.5yr for 45y/o for each model); (b) Graph showing the percentage of women in the lowest (<1/2 relative HR) and highest (2+ relative HR) risk groups at each 6 month period.

A.VI Comparison of observed relative risk distributions for most recent density and longitudinal density (50-60yr)



(a) Histograms showing the distribution of relative hazard ratios (HRs) for each model at 0.5yr, 5yr, 10yr and 15yr (HRs relative to the average HR at 0.5yr for 55y/o for each model); (b) Graph showing the percentage of women in the lowest (<1/2 relative HR) and highest (2+ relative HR) risk groups at each 6 month period.

A.VII Comparison of observed relative risk distributions for most recent density and longitudinal density (60yr+)



(a) Histograms showing the distribution of relative hazard ratios (HRs) for each model at 0.5yr, 5yr and 10yr (HRs relative to the average HR at 0.5yr for 65y/o for each model); (b) Graph showing the percentage of women in the lowest (<1/2 relative HR) and highest (2+ relative HR) risk groups at each 6 month period.

A.VIII Comparison of statistical information (proportional-hazards Cox model fit) in different breast density measures (body mass index not imputed)

Model	BMI	Density	df	LR-χ^2 (model)	AIC (model)	df	ΔLR-χ^2 (density)	ΔAIC (density)
1	Baseline	Baseline	5	601.8	57,503.4	3	294.4	288.4
2	Most recent	Most recent	5	604.8	57,500.5	3	291.8	285.8
3	Most recent	Longitudinal	4	692.3	57,410.9	2	379.3	375.3

Body mass index (BMI) windsorised; clock starts at new baseline mammogram; ΔLR-χ^2 represents the difference in likelihood ratio statistics (LR-χ^2) between a model fit to age at baseline and most recent BMI and a model additionally incorporating the density term(s); ΔAIC represents the difference in Akaike Information Criterion (AIC) between a model fit to age and BMI and a model additionally incorporating the density term(s); n=129,748 women.

A.IX Comparison of statistical information (proportional-hazards Cox model fit) in different breast density measures (no screen-detected mammograms)

Model	BMI	Density	df	LR-χ^2 (model)	AIC (model)	df	ΔLR-χ^2 (density)	ΔAIC (density)
1	Baseline	Baseline	5	604.8	58,394.3	3	296.0	290.0
2	Most recent	Most recent	5	628.0	58,371.1	3	314.5	308.5
3	Most recent	Longitudinal	4	701.0	58,296.1	2	387.6	383.6

Mammograms removed if up to 6 months before event; clock starts at new baseline mammogram; ΔLR-χ^2 represents the difference in likelihood ratio statistics (LR-χ^2) between a model fit to age at baseline and most recent body mass index (BMI) and a model additionally incorporating the density term(s); ΔAIC represents the difference in Akaike Information Criterion (AIC) between a model fit to age and BMI and a model additionally incorporating the density term(s); includes all women since everyone had at least a baseline mammogram and there were no screen-detected baseline mammograms (women excluded if breast cancer event occurred <0.5yr after start of follow-up).

A.X Comparison of statistical information (proportional-hazards Cox model fit) in different breast density measures (by age at baseline group)

Age at baseline (yr)	Density measure	ΔLR-χ^2 (density)				
		Time (yr)				
		All	0-5	5-10	10-15	15+
All	Baseline	296.2	170.4	106.9	39.6	10.1
	Most recent	307.7	187.2	98.7	31.1	13.9
	Longitudinal	379.6	221.7	133.6	43.6	8.3
40-44	Baseline	50.0	14.4	18.0	8.5	2.6
	Most recent	36.9	13.6	12.6	12.3	2.1
	Longitudinal	45.4	13.3	21.9	10.4	1.2
45-49	Baseline	62.9	24.2	21.0	14.7	4.9
	Most recent	65.2	31.2	27.3	11.0	5.6
	Longitudinal	77.4	37.4	31.6	20.6	3.1
50-54	Baseline	50.5	27.6	18.8	10.0	5.4
	Most recent	59.5	36.1	24.2	11.3	4.1
	Longitudinal	71.5	39.0	28.3	11.0	2.8
55-59	Baseline	54.6	51.5	16.3	5.7	3.3
	Most recent	55.9	39.0	20.6	2.7	8.0
	Longitudinal	64.7	57.3	21.8	2.9	3.5
60-64	Baseline	39.6	23.6	23.8	10.0	-
	Most recent	56.2	32.6	22.9	6.8	-
	Longitudinal	70.1	39.1	24.5	10.9	-
65+	Baseline	43.9	39.4	13.5	-	-
	Most recent	49.8	47.2	8.1	-	-
	Longitudinal	61.1	51.7	10.8	-	-

ΔLR-χ^2 represents the difference in likelihood ratio statistics (LR-χ^2) between a model fit to age at baseline and most recent body mass index (BMI) and a model additionally incorporating the density term(s); age at baseline fitted in the model is the age at the start of each time period.

A.XI Comparison of statistical information (proportional-hazards Cox model fit) in different breast density measures (by menopausal status at baseline)

Menopausal status	Density measure	ΔLR-χ^2 (density)				
		Time (yr)				
		All	0-5	5-10	10-15	15+
Premenopausal	Baseline	87.9	25.3	43.6	18.1	8.5
	Most recent	95.6	33.7	44.8	16.7	10.9
	Longitudinal	108.0	33.3	59.0	21.0	6.1
Postmenopausal	Baseline	173.6	122.4	49.7	21.7	3.6
	Most recent	166.8	130.4	41.4	10.1	6.7
	Longitudinal	222.5	160.8	62.3	19.6	3.1

ΔLR-χ^2 represents the difference in likelihood ratio statistics (LR-χ^2) between a model fit to age at baseline and most recent body mass index (BMI) and a model additionally incorporating the density term(s).

A.XII Comparison of statistical information (proportional-hazards Cox model fit) in different breast density measures (in women with the longest follow-up (<60yr at baseline and before 2000))

Subgroup	Density measure	ΔLR-χ^2 (density)				
		Time (yr)				
		All	0-5	5-10	10-15	15+
All	Baseline	172.9	80.6	61.6	29.9	10.1
	Most recent	160.0	86.7	61.8	21.3	13.9
	Longitudinal	199.2	100.6	87.2	33.1	8.3

ΔLR-χ^2 represents the difference in likelihood ratio statistics (LR-χ^2) between a model fit to age at baseline and most recent body mass index (BMI) and a model additionally incorporating the density term(s).

A.XIII Comparison of statistical information (proportional-hazards Cox model fit) in different breast density measures (by menopausal status at baseline, in women with the longest follow-up (<60yr at baseline and before 2000))

Menopausal status	Density measure	ΔLR-χ^2 (density)				
		Time (yr)				
		All	0-5	5-10	10-15	15+
Premenopausal	Baseline	69.7	18.0	32.9	13.4	8.5
	Most recent	63.2	20.4	31.4	10.9	10.9
	Longitudinal	72.0	21.2	45.4	14.6	6.1
Postmenopausal	Baseline	92.8	53.0	31.7	18.5	3.6
	Most recent	80.2	60.1	27.0	7.0	6.7
	Longitudinal	112.6	73.9	45.6	14.5	3.1

ΔLR-χ^2 represents the difference in likelihood ratio statistics (LR-χ^2) between a model fit to age at baseline and most recent body mass index (BMI) and a model additionally incorporating the density term(s).

A.XIV Comparison of statistical information (proportional-hazards Cox model fit) in different breast density measures (by film/digital at baseline)

Mammogram type	Density measure	ΔLR-χ^2 (density)
		Time (yr)
		0-5
Film	Baseline	161.9
	Most recent	171.3
	Longitudinal	211.0
Digital	Baseline	14.3
	Most recent	21.3
	Longitudinal	24.5

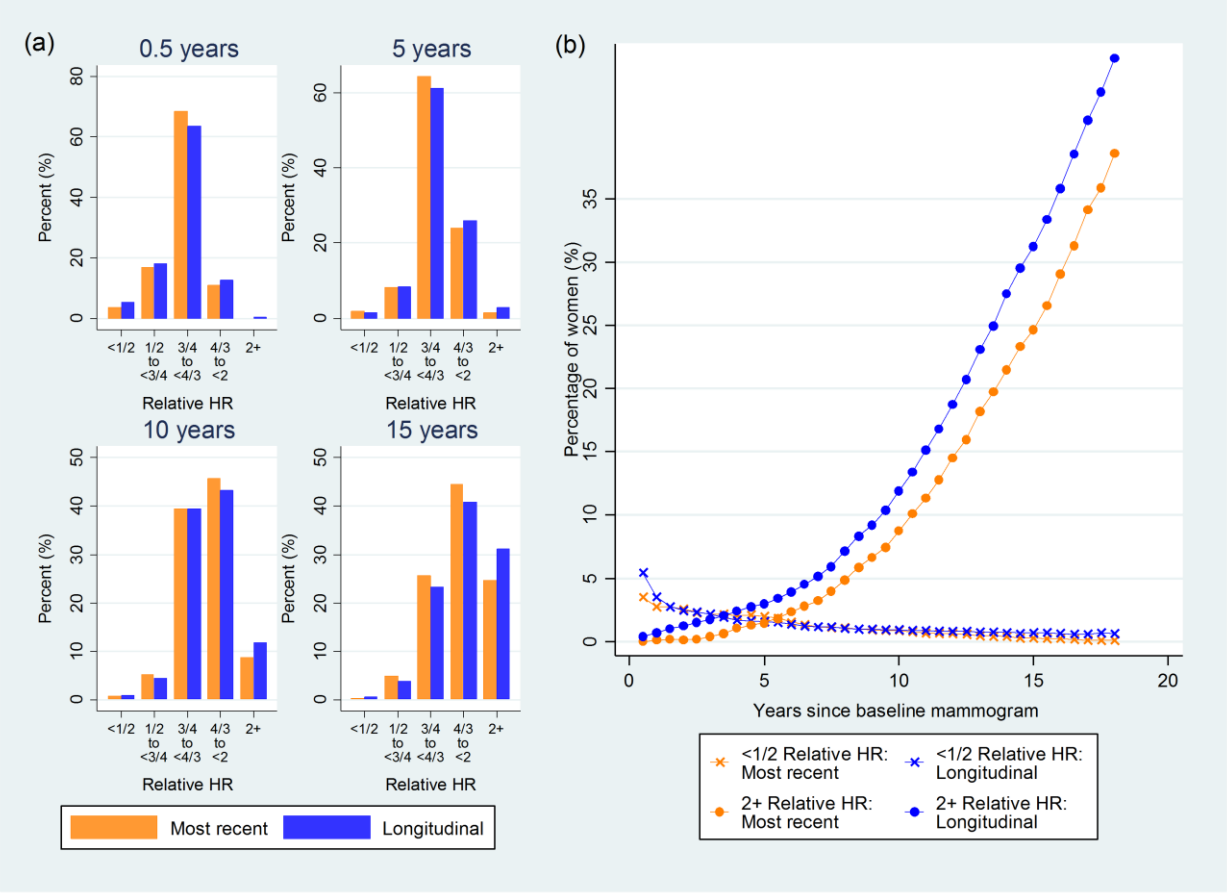
ΔLR-χ^2 represents the difference in likelihood ratio statistics (LR-χ^2) between a model fit to age at baseline and most recent body mass index (BMI) and a model additionally incorporating the density term(s); Film=baseline year <2007, digital= baseline year ≥2007; only assessing 0-5yr to reduce overlap of follow-up from film to digital.

A.XV Comparison of statistical information (proportional-hazards Cox model fit) in different breast density measures (by BI-RADS 3rd/4th at baseline)

Lexicon	Density measure	ΔLR-χ^2 (density)
		Time (yr)
		0-5
BI-RADS 3rd	Baseline	143.9
	Most recent	147.3
	Longitudinal	183.3
BI-RADS 4th	Baseline	52.7
	Most recent	56.4
	Longitudinal	69.3

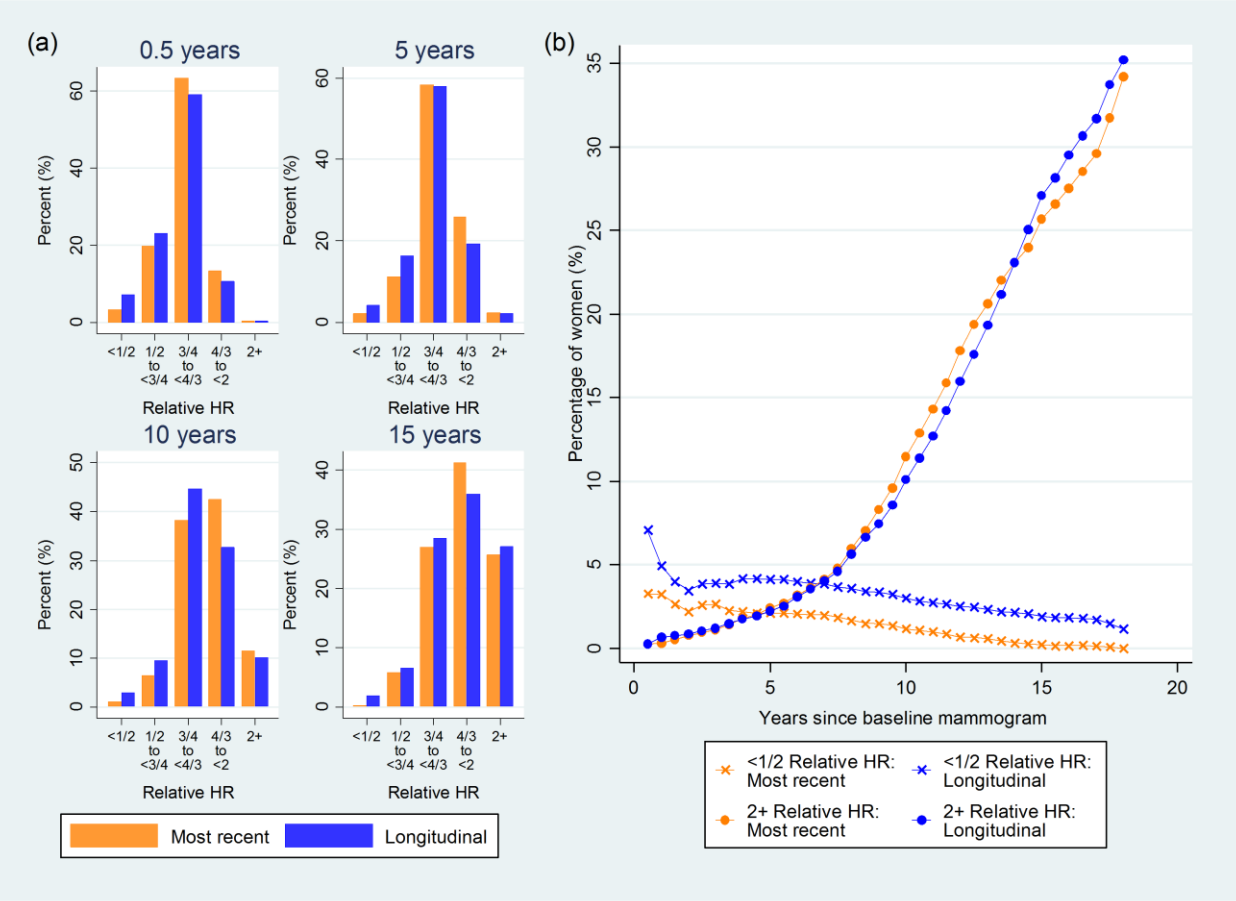
ΔLR-χ^2 represents the difference in likelihood ratio statistics (LR-χ^2) between a model fit to age at baseline and most recent body mass index (BMI) and a model additionally incorporating the density term(s); BI-RADS 3rd lexicon=baseline year <2003, BI-RADS 4th lexicon=baseline year ≥2003; only assessing 0-5yr to reduce overlap of follow-up from BI-RADS 3rd lexicon to BI-RADS 4th lexicon.

A.XVI Comparison of observed relative risk distributions for most recent density and longitudinal density (women with at least 3 mammograms, 40-50yr)



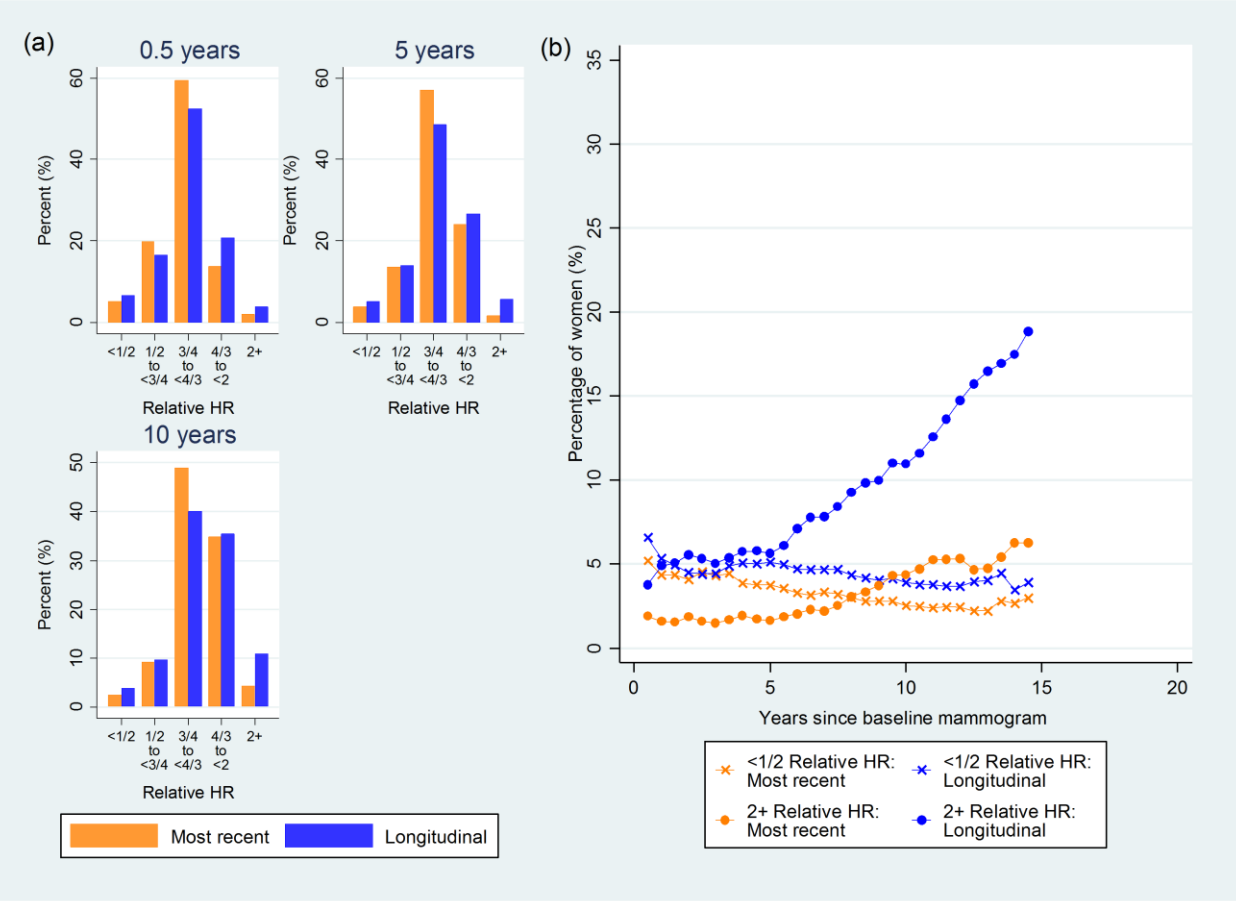
(a) Histograms showing the distribution of relative hazard ratios (HRs) for each model at 0.5yr, 5yr, 10yr and 15yr (HRs relative to the average HR at 0.5yr for 45y/o for each model); (b) Graph showing the percentage of women in the lowest (<1/2 relative HR) and highest (2+ relative HR) risk groups a t each 6 month period.

A.XVII Comparison of observed relative risk distributions for most recent density and longitudinal density (women with at least 3 mammograms, 50-60yr)



(a) Histograms showing the distribution of relative hazard ratios (HRs) for each model at 0.5yr, 5yr, 10yr and 15yr (HRs relative to the average HR at 0.5yr for 55y/o for each model); (b) Graph showing the percentage of women in the lowest (<1/2 relative HR) and highest (2+ relative HR) risk groups a t each 6 month period.

A.XVIII Comparison of observed relative risk distributions for most recent density and longitudinal density (women with at least 3 mammograms, 60yr+)



(a) Histograms showing the distribution of relative hazard ratios (HRs) for each model at 0.5yr, 5yr and 10yr (HRs relative to the average HR at 0.5yr for 65y/o for each model); (b) Graph showing the percentage of women in the lowest (<1/2 relative HR) and highest (2+ relative HR) risk groups at each 6 month period.

Appendix B: Supplementary material for Chapter 4

B.I Search: CENTRAL

#1 MeSH descriptor: [Selective Estrogen Receptor Modulators] explode all trees

#2 MeSH descriptor: [Aromatase Inhibitors] explode all trees

#3 MeSH descriptor: [Tamoxifen] explode all trees

#4 tamoxifen

#5 MeSH descriptor: [Raloxifene Hydrochloride] explode all trees

#6 raloxifene or lasofoxifene or arzoxifene or droloxifene or bazedoxifene or fulvestrant or anastrozole or letrozole or exemestane

#7 #1 or #2 or #3 or #4 or #5 or #6

#8 MeSH descriptor: [Breast Density] explode all trees

#9 (mammogr* or breast or mammary) near dens*

#10 MeSH descriptor: [Mammography] explode all trees

#11 MeSH descriptor: [Mammary Glands, Human] explode all trees

#12 dens*

#13 (#10 or #11) and #12

#14 #8 or #9 or #13

#15 #7 and #14

B.II Search: MEDLINE via OvidSP

1. exp Selective Estrogen Receptor Modulators/
2. exp Aromatase Inhibitors/
3. exp TAMOXIFEN/
4. tamoxifen.mp.
5. exp Raloxifene Hydrochloride/
6. raloxifene.mp.
7. lasofoxifene.mp.
8. arzoxifene.mp.
9. droloxifene.mp.
10. bazedoxifene.mp.
11. fulvestrant.mp.
12. anastrozole.mp.
13. letrozole.mp.
14. exemestane.mp.
15. or/1-14
16. exp Breast Density/

17. exp MAMMOGRAPHY/
18. exp Mammary Glands, Human/
19. ((mammogr* or breast or mammary) adj6 dens*).tw.
20. dens*.tw.
21. (17 or 18) and 20
22. 16 or 19 or 21
23. 15 and 22
24. Animals/ not Humans/
25. 23 not 24
26. limit 25 to yr="1996 -Current"

B.III Search: Embase via OvidSP

1. exp selective estrogen receptor modulator/
2. exp aromatase inhibitor/
3. exp tamoxifen/
4. tamoxifen.ti,ab.
5. exp raloxifene/
6. raloxifene.ti,ab.
7. exp lasofoxifene/
8. lasofoxifene.ti,ab.
9. exp arzoxifene/
10. arzoxifene.ti,ab.
11. exp droloxifene/
12. droloxifene.ti,ab.
13. exp bazedoxifene/
14. bazedoxifene.ti,ab.
15. exp fulvestrant/
16. fulvestrant.ti,ab.
17. exp anastrozole/
18. anastrozole.ti,ab.
19. exp letrozole/
20. letrozole.ti,ab.
21. exp exemestane/
22. exemestane.ti,ab.
23. or/1-22
24. exp breast density/
25. ((mammogr$ or breast or mammary) adj6 dens$).ti,ab.
26. dens$.ti,ab.

27. exp mammography/

28. exp mammary gland/

29. 26 and (27 or 28)

30. 24 or 25 or 29

31. 23 and 30

32. limit 31 to (human and (conference abstracts or embase) and yr="1996 -Current")

B.IV Search: WHO ICTRP

Basic search:

1. breast density OR mammographic density

Advanced search:

Title: density

Condition: breast cancer

Intervention: selective oestrogen receptor modulator OR serm OR aromatase inhibitor OR tamoxifen OR raloxifene OR lasofoxifene OR arzoxifene OR droloxifene OR bazedoxifene OR fulvestrant OR anastrozole OR letrozole OR exemestane

Recruitment status: ALL

B.V Search: ClinicalTrials.gov

Advanced search:

Condition or disease: breast cancer

Other terms: breast density OR mammographic density

Study type: All studies

Study results: All studies

Sex: All

Intervention/treatment: selective oestrogen receptor modulator OR serm OR aromatase inhibitor OR tamoxifen OR raloxifene OR lasofoxifene OR arzoxifene OR droloxifene OR bazedoxifene OR fulvestrant OR anastrozole OR letrozole OR exemestane

B.VI Data capture forms

B.VI.i Data capture forms – Options

New blocks can be added for any of the below (or just enter one as appropriate)

Subgroup

By DCIS/Invasive

By stage

By chemotherapy

By targeted therapy

By HRT

By pre/peri/post menopausal

By age group (e.g. <50/≥50yr)

By BMI group (e.g. <25, 25 to <30, 30 to <35, ≥35kg/m^2)

By baseline density

Endpoint

Treatment

Breast cancer mortality (time to death caused by breast cancer)

Rate of all serious adverse events

Recurrence

Incidence of a secondary primary breast cancer (e.g. in the contralateral breast)

Any recurrence or any death (disease-free survival)

Distant metastases

Death from all causes (all-cause mortality)

Recurrence of invasive cancer only

Recurrence of DCIS cancer only

Troublesome but not serious side effects observed for SERMs and AIs

Prevention

Incidence of invasive breast cancer and DCIS

Rate of all serious adverse events

Incidence of invasive cancer only

Incidence of DCIS cancer only

Troublesome but not serious side effects observed for SERMs and AIs

B.VI.ii Data capture form - Cuzick 2011

Area	Field	Data
Study design	**Type of study**	Nested case-control within a randomised trial
	Matching	None
	Prognostic, predictive or both	Both (prognostic phase II)
	Control group	558 randomised to receive placebo
	Prevention or treatment	Prevention
	Intervention(s)	Tamoxifen 20 mg/day
	Follow-up time period	96 months mentioned for follow-up of whole IBIS-I trial but no detail given for this sub-study
Setting	**Country**	Controls from UK, cases from UK and Finland
	High-risk clinic?	No
	Treatment clinic?	No
	Time period	Diagnosis before October 1, 2007. Recruitment April 1992 – March 2001.
	Urban/rural	Not stated
Participants (and characteristics at baseline)	**No. of participants**	7152 in trial, 126/224 with breast cancer to October 2007, 942 + 123 = 1065 in this study, Tamoxifen cases=51, Tamoxifen controls=456, Placebo cases=72, Placebo controls=486
	Age (yr)	Mean(SD)/Median(IQR): Tamoxifen cases=52(6)/51(48-54), Tamoxifen controls=51(6)/50(42-46), Placebo cases=51(6)/50(46-56), Placebo controls=51(6)/49(46-54)
	Age <50 or ≥50 (yr)	<50/≥50: Tamoxifen cases=26(51%)/25(49%), Tamoxifen controls=269(59%)/187(41%), Placebo cases=40(56%)/32(44%), Placebo controls=283(58%)/203(42%)
	BMI (kg/m^2)	Mean(SD)/Median(IQR): Tamoxifen cases=27(5)/26(24–31), Tamoxifen controls=27(5)/ 26(23–30), Placebo cases=27(5)/26(24–28), Placebo controls=27(5)/26(23–29)
	BMI < 25, 25 to < 30, 30 to < 35, ≥35 (kg/m^2)	≤25/26 to ≤30/≥30: Tamoxifen cases=20(39%)/15(29%)/13(25%) 3 missing, Tamoxifen controls=203(45%)/137(30%)/109(24%) 7 missing, Placebo cases=26(36%)/32(44%)/14(19%) 0 missing, Placebo controls=207(43%)/171(35%)/102(21%) 6 missing
	Ethnicity	Not stated
	Education	Not stated
	Baseline risk (%)	Approximately twice population risk
	Post/peri/pre-menopausal	Pre/Post: cases=58/62, controls=496/433
	Distribution of density at baseline	Mean(SD)/Median(IQR): Tamoxifen cases=47(32)/45(20-80), Tamoxifen controls=44(30)/40(17-70), Placebo cases=53(30)/63(25-80), Placebo controls=44(30)/43(15-70). Categories: 0%/1-10%/11-25%/26-50%/51-75%/76-100%: Tamoxifen cases=5(10%)/5(10%)/7(14%)/10(20%)/11(21%)/13(25%), Tamoxifen controls=53(12%)/40(9%)/55(12%)/116(25%)/103(23%)/89(19%), Placebo cases=3(4%)/5(7%)/10(14%)/13(18%)/19(26%)/22(31%), Placebo controls=53(11%)/53(11%)/61(13%)/109(22%)/111(23%)/99(20%).
	Invasive/DCIS at baseline	NA
	Stage (percentage regional spread) at baseline	NA
Cointerventions	**HRT use**	Never/Previous/Current: Tamoxifen cases=27(53%)/10(20%)/14(27%), Tamoxifen controls=398(65%)/63(14%)/95(21%), Placebo

		cases=46(64%)/12(17%)/14(19%), Placebo controls=316(65%)/62(13%)/108(22%)
	Chemotherapy?	NA
	Targeted therapy?	NA
	Radiotherapy?	NA
	Neoadjuvant endocrine therapy?	NA
Timing	**Time between baseline mammogram and diagnosis**	NA
	Time between diagnosis and start of endocrine therapy (or study entry)	NA
	Time between start of endocrine therapy (or study entry) and the follow-up mammogram	At least 12m after randomisation, median=18m, IQR=16-19m
	Time between baseline mammogram and start of endocrine therapy (or study entry)	At or up to 12m before randomisation
	Time between baseline mammogram and the follow-up mammogram	Median=19m, IQR=18-23m
Biomarker	**Film (digitised for density or not)/FFDM**	Film (Finnish mammograms were digitised films, UK mammograms were original films)
	Pre-processing for quality control of mammographic density?	None stated (but original film used)
	Density measure(s)	Percentage (i) visual assessment to nearest 5% by expert, contralateral MLO
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Tamoxifen arm
	Endpoint	Incidence of invasive breast cancer and DCIS
	Measure	OR
	n total in analysis	507
	n events / cases in analysis	51
	Data	Change density: increase, no change, reduction 5%, reduction ≥10%: cases: 4, 20, 12, 15; controls 16, 141, 82, 217
	Adjustment	Age at entry, breast density at baseline, history of atypical hyperplasia or lobular carcinoma in situ, and body mass index
	Point estimate	2.13, REF, 0.90, 0.32
	SD	-
	SE	-
	95% CI	(0.64 to 7.20), REF, (0.40 to 2.04), (0.14 to 0.72)
	p-value	P trend=0.001
	Comment on statistical method	Logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Placebo arm

	Endpoint	Incidence of invasive breast cancer and DCIS
	Measure	OR
	n total in analysis	558
	n events / cases in analysis	72
	Data	Change density: increase, no change, reduction 5%, reduction ≥10%: cases: 9, 27, 21, 15; controls 57, 206, 98, 125
	Adjustment	Age at entry, breast density at baseline, history of atypical hyperplasia or lobular carcinoma in situ, and body mass index
	Point estimate	1.23, REF, 1.35, 0.69
	SD	-
	SE	-
	95% CI	(0.54 to 2.81), REF, (0.71 to 2.58), (0.34 to 1.41)
	p-value	P trend=0.51
	Comment on statistical method	Logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Tamoxifen arm (prognostic marker) - worked out from raw data
	Endpoint	Incidence of invasive breast cancer and DCIS
	Measure	OR
	n total in analysis	497
	n events / cases in analysis	48
	Data	<10% reduction, ≥10% reduction: cases: 35, 13; controls: 234, 215
	Adjustment	Age at entry, breast density at baseline, history of atypical hyperplasia or lobular carcinoma in situ, and body mass index
	Point estimate	REF, 0.32
	SD	-
	SE	-
	95% CI	REF, (0.15 to 0.66)
	p-value	REF, 0.002
	Comment on statistical method	Logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	All (predictive marker) - worked out from paper
	Endpoint	Incidence of invasive breast cancer and DCIS
	Measure	OR interaction
	n total in analysis	1065
	n events / cases in analysis	123
	Data	<10% reduction, ≥10% reduction: cases: 93, 30; controls: 600, 342
	Adjustment	None – this test was not reported in the original paper – conducted post hoc based on available data
	Point estimate	0.60
	SD	-
	SE	-
	95% CI	(0.25 to 1.45)
	p-value	0.74 (DLR=1.28)
	Comment on statistical method	Only unadjusted possible from the data reported in the paper. Logistic regression using the reported data.
	Other	Other point estimates: tamoxifen vs. placebo 0.95, density change

		0.76
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	All (predictive marker) - worked out from raw data
	Endpoint	Incidence of invasive breast cancer and DCIS
	Measure	OR interaction
	n total in analysis	1049
	n events / cases in analysis	120
	Data	<10% reduction, ≥10% reduction: cases: 92, 28; controls: 591, 338
	Adjustment	Age at entry, breast density at baseline, history of atypical hyperplasia or lobular carcinoma in situ, and body mass index
	Point estimate	0.53
	SD	-
	SE	-
	95% CI	(0.21 to 1.32)
	p-value	0.17
	Comment on statistical method	Logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Tamoxifen arm
	Endpoint	Incidence of invasive breast cancer and DCIS
	Measure	OR compared with placebo
	n total in analysis	1049
	n events / cases in analysis	120
	Data	35 cases <10% reduction, 13 cases ≥10%
	Adjustment	Age at entry, breast density at baseline, history of atypical hyperplasia or lobular carcinoma in situ, and body mass index
	Point estimate	<10% 1.13, ≥10% 0.37
	SD	-
	SE	-
	95% CI	(0.72 to 1.77), (0.20 to 0.69)
	p-value	Not stated
	Comment on statistical method	Logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Baseline breast density ≤10%, Tamoxifen arm
	Endpoint	Incidence of invasive breast cancer and DCIS
	Measure	OR compared with placebo
	n total in analysis	218
	n events / cases in analysis	18
	Data	10 cases <10% reduction, 0 cases ≥10% reduction
	Adjustment	Age at entry, history of atypical hyperplasia or lobular carcinoma in situ, and body mass index
	Point estimate	All 1.36, <10% 1.45, ≥10% NA
	SD	-
	SE	-

	95% CI	(0.51 to 3.66), (0.54 to 3.88)
	p-value	Not stated
	Comment on statistical method	Logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Baseline breast density 11-50%, Tamoxifen arm
	Endpoint	Incidence of invasive breast cancer and DCIS
	Measure	OR compared with placebo
	n total in analysis	377
	n events / cases in analysis	37
	Data	11 cases <10% reduction, 3 cases ≥10% reduction
	Adjustment	Age at entry, history of atypical hyperplasia or lobular carcinoma in situ, and body mass index
	Point estimate	All 0.55, <10% 0.97, ≥10% 0.21
	SD	-
	SE	-
	95% CI	(0.27 to 1.13), (0.43 to 2.14), (0.06 to 0.75)
	p-value	Not stated
	Comment on statistical method	Logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Baseline breast density 51-100%, Tamoxifen arm
	Endpoint	Incidence of invasive breast cancer and DCIS
	Measure	OR compared with placebo
	n total in analysis	454
	n events / cases in analysis	65
	Data	14 cases <10% reduction, 10 cases ≥10% reduction
	Adjustment	Age at entry, history of atypical hyperplasia or lobular carcinoma in situ, and body mass index
	Point estimate	All 0.68, <10% 1.09, ≥10% 0.44
	SD	-
	SE	-
	95% CI	(0.39 to 1.18), (0.55 to 2.15), (0.21 to 0.93)
	p-value	Not stated
	Comment on statistical method	Logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Premenopausal, Tamoxifen arm
	Endpoint	Incidence of invasive breast cancer and DCIS
	Measure	OR compared with placebo
	n total in analysis	554
	n events / cases in analysis	58
	Data	14 cases <10% reduction, 6 cases ≥10% reduction
	Adjustment	Age at entry, breast density at baseline, history of atypical

		hyperplasia or lobular carcinoma in situ, and body mass index
	Point estimate	All 0.59, <10% 1.18, ≥10% 0.27
	SD	-
	SE	-
	95% CI	(0.33 to 1.06), (0.60 to 2.32), (0.11 to 0.66)
	p-value	Not stated
	Comment on statistical method	Logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Postmenopausal, Tamoxifen arm
	Endpoint	Incidence of invasive breast cancer and DCIS
	Measure	OR compared with placebo
	n total in analysis	495
	n events / cases in analysis	62
	Data	21 cases <10%reduction, 7 cases ≥10% reduction
	Adjustment	Age at entry, breast density at baseline, history of atypical hyperplasia or lobular carcinoma in situ, and body mass index
	Point estimate	All 0.87, <10% 1.10, ≥10% 0.53
	SD	-
	SE	-
	95% CI	(0.51 to 1.50), (0.61 to 2.01), (0.22 to 1.28)
	p-value	Not stated
	Comment on statistical method	Logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	HRT use – never, Tamoxifen arm
	Endpoint	Incidence of invasive breast cancer and DCIS
	Measure	OR compared with placebo
	n total in analysis	731
	n events / cases in analysis	83
	Data	21 cases <10% reduction, 9 cases ≥10% reduction
	Adjustment	Age at entry, breast density at baseline, history of atypical hyperplasia or lobular carcinoma in situ, and body mass index
	Point estimate	All 0.60, <10% 0.97, ≥10% 0.31
	SD	-
	SE	-
	95% CI	(0.37 to 0.98), (0.55 to 1.71), (0.15 to 0.67)
	p-value	Not stated
	Comment on statistical method	Logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	HRT use – ever, Tamoxifen arm
	Endpoint	Incidence of invasive breast cancer and DCIS
	Measure	OR compared with placebo
	n total in analysis	318

	n events / cases in analysis	37
	Data	14 cases <10% reduction, 4 cases ≥10% reduction
	Adjustment	Age at entry, breast density at baseline, history of atypical hyperplasia or lobular carcinoma in situ, and body mass index
	Point estimate	All 1.08, <10% 1.54, ≥10% 0.53
	SD	-
	SE	-
	95% CI	(0.54 to 2.18), (0.72 to 3.23), (0.17 to 1.66)
	p-value	Not stated
	Comment on statistical method	Logistic regression
	Other	
Sources of funding and stated conflicts of interest	**Funding**	Cancer Research UK program grant (C569/A10404 to J.C.) for research on the prevention of hormonally related cancers.
	Conflict of interest	J. Cuzick and A. Howell have served as occasional consultants to and advisory board members for AstraZeneca, the maker of tamoxifen. J. F. Forbes received honorarium payments for educational lectures from Astra Zeneca. J. Cuzick is the principal investigator for trials for which his institution (Queen Mary University of London) receives funding from AstraZeneca. The study sponsor had no role in the study design, collection of the data, interpretation of the results, preparation of the manuscript, or the decision to submit the manuscript for publication.

B.VI.iii Data capture form - Kim 2012

Area	Field	Data
Study design	**Type of study**	Retrospective cohort
	Matching	None
	Prognostic, predictive or both	Prognostic (phase II)
	Control group	None
	Prevention or treatment	Treatment
	Intervention(s)	Tamoxifen for up to 5yr, anastrozole and/or letrozole for up to 5yr, anastrozole and/or letrozole for up to 5yr after 2-3yr tamoxifen, tamoxifen for 5yr then an AI but no mention of duration on AI. No dose information or intake frequency. All women had at least 2yr treatment (however, duration of treatment in table 1: min=0.9yr, max=7.9yr).
	Follow-up time period	Abstract says median follow-up 68.8m (text says 67.7m).
Setting	**Country**	South Korea
	High-risk clinic?	No
	Treatment clinic?	Yes
	Time period	Initial diagnosis October 2003-December 2006
	Urban/rural	Seoul National University Hospital
Participants (and characteristics at baseline)	**No. of participants**	Total 1065: Tamoxifen 5yr = 657, Tamoxifen 2-3yr + AI (total 5yr) = 41, Tamoxifen 5yr + AI (unknown total time) = 192, no mention of AI 5yr but by deduction = 175
	Age (yr)	Mean = 49.0 (40.1 in text), SD = 9.3, min = 24, max = 77
	Age <50 or ≥50 (yr)	≤50yr/>50yr = 680(64%)/385(36%)
	BMI (kg/m^2)	Not stated
	BMI < 25, 25 to < 30, 30 to < 35, ≥35 (kg/m^2)	Not stated
	Ethnicity	South Korean institution, otherwise not reported
	Education	Not stated
	Baseline risk (%)	NA
	Post/peri/pre-menopausal	Not stated
	Distribution of density at baseline	Mean = 35.8%, (SD = 14.0%), min = 5.4%, max = 82.2%. Categories <10%/10-25%/25-50%/≥50%: 26(2.4%)/223(20.9%)/641(60.2%)/175(16.4%).
	Invasive/DCIS at baseline	127 (12%) DCIS, 938 (88%) invasive
	Stage (percentage regional spread) at baseline	Stage not reported, but lymph node+/-: 359(34%)/706(66%); >2cm/≤2cm: 427(40%)/638(60%)
Cointerventions	**HRT use**	Not stated
	Chemotherapy?	Neoadjuvant: No=1017(96%), Yes=48(5%), adjuvant: No=247(23%), Yes=818(77%)
	Targeted therapy?	Not stated
	Radiotherapy?	Yes 657 (62%), No 408 (38%)
	Neoadjuvant endocrine therapy?	Not stated
Timing	**Time between baseline mammogram and diagnosis**	2 weeks pre-surgery
	Time between diagnosis and start of endocrine therapy (or study entry)	Not stated
	Time between start of endocrine therapy (or study entry) and the follow-up mammogram	'Average' 13.1m, range 8-20m
	Time between baseline mammogram and start of	Not stated

	endocrine therapy (or study entry)	
	Time between baseline mammogram and the follow-up mammogram	Not stated
Biomarker	**Film (digitised for density or not)/FFDM**	Digital mammograms
	Pre-processing for quality control of mammographic density?	Not stated
	Density measure(s)	Percentage (iv) semi-automated thresholding software (CUMULUS) by one experienced reader, contralateral CC
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	All
	Endpoint	Recurrence (Recurrence-free survival)
	Measure	Hazard ratio, continuous density reduction, per 1%?
	n total in analysis	1065
	n events / cases in analysis	80
	Data	No data presented on recurrence vs. density change
	Adjustment	Age, Size, LN, Grade, Chemotherapy, Ki-67
	Point estimate	0.95
	SD	-
	SE	-
	95% CI	(0.92 to 0.99)
	p-value	0.005
	Comment on statistical method	Cox regression, not clear when clock starts, nor reasons for censoring, loss to follow-up etc.
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	All
	Endpoint	Recurrence (Recurrence-free survival)
	Measure	Hazard ratio, <5% vs. ≥5% density reduction
	n total in analysis	1065, MDR: <5%, ≥5%: 505(47%), 560(53%)
	n events / cases in analysis	80
	Data	No data presented on recurrence vs. density change
	Adjustment	Size, LN, Ki67 (Forward stepwise selection)
	Point estimate	1.67 (equivalently 0.60 for ≥5% density reduction)
	SD	-
	SE	-
	95% CI	(1.07 to 2.63)
	p-value	0.025
	Comment on statistical method	Cox regression, not clear when clock starts, nor reasons for censoring, loss to follow-up etc.
	Other	Sup table 4: Size, LN, Ki67 (Forward stepwise selection), but in text: "adjusted for age and preMD by forward selection stepwise analysis"
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	All
	Endpoint	Recurrence (Recurrence-free survival)
	Measure	Hazard ratio, <0%, 0-5%, 5-10%, ≥10%

	n total in analysis	1065, MDR: <0%, 0-5%, 5-10%, ≥10%: 190(18%), 314(30%), 276(26%), 285(27%)
	n events / cases in analysis	80
	Data	No data presented on recurrence vs. density change
	Adjustment	Size, LN, Grade, Chemotherapy, Ki67
	Point estimate	≥10% (REF), 5-10% 1.33, 0-5% 1.92, <0% 2.26
	SD	-
	SE	-
	95% CI	REF, (0.67 to 2.65), (1.01 to 3.64), (1.10 to 4.64)
	p-value	REF, 0.413, 0.048, 0.027
	Comment on statistical method	Cox regression, not clear when clock starts, nor reasons for censoring, loss to follow-up etc.
	Other	Table 4: Also adjusted for age?
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Age ≤50yr
	Endpoint	Recurrence (Recurrence-free survival)
	Measure	Hazard ratio, <5% vs. ≥5% density reduction
	n total in analysis	680
	n events / cases in analysis	not stated
	Data	not stated
	Adjustment	not stated
	Point estimate	1.13
	SD	-
	SE	-
	95% CI	(0.62 to 2.04)
	p-value	0.7
	Comment on statistical method	A lot not reported, Cox regression, not clear when clock starts, nor reasons for censoring, loss to follow-up etc.
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Age >50yr
	Endpoint	Recurrence (Recurrence-free survival)
	Measure	Hazard ratio, <5% vs. ≥5% density reduction
	n total in analysis	385
	n events / cases in analysis	not stated
	Data	not stated
	Adjustment	not stated
	Point estimate	3.11
	SD	-
	SE	-
	95% CI	(1.19 to 8.14)
	p-value	0.02
	Comment on statistical method	A lot not reported, Cox regression, not clear when clock starts, nor reasons for censoring, loss to follow-up etc.
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Tamoxifen at entry
	Endpoint	Recurrence (Recurrence-free survival)
	Measure	Hazard ratio, <5% vs. ≥5% density reduction
	n total in analysis	890 (assumed)

	n events / cases in analysis	not stated
	Data	not stated
	Adjustment	not stated
	Point estimate	1.52
	SD	-
	SE	-
	95% CI	(0.92 to 2.51)
	p-value	0.11
	Comment on statistical method	A lot not reported, Cox regression, not clear when clock starts, nor reasons for censoring, loss to follow-up etc.
	Other	Tamoxifen 5yr = 657, Tamoxifen 2-3yr + AI (total 5yr) = 41, Tamoxifen 5yr + AI (unknown total time) = 192, no mention of AI 5yr but by deduction = 175. Unclear which groups are reported in the subgroup analysis.
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	AI at entry
	Endpoint	Recurrence (Recurrence-free survival)
	Measure	Hazard ratio, <5% vs. ≥5% density reduction
	n total in analysis	175 (assumed)
	n events / cases in analysis	not stated
	Data	not stated
	Adjustment	not stated
	Point estimate	7.11
	SD	-
	SE	-
	95% CI	(0.90 to 56.37)
	p-value	0.06
	Comment on statistical method	A lot not reported, Cox regression, not clear when clock starts, nor reasons for censoring, loss to follow-up etc.
	Other	Tamoxifen 5yr = 657, Tamoxifen 2-3yr + AI (total 5yr) = 41, Tamoxifen 5yr + AI (unknown total time) = 192, no mention of AI 5yr but by deduction = 175. Unclear which groups are reported in the subgroup analysis.
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Chemotherapy - no
	Endpoint	Recurrence (Recurrence-free survival)
	Measure	Hazard ratio, <5% vs. ≥5% density reduction
	n total in analysis	Unclear if neoadjuvant, adjuvant or both
	n events / cases in analysis	not stated
	Data	not stated
	Adjustment	not stated
	Point estimate	2.20
	SD	-
	SE	-
	95% CI	(0.54 to 8.88)
	p-value	0.27
	Comment on statistical method	A lot not reported, Cox regression, not clear when clock starts, nor reasons for censoring, loss to follow-up etc.
	Other	
Results (add new rows each subgroup and endpoint	**Subgroup**	Chemotherapy - yes

combo)		
	Endpoint	Recurrence (Recurrence-free survival)
	Measure	Hazard ratio, <5% vs. ≥5% density reduction
	n total in analysis	Unclear if neoadjuvant, adjuvant or both
	n events / cases in analysis	not stated
	Data	not stated
	Adjustment	not stated
	Point estimate	1.69
	SD	-
	SE	-
	95% CI	(1.02 to 2.80)
	p-value	0.04
	Comment on statistical method	A lot not reported, Cox regression, not clear when clock starts, nor reasons for censoring, loss to follow-up etc.
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	All
	Endpoint	Recurrence (Recurrence-free survival)
	Measure	Hazard ratio, <5% vs. ≥5% density reduction
	n total in analysis	Unclear
	n events / cases in analysis	not stated
	Data	not stated
	Adjustment	not stated
	Point estimate	1.74
	SD	-
	SE	-
	95% CI	(1.09-2.78)
	p-value	0.02
	Comment on statistical method	A lot not reported, Cox regression, not clear when clock starts, nor reasons for censoring, loss to follow-up etc.
	Other	Different from earlier result
		N.B. Also reported results for relative mammographic density reduction
Sources of funding and stated conflicts of interest	**Funding**	This work was supported by a National Research Foundation of Korea (NRF) Grant funded by the Korean Government (20110005753 and 20110031417)
	Conflict of interest	The authors declare that they have no competing interests

B.VI.iv Data capture form - Knight 2018

Area	Field	Data
Study design	**Type of study**	Case-control
	Matching	Cases were diagnosed with a second primary invasive CBC at least 2 years later with no intervening cancer diagnosis, other than a non-melanoma skin cancer or cervical carcinoma in situ. UBC controls had no history of subsequent cancer diagnosis except for nonmelanoma skin cancer or cervical carcinoma in situ up to their reference date. Matched 1:1 on follow-up time, living in the same study area from first breast cancer to reference date, year of birth (5yr strata), year diagnosis (4yr strata), race / ethnicity
	Prognostic, predictive or both	Prognostic (phase II) (although analysis not specific to any endocrine treatment i.e. ORs are adjusted for tamoxifen use)
	Control group	N/A
	Prevention or treatment	Treatment
	Intervention(s)	Mainly tamoxifen and "study time period and age distribution means that few women received aromatase inhibitors", but specific treatments, doses and intake frequency not reported
	Follow-up time period	Mean (SD): Cases=7.9yr (3.8), Controls=8.0yr (4.0)
Setting	**Country**	3 US sites (Northern California (179 (49%) cases, 213 (53%) controls), Seattle (82 (23%) cases, 80 (20%) controls), Iowa (49 (14%) cases, 46 (11%) controls), one Canada (Ontario (52 (14%) cases, 64 (16%) controls))).
	High-risk clinic?	No
	Treatment clinic?	Yes
	Time period	First diagnosis 1990-2008. Only women in WECARE2 included (second stage of WECARE, recruited 2009-2012).
	Urban/rural	Not stated
Participants (and characteristics at baseline)	**No. of participants**	Mammograms at both time points = 467 women (224 out of 362 cases, 243 out of 403 controls)
	Age (yr)	Mean (SD) age at mammogram before/at 1st diagnosis: Cases=46 (6), Controls=46 (6). Mean (SD) age at mammogram after 1st diagnosis: Cases=47 (6), Controls=47 (6). Diagnosis <55yr for all.
	Age <50 or ≥50 (yr)	Baseline <45/45-<50/50-54: Cases=90(36%)/82(32%)/81(32%), Controls=100(37%)/93(35%)/76(28%). Follow-up <45/45-<50/50-54: Cases=123(37%)/110(33%)/100(30%), Controls=137(36%)/127(34%)/113(30%).
	BMI (kg/m^2)	Mean (SD) BMI at 1st diagnosis: Cases=25.2 (5.5), Controls=25.2 (5.7). Mean (SD) BMI at mammogram after 1st diagnosis: Cases=25.2 (5.4), Controls=25.8 (5.7).
	BMI < 25, 25 to < 30, 30 to < 35, ≥35 (kg/m^2)	Not stated
	Ethnicity	Non-Hispanic white/Other: Cases=295 (83%)/67 (17%), Controls=334 (81%)/69 (19%)
	Education	Not stated
	Baseline risk (%)	Family history: yes/no/adopted or missing: cases=123(34%)/234(65%)/5(1%), controls=89(22%)/306(76%)/8(2%)
	Post/peri/pre-menopausal	Baseline Pre/Post: Case=184 (73%)/68 (27%), Control=211 (79%)/57 (21%). Follow-up Pre/Post: Case=130 (39%)/202(61%), Control=147 (39%)/228 (61%)
	Distribution of density at baseline	Mean (SD) Percent mammographic density before/at 1st diagnosis: Cases=37.6% (18.1)/Controls=35.8% (18.3%); <25%/25% to <50%/≥50%: Cases=67 (26%)/125 (49%)/61 (24%), Controls=81 (30%), 130 (48%), 58 (22%).

	Invasive/DCIS at baseline	100% invasive
	Stage (percentage regional spread) at baseline	Local/Regional/Missing: Cases=251 (69%)/106 (29%)/ 5(1%), Controls=254 (63%)/143 (35%)/6 (1%); ER+/ER-/Missing: Cases=213(59%)/129(36%)/20(6%), Controls=273(68%)/105(26%)/25(6%)
Cointerventions	**HRT use**	Not stated
	Chemotherapy?	Yes/No: Cases=236 (65%)/126 (35%), Controls=272 (67%)/131 (33%)
	Targeted therapy?	Tamoxifen cases=158 (44%), UNK=221 (55%)
	Radiotherapy?	Yes/No: Cases=251 (69%)/111 (31%), Controls=179 (69%)/124 (31%)
	Neoadjuvant endocrine therapy?	Not stated
Timing	**Time between baseline mammogram and diagnosis**	Prior to/at diagnosis mammogram (3yr prior to diagnosis - 1 month post diagnosis (as close as possible to 12 months prior to diagnosis - 1 month post diagnosis))
	Time between diagnosis and start of endocrine therapy (or study entry)	Not stated
	Time between start of endocrine therapy (or study entry) and the follow-up mammogram	Not stated, but diagnosis to follow-up mammogram (>6 months - 4yr (as close as possible to >6 months - 18 months))
	Time between baseline mammogram and start of endocrine therapy (or study entry)	Not stated
	Time between baseline mammogram and the follow-up mammogram	Median=1yr
Biomarker	**Film (digitised for density or not)/FFDM**	Digitised film – digital mammograms excluded (5 cases, 6 controls prior to diagnosis, 39 cases and 41 controls post diagnosis)
	Pre-processing for quality control of mammographic density?	Excluded when visually assessed to be poor image quality (4 cases and 4 controls prior to diagnosis, 11 cases and 6 controls post diagnosis)
	Density measure(s)	Percentage (iv) semi-automated thresholding software (CUMULUS) by experienced reader, contralateral CC
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	All
	Endpoint	Incidence of a secondary primary breast cancer (e.g. in the contralateral breast)
	Measure	OR, ≥10% vs. <10% density reduction
	n total in analysis	435
	n events / cases in analysis	210
	Data	<10%/≥10% reduction: Cases=150 (71%)/60 (29%), Controls=144 (64%)/81 (36%)
	Adjustment	Change in age, estimated body mass index, and menopausal status between prior to/at first diagnosis and post-diagnosis mammograms, and for initial %MD, study centre, race (non-Hispanic white vs. other), age at first diagnosis, age at menarche, number of full-term pregnancies, histologic type, stage, and oestrogen receptor status of first diagnosis, chemotherapy, radiation, and tamoxifen use after first diagnosis.
	Point estimate	0.63

	SD	-
	SE	-
	95% CI	(0.40 to 1.01)
	p-value	Not reported (but p>0.05)
	Comment on statistical method	Unconditional logistic regression, Excluding those with missing menopausal status information and those with an increase ≥10%, note the adjustment for tamoxifen.
	Other	Note that the analysis group did not all receive tamoxifen, not all were ER+ at first diagnosis.
Sources of funding and stated conflicts of interest	**Funding**	This research was supported by the US National Institutes of Health (grant numbers U01 CA83178, R01 CA97397, R01 CA129639, R01 CA114236, P30 CA008748, and R01 CA168339). The funding body had no role in the design of the study, the collection, analysis, and interpretation of the data, or in writing the manuscript.
	Conflict of interest	The authors declare that they have no competing interests.

B.VI.v Data capture form - Ko 2013

Area	Field	Data
Study design	**Type of study**	Retrospective cohort
	Matching	None
	Prognostic, predictive or both	Prognostic (phase II)
	Control group	None
	Prevention or treatment	Treatment
	Intervention(s)	Tamoxifen. All women had at least 2yr treatment. No dose information or intake frequency.
	Follow-up time period	Mean follow-up 59m (SD=17.6), range (26–114m), but text says mean follow-up 61 months.
Setting	**Country**	South Korea
	High-risk clinic?	No
	Treatment clinic?	Yes
	Time period	January 2003 – December 2008
	Urban/rural	National Cancer Center, Goyang
Participants (and characteristics at baseline)	**No. of participants**	n=1066, 67 with total recurrence: 48 systemic, 16(17 also mentioned) loco-regional, 4 contralateral (numbers do not add up)
	Age (yr)	Total: mean (SD)=45.3 (7.6), range=25-78. MDR+: mean (SD)=44 (5.9), range=28-68. MDR-: mean (SD)=46 (8.1), range=25-78
	Age <50 or ≥50 (yr)	Total: ≤50yr/>50yr = 888(83%)/178(17%). MDR+: ≤50yr/>50yr = 308(91%)/30(9%). MDR-: ≤50yr/>50yr = 580(80%)/148(20%)
	BMI (kg/m^2)	Total: mean (SD)=23.4(3.2), range=15.6-50.2. MDR+: mean (SD)=22.9(3.1), range=17.5-35.5. MDR-: mean (SD)=23.1(3.2), range=15.6-50.2.
	BMI < 25, 25 to < 30, 30 to < 35, ≥35 (kg/m^2)	Not stated
	Ethnicity	South Korean institution, otherwise not reported
	Education	Not stated
	Baseline risk (%)	NA
	Post/peri/pre-menopausal	Unclear, title says premenopausal women but age range 25-78 and subgroup analysis of ≤50yr/>50yr used as a proxy for menopausal status and postmenopausal women mentioned in results
	Distribution of density at baseline	BIRAD 1&2 n=141, BIRAD 3 n=503, BIRAD 4 n=422
	Invasive/DCIS at baseline	134 (13%) DCIS, 932 (87%) invasive (implied)
	Stage (percentage regional spread) at baseline	Histologic grade (1/2 vs. 3) = 840(78.8%)/226(21.2%). Lymph node-/+ = 666(61.5%)/410(38.5%).
Cointerventions	**HRT use**	Not stated
	Chemotherapy?	No=303(28.4%), Yes (adjuvant)=588(55.5%), Yes (neoadjuvant)=175(16.4%)
	Targeted therapy?	Not stated
	Radiotherapy?	No=173, Yes=893
	Neoadjuvant endocrine therapy?	Not stated
Timing	**Time between baseline mammogram and diagnosis**	Before surgery but no mention of timeframe.
	Time between diagnosis and start of endocrine therapy (or study entry)	Not stated

	Time between start of endocrine therapy (or study entry) and the follow-up mammogram	Range=10–34 months in text (10-36 months in results), median=19 months
	Time between baseline mammogram and start of endocrine therapy (or study entry)	Not stated
	Time between baseline mammogram and the follow-up mammogram	Not stated
Biomarker	**Film (digitised for density or not)/FFDM**	Digital mammograms
	Pre-processing for quality control of mammographic density?	Exclusion of women if digital mammogram not appropriate for evaluation, but no explanation as to what this means. Reliability not assessed ("We relied on a single radiologist who is a specialist in breast imaging studies, thereby eliminating interobserver variability. We did not seek to measure reproducibility as the BI-RADS density classifications are standardized.").
	Density measure(s)	Categorical (i) BI-RADS (qualitative & quantitative version) by experienced reader, no mention if contralateral or view
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	All
	Endpoint	Recurrence (Recurrence-free survival)
	Measure	Hazard ratio (MDR+ vs. MDR-)
	n total in analysis	1066
	n events / cases in analysis	67
	Data	MDR+=10/338, MDR-=57/728
	Adjustment	Age, BMI, tumor size, lymph node positivity, high histologic grade, HER2 positivity and Ki-67≥14%
	Point estimate	0.35
	SD	-
	SE	-
	95% CI	(0.17 to 0.68)
	p-value	0.002
	Comment on statistical method	Age, BMI, tumour size continuous variables
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	All
	Endpoint	Recurrence (Recurrence-free survival)
	Measure	Hazard ratio (MDR+ vs. MDR-)
	n total in analysis	1066
	n events / cases in analysis	67
	Data	MDR+=10/338, MDR-=57/728
	Adjustment	Age, BMI, tumor size, lymph node status, high ER score, high PgR score and HER2 positivity
	Point estimate	0.36

	SD	
	SE	
	95% CI	(0.18 to 0.70)
	p-value	0.003
	Comment on statistical method	Age, BMI, tumour size continuous variables
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	All
	Endpoint	Recurrence (Recurrence-free survival): systemic recurrence
	Measure	Hazard ratio (MDR+ vs. MDR-)
	n total in analysis	1046
	n events / cases in analysis	48
	Data	MDR+=9/337, MDR-=39/709
	Adjustment	Age, BMI, tumor size, lymph node status, high ER score, high PgR score and HER2 positivity
	Point estimate	0.48
	SD	-
	SE	-
	95% CI	(0.23 to 0.99)
	p-value	0.048
	Comment on statistical method	Age, BMI, tumour size continuous variables
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	All
	Endpoint	Recurrence (Recurrence-free survival): loco-regional recurrence
	Measure	Hazard ratio (MDR+ vs. MDR-)
	n total in analysis	1014
	n events / cases in analysis	16 (17 also mentioned in text)
	Data	MDR+=1/329, MDR-=15/685
	Adjustment	Age, BMI, tumor size, lymph node status, high ER score, high PgR score and HER2 positivity
	Point estimate	0.13
	SD	-
	SE	-
	95% CI	(0.02 to 0.96)
	p-value	0.045
	Comment on statistical method	Age, BMI, tumour size continuous variables
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	≤50 years
	Endpoint	Recurrence (Recurrence-free survival)
	Measure	Hazard ratio (MDR+ vs. MDR-)
	n total in analysis	888
	n events / cases in analysis	Not stated

	Data	Not stated
	Adjustment	Age, BMI, tumor size, lymph node status, high ER score, high PgR score and HER2 positivity
	Point estimate	0.37
	SD	-
	SE	-
	95% CI	(0.18 to 0.76)
	p-value	0.007
	Comment on statistical method	Age, BMI, tumour size continuous variables
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	>50 years
	Endpoint	Recurrence (Recurrence-free survival)
	Measure	Hazard ratio (MDR+ vs. MDR-)
	n total in analysis	178
	n events / cases in analysis	Not stated
	Data	Not stated
	Adjustment	Age, BMI, tumor size, lymph node status, high ER score, high PgR score and HER2 positivity
	Point estimate	0.41
	SD	-
	SE	-
	95% CI	(0.52 to 3.20)
	p-value	0.4
	Comment on statistical method	Age, BMI, tumour size continuous variables
	Other	
Sources of funding and stated conflicts of interest	**Funding**	This work was supported by grant from the National Cancer Center Korea (1211200-1).
	Conflict of interest	Authors declare none.

B.VI.vi Data capture form - Li 2013

Area	Field	Data
Study design	**Type of study**	Cohort study
	Matching	None
	Prognostic, predictive or both	Prognostic (phase II)
	Control group	Group of women not treated with tamoxifen included (might be on other endocrine treatment), no interaction tested in paper (could ask them to do this).
	Prevention or treatment	Treatment
	Intervention(s)	Daily tamoxifen: 231 on 20mg, 123 on 40mg, 108 on 20+40mg, 12 on another dose. "Further adjustment for surgery (i.e., lumpectomy or mastectomy) and tamoxifen dosage, which ranged between 20 and 40 mg per day, did not appreciably change the results".
	Follow-up time period	Median 14.2 yr (range=1.0 to 15.3 yr)
Setting	**Country**	Sweden
	High-risk clinic?	No
	Treatment clinic?	Yes
	Time period	Breast cancer 1993-1995, follow-up to December 31, 2008
	Urban/rural	Not stated
Participants (and characteristics at baseline)	**No. of participants**	No tamoxifen censored=454 (90.8%), no tamoxifen event=46(9.2%), tamoxifen censored=399(84.2%), tamoxifen event=75(15.8%)
	Age (yr)	At diagnosis: 50-59/60-69/≥70: no tamoxifen=207(41.4%)/227(45.4%)/66(13.2%) median (IQR)=62(10), tamoxifen=179(37.8%)/219(46.2%)/76(16.0%) median (IQR)=63(11). Censored median (IQR)=62(10), event median (IQR)=61(10). At baseline: 49-59/60-69/≥70: no tamoxifen=207(41.4%)/230(46%)/63(12.6%) median (IQR)=61(10), tamoxifen=181(38.2)/223(47%)/70(14.8%) median (IQR)=63(10). Censored median (IQR)=62(10), event median (IQR)=62(11).
	Age <50 or ≥50 (yr)	All ≥50 years at diagnosis
	BMI (kg/m^2)	At diagnosis: <25/25-29.9/30-34.9/≥35: no tamoxifen=251(50.2%)/197(39.4%)/45(9%)/7(1.4%) median (IQR)=25(4.6), tamoxifen=246(51.9%)/168(35.4%)/51(10.8%)/9(1.9%) median (IQR)=24.8(5). Censored median (IQR)=24.8(4.8), event median (IQR)=26(4.8).
	BMI < 25, 25 to < 30, 30 to < 35, ≥35 (kg/m^2)	At diagnosis: <25/25-29.9/30-34.9/≥35: no tamoxifen=251(50.2%)/197(39.4%)/45(9%)/7(1.4%), tamoxifen=246(51.9%)/168(35.4%)/51(10.8%)/9(1.9%).
	Ethnicity	Not stated
	Education	Not stated
	Baseline risk (%)	NA
	Post/peri/pre-menopausal	All postmenopausal
	Distribution of density at baseline	0-10/11-25/26-50/51-75/>75: no tamoxifen=0(0%)/229(45.8%)/189(37.8%)/59(11.8%)/23(4.6%) median (IQR)=26.4cm^2(22), tamoxifen=0(0%)/198(41.8%)/192(40.5%)/64(13.5%)/20(4.2%) median (IQR)=28.4cm^2(23.4). Censored median (IQR)=27.3cm^2(22.9), event median (IQR)=27.7cm^2(24.1). DA on the baseline mammogram ranged from 10.8 to 135.4 cm^2 with a median of 27.4 cm^2.
	Invasive/DCIS at baseline	All invasive
	Stage (percentage	Metastatic nodes: none/1-3/4-9/>9: no

	regional spread) at baseline	tamoxifen=471(94.2%)/18(3.6%)/6(1.2%)/5(1%), tamoxifen=247(52.1%)/170(35.9%)/45(9.5%)/12(2.5%), censored=667(78.2%)/148(17.4%)/33(3.9%)/5(0.6%), event=51(42.1%)/40(33.1%)/18(14.9%)/12(9.9%).
Cointerventions	**HRT use**	No/Yes: no tamoxifen=230(46%)/270(54%), tamoxifen=235(49.6%)/239(50.4%), censored=393(46.1%)/460(53.9%), event=72(59.5%)/49(40.5%).
	Chemotherapy?	No/Yes: no tamoxifen=471(94.2%)/29(5.8%), tamoxifen=440(92.8%)/34(7.2%), censored=810(95%)/43(5%), event=101(83.5%)/20(16.5%).
	Targeted therapy?	Not stated
	Radiotherapy?	No/Yes: no tamoxifen=207(41.4%)/293(58.6%), tamoxifen=296(62.4%)/178(37.6%), censored=419(49.1%)/434(50.9%), event=84(69.4%)/37(30.6%).
	Neoadjuvant endocrine therapy?	Not stated
Timing	**Time between baseline mammogram and diagnosis**	At most 1yr before start of treatment or diagnosis date (no tamoxifen group)
	Time between diagnosis and start of endocrine therapy (or study entry)	Median 45 days after diagnosis
	Time between start of endocrine therapy (or study entry) and the follow-up mammogram	6-36 months after start of treatment or diagnosis date (no tamoxifen group)
	Time between baseline mammogram and start of endocrine therapy (or study entry)	At most 1yr before start of treatment or diagnosis date (no tamoxifen group)
	Time between baseline mammogram and the follow-up mammogram	No tamoxifen mean(SD)=1.39yr (0.48), tamoxifen mean(SD)=1.42yr (0.48). No more than 3yr.
Biomarker	**Film (digitised for density or not)/FFDM**	Digitised film
	Pre-processing for quality control of mammographic density?	Deleted bad quality mammograms (Li 2012: High-throughput mammographic-density measurement: a tool for risk prediction of breast cancer).
	Density measure(s)	Absolute (ii) automated area-based method (ImageJ), contralateral MLO
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	All (tamoxifen treated)
	Endpoint	Breast cancer mortality (time to death caused by breast cancer)
	Measure	Hazard ratio for breast density change (relative change in absolute area)
	n total in analysis	474
	n events / cases in analysis	75
	Data	HR for ≥10% change (n=113), no change (-10% to 9%; n=89); 11-20% reduction (n=55), >20% reduction (n=217)
	Adjustment	Unadjusted
	Point estimate	0.66, REF, 0.73, 0.48
	SD	-

	SE	-
	95% CI	(0.35 to 1.24), (REF), (0.35 to 1.56), (0.27 to 0.85)
	p-value	0.110 (trend: density change treated as an ordinal variable)
	Comment on statistical method	Delayed-entry Cox proportional-hazards model
	Other	The prognostic value of DA assessed from the baseline and follow-up mammogram is informative up to 15 years past diagnosis.
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	All (tamoxifen treated)
	Endpoint	Breast cancer mortality (time to death caused by breast cancer)
	Measure	Hazard ratio for breast density change (relative change in absolute area)
	n total in analysis	474
	n events / cases in analysis	75
	Data	HR for ≥10% change (n=113), no change (-10% to 9%; n=89); 11-20% reduction (n=55), >20% reduction (n=217)
	Adjustment	Time interval between baseline and follow-up mammograms (years), age at baseline mammogram (years), ever hormone replacement therapy use (yes/no), body mass index at interview (quartiles), time since menopause at baseline mammogram (years), oestrogen receptor status (positive, negative, or missing), tumour size (<10, 10-19, 20-29, 30-39, 40-49 or ≥50 mm), number of metastatic nodes (none, 1-3, 4-9 or >9), grade (well differentiated, moderately differentiated, poorly differentiated, or missing), radiotherapy treatment (yes/no), chemotherapy treatment (yes/no), change in absolute non-dense area (quartiles) and duration of tamoxifen treatment (months).
	Point estimate	0.99, REF, 0.90, 0.50
	SD	-
	SE	-
	95% CI	(0.50 to 1.94), (REF), (0.40 to 2.04), (0.27 to 0.93)
	p-value	0.017 (trend: density change treated as an ordinal variable)
	Comment on statistical method	Delayed-entry Cox proportional-hazards model
	Other	Further adjustment for surgery (i.e., lumpectomy or mastectomy) and tamoxifen dosage, which ranged between 20 and 40 mg per day, did not appreciably change the results. The prognostic value of DA assessed from the baseline and follow-up mammogram is informative up to 15 years past diagnosis.
		N.B. these results are also presented graphically. Also, there are more data on absolute dense area (absolute and relative density change measure, by quartiles) and percent density (absolute and relative density change measure, by quartiles) in supplement not included here, could be extracted.
Sources of funding and stated conflicts of interest	**Funding**	Supported by Marit and Hans Rausing's Initiative Against Breast Cancer, and by the Agency for Science, Technology and Research, Singapore (J.L.); by Grants No. W81XWH-05-1-0314 (Innovator award) from the US Department of Defense Breast Cancer Research Program, No. 523-2006-972 from the Swedish Research Council and the Swedish E-Science Research Centre (K.H.), No. 5128-B07-01PAF from the Swedish Cancer Society (K.C.), and by a postdoctoral grant from Svenska Sällskapet för Medicinsk Forskning (G.E.).
	Conflict of interest	The author(s) indicated no potential conflicts of interest

B.VI.vii Data capture form - Nyante 2015

Area	Field	Data
Study design	**Type of study**	Case-control (nested design – from larger cohort)
	Matching	Cases died of breast cancer, control patients were selected from breast cancer patients who were alive at the last tumour registry follow-up or who died from causes other than breast cancer (had at least as much follow-up time as matched cases). Matched 2:1 on age at diagnosis (≤50, 51-60, 61-70, >70years), year of diagnosis (1990-1993, 1994-1998, 1999-2002, 2003-2008) and disease stage (localised/regional spread). Mammograms were available for 61 additional controls that were matched to cases without available mammograms. To increase statistical power, these controls were rematched to eligible cases.
	Prognostic, predictive or both	Prognostic (phase II)
	Control group	None
	Prevention or treatment	Treatment
	Intervention(s)	Tamoxifen, not stated explicitly what dose but should be standard 20mg daily. At least 1 tamoxifen prescription started within 1yr of diagnosis.
	Follow-up time period	Not stated
Setting	**Country**	USA, Kaiser Permanente Northwest health plan (Portland, Oregon)
	High-risk clinic?	No
	Treatment clinic?	Yes
	Time period	Primary invasive breast cancer (study entry) between 1990 and 2008. Recruitment (follow-up) between January 1, 1991 and December 31, 2010. Checking for mammograms between January 1, 1988 and December 31, 2010. Prescription records checked between 1986 and 2010.
	Urban/rural	Kaiser Permanente Northwest health plan
Participants (and characteristics at baseline)	**No. of participants**	n=349 (97 who died from breast cancer, 252 controls)
	Age (yr)	Age at diagnosis: mean=59yr. Range: 32-87yr. ≤50/51-60/61-70/>70: cases: 29(29.9%)/22(22.7%)/28(28.9%)/18(18.6%), controls: 73(29.0%)/66(26.2%)/77(30.6%)/36(14.3%).
	Age <50 or ≥50 (yr)	≤50yr/>50yr: cases: 29(29.9%)/68(70.1%), controls: 73(29.0%)/179(71%)
	BMI (kg/m^2)	At baseline: <25/25-29/30-34/≥35/missing: cases: 22(26.2%)/29(34.5%)/17(20.2%)/16(19.1%)/13, controls: 76(33.6%)/76(33.6%)/44(19.5%)/30(13.3%)/26
	BMI < 25, 25 to < 30, 30 to < 35, ≥35 (kg/m^2)	At baseline: <25/25-29/30-34/≥35/missing: cases: 22(26.2%)/29(34.5%)/17(20.2%)/16(19.1%)/13, controls: 76(33.6%)/76(33.6%)/44(19.5%)/30(13.3%)/26
	Ethnicity	White/Non-white/missing: cases: 96(99%)/1(1%)/0, controls: 244(97.2%)/7(2.8%)/1
	Education	Not stated
	Baseline risk (%)	NA
	Post/peri/pre-menopausal	Lacked information on menopausal status
	Distribution of density at baseline	Percent density (%): Cases: mean (SD)=26.2 (16.3), median=23.7, range=3.4-80.8, ≤15/>15to23/>23to31/>31to43/>43: 29(29.9%)/19(19.6%)/16(16.5%)/19(19.6%)/14(14.4%). Controls: mean (SD)=30.0 (17.4), median 27.7, range=0.8-79.4, ≤15/>15to23/>23to31/>31to43/>43: 51(20.2%)/50(19.8%)/51(20.2%)/47(18.7%)/53(21.0%). Absolute dense area (cm^2): Cases: mean (SD)=36.5 (21.5), median 34.1, range=4.6-106.9, ≤21/>21to30/>30to42/>42to57/>57:

		29(29.9%)/17(17.5%)/18(18.6%)/19(19.6%)/14(14.4%).Controls: mean (SD)=41.0 (27.9), median 35.2, range=2.0-236.5,≤21/>21to30/>30to42/>42to57/>57: 50(19.8%)/51(20.2%)/50(19.8%)/51(20.2%)/50(19.8%).
	Invasive/DCIS at baseline	Invasive only (and ER-positive)
	Stage (percentage regional spread) at baseline	Localised/regional spread: cases: 41(42.3%)/56(57.7%), controls: 112(44.4%)/140(55.6%)
Cointerventions	**HRT use**	Nonuser/former/current: cases: 47(48.5%)/17(17.5%)/33(34.0%), controls: 104(41.3%)/37(14.7%)/111(44.1%)
	Chemotherapy?	No/Yes/missing: cases: 41(42.3%)/56(57.7%)/0, controls: 121(48.2%)/130(51.8%)/1
	Targeted therapy?	Not stated
	Radiotherapy?	No/Yes: cases: 37(38.1%)/60(61.9%), controls: 88(34.9%)/164(65.1%).
	Neoadjuvant endocrine therapy?	Not stated
Timing	**Time between baseline mammogram and diagnosis**	Baseline mammogram ≤720 days before diagnosis
	Time between diagnosis and start of endocrine therapy (or study entry)	Tamoxifen started ≤1yr after diagnosis
	Time between start of endocrine therapy (or study entry) and the follow-up mammogram	Follow-up mammogram 90 - 820 days after start of tamoxifen (and within 90 days of a current tamoxifen prescription, closest to 365 days if multiple mammograms obtained), mean=12 months, range=3-26 months
	Time between baseline mammogram and start of endocrine therapy (or study entry)	Baseline mammogram before start of treatment (or study entry for controls), mean=6 months, range=9-47 months
	Time between baseline mammogram and the follow-up mammogram	Mean 18 months, 23 (24%) cases and 58 (23%) controls more than 24 months; 40 (41%) cases and 115 (46%) controls within 12 months
Biomarker	**Film (digitised for density or not)/FFDM**	Digitised film
	Pre-processing for quality control of mammographic density?	Not stated
	Density measure(s)	Percentage (iv) semi-automated thresholding software (CUMULUS) by a single reader, contralateral CC
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	All
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	349
	n events / cases in analysis	97
	Data	OR in tertiles (controls). T1: > -0.5 (REF), T2: -8.7 to -0.5, T3: < -8.7
	Adjustment	Matching factors
	Point estimate	REF, 1.36, 0.44
	SD	-
	SE	-
	95% CI	REF, (0.79 to 2.34), (0.22 to 0.88)
	p-value	0.005 (heterogeneity test, df=2)
	Comment on statistical	Conditional logistic regression

	method	
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Baseline density <20%
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	124
	n events / cases in analysis	38
	Data	OR in tertiles (controls). T1: > -0.5 (REF), T2: -8.7 to -0.5, T3: < -8.7
	Adjustment	Matching factors
	Point estimate	REF, 2.23, NA (0 cases, 11 control)
	SD	-
	SE	-
	95% CI	REF, (0.99 to 5.03), NA
	p-value	NA
	Comment on statistical method	Unconditional logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Baseline density 20% to <37%
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	121
	n events / cases in analysis	37
	Data	OR in tertiles (controls). T1: > -0.5 (REF), T2: -8.7 to -0.5, T3: < -8.7
	Adjustment	Matching factors
	Point estimate	REF, 0.69, 0.35
	SD	-
	SE	-
	95% CI	REF, (0.28 to 1.71), (0.12 to 1.02)
	p-value	0.16 (heterogeneity test, df=2)
	Comment on statistical method	Unconditional logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Baseline density ≥37%
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	104
	n events / cases in analysis	22
	Data	OR in tertiles (controls). T1: > -0.5 (REF), T2: -8.7 to -0.5, T3: < -8.7
	Adjustment	Matching factors
	Point estimate	REF, 1.32,0.60
	SD	-
	SE	-
	95% CI	REF, (0.35 to 4.94), (0.17 to 2.12)
	p-value	0.38 (heterogeneity test, df=2)

	Comment on statistical method	Unconditional logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	All
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	349
	n events / cases in analysis	97
	Data	OR in tertiles (controls). T1: > -0.5 (REF), T2: -8.7 to -0.5, T3: < -8.7
	Adjustment	Matching factors plus baseline density
	Point estimate	REF, 1.38, 0.49
	SD	-
	SE	-
	95% CI	REF, (0.80 to 2.40), (0.23 to 1.02)
	p-value	0.01 (heterogeneity test, df=2)
	Comment on statistical method	Conditional logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Baseline density <20%
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	124
	n events / cases in analysis	38
	Data	OR in tertiles (controls). T1: > -0.5 (REF), T2: -8.7 to -0.5, T3: < -8.7
	Adjustment	Matching factors plus baseline density
	Point estimate	REF, 2.82, NA (0 cases, 11 control)
	SD	-
	SE	-
	95% CI	REF, (1.17 to 6.76), NA
	p-value	NA
	Comment on statistical method	Unconditional logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Baseline density 20% to <37%
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	121
	n events / cases in analysis	37
	Data	OR in tertiles (controls). T1: > -0.5 (REF), T2: -8.7 to -0.5, T3: < -8.7
	Adjustment	Matching factors plus baseline density
	Point estimate	REF, 0.70, 0.35
	SD	-
	SE	-
	95% CI	REF, (0.28 to 1.72), (0.12 to 1.02)

	p-value	0.16 (heterogeneity test, df=2)
	Comment on statistical method	Unconditional logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Baseline density ≥37%
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	104
	n events / cases in analysis	22
	Data	OR in tertiles (controls). T1: > -0.5 (REF), T2: -8.7 to -0.5, T3: < -8.7
	Adjustment	Matching factors plus baseline density
	Point estimate	REF, 1.34,0.59
	SD	-
	SE	-
	95% CI	REF, (0.36 to 5.02), (0.17 to 2.11)
	p-value	0.38 (heterogeneity test, df=2)
	Comment on statistical method	Unconditional logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	All
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	349
	n events / cases in analysis	97
	Data	OR in tertiles (controls). T1: > -0.5 (REF), T2: -8.7 to -0.5, T3: < -8.7
	Adjustment	Matching factors plus baseline density plus tamoxifen duration
	Point estimate	REF, 1.27, 0.47
	SD	-
	SE	-
	95% CI	REF, (0.71 to 2.25), (0.21 to 1.03)
	p-value	0.04 (heterogeneity test, df=2)
	Comment on statistical method	Conditional logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Baseline density <20%
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	124
	n events / cases in analysis	38
	Data	OR in tertiles (controls). T1: > -0.5 (REF), T2: -8.7 to -0.5, T3: < -8.7
	Adjustment	Matching factors plus baseline density plus tamoxifen duration
	Point estimate	REF, 2.22, NA (0 cases, 11 control)
	SD	-
	SE	-

	95% CI	REF, (0.88 to 5.62), NA
	p-value	NA
	Comment on statistical method	Unconditional logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Baseline density 20% to <37%
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	121
	n events / cases in analysis	37
	Data	OR in tertiles (controls). T1: > -0.5 (REF), T2: -8.7 to -0.5, T3: < -8.7
	Adjustment	Matching factors plus baseline density plus tamoxifen duration
	Point estimate	REF, 0.78, 0.38
	SD	-
	SE	-
	95% CI	REF, (0.31 to 1.96), (0.13 to 1.15)
	p-value	0.23 (heterogeneity test, df=2)
	Comment on statistical method	Unconditional logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Baseline density ≥37%
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	104
	n events / cases in analysis	22
	Data	OR in tertiles (controls). T1: > -0.5 (REF), T2: -8.7 to -0.5, T3: < -8.7
	Adjustment	Matching factors plus baseline density plus tamoxifen duration
	Point estimate	REF, 1.31, 0.57
	SD	-
	SE	-
	95% CI	REF, (0.30 to 5.68), (0.14 to 2.38)
	p-value	0.43 (heterogeneity test, df=2)
	Comment on statistical method	Unconditional logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	All
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	349
	n events / cases in analysis	97
	Data	OR by 10% cut-off: <10% reduction, ≥10% reduction. 14 cases and 70 controls with ≥10% reduction
	Adjustment	Matching factors
	Point estimate	REF, 0.42
	SD	-

	SE	-
	95% CI	REF, (0.22 to 0.80)
	p-value	0.009
	Comment on statistical method	Conditional logistic regression, relatively small numbers in ≥10% group
	Other	Reported in supplementary material, tertiles was primary analysis (but this is cut-point used by others)
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	All
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	349
	n events / cases in analysis	97
	Data	OR by 10% cut-off: <10% reduction, ≥10% reduction. 14 cases and 70 controls with ≥10% reduction
	Adjustment	Matching factors plus baseline density
	Point estimate	REF, 0.47
	SD	-
	SE	-
	95% CI	REF, (0.23 to 0.94)
	p-value	0.03
	Comment on statistical method	Conditional logistic regression, relatively small numbers in ≥10% group
	Other	Reported in supplementary material, tertiles was primary analysis (but this is cut-point used by others)
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Baseline density <20%
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	124
	n events / cases in analysis	38
	Data	OR by 10% cut-off: <10% reduction, ≥10% reduction. 0 cases and 6 controls with ≥10% reduction
	Adjustment	Matching factors
	Point estimate	REF, NA (0 cases 6 controls≥ 10% change)
	SD	-
	SE	-
	95% CI	NA
	p-value	NA
	Comment on statistical method	Unconditional logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Baseline density <20%
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	124
	n events / cases in analysis	38
	Data	OR by 10% cut-off: <10% reduction, ≥10% reduction. 0 cases and 6 controls with ≥10% reduction

	Adjustment	Matching factors plus baseline density
	Point estimate	REF, NA (0 cases 6 controls≥ 10% change)
	SD	-
	SE	-
	95% CI	NA
	p-value	NA
	Comment on statistical method	Unconditional logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Baseline density 20-37%
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	121
	n events / cases in analysis	37
	Data	OR by 10% cut-off: <10% reduction, ≥10% reduction. 7 cases and 23 controls with ≥10% reduction
	Adjustment	Matching factors
	Point estimate	REF, 0.60
	SD	-
	SE	-
	95% CI	REF, (0.22 to 1.59)
	p-value	0.3
	Comment on statistical method	Unconditional logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Baseline density 20-37%
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	121
	n events / cases in analysis	37
	Data	OR by 10% cut-off: <10% reduction, ≥10% reduction. 7 cases and 23 controls with ≥10% reduction
	Adjustment	Matching factors plus baseline density
	Point estimate	REF, 0.59
	SD	-
	SE	-
	95% CI	REF, (0.21 to 1.60)
	p-value	0.3
	Comment on statistical method	Unconditional logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Baseline density ≥37%
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	104
	n events / cases in analysis	22
	Data	OR by 10% cut-off: <10% reduction, ≥10% reduction. 7 cases and

		41 controls with ≥10% reduction
	Adjustment	Matching factors
	Point estimate	REF, 0.40
	SD	-
	SE	-
	95% CI	REF, (0.14 to 1.14)
	p-value	0.09
	Comment on statistical method	Unconditional logistic regression
	Other	
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Baseline density ≥37%
	Endpoint	Breast cancer mortality
	Measure	Change in percentage density
	n total in analysis	104
	n events / cases in analysis	22
	Data	OR by 10% cut-off: <10% reduction, ≥10% reduction. 7 cases and 41 controls with ≥10% reduction
	Adjustment	Matching factors plus baseline density
	Point estimate	REF, 0.39
	SD	-
	SE	-
	95% CI	REF, (0.14 to 1.14)
	p-value	0.09
	Comment on statistical method	Unconditional logistic regression
	Other	
		N.B there are more data on absolute dense area, relative percent density and relative absolute dense area in supplement not included here. Also age subgroup results and supplementary subgroups for treatment duration etc. presented graphically, could be extracted.
Sources of funding and stated conflicts of interest	**Funding**	This work was supported by the Intramural Research Program of the National Cancer Institute at the National Institutes of Health.
	Conflict of interest	None stated in paper (not present).

B.VI.viii Data capture form - Sandberg 2013

Area	Field	Data
Study design	**Type of study**	Case-control
	Matching	Cases had invasive CBC diagnosed more than 1 year after the first invasive cancer and with an available mammogram close to the first diagnosis, controls had invasive unilateral breast cancer in the same register (no CBC). Calendar period (+/- 2yr) of first breast cancer diagnosis, age at diagnosis (+/- 2yr), adjuvant therapy, follow-up time (control survived without distant metastasis or CBC at least as long as time between first and subsequent cancer).
	Prognostic, predictive or both	Prognostic (phase II)
	Control group	None
	Prevention or treatment	Treatment
	Intervention(s)	None (radiotherapy, endocrine therapy and/or chemotherapy (or none) administered for breast cancer treatment). Specific treatments, doses and intake frequency not reported.
	Follow-up time period	Mean 8.25yr in both cases and controls
Setting	**Country**	Sweden
	High-risk clinic?	No
	Treatment clinic?	Yes
	Time period	1976 to 2005
	Urban/rural	Stockholm-Gotland health-care region
Participants (and characteristics at baseline)	**No. of participants**	n=422 (211 cases, 211 controls)
	Age (yr)	≤45yr/45-55yr/55-65yr/≥65yr: Cases=37(18%) /68(32%)/56(27%)/50(24%), Controls=37(18%)/68(32%)/56(27%)/50(24%), same proportion in cases and controls by design
	Age <50 or ≥50 (yr)	≤45yr/45-55yr/55-65yr/≥65yr: Cases=37(18%) /68(32%)/56(27%)/50(24%), Controls=37(18%)/68(32%)/56(27%)/50(24%), same proportion in cases and controls by design
	BMI (kg/m^2)	Not available, but fat area used as proxy: Q1(≤67cm^2)/Q2(67-93cm^2)/Q3(93-127cm^2)/Q4(≥127cm^2): Cases=42(20%)/58(27%)/50(24%)/61(29%), Controls=60(28%)/43(20%)/55(26%)/53(25%)
	BMI < 25, 25 to < 30, 30 to < 35, ≥35 (kg/m^2)	Not stated
	Ethnicity	Not stated
	Education	Not stated
	Baseline risk (%)	NA
	Post/peri/pre-menopausal	Pre/Post: Cases=89(42%)/119(56%), Controls=84(40%)/124(59%). Six patients had uncertain menopause status (for example, hysterectomy).
	Distribution of density at baseline	Mean PDA at baseline=28%. PDA Q1(≤5%)/Q2(5-25%)/Q3(25-50%)/Q4(≥50%): Cases=13(6%)/87(41%)/97(46%)/14(7%), Controls=11(5%)/87(41%)/87(41%)/26(12%). DA Q1(≤20cm^2)/Q2(20-34cm^2)/Q3(34-53cm^2)/Q4(≥53cm^2): Cases=55(26%)/44(21%)/56(27%)/56(27%), Controls=55(26%)/56(27%)/49(23%)/51(24%)
	Invasive/DCIS at baseline	All invasive
	Stage (percentage regional spread) at baseline	Tumour-node metastasis stage (cases and controls combined): 1/2/3/unknown: 244/157/16/5. Note 53 ER-negative, 295 ER-positive.
Cointerventions	**HRT use**	Current use at diagnosis/no current use at

		diagnosis/postmenopausal unknown HRT status: 51(21%)/127(52%)/65(27%)
	Chemotherapy?	(W/ or w/o radiotherapy and/or endocrine therapy): cases=28(13% of adjuvant therapies), controls=28(13%)
	Targeted therapy?	Not stated
	Radiotherapy?	(Radiotherapy only): cases=57(27% of adjuvant therapies), controls=57(27%)
	Neoadjuvant endocrine therapy?	Not stated
Timing	**Time between baseline mammogram and diagnosis**	First available mammogram up to 1yr prior to diagnosis, and up to 2 weeks after diagnosis
	Time between diagnosis and start of endocrine therapy (or study entry)	Not stated
	Time between start of endocrine therapy (or study entry) and the follow-up mammogram	Not stated, but time from diagnosis to first available follow-up mammogram (1-5yr after diagnosis): mean=1.6yr, 90% between 1 and 2.2yr (cases and controls combined), cases mean(SD)=1.56yr(0.59), controls mean(SD)=1.54yr(0.57)
	Time between baseline mammogram and start of endocrine therapy (or study entry)	Not stated
	Time between baseline mammogram and the follow-up mammogram	Not stated
Biomarker	**Film (digitised for density or not)/FFDM**	Digitised film
	Pre-processing for quality control of mammographic density?	Yes, poor-quality excluded (88 cases excluded for this reason)
	Density measure(s)	Percentage (v) fully-automated (based on area of density) (ImageJ) and absolute (ii) automated area-based methods (ImageJ), 86% MLO & 14% CC (same view at baseline and follow-up mammograms and same view in matched cases and controls
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Entire sample (minus 66 with percentage density <10% or >90% at baseline since they could not move into extreme density change groups)
	Endpoint	Incidence of a secondary primary breast cancer (e.g. in the contralateral breast)
	Measure	Percentage density (≥10% reduction/<10% reduction to <10% increase/increase ≥10%)
	n total in analysis	356
	n events / cases in analysis	178
	Data	≥10% reduction/<10% reduction to <10% increase/increase ≥10%: 96/243/17
	Adjustment	Matching factors
	Point estimate	≥10% reduction OR=0.49, REF, increase ≥10% OR=0.74
	SD	-
	SE	-
	95% CI	(0.28-0.85), REF, (0.23-2.40)
	p-value	0.04 (P-trend)
	Comment on statistical method	Conditional logistic regression
	Other	This analysis does not focus on endocrine treated group, but is the main analysis in the paper, number of events not stated but calculated from 1:1 matching and use of conditional logistic

		regression.
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Entire sample (minus 84 with area density <10cm^2 or 70cm^2 at baseline since they could not move into extreme density change groups)
	Endpoint	Incidence of a secondary primary breast cancer (e.g. in the contralateral breast)
	Measure	Area density (≥10cm^2 reduction/<10cm^2 reduction to <10cm^2 increase/increase ≥10cm^2)
	n total in analysis	338
	n events / cases in analysis	169
	Data	≥10cm^2 reduction/<10cm^2 reduction to <10cm^2 increase/increase ≥10cm^2: 108/197/33
	Adjustment	Matching factors
	Point estimate	≥10cm^2 reduction OR=0.67, REF, increase ≥10cm^2 OR=0.79
	SD	-
	SE	-
	95% CI	(0.38-1.16), REF, (0.35-1.78)
	p-value	0.35 (P-trend)
	Comment on statistical method	Conditional logistic regression
	Other	This analysis does not focus on endocrine treated group, but is the main analysis in the paper, number of events not stated but calculated from 1:1 matching and use of conditional logistic regression.
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Entire sample (minus 66 with percentage density <10% or >90% at baseline since they could not move into extreme density change groups)
	Endpoint	Incidence of a secondary primary breast cancer (e.g. in the contralateral breast)
	Measure	Percentage density (≥10% reduction/<10% reduction to <10% increase/increase ≥10%)
	n total in analysis	356
	n events / cases in analysis	178
	Data	≥10% reduction/<10% reduction to <10% increase/increase ≥10%: 96/243/17
	Adjustment	Matching factors plus percentage density and non-dense area at first mammogram (both categorised in quartiles)
	Point estimate	≥10% reduction OR=0.45, REF, increase ≥10% OR=0.83
	SD	-
	SE	-
	95% CI	(0.24-0.84), REF, (0.24-2.87)
	p-value	0.04 (P-trend)
	Comment on statistical method	Conditional logistic regression
	Other	This analysis does not focus on endocrine treated group, but is the main analysis in the paper, number of events not stated but calculated from 1:1 matching and use of conditional logistic regression.
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Entire sample (minus 84 with area density <10cm^2 or 70cm^2 at baseline since they could not move into extreme density change groups)
	Endpoint	Incidence of a secondary primary breast cancer (e.g. in the contralateral breast)

	Measure	Area density (≥10cm^2 reduction/<10cm^2 reduction to <10cm^2 increase/increase ≥10cm^2)
	n total in analysis	338
	n events / cases in analysis	169
	Data	≥10cm^2 reduction/<10cm^2 reduction to <10cm^2 increase/increase ≥10cm^2: 108/197/33
	Adjustment	Matching factors plus non-dense area at first mammogram (and dense area since indicated in methods but not table legend?) (both categorised in quartiles)
	Point estimate	≥10cm^2 reduction OR=0.54, REF, increase ≥10cm^2 OR=0.71
	SD	-
	SE	-
	95% CI	(0.30-0.99), REF, (0.30-1.69)
	p-value	0.13 (P-trend)
	Comment on statistical method	Conditional logistic regression
	Other	This analysis does not focus on endocrine treated group, but is the main analysis in the paper, number of events not stated but calculated from 1:1 matching and use of conditional logistic regression.
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Endocrine therapy only (w/ or w/o radiotherapy)
	Endpoint	Incidence of a secondary primary breast cancer (e.g. in the contralateral breast)
	Measure	Percentage density (≥10% reduction/<10% reduction to <10% increase/increase ≥10%)
	n total in analysis	Cases=87(41% were on endocrine therapy), controls=87(41% were on endocrine therapy) - but not clear if all of these women were included in the analysis
	n events / cases in analysis	Cases=87(41% were on endocrine therapy), controls=87(41% were on endocrine therapy) - but not clear if all of these women were included in the analysis
	Data	Not stated
	Adjustment	Unclear - matched factors (plus percentage density and non-dense area?)
	Point estimate	≥10% reduction OR=0.52, REF, increase ≥10% not stated
	SD	-
	SE	-
	95% CI	(0.18-1.51), REF, not stated
	p-value	Not stated (but p>0.05)
	Comment on statistical method	Conditional logistic regression. Numbers, adjustments and results unclear.
	Other	This is relevant for the review.
		N.B. other results for the entire sample adjusted for HRT, and by menopausal status and mammographic view, but these are not in the endocrine therapy only group so are not relevant for this review.
Sources of funding and stated conflicts of interest	**Funding**	This work was supported by the Swedish Research Council [grant no: 521-2008-2728]; Swedish Cancer Society [grant no: CAN 2010/807]; Cancer Research UK [grant no: C405/A8406] A*STAR Graduate Scholarship to JL; the Swedish Research Council [grant no: 523-2006-97 to KH]; and the Swedish Cancer Society [grant no: 5128 B07-01PAF to KC].
	Conflict of interest	The authors declare that they have no competing interests.

B.VI.ix Data capture form - van Nes 2015

Area	Field	Data
Study design	**Type of study**	Sub-cohort within a randomised trial
	Matching	None
	Prognostic, predictive or both	Prognostic (phase II)
	Control group	None
	Prevention or treatment	Treatment
	Intervention(s)	Tamoxifen for 2–3 years followed by exemestane for 3–2 years (totalling five years) or exemestane alone for 5 years. Tamoxifen: 20 mg once a day, orally; Exemestane: 25 mg once a day, orally.
	Follow-up time period	Median 6yr (range 0-9yr)
Setting	**Country**	Netherlands
	High-risk clinic?	No
	Treatment clinic?	Yes
	Time period	Not stated - TEAM trial enrolment in 2001 but period of this study not reported
	Urban/rural	13 hospitals out of 76 included in the TEAM trial (92 Dutch hospitals in total)
Participants (and characteristics at baseline)	**No. of participants**	Exemestane n=197, sequential n=181
	Age (yr)	Median (range): Sequential = 63 years (48–91), Exemestane = 62 years (45–86). <50/50-59/60-69/≥70: Sequential = 3(2%)/66(36%)/69(38%)/43(24%), Exemestane = 8(4%)/67(34%)/66(34%)/56(28%).
	Age <50 or ≥50 (yr)	<50/≥50: Sequential = 3(2%)/178(98%), Exemestane = 8(4%)/189(96%).
	BMI (kg/m^2)	<25/25-30/≥30: Sequential = 61(36%)/68(40%)/39(23%), Exemestane = 79(42%)/70(37%)/38(20%)
	BMI < 25, 25 to < 30, 30 to < 35, ≥35 (kg/m^2)	<25/25-30/≥30: Sequential = 61(36%)/68(40%)/39(23%), Exemestane = 79(42%)/70(37%)/38(20%)
	Ethnicity	Not stated
	Education	Not stated
	Baseline risk (%)	NA
	Post/peri/pre-menopausal	All postmenopausal
	Distribution of density at baseline	Given as figure (by radiologists' scores), can be extracted
	Invasive/DCIS at baseline	All invasive
	Stage (percentage regional spread) at baseline	pT1/pT2/pT3 or 4: Sequential = 79(44%)/86(48%)/15(8%), Exemestane = 91(47%)/97(50%)/7(4%). Nodal status -/+: Sequential = 59(33%)/122(67%), Exemestane = 61(31%)/136(69%)
Cointerventions	**HRT use**	Not stated
	Chemotherapy?	No/Yes: Sequential = 199(66%)/62(34%), Exemestane = 142(72%)/55(28%)
	Targeted therapy?	Not stated
	Radiotherapy?	No/Yes: Sequential = 69(38%)/112(62%), Exemestane = 70(36%)/126(64%)
	Neoadjuvant endocrine therapy?	Not stated
Timing	**Time between baseline mammogram and diagnosis**	(Baseline mammograms were preoperative)
	Time between diagnosis and start of endocrine therapy (or study entry)	Not stated
	Time between start of	T1 (range 6–18 months), T2 (range 18–30 months), and T3

	endocrine therapy (or study entry) and the follow-up mammogram	(range 30–42 months)
	Time between baseline mammogram and start of endocrine therapy (or study entry)	Not stated
	Time between baseline mammogram and the follow-up mammogram	Not stated
Biomarker	**Film (digitised for density or not)/FFDM**	Film (FFDM excluded)
	Pre-processing for quality control of mammographic density?	Not stated
	Density measure(s)	Percentage (ii) visual assessment by three experienced radiologists in 20% bands (Boyd categories), contralateral CC
Results (add new rows each subgroup and endpoint combo)	**Subgroup**	Both treatment arms combined
	Endpoint	Recurrence and Incidence of a secondary primary breast cancer (e.g. in the contralateral breast) combined (all of the study endpoints were included as the outcome: loco-regional recurrence, distance recurrence or contralateral breast cancer)
	Measure	"Change in breast density"
	n total in analysis	378 (Table IB: "Included in the current analysis" left-hand column is the column to use)
	n events / cases in analysis	Loco-regional recurrence cases (4 in the sequential arm, 5 in the exemestane arm), distance recurrence cases (28 in the sequential arm, 20 in the exemestane arm), and contralateral breast cancer cases (4 in the sequential arm, 3 in the exemestane arm)
	Data	Not stated
	Adjustment	Not stated
	Point estimate	"No association between change in breast density and the occurrence of an event"
	SD	-
	SE	-
	95% CI	Not stated
	p-value	Not stated
	Comment on statistical method	Cox regression?
	Other	No analysis of change in density reported.
Source of funding, conflicts	**Funding**	Pfizer, unrestricted research grant.
	Conflict of interest	The authors report no conflicts of interest.

B.VII Risk of bias tables

B.VII.i Risk of bias table - Cuzick 2011

Biases	Issues to consider for judging overall rating of risk of bias	Study Methods & Comments	Rating of reporting (adequacy of reporting: "yes", "partial", "no" or "unsure")	Rating of Risk of bias ("High", "Moderate", or "Low")
Instructions to assess the risk of each potential bias	These issues will guide your thinking and judgement about the overall risk of bias within each of the six domains. These issues are taken together to inform the overall judgement of potential bias for each of the six domains			
1. Study participation	Goal: to judge the risk of selection bias (likelihood that relationship between density reductions and outcome is different for participants and eligible non-participants)			
Source of target population	The source population or population of interest is adequately described for: a) treatment: (i) proportion with DCIS, (ii) cointerventions (chemotherapy/targeted therapy), (iii) severity of cancer at baseline (stage, % regional spread); b) prevention: (i) level of risk in population, including whether some or all are BRCA1/2 mutation carriers, (ii) prior hormone replacement therapy use, (iii) cointerventions such as diet or exercise regimens, or both	The source population is described, based on entry criteria to the IBIS-I trial. "To be eligible for IBIS-I, a woman had to be between 35 and 70 years old and have at least twice the average risk of a 50-year-old woman of developing breast cancer (14). Typically, therefore, an IBIS-I participant would have either a history of benign proliferative breast disease or a strong family history of breast cancer (i.e., a mother or sister who developed breast cancer before age 50 years). The absolute observed 10-year risk of developing breast cancer in the placebo arm of the main study was 6.4% (15).", although no mention of BRCA 1/2 mutation carriers. Prior hormone therapy use is included. There were no cointerventions in the trial.	Yes	
Method used to identify population	The sampling frame and recruitment are adequately described, including methods to identify the sample sufficient to limit potential bias	The sampling frame is well described. "To minimize the administrative workload, control subjects were selected only from the major participating UK centers in Aberdeen, Bristol, Cardiff, Edinburgh, London, Manchester, Nottingham, and Southampton. We identified 1064 potential control subjects (women who had completed 5 years of treatment with full compliance and had not	Yes	

		developed breast cancer) and requested the mammograms for these women. A total of 942 complete sets of mammograms (baseline and first follow-up) were recovered. The missing mammograms had either been lost or destroyed in accordance with the local archiving policy". UK and Finnish cases.		
Recruitment period	Period of recruitment is adequately described	Well described (see data capture form).	Yes	
Place of recruitment	Place of recruitment (setting and geographic location) are adequately described	Well described (see data capture form).	Yes	
Inclusion and exclusion criteria	Inclusion and exclusion criteria are adequately described	Well described e.g. three cases who were diagnosed within the first 12 months of treatment were excluded from the analysis: "Three of these women had been diagnosed within the first 12 months on study and were excluded from the analysis".	Yes	
Adequate study participation	There is adequate participation in the study by eligible individuals	942 of 1064 controls; 126 of 224 cases from the centres. Films were obtainable for 55% of the IBIS-I cases from the UK and Finland. No data to compare included and excluded samples, but reported that this was examined. "The control subjects who were selected for this case–control study did not differ with respect to demographic factors from the IBIS-I control subjects who were not selected (data not shown)" and "These women did not differ from the IBIS-I case subjects who were not selected for this case–control study with respect to demographic factors or tumor characteristics (data not shown)".	Yes	
Baseline characteristics	The baseline study sample (i.e. individuals entering the study) is adequately described for (treatment and prevention) age, menopausal status, cointerventions; (treatment) % DCIS, disease severity; (prevention) breast cancer risk, prior hormone replacement therapy use	Well described (see data capture form).	Yes	
Summary study participation	The study sample represents the population of interest on key characteristics, sufficient to limit potential bias of the observed relationship between density change and outcome			Low
2. Study attrition	Goal: to judge the risk of attrition bias (likelihood that relationship between density reductions and outcome are different for			

	completing and non-completing participants)			
Proportion of baseline sample available for analysis	Response rate (i.e. proportion of study sample allocated treatment who received treatment) is adequate	All controls complied over the full 5yr follow-up period and were not censored due to loss to follow-up: "women who had completed 5 years of treatment with full compliance and had not developed breast cancer".	Yes	
Attempts to collect information on participants who dropped out	Attempts to collect information on participants who dropped out of the study are described	Not described, but withdrawals only reported n=44 Australian women in *Cuzick 2015; Lancet Oncology; 16(1): 67-75* (and not included in density sub-study).	Partial	
Reasons and potential impact of subjects lost to follow-up	Reasons for loss to follow-up are provided	Not described, but withdrawals only reported n=44 Australian women in *Cuzick 2015; Lancet Oncology; 16(1): 67-75* (and not included in density sub-study).	Partial	
Outcome and prognostic factor information on those lost to follow-up	Participants lost to follow-up are adequately described for age at entry and cointerventions (if any), and for a) treatment: (i) DCIS, (ii) disease severity; b) prevention: (i) risk of breast cancer including BRCA1/2 carriers and testing. Whether loss to follow-up or inability to retrieve mammograms, or both, was likely related to the study outcome	NA	Yes	
Study attrition summary	There are no important differences between these characteristics in participants who completed the study and those who did not. Loss to follow-up (from baseline sample to study population analysed) is not associated with key characteristics (i.e. the study data adequately represent the sample) sufficient to limit potential bias to the observed relationship between density change and outcome			Low
3. Prognostic factor measurement	Goal: to judge the risk of measurement bias related to how mammographic density was measured (differential measurement of mammographic density related to the level of outcome)			
Definition of the prognostic factor	A clear definition or description of mammographic density is provided (e.g. including the method of measurement, if subjective then who undertook it, if treatment then whether contralateral breast assessed)	Clear description provided: "visually estimated the proportion of the total breast area that was composed of dense tissue (to the nearest 5%)", expert radiologist: "Mammographic density was assessed visually by one radiologist (R. M. L. Warren)", contralateral breast: "The assessment of mammographic density for both case subjects and	Yes	

		control subjects was based on a composite assessment of both the left and right mediolateral-oblique views, except for the 13 case subjects who were diagnosed at the first follow-up mammogram; those assessments were made using only the film for the contralateral breast".		
Valid and reliable measurement of prognostic factor	Method of mammographic density change measurement is adequately valid and reliable to limit misclassification bias (e.g. may include relevant outside sources of information on measurement properties; also characteristics, such as measurement blinded to case status)	High inter and intra-reader correlation: "correlation between the original and repeat mammographic breast density assessment readings was very high: the Pearson correlation coefficient for the baseline mammograms was 0.98 (95% CI = 0.96 to 0.99), for the follow-up mammograms was 0.97 (95% CI = 0.97 to 0.99), and for the density change over 12–18 months was 0.78 (95% CI = 0.63 to 0.87)" and "correlation between the original mammographic breast density assessment by R.M. L. Warren and the subsequent mammographic breast density assessments by the other four readers was also very high for the baseline mammograms (r ranged from 0.86 to 0.90) and for the follow-up mammograms (r ranged from 0.87 to 0.91) but was only moderate for breast density change over 12–18 months (r ranged from 0.48 to 0.67)", blinded: "original reader was blinded with regard to treatment allocation (tamoxifen or placebo) but not with regard to case–control status. However, the fully blinded rereading of the original films for 40 control subjects and eight case subjects indicated high reproducibility of the original breast density assessment".	Partial	
	Continuous variables are reported or appropriate cut-points (i.e. not data-dependent (except for percentiles)) are used	"The 10% cut point was chosen because it was the minimum change that could be reproducibly detected".	Partial	
Method and setting of prognostic factor measurement	The method and setting of measurement of mammographic density is the same for all study participants. The same mammogram type (film/digital) is used for both baseline and follow-up. The time at which baseline and follow-up mammograms have low variability between participants	Method and setting same ("Mammographic density was assessed visually by one radiologist (R. M. L. Warren)"), same mammogram type (film). Median (IQR) baseline to follow-up=19 (18-23) months.	Yes	
Proportion of data on	Adequate proportion of the study sample has complete data for the	All have complete data by definition of study design.	Yes	

prognostic factor available for analysis	change in mammographic density variable			
Method used for missing data	Appropriate methods of imputation are used for missing mammographic density data	NA	Yes	
Summary	Prognostic factor is adequately measured in study participants to sufficiently limit potential bias			Moderate
4. Outcome measurement	Goal: to judge the risk of bias related to the measurement of outcome (differential measurement of outcome related to the density reductions)			
Definition of the outcome	A clear definition of outcome is provided, including duration of follow-up and level and extent of the outcome construct	Clear definition, outcome is breast cancer diagnosis: "to examine associations between change in mammographic density and the risk of breast cancer".	Yes	
Valid and reliable measurement of outcome	The method of outcome measurement used is adequately valid and reliable to limit misclassification bias	From IBIS-I trial database.	Yes	
Method and setting of outcome measurement	The method and setting of outcome measurement is the same for all study participants, including by age and obesity groups	Yes.	Yes	
Outcome measurement summary	Outcome of interest is adequately measured in study participants to sufficiently limit potential bias			Low
5. Study confounding	Goal: to judge the risk of bias due to confounding (i.e. the effect of density reductions is distorted by another factor that is related to density reductions and the outcome)			
Important confounders measured	Age, BMI, or another measure of adiposity are measured	Age and BMI measured.	Yes	
Definition of the confounding factor	Clear definitions are provided	"Age […] at entry to IBIS-I, body mass index (as a continuous variable)".	Yes	
Valid and reliable measurement of confounders	Measurement of all important confounders is adequately valid and reliable	Measurement is adequately valid.	Yes	
Method and setting of confounding measurement	The method and setting of confounding measurement are the same for all study participants	Yes.	Yes	
Method used for missing data	Appropriate methods are used if imputation is used for missing confounder data	1065 women, 1049 in main result (16 missing BMI) so no imputation used.	Yes	
Appropriate accounting for confounding	The primary analysis will be adjusted for at least age, either through the study design and analysis, or through adjustment in the analysis only; and other prognostic factors	Yes, although no adjustment for change in BMI.	Yes	
Study confounding summary	Important potential confounders are appropriately accounted for, limiting potential bias with respect to the relationship between prognostic factor and outcome			Low

6. Statistical analysis and reporting	Goal: to judge the risk of bias related to the statistical analysis and presentation of results			
Presentation of analytical strategy, model development strategy	There is sufficient presentation of data to assess the adequacy of the analysis	Yes.	Yes	
Model development strategy	The strategy for model building (i.e. inclusion of variables in the statistical model) is appropriate and is based on a conceptual framework or model	Conceptual framework – adjusted density change analysis for other prognostic factors or those associated with density. Does not appear to be variable selection.	Yes	
Reporting of results	The selected statistical model is adequate for the design of the study. There is no selective reporting of results	Yes, adequate. However, an interaction test between density change and treatment arm was not reported.	Partial	
Statistical analysis and presentation summary	The statistical analysis is appropriate for the design of the study, limiting potential for presentation of invalid or spurious results			Low

B.VII.ii Risk of bias table - Kim 2012

Biases	Issues to consider for judging overall rating of risk of bias	Study Methods & Comments	Rating of reporting (adequacy of reporting: "yes", "partial", "no" or "unsure")	Rating of Risk of bias ("High", "Moderate", or "Low")
Instructions to assess the risk of each potential bias	These issues will guide your thinking and judgement about the overall risk of bias within each of the six domains. These issues are taken together to inform the overall judgement of potential bias for each of the six domains			
1. Study participation	Goal: to judge the risk of selection bias (likelihood that relationship between density reductions and outcome is different for participants and eligible non-participants)			
Source of target population	The source population or population of interest is adequately described for: a) treatment: (i) proportion with DCIS, (ii) cointerventions (chemotherapy/targeted therapy), (iii) severity of cancer at baseline (stage, % regional spread); b) prevention: (i) level of risk in population, including whether some or all are BRCA1/2 mutation carriers, (ii) prior hormone replacement therapy use, (iii) cointerventions such as diet or exercise regimens, or both	Source population is not described, only analysis population.	No	
Method used to identify population	The sampling frame and recruitment are adequately described, including methods to identify the sample sufficient to limit potential bias	Well described: "1,542 ER-positive breast cancer patients who underwent curative surgery at Seoul National University Hospital between October 2003 and December 2006".	Yes	
Recruitment period	Period of recruitment is adequately described	Well described (see data capture form).	Yes	
Place of recruitment	Place of recruitment (setting and geographic location) are adequately described	Well described (see data capture form).	Yes	
Inclusion and exclusion criteria	Inclusion and exclusion criteria are adequately described	Adequately described: "Patients were excluded if: 1) they did not receive adjuvant endocrine treatment, such as tamoxifen or an aromatase inhibitor, or were treated for less than 2 years; 2) their digital mammogram images were not available; 3) they had bilateral breast cancer, or 4) distant metastasis was observed before the start of endocrine therapy", but no information on number of exclusions based on each exclusion criteria.	Partial	

Adequate study participation	There is adequate participation in the study by eligible individuals	1065 of 1542 women included. No information on participation based on consent. No information on eligible participants vs. those not eligible (in source population or in 1542 women).	Partial	
Baseline characteristics	The baseline study sample (i.e. individuals entering the study) is adequately described for (treatment and prevention) age, menopausal status, cointerventions; (treatment) % DCIS, disease severity; (prevention) breast cancer risk, prior hormone replacement therapy use	Well described (see data capture form).	Yes	
Summary study participation	The study sample represents the population of interest on key characteristics, sufficient to limit potential bias of the observed relationship between density change and outcome			Moderate
2. Study attrition	Goal: to judge the risk of attrition bias (likelihood that relationship between density reductions and outcome are different for completing and non-completing participants)			
Proportion of baseline sample available for analysis	Response rate (i.e. proportion of study sample allocated treatment who received treatment) is adequate	Only included patients with at least 2yr of endocrine treatment but discrepancy since "mean duration of overall endocrine therapy was 5.1 years (range, 0.9 to 7.9 years)".	Partial	
Attempts to collect information on participants who dropped out	Attempts to collect information on participants who dropped out of the study are described	No information found on drop out or reasons for censoring.	No	
Reasons and potential impact of subjects lost to follow-up	Reasons for loss to follow-up are provided	No information found on loss to follow-up or reasons for censoring.	No	
Outcome and prognostic factor information on those lost to follow-up	Participants lost to follow-up are adequately described for age at entry and cointerventions (if any), and for a) treatment: (i) DCIS, (ii) disease severity; b) prevention: (i) risk of breast cancer including BRCA1/2 carriers and testing. Whether loss to follow-up or inability to retrieve mammograms, or both, was likely related to the study outcome	No information found on loss to follow-up or reasons for censoring.	No	
Study attrition summary	There are no important differences between these characteristics in participants who completed the study and those who did not. Loss to follow-up (from baseline sample to study population analysed) is not associated with key characteristics (i.e. the study data adequately represent the sample) sufficient to limit potential bias to the observed			Moderate

	relationship between density change and outcome			
3. Prognostic factor measurement	Goal: to judge the risk of measurement bias related to how mammographic density was measured (differential measurement of mammographic density related to the level of outcome)			
Definition of the prognostic factor	A clear definition or description of mammographic density is provided (e.g. including the method of measurement, if subjective then who undertook it, if treatment then whether contralateral breast assessed)	Clear definition: "Cumulus software 4.0 (University of Toronto, Toronto, Canada) by a single investigator (JK)". Contralateral breast used.	Yes	
Valid and reliable measurement of prognostic factor	Method of mammographic density change measurement is adequately valid and reliable to limit misclassification bias (e.g. may include relevant outside sources of information on measurement properties; also characteristics, such as measurement blinded to case status)	Valid and reliable method, with estimate of reproducibility - "Intraobserver reproducibility, tested for 10% of randomly selected images (213/2,130), was 0.93 (Pearson correlation coefficient)". Blinded: "blinded to treatment outcome", but no mention of blinding to treatment used.	Partial	
	Continuous variables are reported or appropriate cut-points (i.e. not data-dependent (except for percentiles)) are used	MDR cut-points: "the 5% and 10% absolute reduction cut-offs based on previous findings [5]", but MDR also analysed as a continuous variable. No mention of why cut-points used for MDRR (selective reporting?), and why 0/10/25% used in main text but 15% used in supplementary analysis.	Partial	
Method and setting of prognostic factor measurement	The method and setting of measurement of mammographic density is the same for all study participants. The same mammogram type (film/digital) is used for both baseline and follow-up. The time at which baseline and follow-up mammograms have low variability between participants	Method and setting are the same. "All evaluated images were digital mammograms performed at our institution" and read by a single investigator. No mention of time between baseline and follow-up mammogram.	Partial	
Proportion of data on prognostic factor available for analysis	Adequate proportion of the study sample has complete data for the change in mammographic density variable	All have complete data by definition of study design.	Yes	
Method used for missing data	Appropriate methods of imputation are used for missing mammographic density data	NA	Yes	
Summary	Prognostic factor is adequately measured in study participants to sufficiently limit potential bias			Moderate
4. Outcome measurement	Goal: to judge the risk of bias related to the measurement of outcome (differential measurement of outcome related to the density reductions)			
Definition of	A clear definition of outcome is	Adequate definition "All	Partial	

the outcome	provided, including duration of follow-up and level and extent of the outcome construct	loco-regional or distant disease recurrences were regarded as recurrence events in recurrence-free survival analysis", but no information on start of follow-up or reasons for censoring. Duration of follow-up given (although discrepancy between abstract and text, see data capture form).		
Valid and reliable measurement of outcome	The method of outcome measurement used is adequately valid and reliable to limit misclassification bias	"Clinical and pathologic information on the 1,065 subjects was obtained from the database". Single institution.	Yes	
Method and setting of outcome measurement	The method and setting of outcome measurement is the same for all study participants, including by age and obesity groups	Yes.	Yes	
Outcome measurement summary	Outcome of interest is adequately measured in study participants to sufficiently limit potential bias			Moderate
5. Study confounding	Goal: to judge the risk of bias due to confounding (i.e. the effect of density reductions is distorted by another factor that is related to density reductions and the outcome)			
Important confounders measured	Age, BMI, or another measure of adiposity are measured	Age measured but not BMI.	Partial	
Definition of the confounding factor	Clear definitions are provided	Timing of age not fully described - is age at diagnosis, start of treatment, baseline or follow-up mammogram?	No	
Valid and reliable measurement of confounders	Measurement of all important confounders is adequately valid and reliable	From institution's prospectively maintained web based database?	Yes	
Method and setting of confounding measurement	The method and setting of confounding measurement are the same for all study participants	From institution's prospectively maintained web based database?	Yes	
Method used for missing data	Appropriate methods are used if imputation is used for missing confounder data	No mention of missing data or imputation. Age appears to be non-missing (Table 1).	Partial	
Appropriate accounting for confounding	The primary analysis will be adjusted for at least age, either through the study design and analysis, or through adjustment in the analysis only; and other prognostic factors	Unclear which analyses were adjusted for age. For example, supplementary table 4: Size, LN, Ki67 included (Fwd stepwise selection), but in text: "adjusted for age and preMD by forward selection stepwise analysis". Unclear of adjustments (if any) when analysis separated by tamoxifen and AIs: "When adjusted by age and ET regimen the findings were consistent, showing low MDR as a significant risk factor for recurrence in	Partial	

		patients who had undergone chemotherapy (HR 1.70, 95% CI 1.04 to 2.77, P = 0.033)", but only mentioned for chemotherapy. No adjustment for BMI or change in BMI.		
Study confounding summary	Important potential confounders are appropriately accounted for, limiting potential bias with respect to the relationship between prognostic factor and outcome			High
6. Statistical analysis and reporting	Goal: to judge the risk of bias related to the statistical analysis and presentation of results			
Presentation of analytical strategy, model development strategy	There is sufficient presentation of data to assess the adequacy of the analysis	Details of methods not reported e.g. start of follow-up and censoring, women with <2yr treatment excluded but minimum duration of treatment reported is 0.9yr, errors in tables (e.g. no number of women on AIs in Table 1), mismatching information between text and figures e.g. adjustments used. Description of women in endocrine treatment subgroups and number of women in subgroups not reported (unclear if subgroups are tamoxifen only, AI only or women who switched), therefore we can't separate out treatments - some women might have had cross-over of treatments between mammograms hence affecting density change.	No	
Model development strategy	The strategy for model building (i.e. inclusion of variables in the statistical model) is appropriate and is based on a conceptual framework or model	Partially: "factors with P < 0.05 were considered statistically significant", but unclear which confounders were initially considered for stepwise regression and which (if any were used for the endocrine therapy subgroups relevant for this review). Adjusting factors included in some models and not others with no consistency.	Partial	
Reporting of results	The selected statistical model is adequate for the design of the study. There is no selective reporting of results	Cox model appropriate but delayed-entry perhaps better (taking into account time between mammograms where women would not be at-risk). Unsure of follow-up and results in subgroups relevant for this review. Some selective reporting of results – no consistency with adjustments.	No	

Statistical analysis and presentation summary	The statistical analysis is appropriate for the design of the study, limiting potential for presentation of invalid or spurious results			High

B.VII.iii Risk of bias table - Knight 2018

Biases	Issues to consider for judging overall rating of risk of bias	Study Methods & Comments	Rating of reporting (adequacy of reporting: "yes", "partial", "no" or "unsure")	Rating of Risk of bias ("High", "Moderate", or "Low")
Instructions to assess the risk of each potential bias	These issues will guide your thinking and judgement about the overall risk of bias within each of the six domains. These issues are taken together to inform the overall judgement of potential bias for each of the six domains			
1. Study participation	Goal: to judge the risk of selection bias (likelihood that relationship between density reductions and outcome is different for participants and eligible non-participants)			
Source of target population	The source population or population of interest is adequately described for: a) treatment: (i) proportion with DCIS, (ii) cointerventions (chemotherapy/targeted therapy), (iii) severity of cancer at baseline (stage, % regional spread); b) prevention: (i) level of risk in population, including whether some or all are BRCA1/2 mutation carriers, (ii) prior hormone replacement therapy use, (iii) cointerventions such as diet or exercise regimens, or both	Source population is not described, only analysis population, but information on WECARE can be found elsewhere e.g. Memorial Sloan Kettering website.	Partial	
Method used to identify population	The sampling frame and recruitment are adequately described, including methods to identify the sample sufficient to limit potential bias	Well described: "Each study center identified eligible women through one or more population-based cancer registries". "WECARE study participants were diagnosed prior to age 55 years, between 1990 and 2008, with a first primary local or regional-stage invasive breast cancer. Cases were also diagnosed with a second primary invasive CBC at least 2 years later with no intervening cancer diagnosis, other than a non-melanoma skin cancer or cervical carcinoma in situ". Matched "UBC controls had no history of subsequent cancer diagnosis except for nonmelanoma skin cancer or cervical carcinoma in situ up to their reference date". UBC "controls must not have undergone prophylactic mastectomy of the contralateral breast. All	Yes	

		women had to be alive at the time of contact for interview".		
Recruitment period	Period of recruitment is adequately described	Well described (see data capture form).	Yes	
Place of recruitment	Place of recruitment (setting and geographic location) are adequately described	Well described (see data capture form).	Yes	
Inclusion and exclusion criteria	Inclusion and exclusion criteria are adequately described	Well described: "excluded the few mammograms that were digital from the analysis (5 CBC cases and 6 UBC controls prior to/at first diagnosis, 39 CBC cases and 41 UBC controls post diagnosis)", "excluded a small number of films in which MD could not be read because of poor image quality (4 CBC cases and 4 UBC controls prior to/at first diagnosis, 11 CBC cases and 6 UBC controls post diagnosis)", "excluded mammograms taken more than 36 months prior to or 48 months following first diagnosis", excluded 6% of CBC cases and 7% of UBC controls with a density change of 10% or more (unsure why), and excluded women with missing menopausal status information in the density change analysis.	Yes	
Adequate study participation	There is adequate participation in the study by eligible individuals	The uptake rate is not reported. This is a retrospective design, so requires individuals to still be alive at time of recruitment. This may lead to some selection bias – those who died before this date would not be included, particularly those diagnosed at the start of the period. Could bias against density change. "Women in whom we could not obtain a mammogram in an appropriate time window (see below) were more likely to have an earlier year of first breast cancer diagnosis (65% diagnosed in 1990–1996 vs. 40% in 1990–1996) and to be missing ER status (14% vs. 6%), and were slightly younger (mean age 45 years vs. mean age 46 years). Both groups had similar distributions of histologic type (10% and 13% lobular), stage (68% and 66% local), and, after excluding those with missing status, ER status	Yes	

		(65% and 68% positive). There were also no differences in first-degree family history (27% and 28%)". However, women with mammograms outside final selected timeframe had similar risks as those with mammograms inside timeframe (supplementary tables).		
Baseline characteristics	The baseline study sample (i.e. individuals entering the study) is adequately described for (treatment and prevention) age, menopausal status, cointerventions; (treatment) % DCIS, disease severity; (prevention) breast cancer risk, prior hormone replacement therapy use	These are given, but not for sample where density change analysed.	Partial	
Summary study participation	The study sample represents the population of interest on key characteristics, sufficient to limit potential bias of the observed relationship between density change and outcome			Moderate
2. Study attrition	Goal: to judge the risk of attrition bias (likelihood that relationship between density reductions and outcome are different for completing and non-completing participants)			
Proportion of baseline sample available for analysis	Response rate (i.e. proportion of study sample allocated treatment who received treatment) is adequate	No information found on response rate (i.e. compliance).	No	
Attempts to collect information on participants who dropped out	Attempts to collect information on participants who dropped out of the study are described	No information found on drop out. Registry linkage – no drop out?	No	
Reasons and potential impact of subjects lost to follow-up	Reasons for loss to follow-up are provided	Retrospective study, no information on loss to follow-up (does not include women who died and could not consent / provide questionnaire e.g. "women in whom we could not obtain a mammogram in an appropriate time window (see below) were more likely to have an earlier year of first breast cancer diagnosis (65% diagnosed in 1990–1996 vs. 40% in 1990–1996)" - survival bias whereby women included more likely to have survived at time of interview than wider cohort, we don't know about women who died before the study).	No	
Outcome and prognostic factor information on	Participants lost to follow-up are adequately described for age at entry and cointerventions (if any), and for a) treatment: (i) DCIS, (ii)	No information.	No	

those lost to follow-up	disease severity; b) prevention: (i) risk of breast cancer including BRCA1/2 carriers and testing. Whether loss to follow-up or inability to retrieve mammograms, or both, was likely related to the study outcome			
Study attrition summary	There are no important differences between these characteristics in participants who completed the study and those who did not. Loss to follow-up (from baseline sample to study population analysed) is not associated with key characteristics (i.e. the study data adequately represent the sample) sufficient to limit potential bias to the observed relationship between density change and outcome			High
3. Prognostic factor measurement	Goal: to judge the risk of measurement bias related to how mammographic density was measured (differential measurement of mammographic density related to the level of outcome)			
Definition of the prognostic factor	A clear definition or description of mammographic density is provided (e.g. including the method of measurement, if subjective then who undertook it, if treatment then whether contralateral breast assessed)	Clear definition: "MD measurements were all done in Toronto by one experienced reader (KB) using Cumulus". Contralateral breast used.	Yes	
Valid and reliable measurement of prognostic factor	Method of mammographic density change measurement is adequately valid and reliable to limit misclassification bias (e.g. may include relevant outside sources of information on measurement properties; also characteristics, such as measurement blinded to case status)	Valid and reliable method, with estimate of reproducibility - "We randomly selected 10% of each batch for repeat readings within and between batches. The Pearson correlation was 0.94 for both intra- and inter batch repeats". Blinded: "the reader was blinded to case control status and time sequence of the mammogram", but not reported if blinded to treatment.	Yes	
	Continuous variables are reported or appropriate cut-points (i.e. not data-dependent (except for percentiles)) are used	10% percentage change cut-point used, not justified in text. But noted to be used in other prior studies (particularly Sandbery et al, their ref [12]).	Partial	
Method and setting of prognostic factor measurement	The method and setting of measurement of mammographic density is the same for all study participants. The same mammogram type (film/digital) is used for both baseline and follow-up. The time at which baseline and follow-up mammograms have low variability between participants	Method and setting are the same (KB read all mammograms), "Mammograms were read in batches with both mammograms from the same woman read in the same batch. Mammogram order within each batch was randomized prior to reading" and "The film mammograms were digitized at two locations,	Yes	

		Seattle (all US mammograms) and Toronto (Ontario mammograms), both using a Kodak Lumisys Digital Scanner". Median baseline to follow-up=1yr.		
Proportion of data on prognostic factor available for analysis	Adequate proportion of the study sample has complete data for the change in mammographic density variable	812 CBC and 812 UBC recruited from WECARE II and at least one mammogram obtained from 464 CBC and 500 UBC (potential for bias if ended up with fewer CBC than UBC, so not having mammogram available a risk factor for CBC?). 224/362 CBC and 243/403 UBC with mammograms at both time points but 210 CBC and 225 UBC used in the density change analysis (reason for difference not given explicitly, but could be digital mammogram or quality of image or other exclusions broadly discussed or missing adjusting factors).	Partial	
Method used for missing data	Appropriate methods of imputation are used for missing mammographic density data	Complete case analysis.	Yes	
Summary	Prognostic factor is adequately measured in study participants to sufficiently limit potential bias			Moderate
4. Outcome measurement	Goal: to judge the risk of bias related to the measurement of outcome (differential measurement of outcome related to the density reductions)			
Definition of the outcome	A clear definition of outcome is provided, including duration of follow-up and level and extent of the outcome construct	Clear definition: "We assessed whether change in %MD (defined as the difference between measurements of %MD between the two time points) was associated with CBC in the subset of women who had mammograms at both time points". Date of search of cancer registries (i.e. follow-up) not given.	Partial	
Valid and reliable measurement of outcome	The method of outcome measurement used is adequately valid and reliable to limit misclassification bias	Yes – population registry.	Yes	
Method and setting of outcome measurement	The method and setting of outcome measurement is the same for all study participants, including by age and obesity groups	Yes.	Yes	
Outcome measurement summary	Outcome of interest is adequately measured in study participants to sufficiently limit potential bias			Low
5. Study confounding	Goal: to judge the risk of bias due to confounding (i.e. the effect of density reductions is distorted by			

	another factor that is related to density reductions and the outcome)			
Important confounders measured	Age, BMI, or another measure of adiposity are measured	Age, change in age, BMI, change in estimated BMI measured.	Yes	
Definition of the confounding factor	Clear definitions are provided	Age at first diagnosis, change in age between prior to/at first diagnosis and post-diagnosis mammograms, BMI at first diagnosis, change in estimated BMI between prior to/at first diagnosis and post-diagnosis mammograms (BMI at post-diagnostic mammogram was estimated from the BMI reported at first breast cancer diagnosis and at reference date, using linear interpolation - same for BMI at prior to/at first diagnosis mammogram?)	Yes	
Valid and reliable measurement of confounders	Measurement of all important confounders is adequately valid and reliable	Risk factors for breast cancer were obtained retrospectively by telephone survey - potential for recall bias.	Yes	
Method and setting of confounding measurement	The method and setting of confounding measurement are the same for all study participants	Yes.	Yes	
Method used for missing data	Appropriate methods are used if imputation is used for missing confounder data	Not explicit, but 210/224 CBC and 225/243 UBC used in the density change analysis (could be due to missing adjusting factors).	Partial	
Appropriate accounting for confounding	The primary analysis will be adjusted for at least age, either through the study design and analysis, or through adjustment in the analysis only; and other prognostic factors	Adjusted for age, change in age, and change in estimated BMI (why not BMI at diagnosis as well since between-women BMI also associated with risk?).	Yes	
Study confounding summary	Important potential confounders are appropriately accounted for, limiting potential bias with respect to the relationship between prognostic factor and outcome			Moderate
6. Statistical analysis and reporting	Goal: to judge the risk of bias related to the statistical analysis and presentation of results			
Presentation of analytical strategy, model development strategy	There is sufficient presentation of data to assess the adequacy of the analysis	Characteristics of a wider sample given in Table 1, but not of the sample analysed for density change. "Note that other types of hormonal therapies (e.g., aromatase inhibitors) were not common in this population", but unable to separate effects for this review.	Partial	
Model development strategy	The strategy for model building (i.e. inclusion of variables in the statistical model) is appropriate and is based on a conceptual	Reason for choosing the breast cancer risk factors used not included. Model adjusted for tamoxifen	Partial	

	framework or model	treatment instead of assessing density change in only women on treatment, therefore model does not estimate this review's primary measure. The model includes women who did not receive tamoxifen (but no predictive analysis was done, instead an adjustment was made for tamoxifen use), and ER-negative disease (different prognosis, adjusted for in model instead of separated out).		
Reporting of results	The selected statistical model is adequate for the design of the study. There is no selective reporting of results	Model appears adequate for design of study, but not for this review's analysis. Full model fit not given. Focus of paper was on prognostic ability of mammographic density, change is a secondary aim. Not clear if this is selective reporting. Also not clear why appropriate to exclude 10% or more density increase.	Partial	
Statistical analysis and presentation summary	The statistical analysis is appropriate for the design of the study, limiting potential for presentation of invalid or spurious results			High

B.VII.iv Risk of bias table - Ko 2013

Biases	Issues to consider for judging overall rating of risk of bias	Study Methods & Comments	Rating of reporting (adequacy of reporting: "yes", "partial", "no" or "unsure")	Rating of Risk of bias ("High", "Moderate", or "Low")
Instructions to assess the risk of each potential bias	These issues will guide your thinking and judgement about the overall risk of bias within each of the six domains. These issues are taken together to inform the overall judgement of potential bias for each of the six domains			
1. Study participation	Goal: to judge the risk of selection bias (likelihood that relationship between density reductions and outcome is different for participants and eligible non-participants)			
Source of target population	The source population or population of interest is adequately described for: a) treatment: (i) proportion with DCIS, (ii) cointerventions (chemotherapy/targeted therapy), (iii) severity of cancer at baseline (stage, % regional spread); b) prevention: (i) level of risk in population, including whether some or all are BRCA1/2 mutation carriers, (ii) prior hormone replacement therapy use, (iii) cointerventions such as diet or exercise regimens, or both	Source population is not described, only analysis population.	No	
Method used to identify population	The sampling frame and recruitment are adequately described, including methods to identify the sample sufficient to limit potential bias	Well described: "A total of 2,402 ER-positive breast cancer patients who were enrolled in this study underwent curative surgery at our institution between January 2003 and December 2008". 1,526/2,402 women who received adjuvant tamoxifen for at least 2 years.	Yes	
Recruitment period	Period of recruitment is adequately described	Well described (see data capture form).	Yes	
Place of recruitment	Place of recruitment (setting and geographic location) are adequately described	Well described (see data capture form).	Yes	
Inclusion and exclusion criteria	Inclusion and exclusion criteria are adequately described	Adequately described: 1336/2402 women excluded "if their digital mammograms were not available or not appropriate for evaluation or if they had bilateral breast cancer or occult breast cancer" (n=1066), but no information on number of exclusions based on each exclusion criteria. Women not on tamoxifen should be included in 1336 excluded	Partial	

		women but this is not detailed.		
Adequate study participation	There is adequate participation in the study by eligible individuals	1066 of 1526 tamoxifen-treated patients included. No information on participation based on consent. No information on eligible participants vs. those not eligible (in source population or 1526 women).	Partial	
Baseline characteristics	The baseline study sample (i.e. individuals entering the study) is adequately described for (treatment and prevention) age, menopausal status, cointerventions; (treatment) % DCIS, disease severity; (prevention) breast cancer risk, prior hormone replacement therapy use	Well described (see data capture form).	Yes	
Summary study participation	The study sample represents the population of interest on key characteristics, sufficient to limit potential bias of the observed relationship between density change and outcome			Moderate
2. Study attrition	Goal: to judge the risk of attrition bias (likelihood that relationship between density reductions and outcome are different for completing and non-completing participants)			
Proportion of baseline sample available for analysis	Response rate (i.e. proportion of study sample allocated treatment who received treatment) is adequate	Only included patients with at least 2yr of endocrine treatment.	Yes	
Attempts to collect information on participants who dropped out	Attempts to collect information on participants who dropped out of the study are described	No information found on drop out or reasons for censoring.	No	
Reasons and potential impact of subjects lost to follow-up	Reasons for loss to follow-up are provided	No information found on loss to follow-up or reasons for censoring.	No	
Outcome and prognostic factor information on those lost to follow-up	Participants lost to follow-up are adequately described for age at entry and cointerventions (if any), and for a) treatment: (i) DCIS, (ii) disease severity; b) prevention: (i) risk of breast cancer including BRCA1/2 carriers and testing. Whether loss to follow-up or inability to retrieve mammograms, or both, was likely related to the study outcome	No information found on loss to follow-up or reasons for censoring.	No	
Study attrition summary	There are no important differences between these characteristics in participants who completed the study and those who did not. Loss to follow-up (from baseline sample to study population analysed) is not associated with key characteristics (i.e. the study data adequately represent the			Moderate

	sample) sufficient to limit potential bias to the observed relationship between density change and outcome			
3. Prognostic factor measurement	Goal: to judge the risk of measurement bias related to how mammographic density was measured (differential measurement of mammographic density related to the level of outcome)			
Definition of the prognostic factor	A clear definition or description of mammographic density is provided (e.g. including the method of measurement, if subjective then who undertook it, if treatment then whether contralateral breast assessed)	Method of measurement not completely clear: what is the "computerized system"? Also, "a single radiologist (K. Ko: 10 years of experience in interpreting mammograms) reviewed 2,132 preoperative and postoperative mammograms classified breast density patterns according to BIRADS". No restriction to contralateral breast? E.g. each woman had 2 mammograms (2132 mammograms for 1066 women) so perhaps both breasts were examined in 1 view or one breast was examined in 2 views?	No	
Valid and reliable measurement of prognostic factor	Method of mammographic density change measurement is adequately valid and reliable to limit misclassification bias (e.g. may include relevant outside sources of information on measurement properties; also characteristics, such as measurement blinded to case status)	No test of reliability: "We relied on a single radiologist who is a specialist in breast imaging studies, thereby eliminating interobserver variability. We did not seek to measure reproducibility as the BI-RADS density classifications are standardized", although BI-RADS measures can still have intra-reader variability. Not stated if blinded to patient's identity or breast cancer event status etc.	No	
	Continuous variables are reported or appropriate cut-points (i.e. not data-dependent (except for percentiles)) are used	BI-RADS cut-points, although combining BI-RADS 1 & 2 means that BI-RADS 2 can no longer move down a category so losing information about density change (better to just exclude BI-RADS 1 at baseline as they cannot move down a category?).	Partial	
Method and setting of prognostic factor measurement	The method and setting of measurement of mammographic density is the same for all study participants. The same mammogram type (film/digital) is used for both baseline and follow-up. The time at which baseline and follow-up mammograms have low variability between participants	Method and setting are the same: read by a single radiologist, all women and mammograms from the same institution. No mention of time between baseline and follow-up mammogram.	Partial	
Proportion of data on prognostic	Adequate proportion of the study sample has complete data for the change in mammographic density	All have complete data by definition of study design.	Yes	

factor available for analysis	variable			
Method used for missing data	Appropriate methods of imputation are used for missing mammographic density data	NA	Yes	
Summary	Prognostic factor is adequately measured in study participants to sufficiently limit potential bias			Moderate
4. Outcome measurement	Goal: to judge the risk of bias related to the measurement of outcome (differential measurement of outcome related to the density reductions)			
Definition of the outcome	A clear definition of outcome is provided, including duration of follow-up and level and extent of the outcome construct	Adequate definition: outcome is loco-regional or systemic recurrence, or contralateral breast cancer, but discrepancy with numbers (48+16+4=68, not 67 as stated). "The association of MDR with disease-free survival according to patterns of recurrent disease (loco-regional, systemic, and contralateral recurrence) was analyzed", but no information on start of follow-up or reasons for censoring. Duration of follow-up given (mean 61 months).	Partial	
Valid and reliable measurement of outcome	The method of outcome measurement used is adequately valid and reliable to limit misclassification bias	Single institution: "we collected the clinicopathologic information on 1,066 patients by reviewing the prospective database of our institution and the data of disease recurrence by additional medical record review".	Yes	
Method and setting of outcome measurement	The method and setting of outcome measurement is the same for all study participants, including by age and obesity groups	Yes.	Yes	
Outcome measurement summary	Outcome of interest is adequately measured in study participants to sufficiently limit potential bias			Moderate
5. Study confounding	Goal: to judge the risk of bias due to confounding (i.e. the effect of density reductions is distorted by another factor that is related to density reductions and the outcome)			
Important confounders measured	Age, BMI, or another measure of adiposity are measured	Age and BMI measured.	Yes	
Definition of the confounding factor	Clear definitions are provided	Body mass index (BMI) was calculated as weight/height^2 (kg/m^2), but timing of age and BMI not fully described - are they at diagnosis, start of treatment, baseline or follow-up mammogram?	Partial	
Valid and reliable	Measurement of all important confounders is adequately valid	From institution's prospective database?	Yes	

measurement of confounders	and reliable			
Method and setting of confounding measurement	The method and setting of confounding measurement are the same for all study participants	From institution's prospective database?	Yes	
Method used for missing data	Appropriate methods are used if imputation is used for missing confounder data	No mention of missing data or imputation. Age appears to be non-missing (Table 1: ≤50yr vs. >50yr).	Partial	
Appropriate accounting for confounding	The primary analysis will be adjusted for at least age, either through the study design and analysis, or through adjustment in the analysis only; and other prognostic factors	Analysis reported with adjustment for age and BMI, also other prognostic factors. However, no adjustment for chemotherapy although this was associated with mammographic density reduction. So not clear if tamoxifen-induced, or chemotherapy-induced differences in survival. Also no adjustment for breast density at entry, which is strongly associated with density change in the data. No adjustment for change in BMI.	Partial	
Study confounding summary	Important potential confounders are appropriately accounted for, limiting potential bias with respect to the relationship between prognostic factor and outcome			Moderate
6. Statistical analysis and reporting	Goal: to judge the risk of bias related to the statistical analysis and presentation of results			
Presentation of analytical strategy, model development strategy	There is sufficient presentation of data to assess the adequacy of the analysis	Details of methods not reported e.g. start of follow-up and censoring, errors e.g. total number of events do not add up, unclear adjustments e.g. results from Fig 3 don't match table 6 (are these unadjusted?), title says premenopausal women but age range 25-78 and subgroup analysis of ≤50yr/>50yr used as a proxy for menopausal status and postmenopausal women mentioned in results. "Our institution's guidelines recommend aromatase inhibitors as the first choice endocrine therapy for ER-positive postmenopausal breast cancer patients" - does this mean the postmenopausal women could have been on AIs before their tamoxifen treatment?	No	
Model development strategy	The strategy for model building (i.e. inclusion of variables in the statistical model) is appropriate and is based on a conceptual framework or model	Adjustment for some prognostic factors, but not others that are also associated with mammographic density change (baseline density,	Partial	

		chemotherapy). Not clear how model developed / criteria for building.		
Reporting of results	The selected statistical model is adequate for the design of the study. There is no selective reporting of results	Cox model appropriate but delayed-entry perhaps better (taking into account time between mammograms where women would not be at-risk). Unsure of follow-up and may be some selective reporting of results – why were other adjustments not considered / included?	No	
Statistical analysis and presentation summary	The statistical analysis is appropriate for the design of the study, limiting potential for presentation of invalid or spurious results			High

B.VII.v Risk of bias table - Li 2013

Biases	Issues to consider for judging overall rating of risk of bias	Study Methods & Comments	Rating of reporting (adequacy of reporting: "yes", "partial", "no" or "unsure")	Rating of Risk of bias ("High", "Moderate", or "Low")
Instructions to assess the risk of each potential bias	These issues will guide your thinking and judgement about the overall risk of bias within each of the six domains. These issues are taken together to inform the overall judgement of potential bias for each of the six domains			
1. Study participation	Goal: to judge the risk of selection bias (likelihood that relationship between density reductions and outcome is different for participants and eligible non-participants)			
Source of target population	The source population or population of interest is adequately described for: a) treatment: (i) proportion with DCIS, (ii) cointerventions (chemotherapy/targeted therapy), (iii) severity of cancer at baseline (stage, % regional spread); b) prevention: (i) level of risk in population, including whether some or all are BRCA1/2 mutation carriers, (ii) prior hormone replacement therapy use, (iii) cointerventions such as diet or exercise regimens, or both	Source population is not described, only analysis population, but indication of size given: "all women born in Sweden who were age 50 to 74 years old at first diagnosis of invasive breast cancer in the Swedish Cancer Register were eligible (n=3,979)".	Partial	
Method used to identify population	The sampling frame and recruitment are adequately described, including methods to identify the sample sufficient to limit potential bias	Adequately described e.g."84% (n=3,345) participated by answering a mailed questionnaire".	Yes	
Recruitment period	Period of recruitment is adequately described	Well described (see data capture form).	Yes	
Place of recruitment	Place of recruitment (setting and geographic location) are adequately described	Well described (see data capture form).	Yes	
Inclusion and exclusion criteria	Inclusion and exclusion criteria are adequately described	Well described e.g. Figure 1 & exclusions based on questionnaire data, previous cancer, premenopausal or unknown menopausal status, medical records or registers, non-invasive breast cancer, duplicate records, breast cancer diagnosis before or after study period, non-breast-cancer, no informed consent, mammogram data, no mammograms, no follow-up mammogram, follow-up mammogram >3yr after baseline, baseline density, quintile with smallest dense area, and incomplete covariate	Yes	

		information.		
Adequate study participation	There is adequate participation in the study by eligible individuals	634 women declined participation, 701 were excluded for sensible reasons (including previous cancer, duplicate records, non-invasive breast cancer), 1603 excluded due to mammograms (including 243 in quintile with smallest dense area), 67 incomplete covariates, 974 in analysis. No information on eligible participants vs. those not eligible.	Partial	
Baseline characteristics	The baseline study sample (i.e. individuals entering the study) is adequately described for (treatment and prevention) age, menopausal status, cointerventions; (treatment) % DCIS, disease severity; (prevention) breast cancer risk, prior hormone replacement therapy use	Well described (see data capture form).	Yes	
Summary study participation	The study sample represents the population of interest on key characteristics, sufficient to limit potential bias of the observed relationship between density change and outcome			Moderate
2. Study attrition	Goal: to judge the risk of attrition bias (likelihood that relationship between density reductions and outcome are different for completing and non-completing participants)			
Proportion of baseline sample available for analysis	Response rate (i.e. proportion of study sample allocated treatment who received treatment) is adequate	37 (7.8%) received tamoxifen for <12 months.	Yes	
Attempts to collect information on participants who dropped out	Attempts to collect information on participants who dropped out of the study are described	Registry based, some censoring due to emigration but not stated what amount (but in this population & age group it is likely to be small).	Partial	
Reasons and potential impact of subjects lost to follow-up	Reasons for loss to follow-up are provided	Censored due to: "death, emigration, or end of follow-up (December 31, 2008)".	Yes	
Outcome and prognostic factor information on those lost to follow-up	Participants lost to follow-up are adequately described for age at entry and cointerventions (if any), and for a) treatment: (i) DCIS, (ii) disease severity; b) prevention: (i) risk of breast cancer including BRCA1/2 carriers and testing. Whether loss to follow-up or inability to retrieve mammograms, or both, was likely related to the study outcome	Only reason for loss to follow-up is emigration and these are not described, but likely to be small.	Partial	
Study attrition summary	There are no important differences between these characteristics in participants who completed the study and those who did not. Loss to follow-up (from baseline sample to study population analysed) is not			Low

	associated with key characteristics (i.e. the study data adequately represent the sample) sufficient to limit potential bias to the observed relationship between density change and outcome			
3. Prognostic factor measurement	Goal: to judge the risk of measurement bias related to how mammographic density was measured (differential measurement of mammographic density related to the level of outcome)			
Definition of the prognostic factor	A clear definition or description of mammographic density is provided (e.g. including the method of measurement, if subjective then who undertook it, if treatment then whether contralateral breast assessed)	Clear definition: "All density measurements were obtained by using an automated thresholding method previously described in Li et al.(17). The machine learning method incorporates the knowledge of a trained reader (L.E.) by using segmentations obtained by Cumulus (19) as training data". Contralateral breast: "Only mediolateral oblique views of the breast unaffected by breast cancer were used".	Yes	
Valid and reliable measurement of prognostic factor	Method of mammographic density change measurement is adequately valid and reliable to limit misclassification bias (e.g. may include relevant outside sources of information on measurement properties; also characteristics, such as measurement blinded to case status)	Appears adequately valid: "Externally validated results showed a high correspondence between the automated method and the user-assisted threshold method (Pearson's correlation coefficient 0.872 for DA)".	Yes	
	Continuous variables are reported or appropriate cut-points (i.e. not data-dependent (except for percentiles)) are used	Categories used were chosen "a priori", but no reference of evidence to support this e.g. prior publication using same cut-points or SAP.	Partial	
Method and setting of prognostic factor measurement	The method and setting of measurement of mammographic density is the same for all study participants. The same mammogram type (film/digital) is used for both baseline and follow-up. The time at which baseline and follow-up mammograms have low variability between participants	Method and setting are the same. All film mammograms scanned on same scanner: "Film mammograms were digitized by using an Array 2905HD Laser Film Digitizer (Array Corp, Tokyo, Japan)". Mean 1.4yr (SD 0.5) between mammogram, maximum 3yr.	Yes	
Proportion of data on prognostic factor available for analysis	Adequate proportion of the study sample has complete data for the change in mammographic density variable	All have complete data by definition of study design.	Yes	
Method used for missing data	Appropriate methods of imputation are used for missing mammographic density data	NA	Yes	
Summary	Prognostic factor is adequately measured in study participants to sufficiently limit potential bias			Moderate

4. Outcome measurement	Goal: to judge the risk of bias related to the measurement of outcome (differential measurement of outcome related to the density reductions)			
Definition of the outcome	A clear definition of outcome is provided, including duration of follow-up and level and extent of the outcome construct	Clear definition: "Women were observed from the date of breast cancer diagnosis until death" and "Cause-specific deaths as a result of breast cancer were ascertained by using the cause of death register", duration of follow-up given (see data capture form), extent of outcome construct given from wider eligible population.	Yes	
Valid and reliable measurement of outcome	The method of outcome measurement used is adequately valid and reliable to limit misclassification bias	Yes – population registry.	Yes	
Method and setting of outcome measurement	The method and setting of outcome measurement is the same for all study participants, including by age and obesity groups	Yes.	Yes	
Outcome measurement summary	Outcome of interest is adequately measured in study participants to sufficiently limit potential bias			Low
5. Study confounding	Goal: to judge the risk of bias due to confounding (i.e. the effect of density reductions is distorted by another factor that is related to density reductions and the outcome)			
Important confounders measured	Age, BMI, or another measure of adiposity are measured	Age, BMI and change in non-dense area (proxy for BMI) measured.	Yes	
Definition of the confounding factor	Clear definitions are provided	Age at baseline mammogram (years), body mass index (BMI) at interview (quartiles), quartile of percentage change in non-dense area was used as a proxy for BMI.	Yes	
Valid and reliable measurement of confounders	Measurement of all important confounders is adequately valid and reliable	Measurement is adequately valid.	Yes	
Method and setting of confounding measurement	The method and setting of confounding measurement are the same for all study participants	Yes.	Yes	
Method used for missing data	Appropriate methods are used if imputation is used for missing confounder data	67 excluded from analysis due to missing covariate data so no imputation used. Grade and ER status unknown were coded as such.	Yes	
Appropriate accounting for confounding	The primary analysis will be adjusted for at least age, either through the study design and analysis, or through adjustment in the analysis only; and other prognostic factors	Yes, including adjustments for age, BMI, change in non-dense area and chemotherapy. Adjustment for baseline density included in the relative density change measure.	Yes	

Study confounding summary	Important potential confounders are appropriately accounted for, limiting potential bias with respect to the relationship between prognostic factor and outcome			Low
6. Statistical analysis and reporting	Goal: to judge the risk of bias related to the statistical analysis and presentation of results			
Presentation of analytical strategy, model development strategy	There is sufficient presentation of data to assess the adequacy of the analysis	Yes.	Yes	
Model development strategy	The strategy for model building (i.e. inclusion of variables in the statistical model) is appropriate and is based on a conceptual framework or model	Conceptual framework – adjusted for other prognostic factors or those associated with density. Does not appear to be variable selection.	Yes	
Reporting of results	The selected statistical model is adequate for the design of the study. There is no selective reporting of results	Appears adequate. Not clear though, why appropriate to exclude low density up front – might have been a subgroup analysis chosen as a primary analysis.	Partial	
Statistical analysis and presentation summary	The statistical analysis is appropriate for the design of the study, limiting potential for presentation of invalid or spurious results			Low

B.VII.vi Risk of bias table - Nyante 2015

Biases	Issues to consider for judging overall rating of risk of bias	Study Methods & Comments	Rating of reporting (adequacy of reporting: "yes", "partial", "no" or "unsure")	Rating of Risk of bias ("High", "Moderate", or "Low")
Instructions to assess the risk of each potential bias	These issues will guide your thinking and judgement about the overall risk of bias within each of the six domains. These issues are taken together to inform the overall judgement of potential bias for each of the six domains			
1. Study participation	Goal: to judge the risk of selection bias (likelihood that relationship between density reductions and outcome is different for participants and eligible non-participants)			
Source of target population	The source population or population of interest is adequately described for: a) treatment: (i) proportion with DCIS, (ii) cointerventions (chemotherapy/targeted therapy), (iii) severity of cancer at baseline (stage, % regional spread); b) prevention: (i) level of risk in population, including whether some or all are BRCA1/2 mutation carriers, (ii) prior hormone replacement therapy use, (iii) cointerventions such as diet or exercise regimens, or both	Source population is not described, only analysis population, but indication of size given: "Patients were selected from a cohort of 2315 KPNW members diagnosed with ER-positive primary invasive breast cancer between 1990 and 2008 and treated with adjuvant tamoxifen".	Partial	
Method used to identify population	The sampling frame and recruitment are adequately described, including methods to identify the sample sufficient to limit potential bias	Well described e.g. "Two control patients were matched to each case patient (Figure 1) and sampled to have at least as much follow-up time as the matched case patient" and 61 controls matched to cases without mammograms were "re-matched [...] to eligible cases", 401 women were not included in the sampling due to lack of mammograms.	Yes	
Recruitment period	Period of recruitment is adequately described	Well described (see data capture form).	Yes	
Place of recruitment	Place of recruitment (setting and geographic location) are adequately described	Well described (see data capture form).	Yes	
Inclusion and exclusion criteria	Inclusion and exclusion criteria are adequately described	Well described e.g. after exclusions for bilateral disease (or laterality unknown), prophylactic mastectomy of contralateral breast, death or recurrence within 1yr of initial diagnosis, distant metastases or unstaged diagnosis, never disease-	Yes	

		free, first course of treatment outside KPNW system (n=2142), after exclusions for mammograms not identified (n=1741): cases identified (n=134) of which had mammograms obtainable for digitisation (n=97), and 252 matched controls with mammograms obtainable for digitisation.		
Adequate study participation	There is adequate participation in the study by eligible individuals	37/134 (28%) of cases not included due to lack of mammograms obtainable for digitisation. No information on eligible participants vs. those not eligible.	Partial	
Baseline characteristics	The baseline study sample (i.e. individuals entering the study) is adequately described for (treatment and prevention) age, menopausal status, cointerventions; (treatment) % DCIS, disease severity; (prevention) breast cancer risk, prior hormone replacement therapy use	Well described (see data capture form).	Yes	
Summary study participation	The study sample represents the population of interest on key characteristics, sufficient to limit potential bias of the observed relationship between density change and outcome			Moderate
2. Study attrition	Goal: to judge the risk of attrition bias (likelihood that relationship between density reductions and outcome are different for completing and non-completing participants)			
Proportion of baseline sample available for analysis	Response rate (i.e. proportion of study sample allocated treatment who received treatment) is adequate	34% cases and 68% controls received tamoxifen >52 months, follow-up mammogram within 90 days of prescription.	Yes	
Attempts to collect information on participants who dropped out	Attempts to collect information on participants who dropped out of the study are described	Registry linkage to death, so no drop out.	Yes	
Reasons and potential impact of subjects lost to follow-up	Reasons for loss to follow-up are provided	"Follow-up time was calculated as the time between the first tamoxifen prescription and the earliest of the following: breast cancer death, death from another cause, last tumor registry follow-up, or December 31, 2010".	Yes	
Outcome and prognostic factor information on those lost to follow-up	Participants lost to follow-up are adequately described for age at entry and cointerventions (if any), and for a) treatment: (i) DCIS, (ii) disease severity; b) prevention: (i) risk of breast cancer including BRCA1/2 carriers and testing.	No information on density change vs. death from other causes in controls – those who could be censored before being included as a control. Expect very few but some older women.	Partial	

	Whether loss to follow-up or inability to retrieve mammograms, or both, was likely related to the study outcome			
Study attrition summary	There are no important differences between these characteristics in participants who completed the study and those who did not. Loss to follow-up (from baseline sample to study population analysed) is not associated with key characteristics (i.e. the study data adequately represent the sample) sufficient to limit potential bias to the observed relationship between density change and outcome			Low
3. Prognostic factor measurement	Goal: to judge the risk of measurement bias related to how mammographic density was measured (differential measurement of mammographic density related to the level of outcome)			
Definition of the prognostic factor	A clear definition or description of mammographic density is provided (e.g. including the method of measurement, if subjective then who undertook it, if treatment then whether contralateral breast assessed)	Clear definition: "Absolute dense area (cm^2) and total breast area (cm^2) were measured using Cumulus […] by a single reader". Contralateral breast used.	Partial	
Valid and reliable measurement of prognostic factor	Method of mammographic density change measurement is adequately valid and reliable to limit misclassification bias (e.g. may include relevant outside sources of information on measurement properties; also characteristics, such as measurement blinded to case status)	Valid and reliable method, with estimate of reproducibility - "Reevaluation of 50 randomly selected films yielded intraclass correlation coefficients and coefficients of variation of 0.95 and 8.5% for dense area, 0.99 and 0.5% for total breast area, and 0.96 and 8.5% for percent density". Blinded: "Masked baseline and follow up mammograms from each patient".	Yes	
	Continuous variables are reported or appropriate cut-points (i.e. not data-dependent (except for percentiles)) are used	Tertiles are used as primary analysis, but "assessed absolute change in percent density using the 10% or greater cut-point to assess a comparable level of change as reported in the IBIS-1 study".	Yes	
Method and setting of prognostic factor measurement	The method and setting of measurement of mammographic density is the same for all study participants. The same mammogram type (film/digital) is used for both baseline and follow-up. The time at which baseline and follow-up mammograms have low variability between participants	Method and setting are the same. All film mammograms scanned on same scanner: "digitized using an Array Corporation 2095 Laser Film Digitizer (Roden, the Netherlands; optical density = 4.0)". Time between mammograms mean 18 months, 75% with <24 months – relatively low variability.	Yes	
Proportion of	Adequate proportion of the study	All have complete data by	Yes	

data on prognostic factor available for analysis	sample has complete data for the change in mammographic density variable	definition of study design.		
Method used for missing data	Appropriate methods of imputation are used for missing mammographic density data	NA	Yes	
Summary	Prognostic factor is adequately measured in study participants to sufficiently limit potential bias			Low
4. Outcome measurement	Goal: to judge the risk of bias related to the measurement of outcome (differential measurement of outcome related to the density reductions)			
Definition of the outcome	A clear definition of outcome is provided, including duration of follow-up and level and extent of the outcome construct	Clear definition: "Case patients were defined as patients who died of breast cancer between January 1, 1991 and December 31, 2010", duration of follow-up given (see data capture form). Not clear how extensive in the wider database (including women with mammograms).	Partial	
Valid and reliable measurement of outcome	The method of outcome measurement used is adequately valid and reliable to limit misclassification bias	Yes – population registry.	Yes	
Method and setting of outcome measurement	The method and setting of outcome measurement is the same for all study participants, including by age and obesity groups	Yes.	Yes	
Outcome measurement summary	Outcome of interest is adequately measured in study participants to sufficiently limit potential bias			Low
5. Study confounding	Goal: to judge the risk of bias due to confounding (i.e. the effect of density reductions is distorted by another factor that is related to density reductions and the outcome)			
Important confounders measured	Age, BMI, or another measure of adiposity are measured	Age and BMI measured.	Yes	
Definition of the confounding factor	Clear definitions are provided	"Age […] at diagnosis" and "Body mass index (BMI) was calculated as kg/m^2".	Yes	
Valid and reliable measurement of confounders	Measurement of all important confounders is adequately valid and reliable	Self-reported height and weight from clinical records, obtained within 3 months of both mammograms. 13/97 cases and 26/252 controls missing baseline BMI; 25/97 cases and 59/252 controls missing BMI change.	Yes	
Method and setting of confounding measurement	The method and setting of confounding measurement are the same for all study participants	Yes.	Yes	
Method used for missing	Appropriate methods are used if imputation is used for missing	Missing weight or height values used to calculate	Yes	

data	confounder data	BMI were multiply-imputed using IVEWare.		
Appropriate accounting for confounding	The primary analysis will be adjusted for at least age, either through the study design and analysis, or through adjustment in the analysis only; and other prognostic factors	It is adjusted for age through design, also for baseline density in additional model. Not adjusted for BMI or change in BMI (but "neither baseline BMI nor change in BMI altered the associations"). No evidence of interaction (between density change?) and chemotherapy.	Yes	
Study confounding summary	Important potential confounders are appropriately accounted for, limiting potential bias with respect to the relationship between prognostic factor and outcome			Low
6. Statistical analysis and reporting	Goal: to judge the risk of bias related to the statistical analysis and presentation of results			
Presentation of analytical strategy, model development strategy	There is sufficient presentation of data to assess the adequacy of the analysis	Yes, except for the multiple imputation where little detail is provided. Multiple imputation not used for the main analysis however. 6 women on AIs but "Associations were also similar after excluding […] women treated with aromatase inhibitors".	Partial	
Model development strategy	The strategy for model building (i.e. inclusion of variables in the statistical model) is appropriate and is based on a conceptual framework or model	Yes. "Multivariable models were constructed to assess confounding. Smoking status, tumor size, antidepressant use, and baseline percent density were identified as potential confounders based on literature review and covariable associations with breast cancer death and change in percent density among control patients and included in a preliminary model. Only baseline density was retained in final models after removing variables sequentially and retaining those where removal altered the change in density regression parameter by more than 10%. Tumor size and baseline dense area were assessed similarly in multivariable models for absolute change in dense area, and both were retained".	Yes	
Reporting of results	The selected statistical model is adequate for the design of the study. There is no selective reporting of results	Yes adequate. No selective reporting apparent.	Yes	
Statistical analysis and	The statistical analysis is appropriate for the design of the			Low

presentation summary	study, limiting potential for presentation of invalid or spurious results			

B.VII.vii Risk of bias table - Sandberg 2013

Biases	Issues to consider for judging overall rating of risk of bias	Study Methods & Comments	Rating of reporting (adequacy of reporting: "yes", "partial", "no" or "unsure")	Rating of Risk of bias ("High", "Moderate", or "Low")
Instructions to assess the risk of each potential bias	These issues will guide your thinking and judgement about the overall risk of bias within each of the six domains. These issues are taken together to inform the overall judgement of potential bias for each of the six domains			
1. Study participation	Goal: to judge the risk of selection bias (likelihood that relationship between density reductions and outcome is different for participants and eligible non-participants)			
Source of target population	The source population or population of interest is adequately described for: a) treatment: (i) proportion with DCIS, (ii) cointerventions (chemotherapy/targeted therapy), (iii) severity of cancer at baseline (stage, % regional spread); b) prevention: (i) level of risk in population, including whether some or all are BRCA1/2 mutation carriers, (ii) prior hormone replacement therapy use, (iii) cointerventions such as diet or exercise regimens, or both	Source population is not described, only analysis population, but indication of size given: "Stockholm Breast Cancer Register, a population based register of all breast cancer patients diagnosed since 1976 in the Stockholm-Gotland health-care region (n>30,000)".	Partial	
Method used to identify population	The sampling frame and recruitment are adequately described, including methods to identify the sample sufficient to limit potential bias	Well described: "Women with invasive CBC diagnosed more than one year after the first invasive cancer and with an available mammogram close to the first diagnosis (N=458) were identified as potential cases. Patients with invasive unilateral breast cancer in the same register were identified as potential controls". No metastasis or second primary ipsilateral breast cancer to limit bias of misclassification of outcome.	Yes	
Recruitment period	Period of recruitment is adequately described	Well described (see data capture form).	Yes	
Place of recruitment	Place of recruitment (setting and geographic location) are adequately described	Well described (see data capture form).	Yes	
Inclusion and exclusion criteria	Inclusion and exclusion criteria are adequately described	Well described: "Women with a first primary cancer other than breast cancer and women with distant metastasis at the first or second breast cancer diagnosis were excluded in order to minimize the risk of the CBC being a	Yes	

		misclassified metastasis" and second primary breast cancer in the ipsilateral breast excluded. Women with <10% or >90% PDA (N = 66), or <10 cm^2 or >70 cm^2 DA (N = 84) at baseline were excluded (can't undergo defined changes - why 70cm^2?)		
Adequate study participation	There is adequate participation in the study by eligible individuals	187 (41%) of cases excluded: "For 99 of the 458 eligible CBC-cases we could not locate any follow-up mammogram and for 88 of the CBC-cases either the baseline or the follow-up mammogram could not be used (for example, due to low quality of the mammogram)", therefore "for 271 patients (59%) both the baseline and at least one follow-up mammogram of the unaffected breast from the same view was assessable and could be used". 211 controls with correct side and view so 211 matched case-control pairs. Availability of mammograms "driven by archiving policies, rather than patients not having mammograms taken". Patients excluded due to lack of eligible mammograms did not differ from those included in relation to age at first diagnosis (P-value: 0.23) and calendar period of first diagnosis (P-value: 0.12). More patients included received radiotherapy and endocrine therapy than excluded patients (radiotherapy; 29% vs. 23%, endocrine therapy; 39% vs. 29%, (chemotherapy?)).	Yes	
Baseline characteristics	The baseline study sample (i.e. individuals entering the study) is adequately described for (treatment and prevention) age, menopausal status, cointerventions; (treatment) % DCIS, disease severity; (prevention) breast cancer risk, prior hormone replacement therapy use	Adequately described in Table 1. No breakdown by type of endocrine therapy, but most likely to be tamoxifen according to time frame.	Yes	
Summary study participation	The study sample represents the population of interest on key characteristics, sufficient to limit potential bias of the observed relationship between density change and outcome			Moderate
2. Study attrition	Goal: to judge the risk of attrition bias (likelihood that relationship			

	between density reductions and outcome are different for completing and non-completing participants)			
Proportion of baseline sample available for analysis	Response rate (i.e. proportion of study sample allocated treatment who received treatment) is adequate	41% of cases and 41% of controls (minority of women) on endocrine treatment, but no information found on response rate (i.e. compliance).	No	
Attempts to collect information on participants who dropped out	Attempts to collect information on participants who dropped out of the study are described	No information found on drop out. Registry linkage – no drop out? Exception of informed consent.	No	
Reasons and potential impact of subjects lost to follow-up	Reasons for loss to follow-up are provided	No information found on loss to follow-up.	No	
Outcome and prognostic factor information on those lost to follow-up	Participants lost to follow-up are adequately described for age at entry and cointerventions (if any), and for a) treatment: (i) DCIS, (ii) disease severity; b) prevention: (i) risk of breast cancer including BRCA1/2 carriers and testing. Whether loss to follow-up or inability to retrieve mammograms, or both, was likely related to the study outcome	No information found on loss to follow-up.	No	
Study attrition summary	There are no important differences between these characteristics in participants who completed the study and those who did not. Loss to follow-up (from baseline sample to study population analysed) is not associated with key characteristics (i.e. the study data adequately represent the sample) sufficient to limit potential bias to the observed relationship between density change and outcome			Moderate
3. Prognostic factor measurement	Goal: to judge the risk of measurement bias related to how mammographic density was measured (differential measurement of mammographic density related to the level of outcome)			
Definition of the prognostic factor	A clear definition or description of mammographic density is provided (e.g. including the method of measurement, if subjective then who undertook it, if treatment then whether contralateral breast assessed)	Clear definition: "Mammographic density was measured using our automated thresholding method [24], which incorporates the knowledge of a trained observer by using measurements obtained by an established user-assisted threshold method - Cumulus [25] - as training data". Contralateral breast used.	Yes	
Valid and reliable measurement	Method of mammographic density change measurement is adequately valid and reliable to	Appears adequately valid: "The externally validated results showed a high	Yes	

of prognostic factor	limit misclassification bias (e.g. may include relevant outside sources of information on measurement properties; also characteristics, such as measurement blinded to case status)	correspondence between our automated method and the established user-assisted thresholding method Cumulus (r percent mammographic density) = 0.88 (95% CI: 0.87 to 0.89)".		
	Continuous variables are reported or appropriate cut-points (i.e. not data-dependent (except for percentiles)) are used	Cut-point for percentage density (10%) chosen following IBIS-I: "absolute decrease ≥10%, stable (-10% to +10%, reference level) and absolute increase ≥10%, in agreement with previous literature [20]". Cut-point for area density not stated, but similar proportion to percent measure in different categories (Table 1).	Partial	
Method and setting of prognostic factor measurement	The method and setting of measurement of mammographic density is the same for all study participants. The same mammogram type (film/digital) is used for both baseline and follow-up. The time at which baseline and follow-up mammograms have low variability between participants	Method and setting are the same. "The mammograms were digitized using an Array 2905HD Laser Film Digitizer (Array Corporation, Tokyo, Japan), which covers a range of 0 to 4.7 optical density. The density resolution was set at 12-bit spatial resolution". "90% of the follow-up mammograms were taken between 1 and 2.2 years after diagnosis of the first breast cancer and there was no difference between cases and controls"; some up to 5yr.	Yes	
Proportion of data on prognostic factor available for analysis	Adequate proportion of the study sample has complete data for the change in mammographic density variable	Women with <10% or >90% PDA (N = 66), or <10 cm^2 or >70 cm^2 DA (N = 84) at baseline were excluded (can't undergo defined changes - why 70cm^2?). For the primary interest of this review (women who received endocrine therapy) numbers are not reported.	Partial	
Method used for missing data	Appropriate methods of imputation are used for missing mammographic density data	NA for study sample, unclear for women who received endocrine therapy.	Partial	
Summary	Prognostic factor is adequately measured in study participants to sufficiently limit potential bias			Moderate
4. Outcome measurement	Goal: to judge the risk of bias related to the measurement of outcome (differential measurement of outcome related to the density reductions)			
Definition of the outcome	A clear definition of outcome is provided, including duration of follow-up and level and extent of the outcome construct	Clear definition: "For our main analysis, conditional logistic regression was used for analyzing risk of CBC", duration of follow-up given. Extend of the outcome construct in the wider database not indicated (only number with mammogram).	Partial	
Valid and reliable	The method of outcome measurement used is adequately	Yes – population registry.	Yes	

measurement of outcome	valid and reliable to limit misclassification bias			
Method and setting of outcome measurement	The method and setting of outcome measurement is the same for all study participants, including by age and obesity groups	Yes.	Yes	
Outcome measurement summary	Outcome of interest is adequately measured in study participants to sufficiently limit potential bias			Low
5. Study confounding	Goal: to judge the risk of bias due to confounding (i.e. the effect of density reductions is distorted by another factor that is related to density reductions and the outcome)			
Important confounders measured	Age, BMI, or another measure of adiposity are measured	Age measured, FA at baseline used as a proxy for BMI at baseline (justified by Lokate 2011; Breast Cancer Res; 13:R103).	Yes	
Definition of the confounding factor	Clear definitions are provided	Age at the first breast cancer diagnosis (+/- two years), FA categorized into quartiles.	Yes	
Valid and reliable measurement of confounders	Measurement of all important confounders is adequately valid and reliable	Measurement is adequately valid.	Yes	
Method and setting of confounding measurement	The method and setting of confounding measurement are the same for all study participants	Yes.	Yes	
Method used for missing data	Appropriate methods are used if imputation is used for missing confounder data	Imputation not used.	Yes	
Appropriate accounting for confounding	The primary analysis will be adjusted for at least age, either through the study design and analysis, or through adjustment in the analysis only; and other prognostic factors	Adjusted for age through matching. PDA change model additionally adjusted for FA and PDA baseline (was DA change model also adjusted for FA and DA baseline?). No adjustment for change in BMI (or FA) or chemotherapy-although this group is 'under-represented'. Not clear if subgroup analysis of primary interest to this review (i.e. women on endocrine treatment) was adjusted for other factors besides matching factors.	Partial	
Study confounding summary	Important potential confounders are appropriately accounted for, limiting potential bias with respect to the relationship between prognostic factor and outcome			Moderate
6. Statistical analysis and reporting	Goal: to judge the risk of bias related to the statistical analysis and presentation of results			
Presentation of analytical strategy, model development strategy	There is sufficient presentation of data to assess the adequacy of the analysis	Partially. Odds ratios are presented but not number of cases and controls by density change group. Can't separate out endocrine treatments.	Partial	

Model development strategy	The strategy for model building (i.e. inclusion of variables in the statistical model) is appropriate and is based on a conceptual framework or model	Appropriate per the study design. Not clear if further adjustment for the primary analysis of interest to this review.	Partial	
Reporting of results	The selected statistical model is adequate for the design of the study. There is no selective reporting of results	Appropriate. May be some selective reporting of results, but main focus was not on the comparison of interest for our review.	Partial	
Statistical analysis and presentation summary	The statistical analysis is appropriate for the design of the study, limiting potential for presentation of invalid or spurious results			Moderate

B.VII.viii Risk of bias table - van Nes 2015

Biases	Issues to consider for judging overall rating of risk of bias	Study Methods & Comments	Rating of reporting (adequacy of reporting: "yes", "partial", "no" or "unsure")	Rating of Risk of bias ("High", "Moderate", or "Low")
Instructions to assess the risk of each potential bias	These issues will guide your thinking and judgement about the overall risk of bias within each of the six domains. These issues are taken together to inform the overall judgement of potential bias for each of the six domains			
1. Study participation	Goal: to judge the risk of selection bias (likelihood that relationship between density reductions and outcome is different for participants and eligible non-participants)			
Source of target population	The source population or population of interest is adequately described for: a) treatment: (i) proportion with DCIS, (ii) cointerventions (chemotherapy/targeted therapy), (iii) severity of cancer at baseline (stage, % regional spread); b) prevention: (i) level of risk in population, including whether some or all are BRCA1/2 mutation carriers, (ii) prior hormone replacement therapy use, (iii) cointerventions such as diet or exercise regimens, or both	Source population is TEAM trial (n=2753), although not described, can be found in a referenced paper *van de Velde 2011; Lancet; 377:321-31.*	Yes	
Method used to identify population	The sampling frame and recruitment are adequately described, including methods to identify the sample sufficient to limit potential bias	Sampling frame and recruitment adequately described: 13 hospitals contributing to the TEAM sub-study (supplementary material), "based on adequate inclusion rate, geographical distribution and availability of analogue mammograms over time" (n=774).	Yes	
Recruitment period	Period of recruitment is adequately described	Not described.	No	
Place of recruitment	Place of recruitment (setting and geographic location) are adequately described	Well described (see data capture form).	Yes	
Inclusion and exclusion criteria	Inclusion and exclusion criteria are adequately described	Well described: (203/774 considered ineligible because hospital switched to digital during study period or contralateral mastectomy (n=571 eligible), 129/571 did not have available analogue mammograms (n=442). 219 in sequential arm and 223 in exemestane arm). "Of the 219 patients randomised	Yes	

		to the sequential arm, 28 stopped therapy within one year, five had no preoperative mammogram available, four had no (available) follow-up mammogram, and one did not start study medication, totalling 181 patients for the current analyses. Of the 223 patients randomised to exemestane, 21 stopped therapy within one year, three had no preoperative mammogram available and two had no follow-up mammogram available, leaving 197 patients for the current analyses".		
Adequate study participation	There is adequate participation in the study by eligible individuals	203/774 considered ineligible because hospital switched to digital during study period or contralateral mastectomy (n=571 eligible), 129/571 did not have available analogue mammograms (n=442). 219 in sequential arm and 223 in exemestane arm. After exclusions, sequential n=181 and exemestane n=197. Total included sample (n=378) compared with sampling frame not included (n=774-378=396) in Table IB.	Yes	
Baseline characteristics	The baseline study sample (i.e. individuals entering the study) is adequately described for (treatment and prevention) age, menopausal status, cointerventions; (treatment) % DCIS, disease severity; (prevention) breast cancer risk, prior hormone replacement therapy use	Well described (see data capture form).	Yes	
Summary study participation	The study sample represents the population of interest on key characteristics, sufficient to limit potential bias of the observed relationship between density change and outcome			Low
2. Study attrition	Goal: to judge the risk of attrition bias (likelihood that relationship between density reductions and outcome are different for completing and non-completing participants)			
Proportion of baseline sample available for analysis	Response rate (i.e. proportion of study sample allocated treatment who received treatment) is adequate	"Of the 219 patients randomised to the sequential arm, 28 stopped therapy within one year [...] and one did not start study medication." and "Of the 223 patients randomised to exemestane, 21 stopped therapy within one year".	Yes	
Attempts to collect	Attempts to collect information on participants who dropped out of the	Not described (for 378 women in study sample)	No	

information on participants who dropped out	study are described	but whole trial information should be available.		
Reasons and potential impact of subjects lost to follow-up	Reasons for loss to follow-up are provided	Not described (for 378 women in study sample) but whole trial information should be available.	No	
Outcome and prognostic factor information on those lost to follow-up	Participants lost to follow-up are adequately described for age at entry and cointerventions (if any), and for a) treatment: (i) DCIS, (ii) disease severity; b) prevention: (i) risk of breast cancer including BRCA1/2 carriers and testing. Whether loss to follow-up or inability to retrieve mammograms, or both, was likely related to the study outcome	Not described.	No	
Study attrition summary	There are no important differences between these characteristics in participants who completed the study and those who did not. Loss to follow-up (from baseline sample to study population analysed) is not associated with key characteristics (i.e. the study data adequately represent the sample) sufficient to limit potential bias to the observed relationship between density change and outcome			Moderate
3. Prognostic factor measurement	Goal: to judge the risk of measurement bias related to how mammographic density was measured (differential measurement of mammographic density related to the level of outcome)			
Definition of the prognostic factor	A clear definition or description of mammographic density is provided (e.g. including the method of measurement, if subjective then who undertook it, if treatment then whether contralateral breast assessed)	Clear definition: "visual estimation technique classifying the percentage of mammographic breast density into one of six categories: 0%, <10%, 10–25%, 25–50%, 50–75%, and >75%" by "three independent radiologists being very experienced in reading mammograms". Contralateral breast used.	Yes	
Valid and reliable measurement of prognostic factor	Method of mammographic density change measurement is adequately valid and reliable to limit misclassification bias (e.g. may include relevant outside sources of information on measurement properties; also characteristics, such as measurement blinded to case status)	"The interclass correlation coefficient between the three radiologists (raters) was satisfactory: 0.74". "The patient's identity, date of mammogram and randomisation arm were blinded to the radiologists", but no mention if blinded to case status.	Yes	
	Continuous variables are reported or appropriate cut-points (i.e. not data-dependent (except for percentiles)) are used	Boyd 6-category scale.	Yes	
Method and setting of prognostic	The method and setting of measurement of mammographic density is the same for all study	Method and setting same (three radiologists read all mammograms, although	Partial	

factor measurement	participants. The same mammogram type (film/digital) is used for both baseline and follow-up. The time at which baseline and follow-up mammograms have low variability between participants	whether each one read mammograms per woman etc. not mentioned), same mammogram type (film). Unknown time from baseline to follow-up mammograms.		
Proportion of data on prognostic factor available for analysis	Adequate proportion of the study sample has complete data for the change in mammographic density variable	8 had no preoperative mammogram and 6 had no follow-up mammogram (but not included in final study sample n=378). "Of the total group of 378 patients, 359 mammograms (171 sequential arm, 188 exemestane arm) were reviewed after one year of endocrine therapy, 292 mammograms (123 sequential arm, 169 exemestane arm) after two years, and 116 mammograms (17 of tamoxifen patients and 99 of exemestane patients) after three years of endocrine therapy" - it is unclear which follow-up mammogram(s) were used to calculate density change.	Partial	
Method used for missing data	Appropriate methods of imputation are used for missing mammographic density data	NA	Yes	
Summary	Prognostic factor is adequately measured in study participants to sufficiently limit potential bias			Low
4. Outcome measurement	Goal: to judge the risk of bias related to the measurement of outcome (differential measurement of outcome related to the density reductions)			
Definition of the outcome	A clear definition of outcome is provided, including duration of follow-up and level and extent of the outcome construct	Adequate definition "Time to LRR, DR or CBC was calculated from the start of endocrine therapy up to the date of a LRR, a DR or CBC, respectively", but no reasons for censoring (per protocol so censored when stopped treatment but no other reasons provided).	Partial	
Valid and reliable measurement of outcome	The method of outcome measurement used is adequately valid and reliable to limit misclassification bias	From TEAM trial database.	Yes	
Method and setting of outcome measurement	The method and setting of outcome measurement is the same for all study participants, including by age and obesity groups	Yes.	Yes	
Outcome measurement summary	Outcome of interest is adequately measured in study participants to sufficiently limit potential bias			Low
5. Study confounding	Goal: to judge the risk of bias due to confounding (i.e. the effect of density reductions is distorted by another factor that is related to density reductions and the			

	outcome)			
Important confounders measured	Age, BMI, or another measure of adiposity are measured	Age and BMI measured.	Yes	
Definition of the confounding factor	Clear definitions are provided	No definition on how BMI measured.	No	
Valid and reliable measurement of confounders	Measurement of all important confounders is adequately valid and reliable	Appears adequate (from trial).	Yes	
Method and setting of confounding measurement	The method and setting of confounding measurement are the same for all study participants	Yes.	Yes	
Method used for missing data	Appropriate methods are used if imputation is used for missing confounder data	23 missing BMI, not stated if imputed.	No	
Appropriate accounting for confounding	The primary analysis will be adjusted for at least age, either through the study design and analysis, or through adjustment in the analysis only; and other prognostic factors	Does not appear to be adjusted.	No	
Study confounding summary	Important potential confounders are appropriately accounted for, limiting potential bias with respect to the relationship between prognostic factor and outcome			Moderate
6. Statistical analysis and reporting	Goal: to judge the risk of bias related to the statistical analysis and presentation of results			
Presentation of analytical strategy, model development strategy	There is sufficient presentation of data to assess the adequacy of the analysis	Not sufficient presentation of data. No data presented on density change. Both treatment arms combined.	No	
Model development strategy	The strategy for model building (i.e. inclusion of variables in the statistical model) is appropriate and is based on a conceptual framework or model	No model building done.	Yes	
Reporting of results	The selected statistical model is adequate for the design of the study. There is no selective reporting of results	Statistical model for density change analysis unclear, no results presented for density change.	No	
Statistical analysis and presentation summary	The statistical analysis is appropriate for the design of the study, limiting potential for presentation of invalid or spurious results			High

B.VII.ix ROBINS-I tool for Cuzick 2011

ROBINS-I tool (Stage I): At protocol stage

Specify the review question

Participants	See review 'Types of participants' section
Experimental intervention	See review 'Interventions' section
Comparator	See review 'Comparators' section
Outcomes	See review 'Types of outcome measures' section

List the confounding domains relevant to all or most studies

Age
Menopausal status
Body mass index
Family history of disease
Hormone replacement therapy use
Benign breast disease
Previous cancer other than breast cancer
Ethnicity

List co-interventions that could be different between intervention groups and that could impact on outcomes

Hormone replacement therapy
Risk-reducing surgery

ROBINS-I tool (Stage II): For each study

Specify a target randomized trial specific to the study

Design	~~Individually randomized~~/Cluster randomized/~~Matched (e.g. cross-over)~~ (block randomisation (permuted block sizes of six, eight or ten))
Participants	IBIS-I participants: 35 - 70 years old with at least twice the average risk of a 50-year-old woman of developing breast cancer
Experimental intervention	Oral Tamoxifen 20mg/daily
Comparator	Oral placebo/daily

Is your aim for this study…?

☐ to assess the effect of *assignment to* intervention

☒ to assess the effect of *starting and adhering to* intervention

Specify the outcome

Specify which outcome is being assessed for risk of bias (typically from among those earmarked for the Summary of Findings table). Specify whether this is a proposed benefit or harm of intervention.

Proposed benefit of intervention - Prevention: incidence of invasive breast cancer and DCIS

Specify the numerical result being assessed

In case of multiple alternative analyses being presented, specify the numeric result (e.g. RR = 1.52 (95% CI 0.83 to 2.77) and/or a reference (e.g. to a table, figure or paragraph) that uniquely defines the result being assessed.

Predictive biomarker worked out from raw data: OR=0.53 (95% CI, 0.21 to 1.32), p=0.17

Preliminary consideration of confounders

Complete a row for each important confounding domain (i) listed in the review protocol; and (ii) relevant to the setting of this particular study, or which the study authors identified as potentially important.

"Important" confounding domains are those for which, in the context of this study, adjustment is expected to lead to a clinically important change in the estimated effect of the intervention. "Validity" refers to whether the confounding variable or variables fully measure the domain, while "reliability" refers to the precision of the measurement (more measurement error means less reliability).

(i) Confounding domains listed in the review protocol				
Confounding domain	Measured variable(s)	Is there evidence that controlling for this variable was unnecessary?*	Is the confounding domain measured validly and reliably by this variable (or these variables)?	OPTIONAL: Is failure to adjust for this variable (alone) expected to favour the experimental intervention or the comparator?
Age	Age at entry (years)	No	Yes / ~~No / No information~~	Favour experimental / Favour comparator / No information
Menopausal status	Menopausal status at entry (Premenopausal/Postmenopausal)	No	Yes / ~~No / No information~~	
Body mass index	Body mass index at entry (kg/m^2)	No	Yes / ~~No / No information~~	
Family history of disease	Extensive family history collected as part of Tyrer-Cuzick risk model at entry (%)	No	Yes / ~~No / No information~~	
Hormone replacement therapy use	Use of hormone replacement therapy during study (Never/Previous/Current)	No	Yes / ~~No / No information~~	
Benign breast disease	Atypical hyperplasia or LCIS at entry (No/Yes)	No	Yes / ~~No / No information~~	
Previous cancer other than breast cancer	Women with a history of any invasive cancer (excluding skin cancer) were excluded	Yes - women with a history of any invasive cancer (excluding skin cancer) were excluded from the trial.	Yes / ~~No / No information~~	
Ethnicity	Mixed	No	~~Yes / No~~ / No information	

* In the context of a particular study, variables can be demonstrated not to be confounders and so not included in the analysis: (a) if they are not predictive of the outcome; (b) if they are not predictive of intervention; or (c) because adjustment makes no or minimal difference to the estimated effect of the primary parameter. Note that "no statistically significant association" is not the same as "not predictive".

Preliminary consideration of co-interventions

Complete a row for each important co-intervention (i) listed in the review protocol; and (ii) relevant to the setting of this particular study, or which the study authors identified as important. "Important" co-interventions are those for which, in the context of this study, adjustment is expected to lead to a clinically important change in the estimated effect of the intervention.

(i) Co-interventions listed in the review protocol		
Co-intervention	Is there evidence that controlling for this co-intervention was unnecessary (e.g. because it was not administered)?	Is presence of this co-intervention likely to favour outcomes in the experimental intervention or the comparator
Hormone replacement therapy	No	~~Favour experimental / Favour comparator~~ / No information
Risk-reducing surgery	No	~~Favour experimental / Favour comparator~~ / No information

Risk of bias assessment

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? If N/PN to 1.1: the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	• Analysis adjusted for age, body mass index, and benign breast disease. • Women with a history of any invasive cancer (excluding skin cancer) were excluded from the trial (Cuzick 2015; Lancet Oncology; 16(1): 67-75). • "Adjusting for HRT use had no material impact on the estimate of risk reduction associated with a reduction in breast density". • Although not everyone had a family history of breast cancer, all women were "at least twice the average risk of a 50-year-old woman of developing breast cancer". • Menopausal status: "Overall, tamoxifen was more effective in preventing estrogen receptor–positive breast cancer than it was in preventing estrogen receptor– negative breast cancer and was more effective in women who were premenopausal, had never taken HRT, or who had a previous diagnosis of atypical hyperplasia or LCIS, but there were no statistically significant differences in the odds ratios between the subgroup". • There was no adjustment for ethnicity, but all women were living in the UK or Finland.	~~Y / PY~~ / ~~PN~~ / N
Risk of bias judgement		Low

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? If N/PN to 2.1: go to 2.4	No: "To minimize the administrative workload, control subjects were selected only from the major participating UK centers in Aberdeen, Bristol, Cardiff, Edinburgh, London, Manchester, Nottingham, and Southampton". Cases from the UK and Finland.	~~Y / PY~~ / ~~PN~~ / N / ~~NI~~
2.4. Do start of follow-up and start of intervention coincide for most participants?	Yes: start of follow-up at start of treatment.	Y / ~~PY~~ / ~~PN~~ / ~~N~~ / ~~NI~~
Risk of bias judgement		Low

Bias in classification of interventions		
3.1 Were intervention groups clearly defined?	Yes – randomised controlled trial.	Y / ~~PY~~ / ~~PN~~ / ~~N~~ / ~~NI~~
3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Yes – randomised controlled trial.	Y / ~~PY~~ / ~~PN~~ / ~~N~~ / ~~NI~~
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	No – double-blind randomised controlled trial whereby breast cancer event reported to trial by local co-ordinating centres and Office for National Statistics unaware of intervention allocation (Cuzick 2015; Lancet Oncology; 16(1): 67-75).	~~Y / PY~~ / ~~PN~~ / N / ~~NI~~
Risk of bias judgement		Low

Bias due to deviations from intended interventions		
4.3. Were important co-interventions balanced across intervention groups?	Hormone replacement therapy – yes. Risk-reducing surgery – not described, although mammography required so ineligible.	~~Y~~ / PY / ~~PN~~ / ~~N~~ / ~~NI~~
4.4. Was the intervention implemented successfully for most participants?	Yes – all women consented to the trial, withdrawals only reported n=44 Australian women in Cuzick 2015; Lancet Oncology; 16(1): 67-75 (and not included in density sub-study).	Y / ~~PY~~ / ~~PN~~ / ~~N~~ / ~~NI~~
4.5. Did study participants adhere to the assigned intervention regimen?	Yes – not described, but withdrawals only reported n=44 Australian women in Cuzick 2015; Lancet Oncology; 16(1): 67-75 (and not included in density sub-study). Analysis of compliance in cases: "no statistically significant difference between subjects in the tamoxifen arm who experienced a reduction in mammographic density of less than 10% and subjects in the tamoxifen arm who experienced a greater reduction (P = .25)".	Y / ~~PY~~ / ~~PN~~ / ~~N~~ / ~~NI~~
Risk of bias judgement		Low

Bias due to missing data		
5.1 Were outcome data available for all, or nearly all, participants?	Yes, randomised controlled trial database of outcomes, Also, "In the UK, cancers and deaths are also reported to the IBIS-I central office by the Office for National Statistics" (Cuzick 2015; Lancet Oncology; 16(1): 67-75).	Y / ~~PY~~ / ~~PN~~ / ~~N~~ / ~~NI~~
5.2 Were participants excluded due to missing data on intervention status?	No, no missing data on intervention as randomised controlled trial	~~Y~~ / ~~PY~~ / ~~PN~~ / N / ~~NI~~
5.3 Were participants excluded due to missing data on other variables needed for the analysis?	Yes, 1065-1049=16 women missing in main result due to missing BMI	Y / ~~PY~~ / ~~PN~~ / ~~N~~ / ~~NI~~
5.4 If PN/N to 5.1, or Y/PY to 5.2 or 5.3: Are the proportion of participants and reasons for missing data similar across interventions?	Not described but low number of women with missing BMI so unlikely to affect result (comparison between interventions).	~~NA~~ / ~~Y~~ / PY / ~~PN~~ / ~~N~~ / ~~NI~~
5.5 If PN/N to 5.1, or Y/PY to 5.2 or 5.3: Is there evidence that results were robust to the presence of missing data?	Not described but low number of women with missing BMI so unlikely to affect result (comparison between interventions).	~~NA~~ / ~~Y~~ / PY / ~~PN~~ / ~~N~~ / ~~NI~~
Risk of bias judgement		Low

Bias in measurement of outcomes		
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	No, randomised controlled trial database of outcomes. Also, "In the UK, cancers and deaths are also reported to the IBIS-I central office by the Office for National Statistics" (Cuzick 2015; Lancet Oncology; 16(1): 67-75).	~~Y~~ / ~~PY~~ / ~~PN~~ / N / ~~NI~~
6.2 Were outcome assessors aware of the intervention received by study participants?	No, double-blind randomised controlled trial whereby breast cancer event reported to trial by local co-ordinating centres or Office for National Statistics who were unaware of intervention allocation.	~~Y~~ / ~~PY~~ / ~~PN~~ / N / ~~NI~~
6.3 Were the methods of outcome assessment comparable across intervention groups?	Yes, randomised controlled trial database of outcomes. Also, "In the UK, cancers and deaths are also reported to the IBIS-I central office by the Office for National Statistics" (Cuzick 2015; Lancet Oncology; 16(1): 67-75).	Y / ~~PY~~ / ~~PN~~ / ~~N~~ / ~~NI~~
6.4 Were any systematic errors in measurement of the outcome related to intervention received?	No, double-blind randomised controlled trial whereby breast cancer event reported to trial by local co-ordinating centres or Office for National Statistics unaware of intervention allocation.	~~Y~~ / ~~PY~~ / ~~PN~~ / N / ~~NI~~
Risk of bias judgement		Low

Bias in selection of the reported result		
Is the reported effect estimate likely to be selected, on the basis of the results, from...		
7.1. ... multiple outcome *measurements* within the outcome domain?	No, one outcome of incidence of invasive breast cancer and DCIS.	~~Y / PY~~ / ~~PN~~ / N / ~~NI~~
7.2 ... multiple *analyses* of the intervention-outcome relationship?	No, interaction effect not reported but is worked out from the paper (unadjusted) and from the raw data (adjusted).	~~Y / PY~~ / ~~PN~~ / N / ~~NI~~
7.3 ... different *subgroups*?	No, effect estimate calculated in all women, no subgroups for interaction analysis.	~~Y / PY~~ / ~~PN~~ / N / ~~NI~~
Risk of bias judgement		Low

Overall bias		
Risk of bias judgement		Low

Appendix C: Supplementary material for Chapters 5 and 6

C.I Weighting IBIS-I postmenopausal density changes based on IBIS-II age structure (placebo cases).

No. in IBIS-I stratified by density change category and age group *(A)*												
	-30	-25	-20	-15	-10	-5	0	5	10	15	Total	%
35-39	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0	0.00%
40-44	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0	0.00%
45-49	0	0	0	0	0	2	2	0	0	0	4	11.76%
50-54	0	0	1	0	1	3	5	1	0	1	12	35.29%
55-59	0	0	0	0	0	2	7	3	0	0	12	35.29%
60-64	0	0	0	0	0	2	2	0	0	0	4	11.76%
65-69	0	0	0	0	1	0	1	0	0	0	2	5.88%
70-74	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0	0.00%
Total	0	0	1	0	2	9	17	4	0	1	34	100.00%
Density Change Distribution %	0.00%	0.00%	2.94%	0.00%	5.88%	26.47%	50.00%	11.76%	0.00%	2.94%	100.00%	

	Age groups								
	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	Total
IBIS-II distribution % *(B)*	0.00%	0.35%	2.95%	19.23%	28.29%	29.51%	17.77%	1.90%	100.00%
IBIS-I distribution % *(C)*	0.00%	0.00%	11.76%	35.29%	35.29%	11.76%	5.88%	0.00%	100.00%
Weight = B/C *(D)*	N/A	N/A	0.25	0.54	0.80	2.51	3.02	N/A	7.13
No. in IBIS-I *(E)*	0	0	4	12	12	4	2	0	34
E*D	N/A	N/A	1.00	6.54	9.62	10.04	6.04	N/A	33.23
Reweighted IBIS-I distribution % (F)	N/A	N/A	3.02%	19.67%	28.94%	30.19%	18.18%	N/A	100.00%

A*D												
	-30	-25	-20	-15	-10	-5	0	5	10	15	Total	%
35-39	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.00	0.00%
40-44	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.00	0.00%
45-49	0.00	0.00	0.00	0.00	0.00	0.50	0.50	0.00	0.00	0.00	1.00	3.02%
50-54	0.00	0.00	0.54	0.00	0.54	1.63	2.72	0.54	0.00	0.54	6.54	19.67%
55-59	0.00	0.00	0.00	0.00	0.00	1.60	5.61	2.40	0.00	0.00	9.62	28.94%
60-64	0.00	0.00	0.00	0.00	0.00	5.02	5.02	0.00	0.00	0.00	10.04	30.19%
65-69	0.00	0.00	0.00	0.00	3.02	0.00	3.02	0.00	0.00	0.00	6.04	18.18%
70-74	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.00	0.00%
Total	0.00	0.00	0.54	0.00	3.57	8.76	16.87	2.95	0.00	0.54	33.23	100.00%
Density Change Distribution %	0.00%	0.00%	1.64%	0.00%	10.73%	26.35%	50.77%	8.87%	0.00%	1.64%	100.00%	

C.II Weighting IBIS-I postmenopausal density changes based on IBIS-II age structure (placebo controls).

No. in IBIS-I stratified by density change category and age group *(A)*

	-30	-25	-20	-15	-10	-5	0	5	10	15	Total	%
35-39	0	0	0	0	0	0	1	0	0	0	1	0.44%
40-44	0	0	1	0	0	1	2	0	0	0	4	1.78%
45-49	0	1	2	2	3	6	17	4	1	1	37	16.44%
50-54	1	2	3	2	10	19	35	5	0	1	78	34.67%
55-59	0	1	1	5	10	14	27	4	1	1	64	28.44%
60-64	1	1	0	1	2	6	19	3	1	1	35	15.56%
65-69	0	0	0	0	0	1	3	1	1	0	6	2.67%
70-74	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0	0.00%
Total	2	5	7	10	25	47	104	17	4	4	225	100.00%
Density Change Distribution %	0.89%	2.22%	3.11%	4.44%	11.11%	20.89%	46.22%	7.56%	1.78%	1.78%	100.00%	

	Age groups								
	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	Total
IBIS-II distribution % *(B)*	0.00%	0.35%	2.95%	19.23%	28.29%	29.51%	17.77%	1.90%	100.00%
IBIS-I distribution % *(C)*	0.44%	1.78%	16.44%	34.67%	28.44%	15.56%	2.67%	0.00%	100.00%
Weight = B/C *(D)*	0.00	0.20	0.18	0.55	0.99	1.90	6.66	N/A	10.49
No. in IBIS-I *(E)*	1	4	37	78	64	35	6	0	225
E*D	0.00	0.79	6.64	43.26	63.65	66.41	39.98	N/A	220.73
Reweighted IBIS-I distribution % (F)	0.00%	0.36%	3.01%	19.60%	28.84%	30.09%	18.11%	N/A	100.00%

A*D

	-30	-25	-20	-15	-10	-5	0	5	10	15	Total	%
35-39	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00%
40-44	0.00	0.00	0.20	0.00	0.00	0.20	0.39	0.00	0.00	0.00	0.79	0.36%
45-49	0.00	0.18	0.36	0.36	0.54	1.08	3.05	0.72	0.18	0.18	6.64	3.01%
50-54	0.55	1.11	1.66	1.11	5.55	10.54	19.41	2.77	0.00	0.55	43.26	19.60%
55-59	0.00	0.99	0.99	4.97	9.94	13.92	26.85	3.98	0.99	0.99	63.65	28.84%
60-64	1.90	1.90	0.00	1.90	3.79	11.38	36.05	5.69	1.90	1.90	66.41	30.09%
65-69	0.00	0.00	0.00	0.00	0.00	6.66	19.99	6.66	6.66	0.00	39.98	18.11%
70-74	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.00	0.00%
Total	2.45	4.18	3.21	8.34	19.82	43.78	105.75	19.82	9.73	3.63	220.73	100.00%
Density Change Distribution %	1.11%	1.89%	1.46%	3.78%	8.98%	19.84%	47.91%	8.98%	4.41%	1.64%	100.00%	

C.III Weighting IBIS-I postmenopausal density changes based on IBIS-II age structure (anastrozole cases).

No. in IBIS-I stratified by density change category and age group (A)

	-40	-35	-30	-25	-20	-15	-10	-5	0	5	10	50	Total	%
40-44	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
45-49	0	0	0	1	0	1	0	1	1	0	0	0	4	13.79%
50-54	0	0	0	0	0	2	2	5	4	1	0	0	14	48.28%
55-59	0	0	0	0	0	0	0	1	4	0	0	0	5	17.24%
60-64	0	0	0	0	0	1	1	0	2	0	0	0	4	13.79%
65-69	0	0	0	0	0	0	0	1	1	0	0	0	2	6.90%
70-74	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%
Total	0	0	0	1	0	4	3	8	12	1	0	0	29	100.00%
Density Change Distribution %	0.00%	0.00%	0.00%	3.45%	0.00%	13.79%	10.34%	27.59%	41.38%	3.45%	0.00%	0.00%	100.00%	

	Age groups							
	40-44	45-49	50-54	55-59	60-64	65-69	70-74	Total
IBIS-II distribution % (B)	0.35%	2.95%	19.23%	28.29%	29.51%	17.77%	1.90%	100.00%
IBIS-I distribution % (C)	0.00%	13.79%	48.28%	17.24%	13.79%	6.90%	0.00%	100.00%
Weight = B/C (D)	N/A	0.21	0.40	1.64	2.14	2.58	N/A	6.97
No. in IBIS-I (E)	0	4	14	5	4	2	0	29
E*D	N/A	0.86	5.58	8.20	8.56	5.15	N/A	28.35
Reweighted IBIS-I distribution % (F)	N/A	3.02%	19.67%	28.94%	30.19%	18.18%	N/A	100.00%

A*D

	-40	-35	-30	-25	-20	-15	-10	-5	0	5	10	50	Total	%
40-44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00%
45-49	0.00	0.00	0.00	0.21	0.00	0.21	0.00	0.21	0.21	0.00	0.00	0.00	0.86	3.02%
50-54	0.00	0.00	0.00	0.00	0.00	0.80	0.80	1.99	1.59	0.40	0.00	0.00	5.58	19.67%
55-59	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.64	6.56	0.00	0.00	0.00	8.20	28.94%
60-64	0.00	0.00	0.00	0.00	0.00	2.14	2.14	0.00	4.28	0.00	0.00	0.00	8.56	30.19%
65-69	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.58	2.58	0.00	0.00	0.00	5.15	18.18%
70-74	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00%
Total	0.00	0.00	0.00	0.21	0.00	3.15	2.94	6.42	15.23	0.40	0.00	0.00	28.35	100.00%
Density Change Distribution %	0.00%	0.00%	0.00%	0.75%	0.00%	11.11%	10.36%	22.66%	53.71%	1.41%	0.00%	0.00%	100.00%	

C.IV Weighting IBIS-I postmenopausal density changes based on IBIS-II age structure (anastrozole controls).

No. in IBIS-I stratified by density change category and age group *(A)*														
	-40	-35	-30	-25	-20	-15	-10	-5	0	5	10	50	Total	%
40-44	0	0	0	0	0	0	2	0	0	0	0	0	2	0.93%
45-49	0	1	0	2	3	5	8	5	10	0	0	0	34	15.89%
50-54	1	2	1	2	5	5	14	19	28	4	0	1	82	38.32%
55-59	1	1	0	0	1	3	10	13	22	2	1	0	54	25.23%
60-64	0	0	0	0	0	2	7	3	20	0	2	0	34	15.89%
65-69	0	0	0	0	0	0	0	0	5	2	0	0	7	3.27%
70-74	0	0	0	0	0	0	0	0	1	0	0	0	1	0.47%
Total	2	4	1	4	9	15	41	40	86	8	3	1	214	100.00%
Density Change Distribution %	0.93%	1.87%	0.47%	1.87%	4.21%	7.01%	19.16%	18.69%	40.19%	3.74%	1.40%	0.47%	100.00%	

	Age groups							
	40-44	45-49	50-54	55-59	60-64	65-69	70-74	Total
IBIS-II distribution % *(B)*	0.35%	2.95%	19.23%	28.29%	29.51%	17.77%	1.90%	100.00%
IBIS-I distribution % *(C)*	0.93%	15.89%	38.32%	25.23%	15.89%	3.27%	0.47%	100.00%
Weight = B/C *(D)*	0.38	0.19	0.50	1.12	1.86	5.43	4.06	13.54
No. in IBIS-I *(E)*	2	34	82	54	34	7	1	214
E*D	0.75	6.32	41.15	60.54	63.16	38.02	4.06	214.00
Reweighted IBIS-I distribution % (F)	0.35%	2.95%	19.23%	28.29%	29.51%	17.77%	1.90%	100.00%

A*D														
	-40	-35	-30	-25	-20	-15	-10	-5	0	5	10	50	Total	%
40-44	0.00	0.00	0.00	0.00	0.00	0.00	0.75	0.00	0.00	0.00	0.00	0.00	0.75	0.35%
45-49	0.00	0.19	0.00	0.37	0.56	0.93	1.49	0.93	1.86	0.00	0.00	0.00	6.32	2.95%
50-54	0.50	1.00	0.50	1.00	2.51	2.51	7.03	9.53	14.05	2.01	0.00	0.50	41.15	19.23%
55-59	1.12	1.12	0.00	0.00	1.12	3.36	11.21	14.57	24.66	2.24	1.12	0.00	60.54	28.29%
60-64	0.00	0.00	0.00	0.00	0.00	3.72	13.00	5.57	37.15	0.00	3.72	0.00	63.16	29.51%
65-69	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	27.16	10.86	0.00	0.00	38.02	17.77%
70-74	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.06	0.00	0.00	0.00	4.06	1.90%
Total	1.62	2.31	0.50	1.38	4.19	10.52	33.48	30.61	108.95	15.11	4.84	0.50	214.00	100.00%
Density Change Distribution %	0.76%	1.08%	0.23%	0.64%	1.96%	4.91%	15.64%	14.30%	50.91%	7.06%	2.26%	0.23%	100.00%	

C.V Distribution of density changes at different effect sizes of anastrozole.

Cases					Controls				
Density Change	Placebo Distribution %	1 effect size Distribution %	3/4 effect size Distribution %	1/2 effect size Distribution %	Density Change	Placebo Distribution %	1 effect size Distribution %	3/4 effect size Distribution %	1/2 effect size Distribution %
-40	0.00%	0.00%	0.00%	0.00%	-40	0.00%	0.76%	0.57%	0.38%
-35	0.00%	0.00%	0.00%	0.00%	-35	0.00%	1.08%	0.81%	0.54%
-30	0.00%	0.00%	0.00%	0.00%	-30	1.11%	0.23%	0.45%	0.67%
-25	0.00%	0.75%	0.57%	0.38%	-25	1.89%	0.64%	0.96%	1.27%
-20	1.64%	0.00%	0.41%	0.82%	-20	1.46%	1.96%	1.83%	1.71%
-15	0.00%	11.11%	8.34%	5.56%	-15	3.78%	4.91%	4.63%	4.35%
-10	10.73%	10.36%	10.45%	10.54%	-10	8.98%	15.64%	13.98%	12.31%
-5	26.35%	22.66%	23.58%	24.50%	-5	19.84%	14.30%	15.69%	17.07%
0	50.77%	53.71%	52.98%	52.24%	0	47.91%	50.91%	50.16%	49.41%
5	8.87%	1.41%	3.27%	5.14%	5	8.98%	7.06%	7.54%	8.02%
10	0.00%	0.00%	0.00%	0.00%	10	4.41%	2.26%	2.80%	3.34%
15	1.64%	0.00%	0.41%	0.82%	15	1.64%	0.00%	0.41%	0.82%
50	0.00%	0.00%	0.00%	0.00%	50	0.00%	0.23%	0.18%	0.12%
Total	100.00%	100.00%	100.00%	100.00%	Total	100.00%	100.00%	100.00%	100.00%

C.VI Case/control weighted distribution of density changes at different effect sizes of anastrozole.

Estimated distributions based on IBIS-I: placebo: cases=4%, controls=96%, anastrozole: cases=2%, controls=98%.

Cases & Controls				
Density Change	Placebo Distribution %	1 effect size Distribution %	3/4 effect size Distribution %	1/2 effect size Distribution %
-40	0.00%	0.74%	0.56%	0.37%
-35	0.00%	1.06%	0.79%	0.53%
-30	1.06%	0.23%	0.44%	0.66%
-25	1.81%	0.64%	0.95%	1.25%
-20	1.46%	1.92%	1.80%	1.69%
-15	3.61%	5.04%	4.71%	4.37%
-10	9.06%	15.53%	13.90%	12.28%
-5	20.12%	14.48%	15.85%	17.22%
0	48.03%	50.97%	50.22%	49.47%
5	8.98%	6.94%	7.45%	7.96%
10	4.22%	2.21%	2.74%	3.27%
15	1.64%	0.00%	0.41%	0.82%
50	0.00%	0.23%	0.17%	0.11%
Total	100.00%	100.00%	100.00%	100.00%

C.VII Case/control weighted distribution of dichotomised density changes at different effect sizes of anastrozole.

Estimated distributions based on IBIS-I: placebo: cases=4%, controls=96%, anastrozole: cases=2%, controls=98%.

Density Change	Placebo Distribution %	1 effect size Distribution %	3/4 effect size Distribution %	1/2 effect size Distribution %
≥10% reduction	17.01%	25.17%	23.15%	21.14%
<10% reduction	82.99%	74.83%	76.85%	78.86%
Total	100.00%	100.00%	100.00%	100.00%

C.VIII Power calculation for different sample sizes and effect sizes of anastrozole.

Continuous	Effect size		
Sample size per arm	1	3/4	1/2
400	0.616	0.400	0.214
450	0.663	0.432	0.239
500	0.720	0.487	0.257
550	0.752	0.525	0.275
600	0.791	0.552	0.295
<10% vs. ≥10%	**Effect size**		
Sample size per arm	1	3/4	1/2
400	0.810	0.583	0.308
450	0.854	0.639	0.347
500	0.890	0.677	0.380
550	0.907	0.725	0.416
600	0.937	0.763	0.446

C.IX Case/control weighted distribution of 600 women per arm

Estimated distributions based on IBIS-I: placebo: cases=4%, controls=96%, anastrozole: cases=2%, controls=98%.

	Cases	Controls	Total
Placebo	4% of 600=24	96% of 600=576	600
Anastrozole	2% of 600=12	98% of 600=588	600
Total	36	1164	1200

C.X Case/control weighted distribution of 600 women per arm, accounting for baseline density <10% in IBIS-I postmenopausal women.

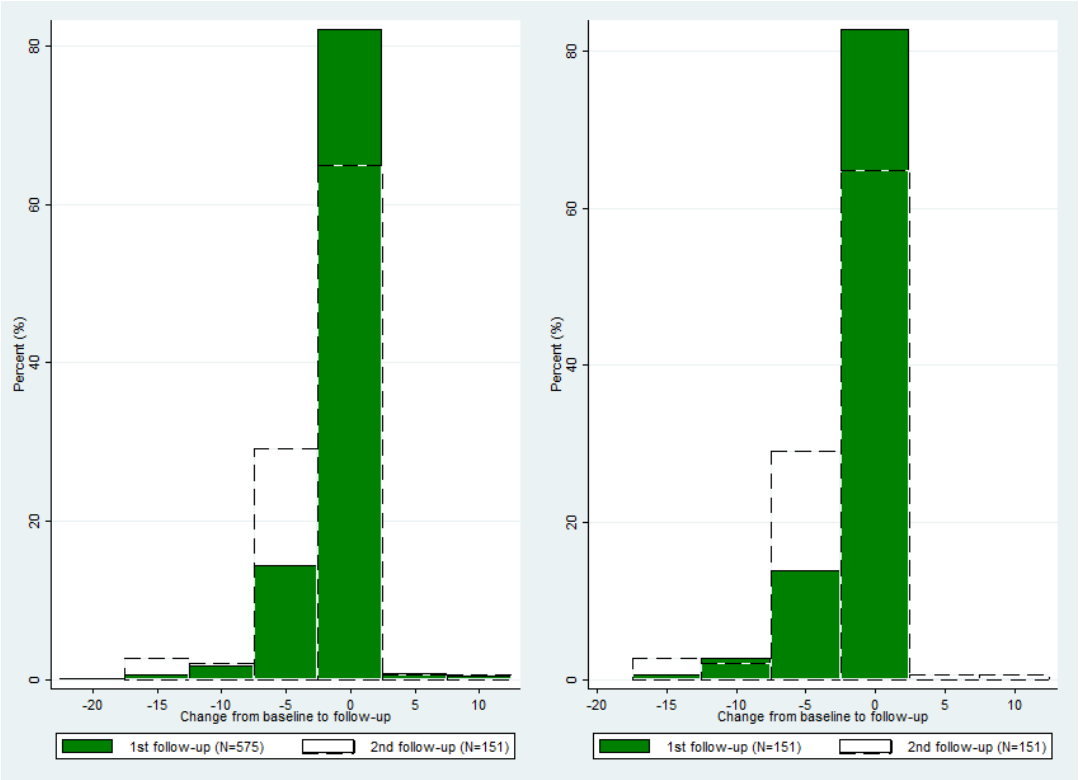
Estimated distributions based on IBIS-I: placebo: cases=4%, controls=96%, anastrozole: cases=2%, controls=98%. There were 81% of postmenopausal cases who had baseline density <10% and 79% of postmenopausal controls who had baseline density <10% in IBIS-I.

	Cases	Controls	Total
Placebo	24/0.81=30	576/0.79=729	759
Anastrozole	12/0.81=15	588/0.79=744	759
Total	44	1473	1518

Because the trial was randomised, there was no reason for variation in these proportions of density by treatment arm at baseline; there were some rounding errors for the total amounts.

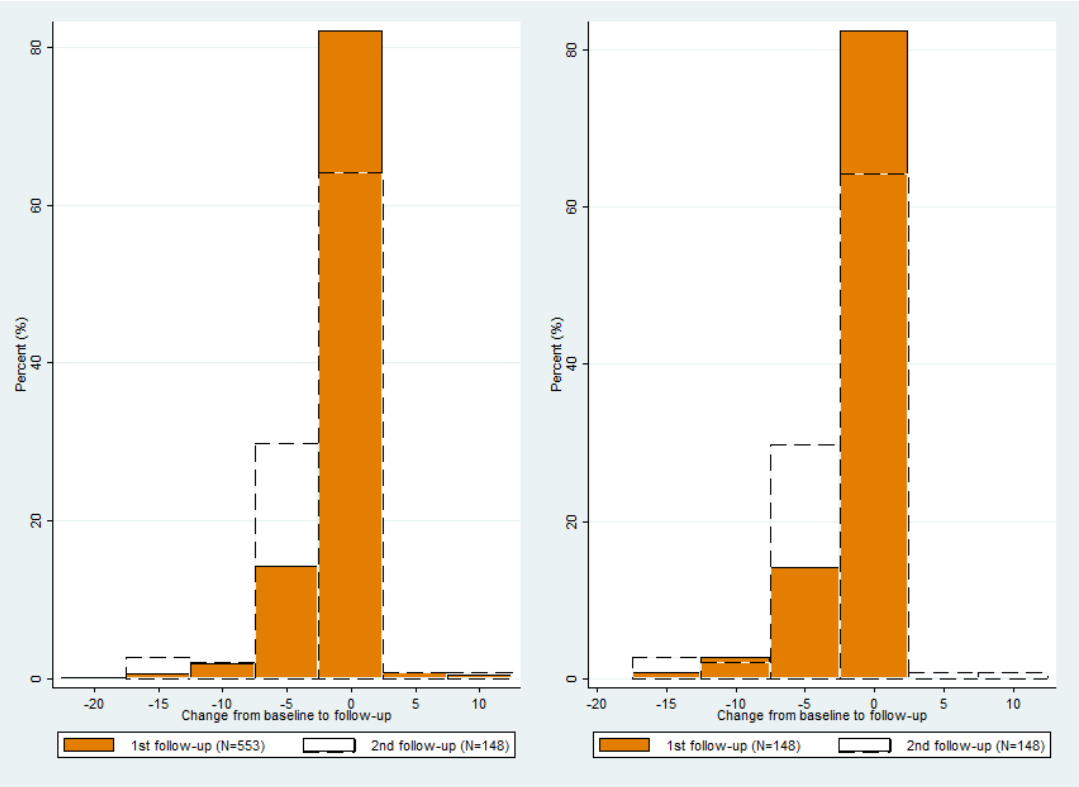
C.XI Density change distribution in cases and controls: 1st follow-up vs. 2nd follow-up (all women and subgroup with available final follow-up)

All women and subgroup with available final follow-up; left and right, respectively:



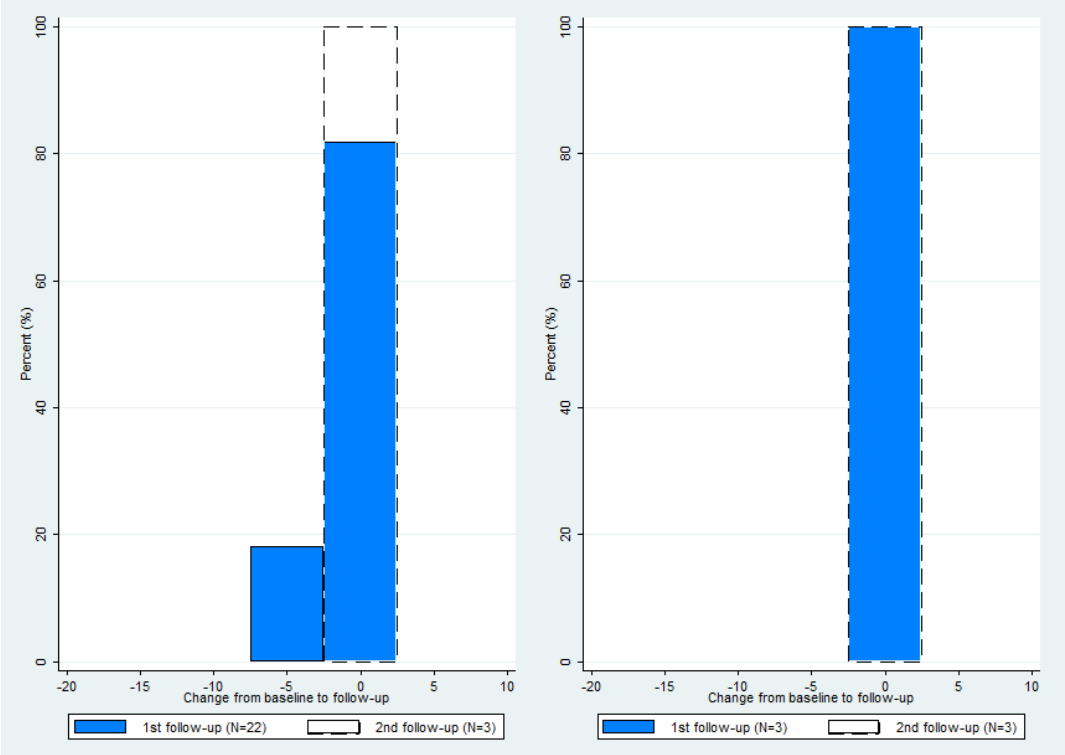
C.XII Density change distribution in controls: 1st follow-up vs. 2nd follow-up (all women and subgroup with available final follow-up)

All women and subgroup with available final follow-up; left and right, respectively:



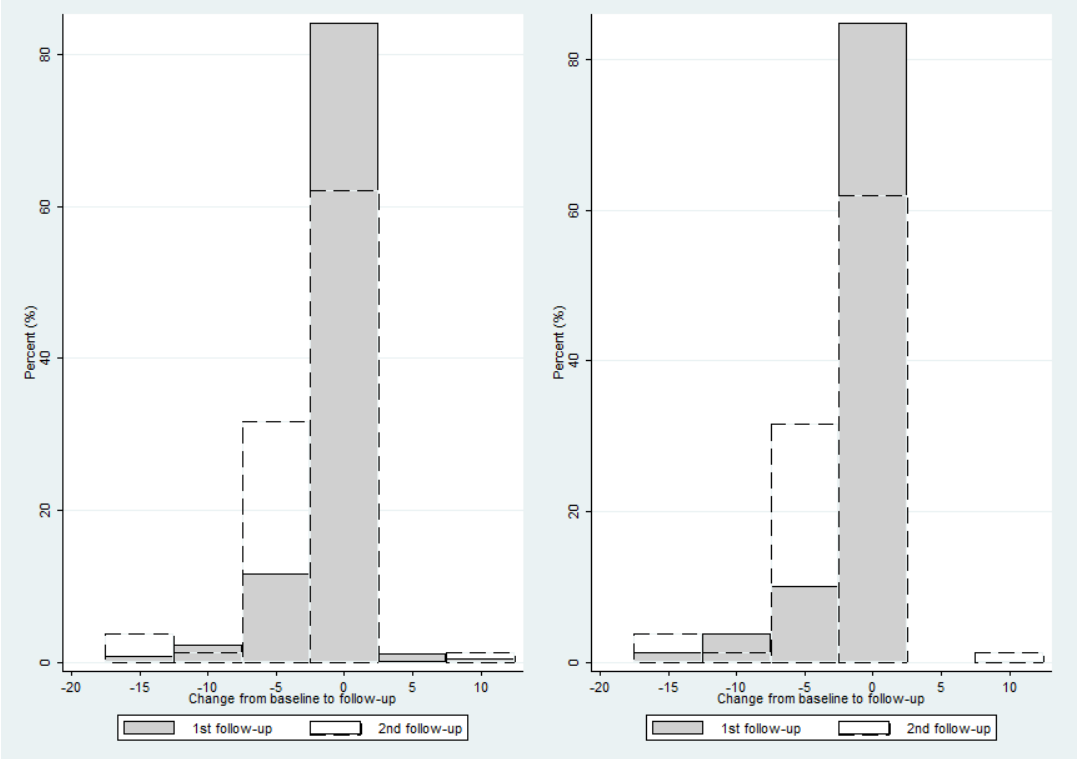
C.XIII Density change distribution in cases: 1st follow-up vs. 2nd follow-up (all women and subgroup with available final follow-up)

All women and subgroup with available final follow-up; left and right, respectively:



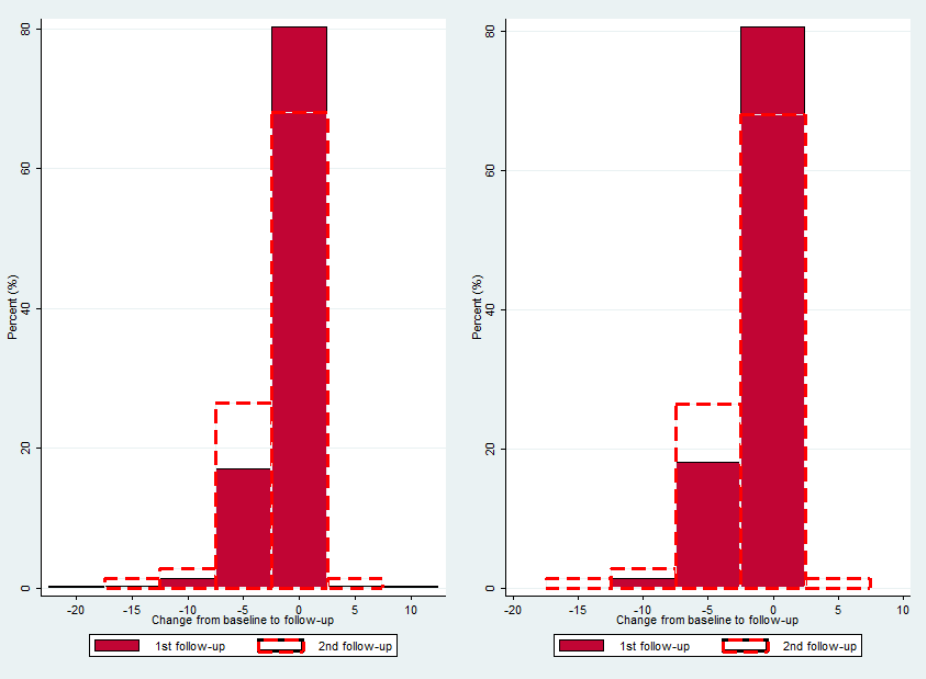
C.XIV Density change distribution in placebo: 1st follow-up vs. 2nd follow-up (all placebo and subgroup with available final follow-up)

All placebo and subgroup with available final follow-up; left and right, respectively:



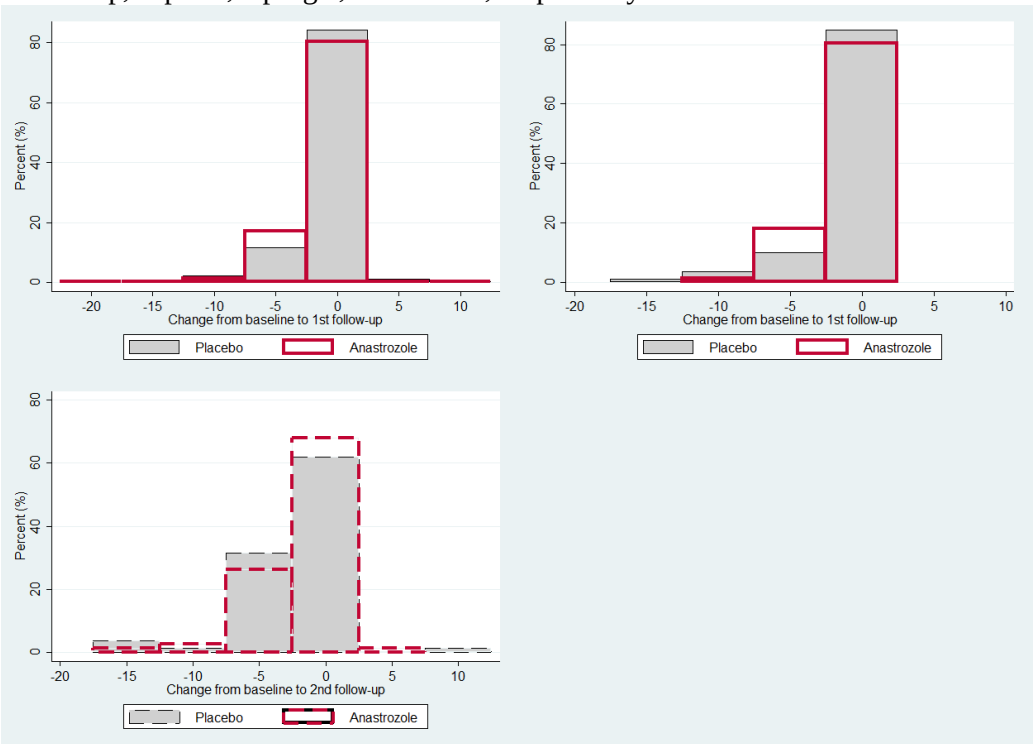
C.XV Density change distribution in anastrozole: 1st follow-up vs. 2nd follow-up (all anastrozole and subgroup with available final follow-up)

All anastrozole and subgroup with available final follow-up; left and right, respectively:



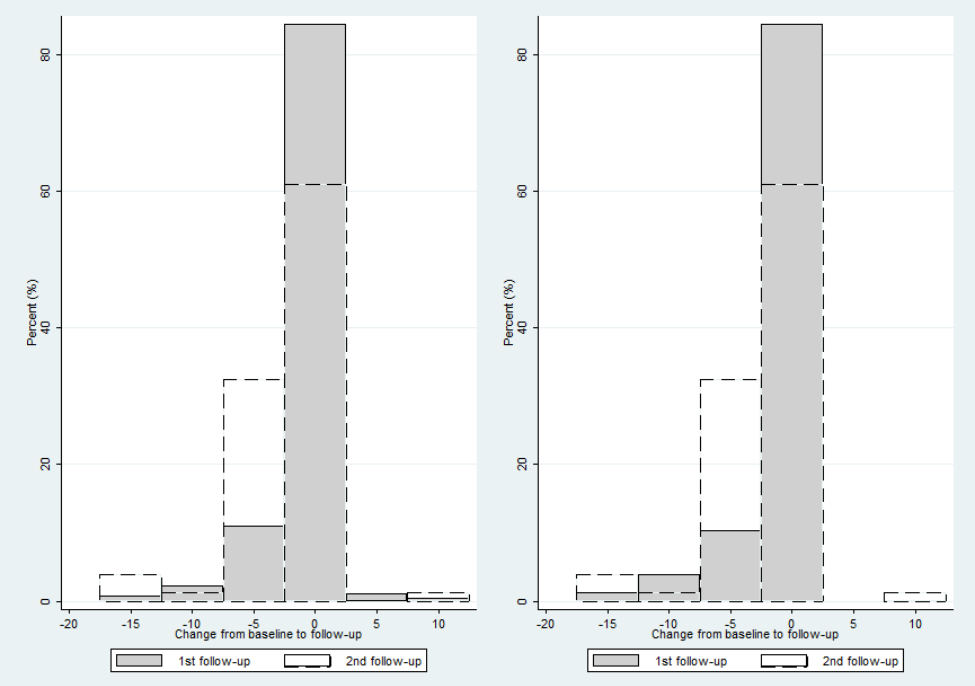
C.XVI Density change distribution in placebo vs. anastrozole (all women at first follow-up, subgroup with available final follow-up at first follow-up and final follow-up)

All women at first follow-up, subgroup with available final follow-up at first follow-up and final follow-up; top left, top right, bottom left, respectively:



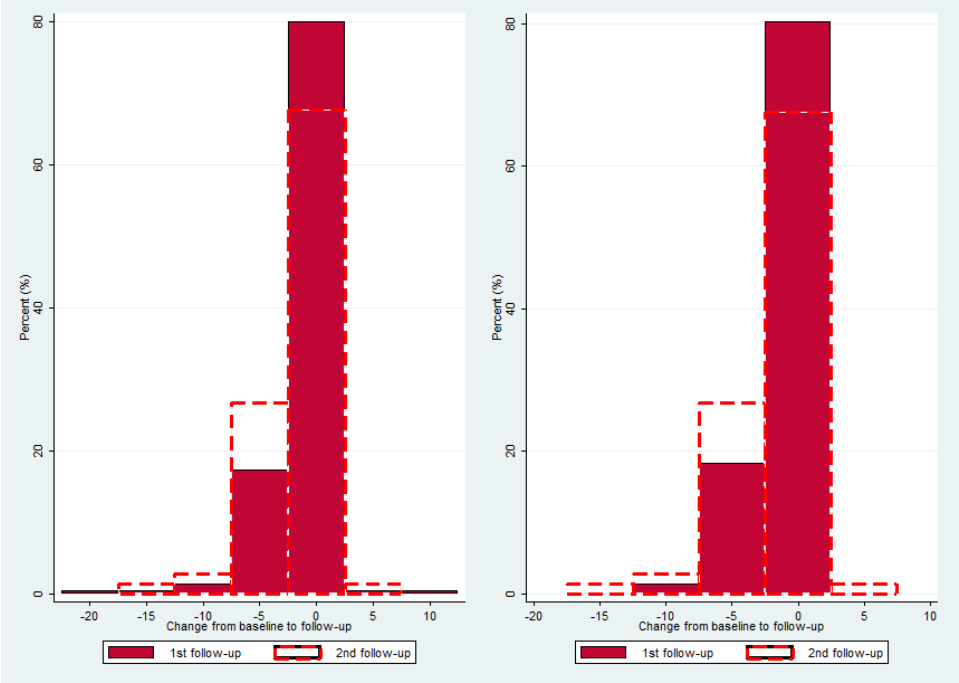
C.XVII Density change distribution in placebo controls: 1st follow-up vs. 2nd follow-up (all placebo controls and subgroup with available final follow-up)

All placebo controls and subgroup with available final follow-up; left and right, respectively:



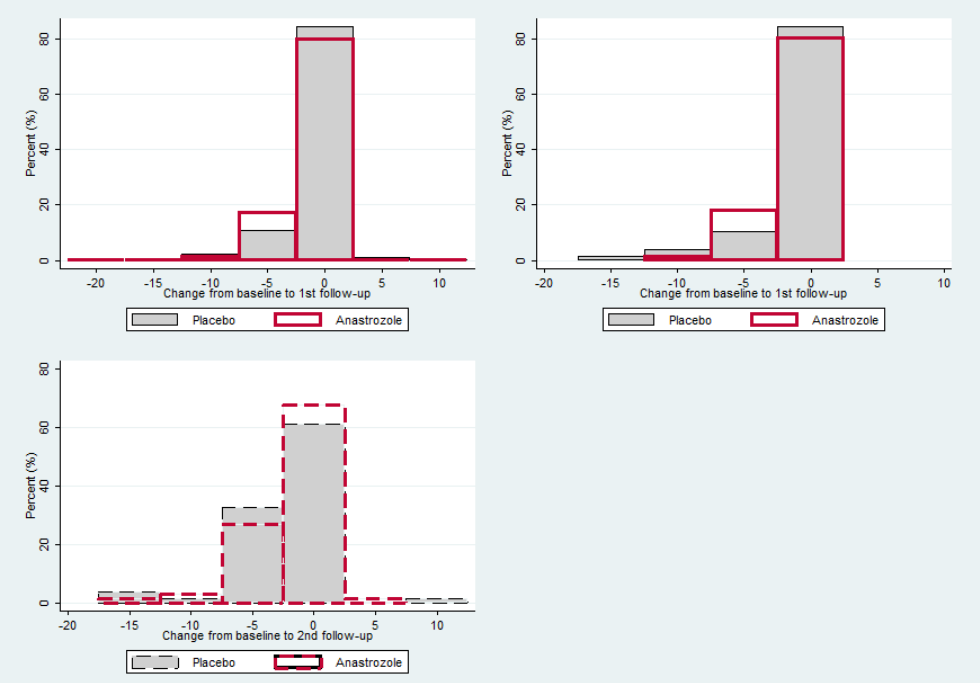
C.XVIII Density change distribution in anastrozole controls: 1st follow-up vs. 2nd follow-up (all anastrozole controls and subgroup with available final follow-up)

All anastrozole controls and subgroup with available final follow-up; left and right, respectively:



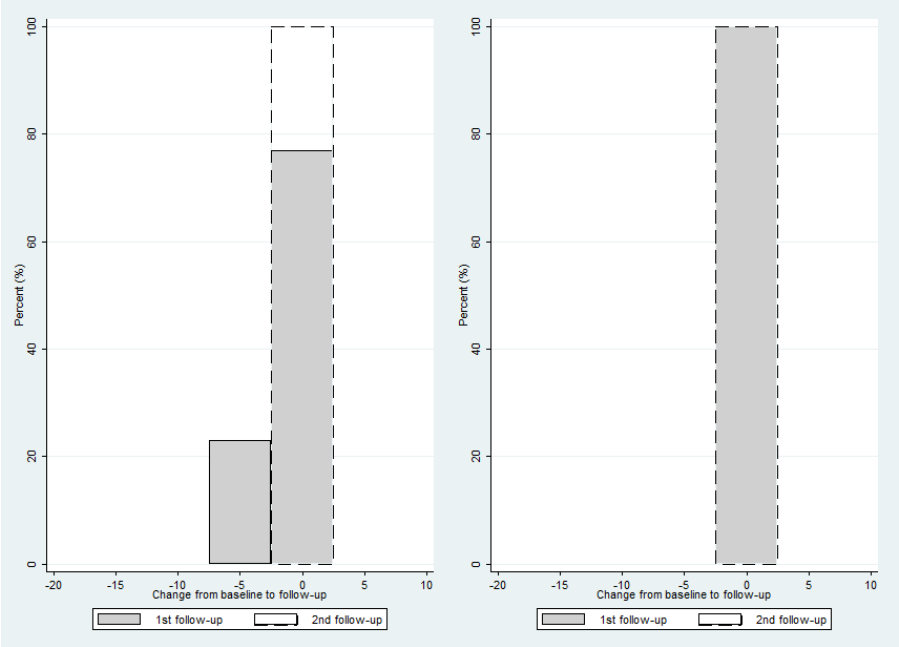
C.XIX Density change distribution in placebo controls vs. anastrozole controls (all women at first follow-up, subgroup with available final follow-up at first follow-up and final follow-up)

All women at first follow-up, subgroup with available final follow-up at first follow-up and final follow-up; top left, top right, bottom left, respectively:



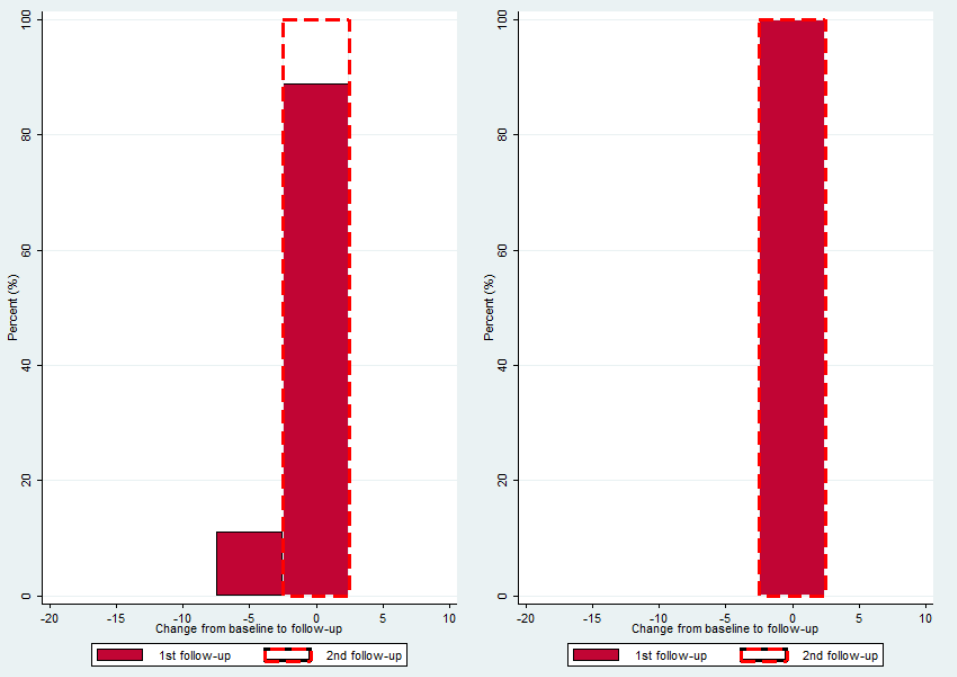
C.XX Density change distribution in placebo cases: 1st follow-up vs. 2nd follow-up (all placebo cases and subgroup with available final follow-up)

All placebo cases and subgroup with available final follow-up; left, right, respectively:



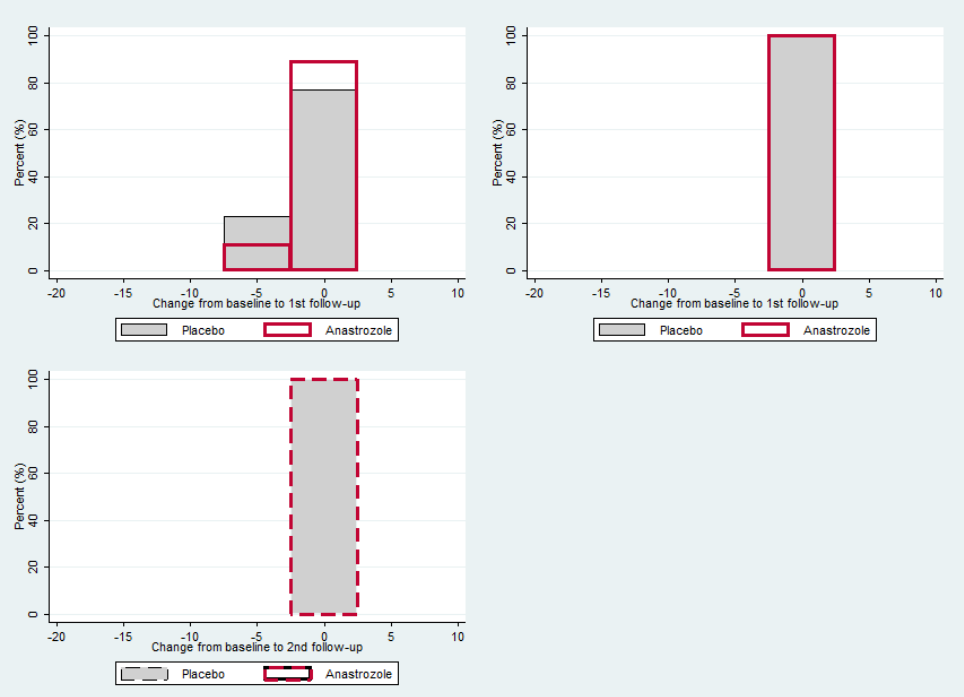
C.XXI Density change distribution in anastrozole cases: 1st follow-up vs. 2nd follow-up (all anastrozole cases and subgroup with available final follow-up)

All anastrozole cases and subgroup with available final follow-up; left, right, respectively:



C.XXII Density change distribution in placebo cases vs. anastrozole cases (all women at first follow-up, subgroup with available final follow-up at first follow-up and final follow-up)

All women at first follow-up, subgroup with available final follow-up at first follow-up and final follow-up; top left, top right, bottom left, respectively:



C.XXIII Statistical Analysis Plan for Chapter 5

1. INTRODUCTION

This document describes the statistical analysis plan for the IBIS-II mammographic density study examining change in density between anastrozole and placebo-treated patients in the IBIS-II Prevention trial.

2. OBJECTIVES OF THE ANALYSIS

2.1 Primary objective

To determine whether women on anastrozole experience different age-adjusted changes in density at first follow-up mammogram than women on placebo in the IBIS-II Prevention trial.

2.2 Secondary objective I

To determine whether women on anastrozole experience different changes in density at first follow-up mammogram than women on placebo in the IBIS-II Prevention trial, after adjustment for age at randomisation, body mass index at randomisation, hormone replacement therapy use up to 12 months before randomisation, age at menopause, image type and time between baseline and first follow-up mammogram.

2.3 Secondary objective II

To determine whether women on anastrozole experience different age-adjusted changes in density at final follow-up mammogram than women on placebo in the IBIS-II Prevention trial.

2.4 Secondary objective III

To determine whether women on anastrozole experience different changes in density at final follow-up mammogram than women on placebo in the IBIS-II Prevention trial, after adjustment for age at randomisation, body mass index at randomisation, hormone replacement therapy use up to 12 months before randomisation, age at menopause, image type and time between baseline and final follow-up mammogram.

2.5 Secondary objective IV

To examine the effect of anastrozole on first density change in subgroups of covariates in the IBIS-II Prevention trial (age at randomisation, body mass index at randomisation, age at menarche, age at menopause, Tyrer-Cuzick 10-year risk, baseline density, age at first birth, oral contraception use, hormone replacement therapy use up to 12 months

before randomisation, smoking status, history of atypical hyperplasia or LCIS, image type, and time between baseline mammogram and follow-up mammogram).

2.6 Secondary objective V

To examine the effect of anastrozole on final density change in subgroups of covariates in the IBIS-II Prevention trial (age at randomisation, body mass index at randomisation, age at menarche, age at menopause, Tyrer-Cuzick 10-year risk, baseline density, age at first birth, oral contraception use, hormone replacement therapy use up to 12 months before randomisation, smoking status, history of atypical hyperplasia or LCIS, image type, and time between baseline mammogram and follow-up mammogram).

3. PERSONNEL

The major statistical analysis will be undertaken by Emma Atakpa at the Centre for Cancer Prevention, Wolfson Institute, Queen Mary University of London, London UK.

4. TIMING OF ANALYSIS

The major statistical analysis will begin in October 2018 (approximate date).

5. STUDY PARTICIPANTS

5.1 Eligible participants

- The primary and secondary (I, IV) statistical analyses will include all randomised women with an appropriate baseline and first follow-up mammogram (within specified timeframes - see below, good quality - as assessed by the experienced radiologist, MLO view only) who are breast cancer-free at the time of their first follow-up mammogram.

- The secondary (II, III, V) statistical analyses will include all randomised women with an appropriate baseline, first follow-up and final follow-up mammogram (within specified timeframes - see below, good quality - as assessed by the experienced radiologist, MLO view only) who are breast cancer-free at the time of their final follow-up mammogram. This will be a subgroup of participants from the primary analysis.

- Baseline mammograms will range from ≥0 months prior to the date of randomisation to <12.5 months prior to the date of randomisation. First follow-up mammograms will range from ≥8.5 months after the date of randomisation to <38.5 months after the date of randomisation. Final follow-up mammograms will range from ≥47.5 months after the date of randomisation to <60.5 months after the date of randomisation. These time

frames are in accordance with analysis from IBIS-I (19) and standard operating procedures for IBIS-II co-ordinating centres.

- Only women with all mammograms of the same image type (i.e. all film or all digital) will be included.

- Only women with ≥10% baseline density will be included.

- Breast cancer-free 'controls' are defined as women who had not been diagnosed with breast cancer at the time of study design. 'Cases' are defined as women who had been clinically diagnosed with breast cancer at the time of study design. Cases will be included if they are diagnosed with breast cancer after their first follow-up mammogram. Final follow-up mammograms for cases will be included if they occur before the breast cancer diagnosis.

- Contralateral mammograms will be used for cases and mammograms from a randomly selected breast side will be used for breast cancer-free controls.

5.2 Sample size calculation

Density change for postmenopausal women in IBIS-I was weighted based on age at randomisation of the IBIS-II cohort, to estimate the expected density change in a cohort with the same age structure as IBIS-II. This was done separately for controls on placebo, controls on tamoxifen, cases on placebo and cases on tamoxifen. By weighting the density change distribution according to the distribution of cases and controls by treatment arm in IBIS-II, an overall density change distribution was formulated. Simulations (10,000 repeats) were conducted to count the number of times there was a significant difference in density change between arms using a Wilcoxon rank sum test and a Pearson chi-squared test (of density change dichotomised into ≥10% reduction and <10% reduction). Simulations were repeated for chosen sample sizes between 400 and 500 women per arm. Different effect sizes (1/2 and 3/4) for tamoxifen were also tested by taking a weighting of the placebo and tamoxifen density change distributions corresponding to the proposed effect size. This was done to allow anastrozole to have a weaker effect on density change than tamoxifen. With 80% power and 3/4 the effect size of tamoxifen, 450 women per arm are required to show a difference in density change from baseline to first follow-up mammogram between the two treatment arms at the 5% type-I error level. In total, 569 anastrozole and 569 placebo-treated women are required, after accounting for exclusions with baseline density <10% based on the number of postmenopausal women with baseline density <10% in IBIS-I. This equates

to approximately 1105 breast cancer-free women and 33 breast-cancer cases. A sample size larger than this is currently impracticable given the resources and number of mammograms received (suitable mammograms, which meet the criteria outlined in section 5.1, have been received for 35 breast-cancer cases and 938 breast cancer-free controls, providing power to detect a difference in density change from baseline to first follow-up mammogram between the two treatment arms at the 5% type-I error level of 85%).

6. OUTCOMES

6.1 Primary outcome

- The primary outcome is defined as the change in density from baseline mammogram to first follow-up mammogram (9-38 months post randomisation). Density will be visually-assessed by an experienced reader (Linda Metaxa) using 5% intervals, following the same method as in IBIS-I. Randomisation of mammograms will be per woman (so that mammograms for each woman will be read in comparison with the other mammograms for that woman), and mammograms will be ordered sequentially. For each woman, density at baseline will be read first, followed by first follow-up mammogram (compared with baseline mammogram) and finally, final follow-up mammogram (compared with both baseline and first follow-up mammogram). First density change will therefore be defined as the difference between baseline density and first follow-up mammogram density; semi-continuously, and dichotomised into <10% or ≥10% absolute reduction. The reader will be blinded to treatment group, case status and risk factors, and images will be appropriately anonymised.

6.2 Secondary outcome

- The secondary outcome is defined as the change in density from baseline mammogram to final follow-up mammogram (48-60 months post randomisation). Density will be visually-assessed by an experienced reader (Linda Metaxa) using 5% intervals, following the same method as in IBIS-I. Randomisation of mammograms will be per woman (so that mammograms for each woman will be read in comparison with the other mammograms for that woman), and mammograms will be ordered sequentially. For each woman, density at baseline will be read first, followed by first follow-up mammogram (compared with baseline mammogram) and finally, final follow-up mammogram (compared with both baseline and first follow-up mammogram). Final density change will therefore be defined as the difference between baseline density and final follow-up mammogram density; semi-continuously, and dichotomised into <10%

or ≥10% absolute reduction. The reader will be blinded to treatment group, case status and risk factors, and images will be appropriately anonymised.

7. STATISTICAL METHODS

7.1 Hypotheses to be tested

7.1.1 Primary hypothesis

- H_0: There is no difference in age-adjusted change in density from baseline to first follow-up mammogram between patients in the anastrozole arm and patients in the placebo arm.
- H_1: Age-adjusted change in density from baseline to first follow-up mammogram is different between patients in the anastrozole arm and patients in the placebo arm.

7.1.2 Secondary hypothesis I

- H_0: There is no difference in change in density from baseline to first follow-up mammogram between patients in the anastrozole arm and patients in the placebo arm, after adjustment for age at randomisation, body mass index at randomisation, hormone replacement therapy use up to 12 months before randomisation, age at menopause, image type, and time between baseline and first follow-up mammogram.
- H_1: Change in density from baseline to first follow-up mammogram is different between patients in the anastrozole arm and patients in the placebo arm, after adjustment for age at randomisation, body mass index at randomisation, hormone replacement therapy use up to 12 months before randomisation, age at menopause, image type, and time between baseline and first follow-up mammogram.

7.1.3 Secondary hypothesis II

- H_0: There is no difference in age-adjusted change in density from baseline to final follow-up mammogram between patients in the anastrozole arm and patients in the placebo arm.
- H_1: Age-adjusted change in density from baseline to final follow-up mammogram is different between patients in the anastrozole arm and patients in the placebo arm.

7.1.4 Secondary hypothesis III

- H_0: There is no difference in change in density from baseline to final follow-up mammogram between patients in the anastrozole arm and patients in the placebo arm, after adjustment for age at randomisation, body mass index at randomisation, hormone

replacement therapy use up to 12 months before randomisation, age at menopause, image type, and time between baseline and first follow-up mammogram.

- H_1: Change in density from baseline to final follow-up mammogram is different between patients in the anastrozole arm and patients in the placebo arm, after adjustment for age at randomisation, body mass index at randomisation, hormone replacement therapy use up to 12 months before randomisation, age at menopause, image type, and time between baseline and first follow-up mammogram.

7.1.5 Secondary hypothesis IV

- H_0: There is no difference in anastrozole-induced change in density from baseline to first follow-up mammogram between subgroups of covariates (age at randomisation, body mass index at randomisation, age at menarche, age at menopause, Tyrer-Cuzick 10-year risk, baseline density, age at first birth, oral contraception use, hormone replacement therapy use up to 12 months before randomisation, smoking status, history of atypical hyperplasia or LCIS, image type, and time between baseline mammogram and follow-up mammogram).
- H_1: Anastrozole-induced change in density from baseline to first follow-up mammogram is different between subgroups of covariates (age at randomisation, body mass index at randomisation, age at menarche, age at menopause, Tyrer-Cuzick 10-year risk, baseline density, age at first birth, oral contraception use, hormone replacement therapy use up to 12 months before randomisation, smoking status, history of atypical hyperplasia or LCIS, image type, and time between baseline mammogram and follow-up mammogram).

7.1.6 Secondary hypothesis V

- H_0: There is no difference in anastrozole-induced change in density from baseline to final follow-up mammogram between subgroups of covariates (age at randomisation, body mass index at randomisation, age at menarche, age at menopause, Tyrer-Cuzick 10-year risk, baseline density, age at first birth, oral contraception use, hormone replacement therapy use up to 12 months before randomisation, smoking status, history of atypical hyperplasia or LCIS, image type, and time between baseline mammogram and follow-up mammogram).
- H_1: Anastrozole-induced change in density from baseline to final follow-up mammogram is different between subgroups of covariates (age at randomisation, body mass index at randomisation, age at menarche, age at menopause, Tyrer-Cuzick 10-year risk, baseline density, age at first birth, oral contraception use, hormone replacement therapy use up to 12 months before randomisation, smoking status, history of atypical

hyperplasia or LCIS, image type, and time between baseline mammogram and follow-up mammogram).

7.2 Analysis methods

All statistical analysis will be conducted in STATA 13. All tests (see below) will be two-sided with a significance level of 5%. Results will be omitted if subgroup numbers are small enough in order to un-blind the statistician.

7.2.1 Baseline characteristics

The following baseline characteristics will be summarised in a frequency table, overall and by treatment. Frequency counts & percentages will be provided for categorical data and means (standard deviation, SD) and medians (interquartile range, IQR) will be provided for continuous data. Two-sample t-tests (*STATA's "ttest" command*) and Wilcoxon rank sum tests (*STATA's "ranksum" command*) will test differences between treatment arms in continuous data and Pearson chi-squared tests (*STATA's "tab, chi2" command*) will test differences between treatment arms in categorical data:

- Age at randomisation (mean (SD), median (IQR))
- Body Mass Index (BMI) at randomisation (mean (SD), median (IQR))
- Age at menarche (mean (SD), median (IQR))
- Age at menopause (mean (SD), median (IQR))
- Tyrer-Cuzick 10-year risk (mean (SD), median (IQR))
- Baseline density (mean (SD), median (IQR))
- Age at first birth (nulliparous/>27/21-27/≤20)
- Oral contraception use (never/previously/currently)
- Hormone Replacement Therapy (HRT) use up to 12 months before randomisation (no/yes)
- Smoking status (never/former/current)
- History of atypical hyperplasia or LCIS (no/yes)
- Image type (film/digital)

7.2.2 Baseline characteristics and baseline density

- Univariate bootstrap linear regression models of baseline density on baseline covariates (excluding baseline density), adjusted for age at randomisation (except age at randomisation): n, β-coefficient, 95% confidence interval, P-value.
  STATA's "regress" command

- Multivariable bootstrap linear regression models of baseline density on baseline covariates (excluding baseline density). All covariates will be included in the multivariable model: n, β-coefficient, 95% confidence interval, P-value.
  STATA's "bootstrap", "estat bootstrap, all" & "regress" commands
- Univariate logistic regression models of baseline density (dichotomised into <50% or ≥50%) on baseline covariates (excluding baseline density), adjusted for age at randomisation (except age at randomisation): n, odds ratio, 95% confidence interval, P-value.
  STATA's "logistic" command
- Multivariable logistic regression models of baseline density (dichotomised into <50% or ≥50%) on baseline covariates (excluding baseline density). All covariates will be included in the multivariable model: n, odds ratio, 95% confidence interval, P-value.
  STATA's "logistic" command

7.2.3 Primary analysis

- Wilcoxon rank sum test comparing density change in the anastrozole and placebo arms: P-value.
  STATA's "summarize, detail" command
  STATA's "tabstat, statistics(iqr)" command
  STATA's "ranksum, by(treatment)" command
- Pearson chi-squared test comparing density change in the anastrozole and placebo arms (density change dichotomised into <10% absolute reduction and ≥10% absolute reduction): P-value.
  STATA's "tabulate, chi2" command
- Bootstrap linear regression model of change in density on treatment arm and age at randomisation: n, β-coefficient, 95% confidence interval, P-value.
  STATA's "bootstrap", "estat bootstrap, all" & "regress" commands
- Logistic regression model of change in density (dichotomised into <10% absolute reduction and ≥10% absolute reduction) on treatment arm and age at randomisation: n, odds ratio, 95% confidence interval, P-value.
  STATA's "logistic" command

7.2.4 Secondary analysis I

- Bootstrap linear regression model of change in density on treatment arm, adjusted for covariates. All covariates will be included in the multivariable model: n, β-coefficient, 95% confidence interval, P-value.
  STATA's "bootstrap", "estat bootstrap, all" & "regress" commands

- Logistic regression model of change in density (dichotomised into <10% absolute reduction and ≥10% absolute reduction) on treatment arm, adjusted for covariates. All covariates will be included in the multivariable model: n, odds ratio, 95% confidence interval, P-value.
  STATA's "logistic" command

7.2.5 Secondary analysis II

- Wilcoxon rank sum test comparing density change in the anastrozole and placebo arms: P-value.
  STATA's "summarize, detail" command
  STATA's "tabstat, statistics(iqr)" command
  STATA's "ranksum, by(treatment)" command
- Pearson chi-squared test comparing density change in the anastrozole and placebo arms (density change dichotomised into <10% absolute reduction and ≥10% absolute reduction): P-value.
  STATA's "tabulate, chi2" command
- Bootstrap linear regression model of change in density on treatment arm and age at randomisation: n, β-coefficient, 95% confidence interval, P-value.
  STATA's "bootstrap", "estat bootstrap, all" & "regress" commands
- Logistic regression model of change in density (dichotomised into <10% absolute reduction and ≥10% absolute reduction) on treatment arm and age at randomisation: n, odds ratio, 95% confidence interval, P-value.
  STATA's "logistic" command

7.2.6 Secondary analysis III

- Bootstrap linear regression model of change in density on treatment arm, adjusted for covariates. All covariates will be included in the multivariable model: n, β-coefficient, 95% confidence interval, P-value.
  STATA's "bootstrap", "estat bootstrap, all" & "regress" commands
- Logistic regression model of change in density (dichotomised into <10% absolute reduction and ≥10% absolute reduction) on treatment arm, adjusted for covariates. All covariates will be included in the multivariable model: n, odds ratio, 95% confidence interval, P-value.
  STATA's "logistic" command

7.2.7 Secondary analysis IV

- Wilcoxon rank-sum test comparing first density change between subgroups of covariates in the anastrozole arm only (for covariates with 2 subgroups): P-value.

STATA's "ranksum, by(subgroup)" command

- Cuzick's trend test comparing first density change between subgroups of covariates in the anastrozole arm only (for covariates with >2 ordered subgroups): P-value.
  STATA's "nptrend, by(subgroup)" command
- Logistic regression to assess the odds of a higher first density reduction (≥10% absolute reduction) in one subgroup compared to the reference subgroup in the anastrozole arm only: Odds ratio, 95% confidence interval, P-value.
  STATA's "logistic" command

7.2.8 Secondary analysis V

- Wilcoxon rank-sum test comparing final density change between subgroups of covariates in the anastrozole arm only (for covariates with 2 subgroups): P-value.
  STATA's "ranksum, by(subgroup)" command
- Cuzick's trend test comparing final density change between subgroups of covariates in the anastrozole arm only (for covariates with >2 ordered subgroups): P-value.
  STATA's "nptrend, by(subgroup)" command
- Logistic regression to assess the odds of a higher final density reduction (≥10% absolute reduction) in one subgroup compared to the reference subgroup in the anastrozole arm only: Odds ratio, 95% confidence interval, P-value.
  STATA's "logistic" command

7.2.9 Adjustment covariates

The following covariates are chosen as potential confounders for density change based on previous literature and significant covariates in the analysis from IBIS-I (203) (Cuzick 2004) and hence will be included in adjusted regression models (7.2.4 & 7.2.6):

- Age at randomisation (continuous)
- BMI at randomisation (continuous)
- Age at menopause (continuous)
- HRT use up to 12 months before randomisation (no/yes)
- Image type (film/digital)
- Time between baseline mammogram and follow-up mammogram (continuous)

A separate model will be conducted, including an adjustment for baseline density (continuous) and age at randomisation only (7.2.4 & 7.2.6).

A separate model will be conducted, including an adjustment for baseline density (continuous) and the covariates above (7.2.4 & 7.2.6).

Time on treatment will not be included in adjustments because an intention-to-treat analysis will be conducted.

7.2.10 Subgroup covariates

The following covariates will be considered in subgroup analyses (7.2.7 & 7.2.8):

- Age at randomisation (<median age, ≥median age)
- BMI at randomisation (<median BMI, ≥median BMI)
- Age at menarche (<median age at menarche, ≥median age at menarche)
- Age at menopause (<median age at menopause, ≥median age at menopause)
- Tyrer-Cuzick 10-year risk (<median risk, ≥median risk)
- Baseline density (<median baseline density, ≥median baseline density)
- Age at first birth (nulliparous/>27/21-27/≤20)
- Oral contraception use (never/previously/currently)
- HRT use up to 12 months before randomisation (no/yes)
- Smoking status (never/former/current)
- History of atypical hyperplasia or LCIS (no/yes)
- Image type (film/digital)
- Time between baseline mammogram and follow-up mammogram (<median time between baseline mammogram and follow-up mammogram, ≥median time between baseline mammogram and follow-up mammogram)

7.3 Tables

Variable	Overall		Placebo		Anastrozole	
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)
Age at randomisation (yr)						
P*						
Body Mass Index (kg/m^2)						
P*						
Age at menarche (yr)						
P*						
Age at menopause (yr)						
P*						
Tyrer-Cuzick 10-year risk (%)						
P*						
Baseline density (%)						
P*						
	N	%	N	%	N	%
Age at first birth (yr)						
Nulliparous						
>27						
21-27						
≤20						
P**						
Oral contraception use						
Never						
Previously						
Currently						
P**						
HRT use up to 12 months before randomisation						
No						
Yes						
P**						
Smoking status						
Never						
Former						
Current						
P**						
History of Atypical Hyperplasia or LCIS						
No						
Yes						
P**						
Image type						
Film						
Digital						
P**						

Table 1: Baseline characteristics overall and by treatment. *P-value from two-sample t-test (corresponding to mean column) and Wilcoxon rank sum test (corresponding to median column), **p-value from Pearson chi-squared test of association.

	Univariate Linear (bootstrap) regression		Multivariable Linear (bootstrap) regression[1]		Univariate Logistic regression		Multivariable Logistic regression[1]	
Variable	β-coefficient (95% CI)***	Bootstrap SE	β-coefficient (95% CI)***	Bootstrap SE	OR (95 % CI)#	P-value #	OR (95 % CI)#	P-value #
Age at randomisation (yr)*								
Body Mass Index (kg/m^2)*								
Age at menarche (yr)*								
Age at menopause (yr)*								
Tyrer-Cuzick 10-year risk (%)*								
Age at first birth (yr)**								
Nulliparous	*Ref*		*Ref*		*Ref*		*Ref*	
>27								
21-27								
≤20								
Oral contraception use**								
Never	*Ref*		*Ref*		*Ref*		*Ref*	
Previously								
Currently								
HRT use up to 12 months before randomisation**								
No	*Ref*		*Ref*		*Ref*		*Ref*	
Yes								
Smoking status**								
Never	*Ref*		*Ref*		*Ref*		*Ref*	
Former								
Current								
History of Atypical Hyperplasia or LCIS**								
No	*Ref*		*Ref*		*Ref*		*Ref*	
Yes								
Image type**								
Film	*Ref*		*Ref*		*Ref*		*Ref*	
Digital								

Table 2: Association between baseline covariates and baseline breast density (semi-continuous and dichotomised ≥50% and <50%) in univariate and multivariable models. All covariates adjusted for age at randomisation in univariate models (except for age at randomisation). * β-coefficient represents effect on baseline density per unit increase in covariate, OR represents odds of having ≥50% baseline density per unit increase in covariate. ** β-coefficient represents difference in baseline density from reference category, OR represents odds of having ≥50% baseline density relative to the reference category. *** Percentile 95% CI. # 95% CI and p-value from a Wald test. [1]Model includes all covariates. N=.

Boyd category at entry	Number of women													
	Boyd category at first follow-up						Total	Boyd category at final follow-up						Total
	0 %	1-10 %	11-25 %	26-50 %	51-75 %	76-100 %		0 %	1-10 %	11-25 %	26-50 %	51-75 %	76-100 %	
0%														
1-10%														
11-25%														
26-50%														
51-75%														
76-100%														
Total														

Table 3: Cross tabulation of number of women in each Boyd category at entry to the study with category at first and final follow-up. The first number in each cell is the total number of subjects. Numbers in parentheses are the placebo and anastrozole groups, respectively.

	N	Median	IQR	P-value
First Follow-up				
Placebo				
Anastrozole				
Final Follow-up				
Placebo				
Anastrozole				

Table 4: Semi-continuous density change by treatment arm. P-value from Wilcoxon rank sum test.

	N (% of follow-up)		χ^2	P-value
	<10% reduction	≥10% reduction		
First Follow-up				
Placebo				
Anastrozole				
Final Follow-up				
Placebo				
Anastrozole				

Table 5: Dichotomised density change by treatment arm. P-value from Pearson chi-squared test.

	Univariate		Adjusted[1]		Adjusted[2]		Adjusted[3]	
	β-coefficient (95% CI)***	Bootstrap SE	β-coefficient (95% CI)***	Bootstrap SE	β-coefficient (95% CI)***	Bootstrap SE	β-coefficient (95% CI)***	Bootstrap SE
	First follow-up							
Treatment**#								
Placebo	*Ref*		*Ref*		*Ref*		*Ref*	
Anastrozole								
Age at randomisation (yr)*								
Body Mass Index (kg/m^2)*					-	-		
Age at menopause (yr)*					-	-		
Baseline density (%)*			-	-				
HRT use up to 12 months before randomisation**								
No	*Ref*		*Ref*		-	-	*Ref*	
Yes					-	-		
Image type**								
Film	*Ref*		*Ref*		-	-	*Ref*	
Digital					-	-		
Time between baseline mammogram and first follow-up mammogram (yr)*					-	-		
	Final follow-up							
Treatment**#								
Placebo	*Ref*		*Ref*		*Ref*		*Ref*	
Anastrozole								
Age at randomisation (yr)*								
Body Mass Index (kg/m^2)*					-	-		
Age at menopause					-	-		

(yr)*								
Baseline density (%)*			-	-				
HRT use up to 12 months before randomisation**								
No	*Ref*		*Ref*		-	-	*Ref*	
Yes					-	-		
Image type**								
Film	*Ref*		*Ref*		-	-	*Ref*	
Digital					-	-		
Time between baseline mammogram and final follow-up mammogram (yr)*					-	-		

Table 6: Bootstrap linear regression results for change in density on treatment arm in univariate and adjusted models. * β-coefficient represents effect on density change per unit increase in covariate, ** β-coefficient represents difference in density change from reference category. *** Percentile 95% CI. #Additionally adjusted for age at randomisation in univariate model. [1]Model includes all variables except for baseline density, [2]Model includes treatment, age at randomisation and baseline density, [3]Model includes all covariates. N= for first follow-up, N= for final follow-up.

	Univariate		Adjusted[1]		Adjusted[2]		Adjusted[3]	
	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value
	First follow-up							
Treatment**#								
Placebo	*Ref*		*Ref*		*Ref*		*Ref*	
Anastrozole								
Age at randomisation (yr)*								
Body Mass Index (kg/m^2)*					-	-		
Age at menopause (yr)*					-	-		
Baseline density (%)*			-	-				
HRT use up to 12 months before randomisation**								
No	*Ref*		*Ref*		-	-	*Ref*	
Yes					-	-		
Image type**								
Film	*Ref*		*Ref*		-	-	*Ref*	
Digital					-	-		
Time between baseline mammogram and first follow-up mammogram (yr)*					-	-		

	Final follow-up							
Treatment**#								
Placebo	*Ref*		*Ref*		*Ref*		*Ref*	
Anastrozole								
Age at randomisation (yr)*								
Body Mass Index (kg/m^2)*					-	-		
Age at menopause (yr)*					-	-		
Baseline density (%)*			-	-				
HRT use up to 12 months before randomisation**								
No	*Ref*		*Ref*		-	-	*Ref*	
Yes					-	-		
Image type**								
Film	*Ref*		*Ref*		-	-	*Ref*	
Digital					-	-		
Time between baseline mammogram and final follow-up mammogram (yr)*					-	-		

Table 7: Logistic regression results for change in density (dichotomised into <10% absolute reduction and ≥10% absolute reduction) on treatment arm in univariate and adjusted models. * OR represents odds of ≥10% density reduction per unit increase in covariate, ** OR represents odds of ≥10% density reduction relative to the reference category. #Additionally adjusted for age at randomisation in univariate model. P-values from logistic regression model Wald test of covariate. [1]Model includes all variables except for baseline density, [2]Model includes treatment, age at randomisation and baseline density, [3]Model includes all covariates. N= for first follow-up, N= for final follow-up.

	First follow-up					Final follow-up				
	N	Median (IQR)	P-value#	OR (95% CI)##	P-value###	N	Median (IQR)	P-value#	OR (95% CI)##	P-value###
Age at randomisation (yr)*										
Younger age at randomisation				*Ref*					*Ref*	
Older age at randomisation										
Body Mass Index (kg/m^2)*										
Lower Body Mass Index				*Ref*					*Ref*	
Higher Body Mass Index										

Age at menarche (yr)*										
Younger age at menarche				*Ref*					*Ref*	
Older age at menarche										
Age at menopause (yr)*										
Younger age at menopause				*Ref*					*Ref*	
Older age at menopause										
Tyrer-Cuzick 10-year risk (%)*										
Lower Tyrer-Cuzick 10-year risk				*Ref*					*Ref*	
Higher Tyrer-Cuzick 10-year risk										
Baseline density (%)*										
Lower baseline density				*Ref*					*Ref*	
Higher baseline density										
Age at first birth (yr)										
Nulliparous				*Ref*					*Ref*	
>27										
21-27										
≤20										
Oral contraception use										
Never				*Ref*					*Ref*	
Previously										
Currently										
HRT use up to 12 months before randomisation										
No				*Ref*					*Ref*	
Yes										
Smoking status										
Never				*Ref*					*Ref*	
Former										
Current										

History of Atypical Hyperplasia or LCIS										
No				*Ref*					*Ref*	
Yes										
Image type										
Film				*Ref*					*Ref*	
Digital										
Time between baseline mammogram and follow-up mammogram (yr)*										
Shorter time between mammograms				*Ref*					*Ref*	
Longer time between mammograms										

Table 8: First and final change in density by subgroups of covariates in the anastrozole arm only. *Continuous variables dichotomised by median of variable in all women. #P-value from Wilcoxon rank-sum test for covariates with 2 subgroups, P-value from Cuzick's trend test for covariates with >2 subgroups. ## OR represents odds of ≥10% density reduction relative to the reference category. ###P-value of odds ratio from a univariate logistic regression model Wald test of covariate.

C.XXIV **Proforma for Chapter 5**

Request Originator:	Ms Emma Atakpa PhD student, Centre for Cancer Prevention, Wolfson Institute Supervisors: Professor Jack Cuzick Dr Adam Brentnall
Reason for request:	To complete an analysis comparing mammographic density change between the placebo and anastrozole arms of the IBIS-II Prevention trial.
Background:	Mammographic density (herein referred to as 'density') is one of the strongest known risk factors for breast cancer. Women in the highest density category (≥75%) are at a 4 to 6-fold increased risk of developing breast cancer relative to those with little or no density (79). Whilst postmenopausal hormone replacement therapy (HRT) is associated with an increase in risk of breast cancer (424) and density (188, 193); selective oestrogen receptor modulators (SERMs), such as tamoxifen, decrease risk (198, 199, 425) and density (203-205). Most importantly, high-risk women who experienced ≥10% density reduction after approximately 18 months of prophylactic tamoxifen were shown to be at approximately 68% lower risk of developing breast cancer compared with women who experienced <10% density reduction after the same treatment in the IBIS-I trial (19). Similar to SERMs, aromatase inhibitors (AIs) are an anti-oestrogenic drug given to women in the treatment of breast cancer. In 2014, analysis from IBIS-II showed that anastrozole (an AI) reduced the risk of breast cancer in high-risk postmenopausal women (208). However, studies looking into the relationship between AIs and density have so far shown modest or insignificant results (267-269, 426). Many of these studies lack statistical power due to their small sample size (12 month density change was assessed in only 43 women in Prowell et al. (426), 49 women in Ciglar et al. (268), and 65 women in Ciglar et al. (269)), whilst larger studies such as Vachon et al. (267) are based on adjuvant AI treatment only. It is still unknown whether preventive anastrozole treatment reduces density more than the natural decline which tends to occur with age.
Aims:	*Primary objective*: To determine whether women on anastrozole experience different age-adjusted changes in density at first follow-up mammogram than women on placebo in the IBIS-II Prevention trial. *Primary hypothesis*: Age-adjusted change in density from baseline to first follow-up mammogram is different between patients in the anastrozole arm and patients in the placebo arm. *Secondary objective I*: To determine whether women on anastrozole experience different changes in density at first follow-up mammogram than women on placebo in the IBIS-II Prevention trial, after adjustment for age at randomisation, body mass index at randomisation, hormone replacement therapy use up to 12 months before randomisation, age at menopause, image type and time between baseline and first follow-up mammogram. *Secondary hypothesis I*: Change in density from baseline to first follow-up mammogram is different between patients in the anastrozole arm and patients in

	the placebo arm, after adjustment for age at randomisation, body mass index at randomisation, hormone replacement therapy use up to 12 months before randomisation, age at menopause, image type, and time between baseline and first follow-up mammogram. *Secondary objective II*: To determine whether women on anastrozole experience different age-adjusted changes in density at final follow-up mammogram than women on placebo in the IBIS-II Prevention trial. *Secondary hypothesis II*: Age-adjusted change in density from baseline to final follow-up mammogram is different between patients in the anastrozole arm and patients in the placebo arm. *Secondary objective III*: To determine whether women on anastrozole experience different changes in density at final follow-up mammogram than women on placebo in the IBIS-II Prevention trial, after adjustment for age at randomisation, body mass index at randomisation, hormone replacement therapy use up to 12 months before randomisation, age at menopause, image type and time between baseline and final follow-up mammogram. *Secondary hypothesis III*: Change in density from baseline to final follow-up mammogram is different between patients in the anastrozole arm and patients in the placebo arm, after adjustment for age at randomisation, body mass index at randomisation, hormone replacement therapy use up to 12 months before randomisation, age at menopause, image type, and time between baseline and first follow-up mammogram. *Secondary objective IV*: To examine the effect of anastrozole on first density change in subgroups of covariates in the IBIS-II Prevention trial (age at randomisation, body mass index at randomisation, age at menarche, age at menopause, Tyrer-Cuzick 10-year risk, baseline density, age at first birth, oral contraception use, hormone replacement therapy use up to 12 months before randomisation, smoking status, history of atypical hyperplasia or LCIS, image type, and time between baseline mammogram and follow-up mammogram). *Secondary hypothesis IV*: Anastrozole-induced change in density from baseline to first follow-up mammogram is different between subgroups of covariates (age at randomisation, body mass index at randomisation, age at menarche, age at menopause, Tyrer-Cuzick 10-year risk, baseline density, age at first birth, oral contraception use, hormone replacement therapy use up to 12 months before randomisation, smoking status, history of atypical hyperplasia or LCIS, image type, and time between baseline mammogram and follow-up mammogram). *Secondary objective V*: To examine the effect of anastrozole on final density change in subgroups of covariates in the IBIS-II Prevention trial (age at randomisation, body mass index at randomisation, age at menarche, age at menopause, Tyrer-Cuzick 10-year risk, baseline density, age at first birth, oral contraception use, hormone replacement therapy use up to 12 months before randomisation, smoking status, history of atypical hyperplasia or LCIS, image type, and time between baseline mammogram and follow-up mammogram). *Secondary hypothesis V*: Anastrozole-induced change in density from baseline to final follow-up mammogram is different between subgroups of covariates (age at randomisation, body mass index at randomisation, age at menarche, age at

	menopause, Tyrer-Cuzick 10-year risk, baseline density, age at first birth, oral contraception use, hormone replacement therapy use up to 12 months before randomisation, smoking status, history of atypical hyperplasia or LCIS, image type, and time between baseline mammogram and follow-up mammogram).
Data required:	Suitable mammograms (within the following specified timeframes, before breast cancer diagnosis, MLO view, deemed to be good quality) have been received for 35 breast-cancer cases and 938 breast cancer-free controls. Baseline (up to 12 months before randomisation) and first follow-up (9-38 months post randomisation) mammograms for these 973 women were batched and sent to a radiologist at St Bartholomew's hospital for density scoring. A final (48-60 months post randomisation) follow-up mammogram was also sent for scoring if it was available. With 973 women, the power to detect a difference in density change from baseline to first follow-up mammogram between the two treatment arms at the 5% type-I error level is 85%. This calculation allows for a weaker effect of anastrozole on density change than tamoxifen (3/4 the effect size observed in IBIS-I with tamoxifen). Since the requestor is blinded to treatment allocation, a set of STATA code will be sent to Dr Sestak to run on the un-blinded data for these 973 women.
Describe the specific analyses or tables requested:	Methods: Mammograms were visually assessed by a radiologist (Dr Metaxa) at St Bartholomew's hospital. Each mammogram was scored in 5% increments, following the same method as in IBIS-I. Contralateral mammograms were used for cases and mammograms from a randomly selected breast side were used for breast cancer-free controls. Density change will be defined as the difference between baseline and first follow-up mammogram as well as baseline and final follow-up mammogram. Only women with ≥10% baseline density will be included. Only women with all mammograms of the same image type (i.e. all film or all digital) will be included. Statistical analysis: Baseline characteristics will be summarised by treatment arm using frequency tables with age at randomisation, body mass index at randomisation, age at menarche, age at menopause, Tyrer-Cuzick 10-year risk, baseline density, age at first birth, oral contraception use, hormone replacement therapy use up to 12 months before randomisation, smoking status, history of atypical hyperplasia or LCIS, and image type. An exploratory analysis will also assess the association between baseline covariates and baseline density using univariate (adjusted for age at randomisation) and multivariable linear regression and univariate (adjusted for age at randomisation) and multivariable logistic regression. The primary analysis will compare the density change at first follow-up in anastrozole-treated patients with placebo-treated patients in the IBIS-II

	Prevention trial. A linear regression model will examine the association between treatment arm and change in density, adjusted for age at randomisation. A logistic regression model will also be used to examine the association between treatment arm and change in density (dichotomised into <10% absolute reduction and ≥10% absolute reduction), adjusted for age at randomisation. The secondary analysis (I) will repeat the primary analysis, adjusted for age at randomisation, body mass index at randomisation, hormone replacement therapy use up to 12 months before randomisation, age at menopause, image type and time between baseline and first follow-up mammogram. The secondary analysis (II) will repeat the primary analysis but for final follow-up density change, in a subgroup of women who have an available final mammogram density score. The secondary analysis (III) will repeat the primary analysis but for final follow-up density change, in a subgroup of women who have an available final mammogram density score, adjusted for age at randomisation, body mass index at randomisation, hormone replacement therapy use up to 12 months before randomisation, age at menopause, image type and time between baseline and final follow-up mammogram. The secondary analysis (IV) will use Wilcoxon rank-sum or Cuzick trend tests to assess whether the effect of anastrozole on first density change varies between different covariate subgroups, and logistic regression to assess the odds of a high density reduction (≥10% absolute reduction) in one subgroup relative to another subgroup, in anastrozole treated patients. Covariates to be split into subgroups are: age at randomisation, body mass index at randomisation, age at menarche, age at menopause, Tyrer-Cuzick 10-year risk, baseline density, age at first birth, oral contraception use, hormone replacement therapy use up to 12 months before randomisation, smoking status, history of atypical hyperplasia or LCIS, image type, and time between baseline mammogram and follow-up mammogram. The secondary analysis (V) will use Wilcoxon rank-sum or Cuzick trend tests to assess whether the effect of anastrozole on final density change varies between different covariate subgroups, and logistic regression to assess the odds of a high density reduction (≥10% absolute reduction) in one subgroup relative to another subgroup, in anastrozole treated patients. Covariates to be split into subgroups are: age at randomisation, body mass index at randomisation, age at menarche, age at menopause, Tyrer-Cuzick 10-year risk, baseline density, age at first birth, oral contraception use, hormone replacement therapy use up to 12 months before randomisation, smoking status, history of atypical hyperplasia or LCIS, image type, and time between baseline mammogram and follow-up mammogram. All statistical analysis will be conducted in STATA 13. All tests will be two-sided with a significance level of 5%. Results will be omitted if subgroup numbers are small enough to un-blind the requestor.

What will the data be used for?	The results of the analysis will form a chapter in Ms Emma Atakpa's PhD thesis. The results will also be prepared for publication in a peer-reviewed journal (dependent on results and TSC permitting).
Date required by:	Analysis will begin as soon as possible after responses from the TSC.
Other comments:	
Proposed authorship:	E. Atakpa, A. Brentnall, L. Metaxa, I. Sestak, J.F. Forbes, A. Howell, J. Cuzick
This section to be completed by IBIS-II Trial Steering Committee	
Does the Request have Executive Committee Approval? Please place a X in appropriate box	Yes ☐ No ☐
If no, reason for rejection:	
Date received by IBIS-II CCO:	
Date decision sent to applicant:	

C.XXV Statistical Analysis Plan for Chapter 6

1. INTRODUCTION

This document describes the statistical analysis plan for the IBIS-II mammographic density study examining anastrozole-induced change in density and risk of breast cancer in patients from the IBIS-II Prevention trial.

2. OBJECTIVES OF THE ANALYSIS

2.1 Primary objective

To determine whether women on anastrozole who experience a ≥10% reduction in density at first follow-up mammogram have a different level of age-adjusted risk of breast cancer than women on anastrozole who experience a <10% reduction in density at first follow-up mammogram in the IBIS-II Prevention trial.

2.2 Secondary objective I

To determine whether women on anastrozole who experience a ≥5% reduction in density at first follow-up mammogram have a different level of age-adjusted risk of breast cancer than women on anastrozole who experience a <5% reduction in density at first follow-up mammogram in the IBIS-II Prevention trial.

2.3 Secondary objective II

To determine whether women on anastrozole who experience a ≥10% reduction in density at first follow-up mammogram have a different level of risk of breast cancer than women on anastrozole who experience a <10% reduction in density at first follow-up mammogram in the IBIS-II Prevention trial, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.

2.4 Secondary objective III

To determine whether women on anastrozole who experience a ≥5% reduction in density at first follow-up mammogram have a different level of risk of breast cancer than women on anastrozole who experience a <5% reduction in density at first follow-up mammogram in the IBIS-II Prevention trial, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.

2.5 Secondary objective IV

To determine whether the age-adjusted effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) is different between anastrozole-treated and placebo-treated patients (interaction test for anastrozole-induced density change as a predictive biomarker for breast cancer risk reduction).

2.6 Secondary objective V

To determine whether the age-adjusted effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) is different between anastrozole-treated and placebo-treated patients (interaction test for anastrozole-induced density change as a predictive biomarker for breast cancer risk reduction).

2.5 Secondary objective VI

To determine whether the effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) is different between anastrozole-treated and placebo-treated patients (interaction test for anastrozole-induced density change as a predictive biomarker for breast cancer risk reduction), after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.

2.6 Secondary objective VII

To determine whether the effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) is different between anastrozole-treated and placebo-treated patients (interaction test for anastrozole-induced density change as a predictive biomarker for breast cancer risk reduction), after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.

2.7 Secondary objective VIII

To determine whether women on anastrozole who experience a ≥10% reduction in density at first follow-up mammogram have a different level of age-adjusted risk of breast cancer than women on placebo, and whether women on anastrozole who experience a <10% reduction in density at first follow-up mammogram have a different level of age-adjusted risk of breast cancer than women on placebo in the IBIS-II Prevention trial (test for anastrozole-induced density change as a predictive biomarker for breast cancer risk reduction).

2.8 Secondary objective IX

To determine whether women on anastrozole who experience a ≥5% reduction in density at first follow-up mammogram have a different level of age-adjusted risk of breast cancer than women on placebo, and whether women on anastrozole who experience a <5% reduction in density at first follow-up mammogram have a different level of age-adjusted risk of breast cancer than women on placebo in the IBIS-II Prevention trial (test for anastrozole-induced density change as a predictive biomarker for breast cancer risk reduction).

2.9 Secondary objective X

To determine whether women on anastrozole who experience a ≥10% reduction in density at first follow-up mammogram have a different level of risk of breast cancer than women on placebo, and whether women on anastrozole who experience a <10% reduction in density at first follow-up mammogram have a different level of risk of breast cancer than women on placebo in the IBIS-II Prevention trial (test for anastrozole-induced density change as a predictive biomarker for breast cancer risk reduction), after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.

2.10 Secondary objective XI

To determine whether women on anastrozole who experience a ≥5% reduction in density at first follow-up mammogram have a different level of risk of breast cancer than women on placebo, and whether women on anastrozole who experience a <5% reduction in density at first follow-up mammogram have a different level of risk of breast cancer than women on placebo in the IBIS-II Prevention trial (test for anastrozole-induced density change as a predictive biomarker for breast cancer risk reduction), after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.

2.11 Secondary objective XII

To examine the effect of anastrozole-induced first density reduction of ≥10% on breast cancer risk in different subgroups of covariates in the IBIS-II Prevention trial (subgroups: tumour ER status, age at randomisation, body mass index at randomisation, baseline density, history of atypical hyperplasia or LCIS, hormone replacement use up to 12 months before randomisation, Tyrer-Cuzick 10-year risk, image type, and time between baseline mammogram and first follow-up mammogram).

2.12 Secondary objective XIII

To examine the effect of anastrozole-induced first density reduction of ≥5% on breast cancer risk in different subgroups of covariates in the IBIS-II Prevention trial (subgroups: tumour ER status, age at randomisation, body mass index at randomisation, baseline density, history of atypical hyperplasia or LCIS, hormone replacement use up to 12 months before randomisation, Tyrer-Cuzick 10-year risk, image type, and time between baseline mammogram and first follow-up mammogram).

3. PERSONNEL

The major statistical analysis will be undertaken by Emma Atakpa at the Centre for Cancer Prevention, Wolfson Institute, Queen Mary University of London, London UK.

4. TIMING OF ANALYSIS

The major statistical analysis will begin in October 2018 (approximate date).

5. STUDY PARTICIPANTS

5.1 Eligible participants

- The primary and secondary statistical analyses will include all randomised women with an appropriate baseline and first follow-up mammogram (within specified timeframes - see below, good quality - as assessed by the experienced radiologist, MLO view only) who are breast cancer-free at the time of their first follow-up mammogram.

- Baseline mammograms will range from ≥0 months prior to the date of randomisation to <12.5 months prior to the date of randomisation. First follow-up mammograms will range from ≥8.5 months after the date of randomisation to <38.5 months after the date of randomisation. These time frames are in accordance with analysis from IBIS-I (19) and standard operating procedures for IBIS-II co-ordinating centres.

- Only women with all mammograms of the same image type (i.e. all film or all digital) will be included.

- Only women with ≥10% baseline density will be included.

- Breast cancer-free 'controls' are defined as women who had not been diagnosed with breast cancer at the time of study design. 'Cases' are defined as women who had been clinically diagnosed with breast cancer at the time of study design. Cases will be included if they are diagnosed with breast cancer after their first follow-up mammogram.

- Contralateral mammograms will be used for cases and mammograms from a randomly selected breast side will be used for breast cancer-free controls.

5.2 Sample size calculation

The estimated distribution of IBIS-II cases and controls with <10% and ≥10% density reduction per treatment arm was calculated by weighting these distributions observed in IBIS-I (19) with hazard ratios from IBIS-I (200) and IBIS-II (208). The sample size from IBIS-I (19) was then weighted using chosen multipliers to obtain a variety of distributions for different sample sizes. The chosen sample sizes were 50, 100, 150 and 200 cases, with 3 controls per 1 case. Using a difference of proportions sample size calculation, 247 cases and 1013 controls are required to show a difference in risk between anastrozole-treated patients experiencing ≥10% density change and anastrozole-treated patients experiencing <10% density change from baseline to first follow-up mammogram at the 5% type-I error level and with 80% power. This number also accounts for exclusions with baseline density <10% based on the number of postmenopausal women with baseline density <10% in IBIS-I. The study is currently underpowered since suitable mammograms (meeting the criteria outlined in section 5.1) have been received for only 35 breast-cancer cases and 938 breast cancer-free controls. With 35 cases and 105 controls (1:3 ratio of cases to controls), the power to detect a difference in risk of breast cancer in anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram relative to anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram is 22% at the 5% type-I error level. A sample size larger than this is currently impracticable given the resources and number of mammograms received.

6. COVARIATES OF INTEREST

6.1 Primary covariate of interest

The primary covariate of interest is defined as the change in density from baseline mammogram to first follow-up mammogram (9-38 months post randomisation). Density will be visually-assessed by an experienced reader (Linda Metaxa) using 5% intervals, following the same method as in IBIS-I. Randomisation of mammograms will be per woman (so that mammograms for each woman will be read in comparison with the other mammograms for that woman), and mammograms will be ordered sequentially. For each woman, density at baseline will be read first, followed by first follow-up mammogram (compared with baseline mammogram). Density change will therefore be defined as the difference between baseline density and first follow-up

mammogram density; semi-continuously, dichotomised into <10% or ≥10% absolute reduction and <5% or ≥5% absolute reduction. The reader will be blinded to treatment group, case status and risk factors, and images will be appropriately anonymised.

7. OUTCOMES

7.1 Primary outcome

The primary outcome is defined as the risk of developing breast cancer.

8. STATISTICAL METHODS

8.1 Hypotheses to be tested

8.1.1 Primary hypothesis

- H_0: There is no difference in age-adjusted risk of breast cancer between anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram and anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram
- H_1: Age-adjusted risk of breast cancer in anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram is different to age-adjusted risk of breast cancer in anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram

8.1.2 Secondary hypothesis I

- H_0: There is no difference in age-adjusted risk of breast cancer between anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram and anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram
- H_1: Age-adjusted risk of breast cancer in anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram is different to age-adjusted risk of breast cancer in anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram

8.1.3 Secondary hypothesis II

- H_0: There is no difference in risk of breast cancer between anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram and anastrozole-treated patients who experience a <10% reduction in density at first follow-up

mammogram, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.

- H_1: Risk of breast cancer in anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram is different to risk of breast cancer in anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.

8.1.4 Secondary hypothesis III

- H_0: There is no difference in risk of breast cancer between anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram and anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.
- H_1: Risk of breast cancer in anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram is different to risk of breast cancer in anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.

8.1.5 Secondary hypothesis IV

- H_0: There is no difference in the age-adjusted effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) between anastrozole-treated patients and placebo-treated patients.
- H_1: The age-adjusted effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) in anastrozole-treated patients is different to the age-adjusted effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) in placebo-treated patients.

8.1.6 Secondary hypothesis V

- H_0: There is no difference in the age-adjusted effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) between anastrozole-treated patients and placebo-treated patients.
- H_1: The age-adjusted effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) in anastrozole-treated patients is different to the age-adjusted effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) in placebo-treated patients.

8.1.7 Secondary hypothesis VI

- H_0: There is no difference in the effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) between anastrozole-treated patients and placebo-treated patients, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.
- H_1: The effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) in anastrozole-treated patients is different to the effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) in placebo-treated patients, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.

8.1.8 Secondary hypothesis VII

- H_0: There is no difference in the effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) between anastrozole-treated patients and placebo-treated patients, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.
- H_1: The effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) in anastrozole-treated patients is different to the effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) in placebo-treated patients, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.

8.1.9 Secondary hypothesis VIII

- H_0: There is no difference in age-adjusted risk of breast cancer between anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram and placebo-treated patients, and there is no difference in age-adjusted risk of breast cancer between anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram and placebo-treated patients.
- H_1: Age-adjusted risk of breast cancer in anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram is different to age-adjusted risk of breast cancer in placebo-treated patients, and age-adjusted risk of breast cancer in anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram is different to age-adjusted risk of breast cancer in placebo-treated patients.

8.1.10 Secondary hypothesis IX

- H_0: There is no difference in age-adjusted risk of breast cancer between anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram and

placebo-treated patients, and there is no difference in age-adjusted risk of breast cancer between anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram and placebo-treated patients

- H_1: Age-adjusted risk of breast cancer in anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram is different to age-adjusted risk of breast cancer in placebo-treated patients, and age-adjusted risk of breast cancer in anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram is different to age-adjusted risk of breast cancer in placebo-treated patients.

8.1.11 Secondary hypothesis X

- H_0: There is no difference in risk of breast cancer between anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram and placebo-treated patients, and there is no difference in risk of breast cancer between anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram and placebo-treated patients, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.
- H_1: Risk of breast cancer in anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram is different to risk of breast cancer in placebo-treated patients, and risk of breast cancer in anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram is different to risk of breast cancer in placebo-treated patients, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.

8.1.12 Secondary hypothesis XI

- H_0: There is no difference in risk of breast cancer between anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram and placebo-treated patients, and there is no difference in risk of breast cancer between anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram and placebo-treated patients, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.
- H_1: Risk of breast cancer in anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram is different to risk of breast cancer in placebo-treated patients, and risk of breast cancer in anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram is different to risk of breast cancer in placebo-treated patients, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.

8.1.13 Secondary hypothesis XII

- H_0: There is no difference in breast cancer risk between subgroups of covariates (tumour ER status, age at randomisation, body mass index at randomisation, baseline density, history of atypical hyperplasia or LCIS, hormone replacement use up to 12 months before randomisation, Tyrer-Cuzick 10-year risk, image type, and time between baseline mammogram and first follow-up mammogram) in women who experience an anastrozole-induced ≥10% reduction in density from baseline to first follow-up mammogram.
- H_1: Breast cancer risk in women who experience an anastrozole-induced ≥10% reduction in density from baseline to first follow-up mammogram is different between subgroups of covariates (tumour ER status, age at randomisation, body mass index at randomisation, baseline density, history of atypical hyperplasia or LCIS, hormone replacement use up to 12 months before randomisation, Tyrer-Cuzick 10-year risk, image type, and time between baseline mammogram and first follow-up mammogram).

8.1.14 Secondary hypothesis XIII

- H_0: There is no difference in breast cancer risk between subgroups of covariates (tumour ER status, age at randomisation, body mass index at randomisation, baseline density, history of atypical hyperplasia or LCIS, hormone replacement use up to 12 months before randomisation, Tyrer-Cuzick 10-year risk, image type, and time between baseline mammogram and first follow-up mammogram) in women who experience an anastrozole-induced ≥5% reduction in density from baseline to first follow-up mammogram.
- H_1: Breast cancer risk in women who experience an anastrozole-induced ≥5% reduction in density from baseline to first follow-up mammogram is different between subgroups of covariates (tumour ER status, age at randomisation, body mass index at randomisation, baseline density, history of atypical hyperplasia or LCIS, hormone replacement use up to 12 months before randomisation, Tyrer-Cuzick 10-year risk, image type, and time between baseline mammogram and first follow-up mammogram).

8.2 Analysis methods

All statistical analysis will be conducted in STATA 13. All tests (see below) will be two-sided with a significance level of 5%. Results will be omitted if subgroup numbers are small enough in order to un-blind the statistician.

8.2.1 Baseline characteristics and risk

The following baseline characteristics will be summarised in a frequency table by treatment arm and case status, along with univariate and multivariable odds ratios for risk of developing breast cancer. Frequency counts & percentages will be provided for categorical data and means (standard deviation, SD) and medians (interquartile range,

IQR) will be provided for continuous data. Two-sample t-tests (*STATA's "ttest" command*) and Wilcoxon rank sum tests (*STATA's "ranksum" command*) will test differences between cases and controls in continuous data and Pearson chi-squared tests (*STATA's "tab, chi2" command*) will test differences between cases and controls in categorical data:

- Age at randomisation (mean (SD), median (IQR))
- Body Mass Index (BMI) at randomisation (mean (SD), median (IQR))
- Age at menarche (mean (SD), median (IQR))
- Age at menopause (mean (SD), median (IQR))
- Tyrer-Cuzick 10-year risk (mean (SD), median (IQR))
- Baseline density (mean (SD), median (IQR))
- Age at first birth (nulliparous/>27/21-27/≤20)
- Oral contraception use (never/previously/currently)
- Hormone Replacement Therapy (HRT) use up to 12 months before randomisation (no/yes)
- Smoking status (never/former/current)
- History of atypical hyperplasia or LCIS (no/yes)
- Image type (film/digital)

- Univariate logistic regression models of risk of breast cancer on baseline covariates, adjusted for age at randomisation (except age at randomisation): n, odds ratio, 95% confidence interval, P-value.
  STATA's "logistic" command
- Multivariable logistic regression models of risk of breast cancer on baseline covariates. All covariates will be included in the multivariable model: n, odds ratio, 95% confidence interval, P-value.
  STATA's "logistic" command

8.2.2 Primary analysis

- Logistic regression model of breast cancer risk on change in density (dichotomised into <10% absolute reduction and ≥10% absolute reduction; reference group: <10% absolute reduction) and age at randomisation in anastrozole-treated patients: n, odds ratio, 95% confidence interval, P-value.
  STATA's "logistic" command

8.2.3 Secondary analysis I

- Logistic regression model of breast cancer risk on change in density (dichotomised into <5% absolute reduction and ≥5% absolute reduction; reference group: <5% absolute reduction) and age at randomisation in anastrozole-treated patients: n, odds ratio, 95% confidence interval, P-value.
  STATA's "logistic" command

8.2.4 Secondary analysis II

- Logistic regression model of breast cancer risk on change in density (dichotomised into <10% absolute reduction and ≥10% absolute reduction; reference group: <10% absolute reduction) in anastrozole-treated patients, adjusted for covariates. All covariates will be included in the multivariable model: n, odds ratios, 95% confidence intervals, P-values.
  STATA's "logistic" command

8.2.5 Secondary analysis III

- Logistic regression model of breast cancer risk on change in density (dichotomised into <5% absolute reduction and ≥5% absolute reduction; reference group: <5% absolute reduction) in anastrozole-treated patients, adjusted for covariates. All covariates will be included in the multivariable model: n, odds ratios, 95% confidence intervals, P-values.
  STATA's "logistic" command

8.2.6 Secondary analysis IV

- Logistic regression model of breast cancer risk on change in density (dichotomised into <10% absolute reduction and ≥10% absolute reduction), treatment and an interaction between both, and age at randomisation: n, odds ratios, 95% confidence intervals, P-values.
  STATA's "logistic" command

8.2.7 Secondary analysis V

- Logistic regression model of breast cancer risk on change in density (dichotomised into <5% absolute reduction and ≥5% absolute reduction), treatment and an interaction between both, and age at randomisation: n, odds ratios, 95% confidence intervals, P-values.
  STATA's "logistic" command

8.2.8 Secondary analysis VI

- Logistic regression model of breast cancer risk on change in density (dichotomised into <10% absolute reduction and ≥10% absolute reduction), treatment and an interaction

between both, adjusted for covariates. All covariates will be included in the multivariable model: n, odds ratios, 95% confidence intervals, P-values.
STATA's "logistic" command

8.2.9 Secondary analysis VII

- Logistic regression model of breast cancer risk on change in density (dichotomised into <5% absolute reduction and ≥5% absolute reduction), treatment and an interaction between both, adjusted for covariates. All covariates will be included in the multivariable model: n, odds ratios, 95% confidence intervals, P-values.
  STATA's "logistic" command

8.2.10 Secondary analysis VIII

- Logistic regression model of breast cancer risk on a variable for change in density and treatment (factorised into: anastrozole-induced <10% absolute reduction, anastrozole-induced ≥10% absolute reduction, and placebo; reference category: placebo) and age at randomisation: n, odds ratio, 95% confidence interval, P-value.
  STATA's "logistic" command

8.2.11 Secondary analysis IX

- Logistic regression model of breast cancer risk on a variable for change in density and treatment (factorised into: anastrozole-induced <5% absolute reduction, anastrozole-induced ≥5% absolute reduction, and placebo; reference category: placebo) and age at randomisation: n, odds ratio, 95% confidence interval, P-value.
  STATA's "logistic" command

8.2.12 Secondary analysis X

- Logistic regression model of breast cancer risk on a variable for change in density and treatment (factorised into: anastrozole-induced <10% absolute reduction, anastrozole-induced ≥10% absolute reduction, and placebo; reference category: placebo), adjusted for covariates. All covariates will be included in the multivariable model: n, odds ratio, 95% confidence interval, P-value.
  STATA's "logistic" command

8.2.13 Secondary analysis XI

- Logistic regression model of breast cancer risk on a variable for change in density and treatment (factorised into: anastrozole-induced <5% absolute reduction, anastrozole-induced ≥5% absolute reduction, and placebo; reference category: placebo), adjusted for

covariates. All covariates will be included in the multivariable model: n, odds ratio, 95% confidence interval, P-value.

STATA's "logistic" command

8.2.14 Secondary analysis XII

- Odds ratios to assess the odds of breast cancer for anastrozole-induced ≥10% density reduction in each subgroup (relative to a reference subgroup), in the anastrozole arm only: n, odds ratio, 95% confidence interval, P-value.

  STATA's "logistic" command

8.2.15 Secondary analysis XIII

- Odds ratios to assess the odds of breast cancer for anastrozole-induced ≥5% density reduction in each subgroup (relative to a reference subgroup), in the anastrozole arm only: n, odds ratio, 95% confidence interval, P-value.

  STATA's "logistic" command

8.2.12 Adjustment covariates

The following covariates will be included in adjusted regression models (8.2.4, 8.2.5, 8.2.8 & 8.2.9):

- Age at randomisation (continuous)
- BMI at randomisation (continuous)
- Baseline density (continuous)
- Tyrer-Cuzick 10-year risk (continuous)

Time on treatment will not be included in adjustments because an intention-to-treat analysis will be conducted.

8.2.13 Subgroup covariates

The following covariates will be considered in subgroup analyses (8.2.10 & 8.2.11):

- Tumour ER status (negative/positive)
- Age at randomisation (<median age, ≥median age)
- BMI at randomisation (<median BMI, ≥median BMI)
- Baseline density (<median baseline density, ≥median baseline density)
- History of atypical hyperplasia or LCIS (no/yes)
- HRT use up to 12 months before randomisation (no/yes)
- Tyrer-Cuzick 10-year risk (<median risk, ≥median risk)

- Image type (film/digital)
- Time between baseline and first mammogram (<median time between baseline and first mammogram, ≥median time between baseline and first mammogram)

8.2.14 Sensitivity analysis for compliance

It is possible that the greater reduction in breast cancer risk we might observe in subjects from the anastrozole arm who experience a density reduction of at least 10% or 5% (compared with similar women who experience <10% or <5% density reduction) reflects better treatment compliance and is not a measure of biological response to treatment. To test this, we will use Kaplan–Meier curves (censored ~3 months (90 days) before cancer diagnosis) (*STATA's "sts graph" command*) and log rank tests (*STATA's "sts test, logrank" command*) to assess the difference in time to stopping treatment between cases on anastrozole with ≥10% vs. <10% reduction in density and ≥5% vs. <5% reduction in density.

8.3 Tables

Variable	All						Placebo						Anastrozole					
	Cases		Controls		Univariate OR (95% CI) ***#	Multivariable OR (95% CI) ***##	Cases		Controls		Univariate OR (95% CI) ***#	Multivariable OR (95% CI) ***##	Cases		Controls		Univariate OR (95% CI) ***#	Multivariable OR (95% CI) ***##
	Mean (SD)*	Median (IQR)**	Mean (SD)*	Median (IQR)**			Mean (SD)*	Median (IQR)**	Mean (SD)*	Median (IQR)**			Mean (SD)*	Median (IQR)**	Mean (SD)*	Median (IQR)**		
Age at randomisation+ (yr)																		
P																		
Body Mass Index+ (kg/m2)																		
P																		
Baseline density+ (%)																		
P																		
Tyrer-Cuzick 10-year risk+ (%)																		
	N+++	%	N+++	%			N+++	%	N+++	%			N+++	%	N+++	%		
History of atypical hyperplasia or LCIS++																		
No					*Ref*	*Ref*					*Ref*	*Ref*					*Ref*	*Ref*
Yes																		
P																		
HRT use up to 12 months before randomisation++																		
No					*Ref*	*Ref*					*Ref*	*Ref*					*Ref*	*Ref*
Yes																		
P																		

Table 1: Baseline characteristics by case status and treatment, with odds ratios (ORs) for the risk of developing breast cancer from univariate and multivariable logistic regression models. *P-value from two-sample t-test, **p-value from Wilcoxon rank sum test, ***p-value from logistic regression model Wald test of covariate, +++p-value from Pearson chi-squared test of association. +OR represents odds of breast cancer per unit increase in covariate, ++OR represents odds of breast cancer relative to the reference category. #All univariate models (except that for age at randomisation) are adjusted for age at randomisation. ##Multivariable models include all variables. N=.

	Number of women													
Boyd category at entry	Boyd category at first follow-up: Cases						Total	Boyd category at first follow-up: Controls						Total
	0%	1-10%	11-25%	26-50%	51-75%	76-100%		0%	1-10%	11-25%	26-50%	51-75%	76-100%	
0%														
1-10%														
11-25%														
26-50%														
51-75%														
76-100%														
Total														

Table 2: Cross tabulation of number of women in each Boyd category at entry to the study with category at first follow-up, by case status. The first number in each cell is the total number of subjects. Numbers in parentheses are the placebo and anastrozole groups, respectively.

Change in breast density, No. (%)	Placebo		Anastrozole	
	Cases (N=)	Controls (N=)	Cases (N=)	Controls (N=)
Mean (SD)				
Median (IQR)				
<5% reduction				
≥5% reduction				
<10% reduction				
≥10% reduction				
P trend				

Table 3: Density change by treatment and case status. P-value from Wald test of change in density from a logistic regression model of breast cancer risk on change in density (semi-continuous) in each treatment arm.

Variable	≥5% reduction					≥10% reduction				
	N	Univariate		Adjusted#		N	Univariate		Adjusted#	
		OR (95% CI)	P-value	OR (95% CI)	P-value		OR (95% CI)	P-value	OR (95% CI)	P-value
Density change*+										
<5% reduction		*Ref*	*Ref*	*Ref*	*Ref*		-	-	-	-
≥5% reduction							-	-	-	-
<10% reduction		-	-	-	-		*Ref*	*Ref*	*Ref*	*Ref*
≥10% reduction		-	-	-	-					
Age at randomisation (yr)**										
BMI at randomisation (kg/m^2)**										
Baseline density (%)**										
Tyrer-Cuzick 10-year risk (%)**										

Table 4: Logistic regression results for risk of breast cancer on first change in density (dichotomised into <5% reduction or ≥5% reduction, and <10% reduction or ≥10% reduction) in anastrozole arm only, in univariate and adjusted models. *OR represents odds of breast cancer relative to the reference category, **OR represents odds of breast cancer per unit increase in covariate. #Multivariable models include all variables. +Additionally adjusted for age at randomisation in univariate model. P-values from logistic regression model Wald test. N=.

Variable	<5% or ≥5% reduction					<10% or ≥10% reduction				
	N	Univariate		Adjusted#		N	Univariate		Adjusted#	
		OR (95% CI)	P-value	OR (95% CI)	P-value		OR (95% CI)	P-value	OR (95% CI)	P-value
Density change*+										
<5% reduction							-	-	-	-
≥5% reduction							-	-	-	-
<10% reduction		-	-	-	-					
≥10% reduction		-	-	-	-					
Age at randomisation (yr)**										
BMI at randomisation (kg/m^2)**										
Baseline density (%)**										
Tyrer-Cuzick 10-year risk (%)**										

Table 5: Logistic regression results for risk of breast cancer on first change in density (dichotomised into <5% reduction or ≥5% reduction, and <10% reduction or ≥10% reduction) in anastrozole-treated women relative to all women in the placebo arm, in univariate and adjusted models. *OR represents odds of breast cancer relative to all women in the placebo arm, **OR represents odds of breast cancer per unit increase in covariate. #Multivariable models include all variables. +Additionally adjusted for age at randomisation in univariate model. P-values from logistic regression model Wald test. N=.

Variable	No. of control subjects/No. of case subjects	Anastrozole, density reduction ≥5%		P###	Anastrozole, density reduction ≥10%		P###
		No. of case subjects	OR# (95% CI)		No. of case subjects	OR## (95% CI)	
Overall							
Tumour ER status							
Negative							
Positive							
Age at randomisation (yr)*							
Younger age at randomisation							
Older age at randomisation							
Body Mass Index (kg/m^2)*							
Lower Body Mass Index							
Higher Body Mass Index							
Baseline density*							
Lower baseline density							

Higher baseline density							
History of atypical hyperplasia or LCIS							
No							
Yes							
HRT use up to 12 months before randomisation							
No							
Yes							
Tyrer-Cuzick 10-year risk (%)*							
Lower Tyrer-Cuzick 10-year risk							
Higher Tyrer-Cuzick 10-year risk							
Image type							
Film							
Digital							
Time between baseline mammogram and first follow-up mammogram (yr)*							
Shorter time between mammograms							
Longer time between mammograms							

Table 6: Risk of breast cancer on anastrozole-induced first change in density by subgroups of covariates. *Continuous variables dichotomised by median of variable in all women, #OR represents odds of developing breast cancer for women with ≥5% density reduction (relative to women with <5% density reduction) in different subgroups. ##OR represents odds of developing breast cancer for women with ≥10% density reduction (relative to women with <10% density reduction) in different subgroups. ###P-values from a Wald test.

C.XXVI Proforma for Chapter 6

Request Originator:	Ms Emma Atakpa PhD student, Centre for Cancer Prevention, Wolfson Institute Supervisors: Professor Jack Cuzick Dr Adam Brentnall
Reason for request:	To complete an analysis assessing anastrozole-induced change in mammographic density as a biomarker for breast cancer risk in patients from the IBIS-II Prevention trial.
Background:	Mammographic density (herein referred to as 'density') is one of the strongest known risk factors for breast cancer. Women in the highest density category (≥75%) are at a 4 to 6-fold increased risk of developing breast cancer relative to those with little or no density (79). Whilst postmenopausal hormone replacement therapy (HRT) is associated with an increase in risk of breast cancer (424) and density (188, 193); selective oestrogen receptor modulators (SERMs), such as tamoxifen, decrease risk (198, 199, 425) and density (203-205). Most importantly, high-risk women who experienced ≥10% density reduction after approximately 18 months of prophylactic tamoxifen were shown to be at approximately 68% lower risk of developing breast cancer compared with women who experienced <10% density reduction after the same treatment in the IBIS-I trial (19). This makes tamoxifen-induced density reduction a potential biomarker for risk reduction. Similar to SERMs, aromatase inhibitors (AIs) are an anti-oestrogenic drug given to women in the treatment of breast cancer. In 2014, analysis from IBIS-II showed that anastrozole (an AI) reduced the risk of breast cancer in high-risk postmenopausal women (208). It is not yet known whether anastrozole-induced density reduction can also be used as a biomarker for risk reduction, whereby a ≥10% reduction in density after approximately 18 months of preventive anastrozole treatment would be associated with a lower risk of breast cancer compared with <10% density reduction after the same treatment.
Aims:	• *Primary objective*: To determine whether women on anastrozole who experience a ≥10% reduction in density at first follow-up mammogram have a different level of age-adjusted risk of breast cancer than women on anastrozole who experience a <10% reduction in density at first follow-up mammogram in the IBIS-II Prevention trial. • *Primary hypothesis*: Age-adjusted risk of breast cancer in anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram is different to age-adjusted risk of breast cancer in anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram. • *Secondary objective I*: To determine whether women on anastrozole who experience a ≥5% reduction in density at first follow-up mammogram have a different level of age-adjusted risk of breast cancer than women on anastrozole who experience a <5% reduction in density at first follow-up

	mammogram in the IBIS-II Prevention trial. • *Secondary hypothesis I*: Age-adjusted risk of breast cancer in anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram is different to age-adjusted risk of breast cancer in anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram. • *Secondary objective II*: To determine whether women on anastrozole who experience a ≥10% reduction in density at first follow-up mammogram have a different level of risk of breast cancer than women on anastrozole who experience a <10% reduction in density at first follow-up mammogram in the IBIS-II Prevention trial, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk. • *Secondary hypothesis II*: Risk of breast cancer in anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram is different to risk of breast cancer in anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk. • *Secondary objective III*: To determine whether women on anastrozole who experience a ≥5% reduction in density at first follow-up mammogram have a different level of risk of breast cancer than women on anastrozole who experience a <5% reduction in density at first follow-up mammogram in the IBIS-II Prevention trial, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk. • *Secondary hypothesis III*: Risk of breast cancer in anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram is different to risk of breast cancer in anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk. • *Secondary objective IV*: To determine whether the age-adjusted effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) is different between anastrozole-treated and placebo-treated patients (interaction test for anastrozole-induced density change as a predictive biomarker for breast cancer risk reduction). • *Secondary hypothesis IV*: The age-adjusted effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) in anastrozole-treated patients is different to the age-adjusted effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) in placebo-treated patients. • *Secondary objective V*: To determine whether the age-adjusted effect of ≥5% reduction in density at first follow-up mammogram on breast cancer

	risk (relative to <5% reduction) is different between anastrozole-treated and placebo-treated patients (interaction test for anastrozole-induced density change as a predictive biomarker for breast cancer risk reduction). • *Secondary hypothesis V*: The age-adjusted effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) in anastrozole-treated patients is different to the age-adjusted effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) in placebo-treated patients. • *Secondary objective VI*: To determine whether the effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) is different between anastrozole-treated and placebo-treated patients (interaction test for anastrozole-induced density change as a predictive biomarker for breast cancer risk reduction), after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk. • *Secondary hypothesis VI*: The effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) in anastrozole-treated patients is different to the effect of ≥10% reduction in density at first follow-up mammogram on breast cancer risk (relative to <10% reduction) in placebo-treated patients, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk. • *Secondary objective VII*: To determine whether the effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) is different between anastrozole-treated and placebo-treated patients (interaction test for anastrozole-induced density change as a predictive biomarker for breast cancer risk reduction), after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk. • *Secondary hypothesis VII*: The effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) in anastrozole-treated patients is different to the effect of ≥5% reduction in density at first follow-up mammogram on breast cancer risk (relative to <5% reduction) in placebo-treated patients, after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk. • *Secondary objective VIII*: To determine whether women on anastrozole who experience a ≥10% reduction in density at first follow-up mammogram have a different level of age-adjusted risk of breast cancer than women on placebo, and whether women on anastrozole who experience a <10% reduction in density at first follow-up mammogram have a different level of age-adjusted risk of breast cancer than women on placebo in the IBIS-II Prevention trial (test for anastrozole-induced density change as a predictive biomarker for breast cancer risk reduction). • *Secondary hypothesis VIII*: Age-adjusted risk of breast cancer in anastrozole-treated patients who experience a ≥10% reduction in density at

	first follow-up mammogram is different to age-adjusted risk of breast cancer in placebo-treated patients, and age-adjusted risk of breast cancer in anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram is different to age-adjusted risk of breast cancer in placebo-treated patients. • *Secondary objective IX*: To determine whether women on anastrozole who experience a ≥5% reduction in density at first follow-up mammogram have a different level of age-adjusted risk of breast cancer than women on placebo, and whether women on anastrozole who experience a <5% reduction in density at first follow-up mammogram have a different level of age-adjusted risk of breast cancer than women on placebo in the IBIS-II Prevention trial (test for anastrozole-induced density change as a predictive biomarker for breast cancer risk reduction). • *Secondary hypothesis IX*: Age-adjusted risk of breast cancer in anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram is different to age-adjusted risk of breast cancer in placebo-treated patients, and age-adjusted risk of breast cancer in anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram is different to age-adjusted risk of breast cancer in placebo-treated patients. • *Secondary objective X*: To determine whether women on anastrozole who experience a ≥10% reduction in density at first follow-up mammogram have a different level of risk of breast cancer than women on placebo, and whether women on anastrozole who experience a <10% reduction in density at first follow-up mammogram have a different level of risk of breast cancer than women on placebo in the IBIS-II Prevention trial (test for anastrozole-induced density change as a predictive biomarker for breast cancer risk reduction), after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk. • *Secondary hypothesis X*: Risk of breast cancer in anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram is different to risk of breast cancer in placebo-treated patients, and risk of breast cancer in anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram is different to risk of breast cancer in placebo-treated patients, both after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk. • *Secondary objective XI*: To determine whether women on anastrozole who experience a ≥5% reduction in density at first follow-up mammogram have a different level of risk of breast cancer than women on placebo, and whether women on anastrozole who experience a <5% reduction in density at first follow-up mammogram have a different level of risk of breast cancer than women on placebo in the IBIS-II Prevention trial (test for anastrozole-induced density change as a predictive biomarker for breast cancer risk reduction), after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk.

	• *Secondary hypothesis XI*: Risk of breast cancer in anastrozole-treated patients who experience a ≥5% reduction in density at first follow-up mammogram is different to risk of breast cancer in placebo-treated patients, and risk of breast cancer in anastrozole-treated patients who experience a <5% reduction in density at first follow-up mammogram is different to risk of breast cancer in placebo-treated patients, both after adjustment for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk. • *Secondary objective XII*: To examine the effect of anastrozole-induced first density reduction of ≥10% on breast cancer risk in different subgroups of covariates in the IBIS-II Prevention trial (covariates to be split into subgroups are: tumour ER status, age at randomisation, body mass index at randomisation, baseline density, history of atypical hyperplasia or LCIS, hormone replacement use up to 12 months before randomisation, Tyrer-Cuzick 10-year risk, image type, and time between baseline mammogram and first follow-up mammogram). • *Secondary hypothesis XII*: Breast cancer risk in women who experience an anastrozole-induced ≥10% reduction in density from baseline to first follow-up mammogram is different between subgroups of covariates (tumour ER status, age at randomisation, body mass index at randomisation, baseline density, history of atypical hyperplasia or LCIS, hormone replacement use up to 12 months before randomisation, Tyrer-Cuzick 10-year risk, image type, and time between baseline mammogram and first follow-up mammogram). • *Secondary objective XIII*: To examine the effect of anastrozole-induced first density reduction of ≥5% on breast cancer risk in different subgroups of covariates in the IBIS-II Prevention trial (covariates to be split into subgroups are: tumour ER status, age at randomisation, body mass index at randomisation, baseline density, history of atypical hyperplasia or LCIS, hormone replacement use up to 12 months before randomisation, Tyrer-Cuzick 10-year risk, image type, and time between baseline mammogram and first follow-up mammogram). • *Secondary hypothesis XIII*: Breast cancer risk in women who experience an anastrozole-induced ≥5% reduction in density from baseline to first follow-up mammogram is different between subgroups of covariates (tumour ER status, age at randomisation, body mass index at randomisation, baseline density, history of atypical hyperplasia or LCIS, hormone replacement use up to 12 months before randomisation, Tyrer-Cuzick 10-year risk, image type, and time between baseline mammogram and first follow-up mammogram).
Data required:	Suitable mammograms (within the following specified timeframes, before breast cancer diagnosis, MLO view, deemed to be good quality) have been received for 35 breast-cancer cases and 938 breast cancer-free controls. Baseline (up to 12 months before randomisation) and first follow-up (9-38 months post randomisation) mammograms for these 973 women were batched and sent to a radiologist at St Bartholomew's hospital for density scoring.

	The study is currently underpowered. With 35 cases and 105 controls (1:3 ratio of cases to controls), the power to detect a difference in risk of breast cancer in anastrozole-treated patients who experience a ≥10% reduction in density at first follow-up mammogram relative to anastrozole-treated patients who experience a <10% reduction in density at first follow-up mammogram is 22% at the 5% type-I error level (please see 'Other comments'). Since the requestor is blinded to treatment allocation, a set of STATA code will be sent to Dr Sestak to run on the un-blinded data for these 973 women.
Describe the specific analyses or tables requested:	**Methods:** • Mammograms were visually assessed by a radiologist (Dr Metaxa) at St Bartholomew's hospital. • Each mammogram was scored in 5% increments, following the same method as in IBIS-I. • Contralateral mammograms were used for cases and mammograms from a randomly selected breast side were used for breast cancer-free controls. • Density change will be defined as the difference between baseline and first follow-up mammogram. • Only women with ≥10% baseline density will be included. • Only women with all mammograms of the same image type (i.e. all film or all digital) will be included. **Statistical analysis:** • Baseline characteristics will be summarised by treatment arm and case status using frequency tables with age at randomisation, body mass index at randomisation, age at menarche, age at menopause, Tyrer-Cuzick 10-year risk, baseline density, age at first birth, oral contraception use, hormone replacement therapy use up to 12 months before randomisation, smoking status, history of atypical hyperplasia or LCIS, and image type. An exploratory analysis will also assess the association between risk of developing breast cancer and baseline characteristics using univariate (adjusted for age at randomisation) and multivariable logistic regression. • The primary analysis will use a logistic regression model to examine the association between risk of breast cancer and anastrozole-induced ≥10% first follow-up density reduction (relative to anastrozole-induced <10% absolute reduction), adjusted for age at randomisation, in anastrozole-treated patients. • The secondary analysis (I) will use a logistic regression model to examine the association between risk of breast cancer and anastrozole-induced ≥5% first follow-up density reduction (relative to anastrozole-induced <5% absolute reduction), adjusted for age at randomisation, in anastrozole-treated patients. • The secondary analysis (II) will use a logistic regression model to examine

	the association between risk of breast cancer and anastrozole-induced ≥10% first follow-up density reduction (relative to anastrozole-induced <10% absolute reduction), adjusted for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk; in anastrozole-treated patients. • The secondary analysis (III) will use a logistic regression model to examine the association between risk of breast cancer and anastrozole-induced ≥5% first follow-up density reduction (relative to anastrozole-induced <5% absolute reduction), adjusted for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk; in anastrozole-treated patients. • The secondary analysis (IV) will use a logistic regression model to examine the association between risk of breast cancer and first follow-up density reduction (dichotomised into <10% absolute reduction and ≥10% absolute reduction), treatment and an interaction between density reduction and treatment, adjusted for age at randomisation. • The secondary analysis (V) will use a logistic regression model to examine the association between risk of breast cancer and first follow-up density reduction (dichotomised into <5% absolute reduction and ≥5% absolute reduction), treatment and an interaction between density reduction and treatment, adjusted for age at randomisation. • The secondary analysis (VI) will use a logistic regression model to examine the association between risk of breast cancer and first follow-up density reduction (dichotomised into <10% absolute reduction and ≥10% absolute reduction), treatment and an interaction between density reduction and treatment, adjusted for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk. • The secondary analysis (VII) will use a logistic regression model to examine the association between risk of breast cancer and first follow-up density reduction (dichotomised into <5% absolute reduction and ≥5% absolute reduction), treatment and an interaction between density reduction and treatment, adjusted for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk. • The secondary analysis (VIII) will use a logistic regression model to examine the association between risk of breast cancer and anastrozole-induced ≥10% first follow-up density reduction (relative to placebo-treated women), and risk of breast cancer and anastrozole-induced <10% first follow-up density reduction (relative to placebo-treated women), adjusted for age at randomisation. • The secondary analysis (IX) will use a logistic regression model to examine the association between risk of breast cancer and anastrozole-induced ≥5% first follow-up density reduction (relative to placebo-treated women), and

	risk of breast cancer and anastrozole-induced <5% first follow-up density reduction (relative to placebo-treated women), adjusted for age at randomisation. • The secondary analysis (X) will use a logistic regression model to examine the association between risk of breast cancer and anastrozole-induced ≥10% first follow-up density reduction (relative to placebo-treated women), and risk of breast cancer and anastrozole-induced <10% first follow-up density reduction (relative to placebo-treated women), adjusted for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk. • The secondary analysis (XI) will use a logistic regression model to examine the association between risk of breast cancer and anastrozole-induced ≥5% first follow-up density reduction (relative to placebo-treated women), and risk of breast cancer and anastrozole-induced <5% first follow-up density reduction (relative to placebo-treated women), adjusted for age at randomisation, body mass index at randomisation, baseline density, and Tyrer-Cuzick 10-year risk. • The secondary analysis (XII) will use logistic regression to assess the odds of developing breast cancer in women who have experienced anastrozole-induced ≥10% density reduction in different subgroups of covariates (tumour ER status, age at randomisation, body mass index at randomisation, baseline density, history of atypical hyperplasia or LCIS, hormone replacement use up to 12 months before randomisation, Tyrer-Cuzick 10-year risk, image type, and time between baseline mammogram and first follow-up mammogram). • The secondary analysis (XIII) will use logistic regression to assess the odds of developing breast cancer in women who have experienced anastrozole-induced ≥5% density reduction in different subgroups of covariates (tumour ER status, age at randomisation, body mass index at randomisation, baseline density, history of atypical hyperplasia or LCIS, hormone replacement use up to 12 months before randomisation, Tyrer-Cuzick 10-year risk, image type, and time between baseline mammogram and first follow-up mammogram). • All statistical analysis will be conducted in STATA 13. All tests will be two-sided with a significance level of 5%. • Results will be omitted if subgroup numbers are small enough to un-blind the requestor.
What will the data be used for?	The results of the analysis will form a chapter in Ms Emma Atakpa's PhD thesis. The results will also be prepared for publication in a peer-reviewed journal (dependent on results and TSC permitting).
Date	Analysis will begin as soon as possible after responses from the TSC.

required by:	
Other comments:	There is currently not enough power to complete the primary objective. However, Kim et al. (264) found a provocative result of increased risk of recurrence in oestrogen receptor-positive breast cancer cases on AIs who lost <5% density after 8-20 months of treatment relative to similarly treated women who lost ≥5% density, although results were not significant (HR=7.11; 95% CI: 0.90-56.37; p=0.06). The number of recurrences on AIs was not reported but it is estimated to be 13 from other numbers reported in the paper. Assuming 32% of the 35 cases in this study were on anastrozole (40 anastrozole cases/125 cases in Cuzick et al. (208)) we estimate there to be approximately 11 anastrozole cases. There may therefore be enough power in this study to detect an effect if density change is dichotomised into <5% and ≥5% reduction (secondary objectives).
Proposed authorship:	E. Atakpa, A. Brentnall, L. Metaxa, I. Sestak, J.F. Forbes, A. Howell, J. Cuzick

This section to be completed by IBIS-II Trial Steering Committee

Does the Request have Executive Committee Approval? Please place a X in appropriate box	Yes ☐	No ☐
If no, reason for rejection:		
Date received by IBIS-II CCO:		
Date decision sent to applicant:		