QUEEN MARY, UNIVERSITY OF LONDON

Clinical and epidemiological issues and applications of mammographic density

Valentina Assi

A thesis submitted in partial fulfillment for the
Degree of Doctor of Philosophy

January 2014

Statement of Originality

I, Valentina Assi, 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.

January 2014

Details of collaboration and publications:
1. This dissertation was undertaken in the Centre for Cancer Prevention, Queen Mary, University of London, under the supervision of Professor Stephen Duffy and Dr. Jane Warwick.
2. Chapter 1 was partially published as:
Assi V, Warwick J, Cuzick J, Duffy SW. Clinical and epidemiological issues in mammographic density. *Nat Rev Clin Oncol*, 9(1):33-40, 2011.

The History of every major Galactic Civilization tends to pass through three distinct and recognizable phases, those of Survival, Inquiry and Sophistication, otherwise known as the How, Why and Where phases.
For instance, the first phase is characterized by the question *How can we eat?* the second by the question *Why do we eat?* and the third by the question *Where shall we have lunch?*

Douglas Adams
The Hitchhiker's Guide to the Galaxy, (1979)

QUEEN MARY, UNIVERSITY OF LONDON

Abstract

Barts and The London, School of Medicine and Dentistry
Wolfson Institute of Preventive Medicine

Doctor of Philosophy

by Valentina Assi

Mammographic density, the amount of radiodense tissue on a mammogram, is a strong risk factor for breast cancer, with properties that could be an asset in screening and prevention programmes. Its use in risk prediction contexts is currently limited, however, mainly due to difficulties in measuring and interpreting density.

This research investigates firstly, the properties of density as an independent marker of breast cancer risk and secondly, how density should be measured.

The first question was addressed by analysing data from a chemoprevention trial, a trial of hormonal treatment, and a cohort study of women with a family history of breast cancer . Tamoxifen-induced density reduction was observed to be a good predictor of breast cancer risk reduction in high-risk unaffected subjects. Density and its changes did not predict risk or treatment outcome in subjects with a primary invasive breast tumour. Finally absolute density predicted risk better than percent density and showed a potential to improve existing risk-prediction models, even in a population at enhanced familial risk of breast cancer.

The second part of thesis focuses on density measurement and in particular evaluates two fully-automated volumetric methods, Quantra and Volpara. These two methods are highly correlated and in both cases absolute density (cm^3) discriminated cases from controls better than percent density. Finally, we evaluated and compared different measurement methods. Our findings suggested good reliability of the Cumulus and visual assessments. Quantra volumetric estimates appeared negligibly affected by measurement error, but were less variable than visual bi-dimensional ones, affecting their ability to discriminate cases from controls. Overall, visual assessments showed the strongest association with breast cancer risk in comparison to computerised methods.

Our research supports the hypothesis that density should have a role in personalising screening programs and risk management. Volumetric density measuring methods, though promising, could be improved.

Acknowledgements

Foremost, I express my sincere appreciation and gratitude to my supervisor Professor Stephen Duffy, for his continuous help and support during my PhD research, for his patience and his immense knowledge, and, last but not least, for being an example with his smiling attitude. Likewise, I extend my deepest gratitude to my second supervisor Dr Jane Warwick for her support and guidance over the past years, but mostly for helping me putting things in perspective both in terms of my research and my career, for opening my eyes to the bigger picture.

Sincere thanks to Dr Nicholas Perry, Dr Ralph Highnam and Professor Jack Cuzick, along with my supervisors, for providing most of the data presented in this thesis, and encouraging interesting conversations and thoughts on breast density issues.

I would like to acknowledge all the staff involved in the data collection and all the women enrolled in the studies analysed in this thesis (IBIS-I, ATAC and FH01), and especially staff and patients of the London Breast Institute within the Princess Grace Hospital. Moreover I shall thank Cancer Research UK for supporting me and this work financially.

The staff of the Centre for Cancer Prevention have contributed immensely to my personal and professional time at Queen Mary. The group has been a source of friendships as well as good advice and support. I am especially grateful to Prue, Adam, Sue, Amanda and Dharmishta who proof-read this thesis and enlightened me on the darker sides of British culture. Special thanks go to Adam and Prue who introduced me to English humour (mango/man-go, 7-8-9/7 ate 9).

A PhD in a foreign country implies by definition challenging times and my SGI friends provided me with probably the best support network I could hope for. They became my family in London and had a fundamental role in encouraging me to keep going when all I wanted was to give up. And it happened a lot. It would be hard to name them all but in particular: Pina, Alessandra, Kiri, Kriti, Niharika, Tatiana and Richa.
For similar reasons, as well as supporting my personal growth and being the most amazing hangover cure, I'm grateful to the Comitato Coccole ($\approx$ Cheering-up Commitee), also known as Barbara with her little friend Riccio. We had highs and lows, but what 13 year long friendship has not? It was such great luck to live only 5 minutes away

and to be able to count on her for most of the years spent here in London. She is a great companion for both good and bad times, for parties, Zumba and Yoga classes, Six Nations games, retail-therapy... Not to mention that her tips on how to improve my "graciousness" are priceless. Finally, Italian dinners at her place were one of the most effective remedies against homesickness.
Franz was my other piece of "Home, away from Home". Since our years studying together in Milan, I cannot help smiling when he is around. He is by far my favourite heart-less robot.

Last but not least, I express my deepest gratitude to my family, Katia, Roberto and Alessandro, for loving me unconditionally and always supporting me at their best. To my mother for trying any possible tiramisú recipes, although hers is already the most delicious, and being an inspiration with her tireless attitude and youthful spirit. To my father, because his over-worries about flights and trains are funny to laugh at and just another evidence that he is always there for us. To my brother, the Prince, because, paradoxically, these years apart brought us closer than ever and he has the extraordinary ability of calling me at the most appropriate times. Thank you.

Contents

List of Figures

List of Tables

To Giovanna, Giuditta and Katia

Chapter 1

Review of breast density issues and study plan

1.1 Introduction

This chapter first describes the epidemiological and clinical issues in breast density, then summarises the research for this PhD project. The project itself is in two parts, evaluation of density as a biomarker of risk in breast cancer, and researching automated methods of density estimation which can render it usable in population screening.

1.2 Review of issues in breast density

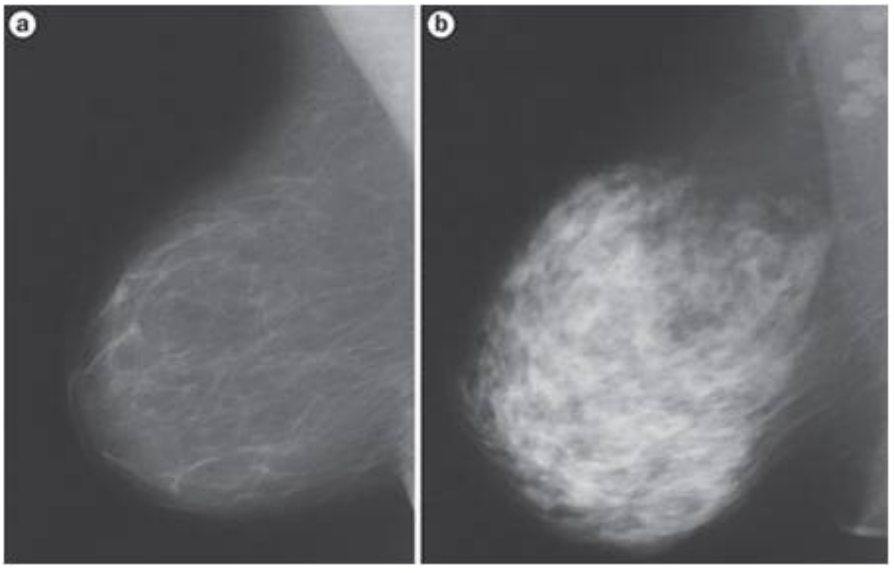
1.2.1 Introduction

Mammographic density is the portion of the breast that appears white on a mammogram (i.e. radiodense) and it identifies the amount of fibroglandular tissue, stroma and epithelium, in the breast. For a better understanding, Figure 1.1 shows mammograms from two breasts with different breast composition: (a) is primarily fatty and has low levels of density, whereas (b) has predominantly dense tissue. As early as 1976, Wolfe observed that high levels of mammographic density (henceforth referred to simply as density) were associated with increasing risk of breast cancer [5]. Since then, the mechanisms that account for this association have been investigated, and the association with risk further quantified. A recent comprehensive review and meta-analysis of the effect of density on breast cancer risk has been performed by McCormack and dos Santos Silva

[6], documenting associations of qualitative and quantitative density measures with risk. The precise mechanism for the effect on risk remains uncertain [7, 8], although recent hypotheses include increased stromal collagen [9], delayed involution [10] and increased expression of aromatase [11]. Other aspects of density have been extensively researched, and it has been found that density is heritable, alterable by weight change or exogenous hormones and has significant associations with other risk factors for breast cancer [12], including age, menopausal status, parity, heritability, exogenous hormones (hormone replacement therapy and tamoxifen) and body mass index (BMI). The influence of density on breast cancer risk persists even after adjustment for these factors [13, 14], and high amounts of fibroglandular tissue were shown to predict higher risk for up to 7-10 years [15, 16]. In addition to this, recent studies of primary chemoprevention and recurrence-prevention with tamoxifen [17–20] suggest that a reduction in density reflects a decrease in breast cancer risk.

In this chapter, the following aspects of density research are reviewed: methods of measuring density, quantification of the independent effect of density on breast cancer risk, the relationship between density and other breast cancer risk factors, the implication of density for breast cancer screening, the alterability of density at individual level, and its potential in monitoring the effects of risk reducing interventions. In addition, some suggestions for the future applications of density as a breast cancer risk biomarker have been included.

FIGURE 1.1: Mammograms of a breast with very low(a) and high density (b)



1.2.2 Measurement of mammographic density

Methods of assessing breast density have varied over the years, since in 1976 Wolfe presented his method based on *parenchymal patterns* [5], and, to date, no ideal method has been developed. Wolfe was the first to suggest a strong association between the radiographic appearance of breast tissue and the risk of breast cancer. He identified four parenchymal patterns and dichotomised these into two main risk categories: patterns N1, predominately fat, and P1, ductal prominence in less than 25% of the breast, which he suggested indicated low risk, and P2, ductal prominence in at least 25% of the breast, and DY, extensive dysplasia, which he concluded conferred higher risk. Note that DY does not necessary imply dysplasia, which is a pathological rather than radiological phenomenon but merely the presence of large uniform areas of density on a mammogram. Observing that the typical radiological appearance of the premenopausal breast contained substantial dense areas, Gram et al [21] proposed the Tabàr classification which has five patterns, based on anatomic-mammographic correlation: I scalloped contours and Cooper's ligaments, evenly scattered terminal ductal lobular units and oval lucent areas corresponding to fatty replacement; II complete fatty replacement; III retroareolar prominent duct pattern and fatty involution; IV extensive nodular and linear densities; and V homogenous fibrosis with convex contour. Tabar patterns II, III, IV and V correspond approximately to Wolfe's patterns N1, P1, P2, and DY respectively; and similarly Tabar patterns IV and V are associated with increased risk [22]. However, some mammograms which would be P2 or DY in Wolfe's classification are classified as Tabar pattern I, since this represents the typical rather dense appearance of the premenopausal breast.

The Tabàr and Wolfe classification methods are categorisations based on both qualitative and quantitative observations, but it has been noted that purely quantitative measures of breast density give better risk prediction [15, 23, 24]. Thus, over time, these semi-quantitative methods have been replaced by largely quantitative methods, with percent density most often assessed visually and divided into categories. In the 1990s the American College of Radiology proposed the Breast Imaging Reporting and Data System (*BIRADS*) lexicon that classifies density as: BIRADS-1, almost entirely fatty breast ("<25 % dense"); BIRADS-2, presence of scattered fibroglandular densities ("26-50% dense"); BIRADS-3, heterogeneously dense breast ("51-75% dense"); and BIRADS-4, extremely dense breast (">75 % dense"). Many researchers have used a 5-category classification ("<10% dense", "10% to 25% dense", "25% to 50% dense", "50% to 75% dense" and ">75% dense"), sometimes augmented to six categories by allowing zero a separate class [25, 26]. Both semi-quantitative methods and BIRADS rely on

visual assessments, and are relatively quick and easy to perform [27]. However, their subjective nature constitutes their major weakness, because intra- and inter-observer variability adversly affects the reproducibility of the results. A standardised lexicon, such as BIRADS, high-quality mammograms and good training for the raters can undoubtedly help to improve reproducibility. In a study on this issue, the inter-observer agreement for Wolfe's patterns, as measured by Kappa statistics, ranged from 0.69 to 0.88, and the intra-observer from 0.69 to 0.87, whereas using BIRADS both intra- and inter-observer agreements were as 0.89 [27].

Visually estimated percent breast density, measured on a 21-point semi-continuous scale (0, 5%, 10%,... 100%), has also been found to predict breast cancer risk [4, 17] but there is nevertheless much interest in fully automated and computer-assisted methods, firstly to render the process less labour-intensive and less subjective, and secondly to allow more comprehensive information (measurements of total breast area, dense area and non-dense area) to be obtained. Technological progress has generated several computer-assisted or automated methods, such as *planimetry* and *interactive thresholding* (*Cumulus*) [28]. The former was developed from Wolfe's expert outlining method where the edge of the breast and area of dense tissue on a mammogram were marked on an acetate sheet and quantified using planimetry [29]. For interactive thresholding, instead, an observer chooses the appropriate threshold level by manipulating an image of the mammogram on a computer screen so that the dense tissue and edge of the breast are identified, and the number of pixels in the dense and non-dense areas automatically calculated [28, 30, 31]. Percent dense area can then be calculated from absolute dense and total breast area. Hence, interactive thresholding is less subjective than visual assessment, provides more comprehensive information (measurements of total breast area, dense area and non-dense area, in addition to a continuous measure of percent dense area) and is similarly predictive of breast cancer risk [30]. Cumulus is the most widely used interactive thresholding tool and is regarded by many as the gold standard. However, this method is time-consuming, so it is not ideal for density assessment in routine population screening (taking an average of two to three minutes per film, whereas purely visual classification allows performing the assessments of Wolfe patterns, Boyd's six-category classification and continuous percent density in less than 1.5 min per mammogram [27]). There is debate as to whether absolute dense area is a more useful measure than percent breast density for independent risk prediction [16, 32, 33]. This will be discussed further in the Breast Density and Breast Cancer Risk section 1.2.3. It is also known that both percent dense area and absolute dense area tend to be lower when assessed from digital mammograms rather than films [34], in part due to the processing of the image, but also to the better delineation of the breast edge on a digital mammogram [35].

Current methods of measuring density, both visual and semi-automated, have two limitations: (1) they require significant input from the human user, which in turn implies a need for training and maintains the labour-intensive nature of density measurement; and (2) they reduce a three-dimensional feature, the breast, to its 2-dimensional projection obtained on a mammogram. Controlling the reproducibility and reliability of density measurements (inter- and intra- reader error) is a major challenge, and one to which there is currently no easy solution as even when there is a high level of agreement on average, for individual mammograms the disagreement could be as much as 30-35% (for percent dense area read visually), or more than a category (for Wolfe patterns or Boyd's six-category classification)[27].

Methods that assess the volume of dense tissue are currently being developed and will hopefully give more precise estimates of the amount of fibroglandular tissue within the breast and therefore be better predictors of breast cancer risk. With the advent of digital mammography, there is particular interest in fully automated volumetric density estimation. Several automated volumetric approaches have been proposed [36–41], including some that are undergoing commercial development, but so far the results suggest that the volumetric methods provide a similar or weaker risk prediction than the area-based measures. Most recently, Shepherd et al [42] found that the categorical risk classification for 20% of women, with or without breast cancer, could be improved by including fibroglandular volume in the risk assessment model. We shall return to this issue in section Future Applications 1.2.7.

Magnetic resonance Imaging (MRI)-based analyses has received great attention because it provides a detailed three dimensional distribution of fibroglandular tissue that is not subject to the tissue overlap problem in mammography [43]. For this reason, this measure is anticipated to be the most accurate estimate of the amount of dense tissue in the breast and several studies have compared it with density measured by mammography [43–49] showing that the MRI measure for the volume fibroglandular tissue is highly correlated to mammographic density, assessed either visually or [44–47] or with a volumetric fully-automated software [48]. Recently Wang and colleagues [50] compared three fully-automated volumetric methods based on FFDM, i.e. SXA [42], Quantra [51] and Volpara [52], to MRI breast density and found them in moderate agreement. Apparently, most cases of disagreement were driven by differences in total breast volume measured for MRI versus mammography [50]. MRI fibroglandular volume is also often used ground truth against which mammogram measurements are compared [53], especially when developing new volumetric methods.

Besides MRI fibroglandular volume, it has also been speculated that another estimates of breast density from MRI, i.e. background parenchymal enhancement, could provide

further relevant information on breast cancer risk. Background parenchymal enhancement is the volume and intensity that fibroglandular tissue "enhances" after intravenous contrast agent administration during MRI, and it appears to have an association with breast cancer risk as strong as mammographic density and which may be especially useful evaluating the effectiveness of a treatment, such as tamoxifen [49, 54]. Although background parenchymal enhancement could promote a better understanding of the relationship between the amount of fibroglandular tissue in the breast and breast cancer risk, the research in this field is still at an early stage.

Another approach adopted by several research groups is to automate 2-dimensional measurements. In Sweden Li *et al.* developed Auto Threshold (v1.10) ImageJ plug-in, aiming to mimic Cumulus readings [55]. Results on 1,498 cases and 1,495 controls indicate a reasonable agreement between fully-automated and computer-assisted thresholding. Auto Threshold is less labour- and time-consuming, and therefore could be introduced into clinical practice more easily. Following poor results from the comparisons between automated volumetric measures and visual assessment, it is believed that, besides the amount of dense tissue, other aspects of mammographic appearance may be informative of breast cancer risk [56]. Thus, in Denmark Nielsen and colleagues have developed a software that extracts textural information from all pixels of segmented breast images, and recognises texture relating to breast cancer status of women [56]. Their findings showed that their mammographic texture resemblance (MTR) marker was independent of mammographic density and more predictive of risk. Likewise Schmidt and colleagues designed CIRRUS, a fully-automated predictor of breast cancer risk, that tries to identify the features in a mammographic image that best predict breast cancer risk [57]. Heine et al. adopted a simpler but still effective approach [58]. Their work focused on the estimation of the "variation measure", i.e. the standard deviation of calibrated pixel values, that appeared to have an association with breast cancer risk at least as strong as percent density assessed with Cumulus in a case-cohort study (2,311 subjects, including 217 cases) and two case-control studies (1,967 and 762 subjects, including 928 and 246 cases, respectively) [58]. These methods are promising but require further external validation.

Electrical impedance tomography is based on the concept that every different tissue, and in particular malignant tumours, reacts differently to the same electrical pulse, and produces reconstructed tomographic 2D or 3D images of the impedance, and therefore of the tissue composition [59]. Likewise, digital breast tomosynthesis provides 3D images of breast composition by using multiple low-dose radiographic exposures taken at different angles [60]. There are high expectations that using these breast imaging methods, it should be possible to have a more accurate estimate of the amount of breast dense tissue,

than those based only on two-view 2-dimensional mammograms. Methods for density estimation from these new 3D breast imaging systems are currently being developed, however the research on this topic has just started moving its first steps [61].

1.2.3 Breast density and breast cancer risk

Increasing density is associated with an increase in risk of breast cancer regardless of the type of assessment (qualitative or quantitative)[6], the population (symptomatic or asymptomatic) [13], whether the density assessment was made on a negative mammogram years prior to the cancer diagnosis (incidence studies) or on the contralateral mammogram at the time of the breast cancer diagnosis [13]. It should be noted, however, that the effect is attenuated with increasing time since measurement [6]. Two comprehensive meta-analyses, conducted more than a decade apart, both found a strong effect of density on risk and both observed that quantitative measures gave superior risk prediction to qualitative ones [6, 62]. The RRs for Wolfe's most dense category (DY) relative to the least dense (N1) were 3.98 (95% CI (2.53-6.27)) and 2.42 (1.98-2.97), respectively in incidence and prevalence studies; for BIRADS category "extremely dense ($\geq$ 75% *dense*)", there was a risk of 4.08 (2.96 – 5.63) relative to the lowest category, "almost entirely fatty ($<$ 25% *dense*)"; whereas the strength of the association between percentage density and breast cancer risk was estimated as 4.64 (3.64-5.91) for 75% density or more compared to less than 5% [6]. Mammographic density is therefore an important risk predictor for breast cancer, although the mechanisms that account for this association are currently still unclear and under investigation [7, 8, 63]. Almost 5% of the female population has extremely dense breasts (dense tissue in 75% or more of the breast), and these women have a 4- to 6-fold increase in risk compared with those with primarily fatty breasts ($<$ 10% dense tissue) [6, 26, 64]. It should be remembered, however, that the mean percent dense area in postmenopausal women is approximately 30% so the relative risk of developing breast cancer for a woman with greater than 75% density compared to the general population is somewhat lower than that typically reported for comparisons of most and least dense.

The masking bias hypothesis was suggested by Egan and Mosteller [65] to explain the relationship between extensive fibroglandular tissue and subsequent breast cancer risk. It is based on the fact that density is associated with failure to detect some breast cancers by mammography [26, 66], since density renders the task of detecting presymptomatic tumours in the breast more difficult [67]. One supposition is that, in primarily dense

breasts, tumors may be masked because they share the same x-ray attenuation properties as fibroglandular tissue, therefore an artificially higher incidence of symptomatic breast cancer is observed in the ensuing years [26, 66, 68]. However, the increased breast cancer risk in asymptomatic women, and the fact that the risk remains increased after excluding cancers in the year following negative mammography, indicate that the effect is not entirely due to masking [6]. Indeed, mammographic density predicts breast cancer risk up to 10 years thereafter [15, 16]. However, although it is clear that the observed effect of density on risk is not completely due to masking, the phenomenon is clearly a potential biasing factor in individual studies. interpretation of results should always bear in ming the potential of masking to inflate associations with symptomatic cancers and attenuate the associations with screen -detected.

Breast density has a high attributable fraction compared with other risk factors for breast cancer [69] and may be reduced by hormonal therapy [17–20] and diet [70]. It is therefore hoped that by reducing particularly high levels of density a substantial proportion of breast cancers might be prevented. Most breast cancer risk factors account for a very small percentage of cases, either because the increased risk associated with the factor is modest or because the prevalence of the factor is very low [71]. Boyd et al [26] found that 16% of all breast cancers and 26% of breast cancers in women aged under 56 were attributable to breast densities greater than 50%. This underlines the potential role of density in breast cancer prevention [17, 72–74].

As noted above, quantitative measures of breast density have been observed to have a stronger association with risk than qualitative measures. However, this is not universally observed, so there remains the question as to whether relative or absolute breast density is a better predictor of risk [8, 16, 33]. Breast density is negatively confounded with two other risk factors, age and body mass index (BMI). Increases in all three are associated with increased risk of breast cancer, but percent breast density reduces with both increasing age and BMI [12]. This has a number of implications for risk estimation. First, the effect of percent density on risk [25], if no adjustment is made for age and BMI, will be underestimated. Secondly, it should be noted that absolute dense area does not show this confounding with BMI [75]. Vachon et al [76] found that absolute dense area had similar risk predictive potential as percent density. Stone et al [33] found similar results and in addition found that absolute area did not need adjustment for non-dense area (as a surrogate for BMI) to attain maximum predictive effect, whereas percent density did require such adjustment. However, Wong et al [75] found that percent density predicted risk better than absolute dense area, albeit in an Asian population. Thus this remains an open question.

The influence of mammographic density on breast cancer persists after adjustment for

other risk factors [14, 76, 77]. This suggests that it can be of use in further improving individual risk models [1, 73, 74, 78, 79] which in turn will aid in identifying high risk groups who might benefit from specific preventive or surveillance interventions. See Future Applications 1.2.7 for further discussion of this.

1.2.3.1 Breast density and breast cancer outcome

Despite being a strong predictor of breast cancers, higher levels of density do not necessarily imply a worse prognosis, in terms of death or recurrence. Firstly, it was speculated that an extensive amount of breast density would be associated with more aggressive tumour characteristics [80], perhaps also because of masking-bias. However, results of a recent meta-analysis [81] contradicted this hypothesis and concluded that subjects with denser breasts are more likely to develop breast cancer, regardless of tumour sub-type. Secondly, two studies [82, 83] comprising women who experienced ductal carcinoma *in situ* , observed an association between breast density and subsequent invasive breast cancer in the contralateral, but not ipsilateral breast, suggesting that dense tissue play a role in carcinogenesis but not in the further development of the tumour. Finally, a Danish study [84] involving 48,052 women, aged 50-69 years, screened between 1991 and 2001, reported that breast tumours were more frequent among subjects with denser breasts, but they were, overall, less severe. The proportion of fatalities was lower than in women with fattier breasts. Gierach and colleagues had similar findings from a cohort of 9,232 women diagnosed with primary invasive breast carcinoma between 1996 and 2005 [85]. As in the Danish study, the data indicated a lack of association between breast density and breast cancer mortality, suggesting that following carcinogenesis, other factors determine how the tumour develops.

1.2.4 Breast density and other breast cancer risk factors

Mammographic density is a representation of the fibroglandular component of breast-tissue, and, although mammograms do not allow the distinction, the dense area comprises both stroma, the connective fibrous tissue, and epithelium, the two layers of epithelial cells covering it. The complex interactions between these two features may provide the biological basis for tumorigenesis [7]: although breast cancer tumours arise mostly from ductal epithelial cells, stroma also plays an important role in the origin and

development of breast cancer tumours, contributing both instructive and permissive signals, and represents, therefore, a potential target for cancer-preventing interventions [86, 87].

1.2.4.1 Age, parity and menopause

Breast density changes throughout a woman's adult lifetime; the proportion of fibroglandular tissue tends to increase until the 3rd decade after menarche (30-40 year old), then declines progressively with increasing age [88]. Further, more pronounced reductions occur subsequent to each pregnancy and following the menopause, conferring on average reductions of 2% and 2.4% in percentage density [89–91]. This suggests a strong connection between mammographic density and the cumulative exposure to hormones and growth factors that stimulate cell division in breast stroma and epithelium, suspected to represent an important factor in the age-specific incidence of breast cancer in the population [92].

1.2.4.2 Confounding with body mass index

The interpretation of mammographic density as a marker of breast cancer risk, particularly when density is measured in relative terms, is complicated by the fact that density is confounded with BMI. A woman with high BMI tends to have larger breasts than a woman with low BMI so that 50% dense area might represent very different absolute amounts of dense breast tissue. Similarly a low amount of relative density may still be associated with a high risk, because of an extensive dense area, surrounded by an even larger area of non-dense tissue. Further, when expressing density with percent scores, its association with breast cancer risk appears weaker in women with larger breasts [93]. Body mass index is an alterable feature. A woman might reduce her BMI with diet and exercise, for example, and thereby reduce her breast cancer risk, yet if her dense breast area remains the same (and her breast size decreases) losing weight will mean that her estimate of risk, based on percent dense area alone, would be increased. A recent review [94] focusing on mammographic density and physical activity, found no evidence of an association between them, although results may be influenced by the heterogeneity of the studies included, in terms of population, definition of physical activity and how this was reported.

In general, classic anthropometric measures of adiposity (body mass index and waist circumference) have been shown to be as good as adiposity measures from whole body dual x-ray absorptiometry and computed tomography for adjusting mammographic density

for this confounding[95].

1.2.4.3 Confounding with age

As already noted, mammographic density is negatively confounded with age, so that density decreases and the risk of developing breast cancer increases with increasing age. This apparent anomaly was first explained by Boyd et al [92] who found that the decline in percent dense area with age corresponds to the cumulative rate of breast tissue aging described in Pike's model [96] and depends more on hormonal history (for example, age at menarche, age at first birth and age at menopause) than chronological age. McCormack et al [91] further showed that there is a high degree of tracking post age 50 (i.e. that a woman's percent density ranking, at age 50, relative to other women in her cohort, will be the same at age 60) so that a single assessment of a woman's mammographic density at age 50 or more might be viewed as a reasonable long-term measure of her breast cancer risk [91]. As Lokate and colleagues [97] observed in their Dutch cohort, cases and controls experienced similar age-related declines in mammographic density, however at all mammographic examinations, density was higher on average for cases than for controls.

1.2.4.4 Exogenous hormones

The hypothesis of a strong association between mammographic density and exogenous sex hormones is supported by evidence from a number of studies. A major meta-analysis [98] showed that oral contraceptives cause a significant increase in breast cancer incidence among current users, relative risk 1.24 (1.15-1.33), and that this effect decreases with time after stopping use, i.e. it is null after 10 or more years, RR 1.01 $(0.96 - 1.05)$. However, findings relating hormonal contraceptive use and percent mammographic density are generally negative [76, 99]. In contrast combined hormone replacement therapy (HRT) with oestrogen and progesterone was found to increase both risk of breast cancer [100–102] and mammographic density [101–105]. In particular oestrogen/progestin combination treatments confer a significant mean increase in density ranging from 3% to 5% [103, 104, 106]. Sala et al [99] found that while current users of HRT had significantly increased density compared to never users, former users did not, which suggests that exogenous oestrogen has only a short term effect on density. More recently, in a multi-ethnic study [102] Hou and colleagues observed that the effect of HRT on breast

cancer risk was modified by mammographic density, BMI and ethnicity. In particular they observed that underweight/normal weight women with dense breasts were found to be more sensitive to the detrimental effects of HRT on breast cancer risk, whereas for black women and overweight/obese women with fatty breasts, HRT use did not affect their risk of breast cancer.

Treatment with tamoxifen, a selective oestrogen receptor modulator (SERM), causes a greater density reduction (7.9% (6.9%-8.9%)) than placebo treatments (3.5% (2.7%-4.3%)), within 18 months [72], and this change seems to reflect a reduction of risk of breast cancer in a high-risk population [17–20]. It is not yet known whether the same will be found with other prospective breast cancer prevention therapies. A number of small studies have looked at the effect of raloxifene, another SERM, on mammographic density and reported no significant change in density with treatment [107, 108] or a change of a similar magnitude to that observed in women on placebo [109]. None are conclusive, however, as they suffer from a number of weaknesses, including a lack of statistical power. Studies of aromatase inibitors have had negative or inconclusive results. Two small studies of letrozole in women who had previously undergone 5 years of tamoxifen treatment [110] and in women with at least 25% mammographic density (including women with previously treated breast cancer) [111] found no reduction in density but again are too small to be considered conclusive. A randomised study of mammographic breast density and exemestane similarly found no significant change in percent density with treatment [111]. More recently Vachon *et al.* [112] investigated the changes in breast composition due to aromatase inhibitors comparing 369 early-stage postmenopausal patients and 369 matched controls. They found no evidence of significant change in percent density due to aromatase inhibitor therapy. However it has been speculated that breast magnetic resonance imaging (MRI) could be more appropriate than mammography examination for recording changes in breast dense tissue due to endocrine treatment [107].

1.2.4.5 Endogenous hormones

Measurements of hormones and growth factors have shown that blood concentrations of insulin-like growth factor IGF-I and prolactin, respectively in pre- and post-menopausal women, are associated with mammographic density, both before and after adjustment for other risk factors [106]. In particular, IGF-I is a known mitogenic agent that strongly influences epithelial cells of the breast, and is thought to be responsible for a proportion of breast tumours [113]. It is also hypothesized that pre-menopausal levels of IGF-I may

be related to risk of post-menopausal breast-cancer [113].

Measurement of hormone levels in ovulating women is complicated by the fact that levels depend heavily on when in the cycle the blood is taken. Nevertheless studies aimed at testing the hypothesis that, even in postmenopausal women, breast density reflects circulating sex hormone levels found no association after adjustment for BMI [114–116]. Additional support for the hypothesis that mammographic density and post-menopausal endogenous oestrogen levels are independent lies in the lack of association between density of the surrounding breast-tissues and the oestrogen receptor status of a breast cancer [80, 117, 118]. High mammographic density has been observed to increase the risk of both ER-positive and ER-negative breast cancers [81], suggesting that mammographic density may influence breast cancer risk through oestrogen-independent pathways [80]. It may also be that endogenous hormone levels are best measured in the breast tissue itself rather than in the blood because the concentrations of oestrogens in the tumours of postmenopausal women are significantly higher than that found in plasma [119]. Furthermore, it has been shown in mice that in situ aromatisation (oestrogen production within the breast itself through the action of the enzyme aromatase) is the key determinant of tumour estradiol levels and tumour growth [120]. Recently Varghese and colleagues [121] investigated the relationship between mammographic density and hormone levels on 1,286 women to verify whether the increased risk of breast cancer caused by endogenous hormones through cell proliferation may be reflected in mammographic density [122]. They observed no association between density, both in absolute and percent terms, and estradiol or testosterone levels, after adjusting for BMI.

1.2.4.6 Heritability

Genetic and family history factors are long-established risk factors for breast cancer [123]. Likewise heritability appears to account for a large proportion of mammographic density variation, with the odds of a woman having dense breasts (BIRADS category 3 or 4) being 17% (9%-25%) greater for women with a first-degree relative who has had breast cancer, 19% (10%-29%) greater if the affected relative is a sister, and 22% (10%-35%) greater if the relative was diagnosed with breast cancer before 50 years of age [124].

Epidemiologic risk factors associated with mammographic density, e.g. menopausal status, age, body mass index (BMI), and number of live births, account for only 20-30% of variation in the trait [12]; on the other hand the contribution of genetic factors, to the observed variability in density, has been estimated as between 29% [125] and

60% [123, 126]. Supporting the hypothesis that genetic factors influence mammographic density, twin studies demonstrated that the correlation of breast dense tissue between monozygotic twins was approximately twice as strong as that between dizygotic twins, 0.63 versus 0.27 [123]. Most of the studies suggest considering also an interaction between genes and environment [123, 127, 128].

Special interest in the genetics lying behind mammographic density arises from the fact that its strong association with breast cancer risk and its high heritability meet the criteria for considering it as a potential intermediate phenotype of breast cancer. It is believed that genes responsible for familial correlation in mammographic density could also affect the incidence of breast cancer in the population and contribute partially to the familial aggregation of the disease [123]. Therefore several studies have been conducted to identify these genes [129–133]. Candidate gene approaches indicated the presence of an association of both COMT Val158Met and IGF1 rs6220 A<G with premenopausal mammographic density, and of ESR1 (Xbal and Pvull) polymorphisms with post-menopausal mammographic density [134].The first large genome-wide linkage study [127] concluded that there may be at least one gene influencing mammographic density on chromosome 5p. Further recent findings from Dumas and Diorio [135] indicated that some single nucleotide polymorphisms (SNPs) located in the CYP181, COMT or HSD1781 genes are associated with mammographic density among nulliparous women, non-hormone users, those with late menarche or women with high BMI, hence modifying effects of estrogen-related factors should be considered when evaluating associations of polymorphisms in estrogen-related genes with premenopausal mammographic density.

Furthermore Tamimi and colleagues [136] conducted a cross-sectional study to investigate the potential association between mammographic density and the 11 SNPs identified by a multistage genome-wide association study (GWAS) [129], but results were inconclusive. However, in 2011 results from another GWAS led by Lindström and colleagues [131] reported an association between percent mammographic density and rs10995190 in ZNF365, also known for being associated with susceptibility to breast cancer. In addition to this, findings from a three-stage GWAS provided additional evidence that common genetic variation contributes to breast tissue composition and identified a novel percent mammographic density locus at 12q24, that they suggested should be the focus of further studies [132]. Recently, Varghese and colleagues [133] investigated the potential shared genetic basis for mammographic density and breast cancer risk, using two GWAS: one to identify a set of SNPs that showed the strongest association with mammographic density, and one to examine their association with breast cancer risk. Their results confirmed Lindstrom findings regarding rs10995190 [131], and then suggested another SNP (rs10509168) also in the ZNF365 gene. Varghese *et al.* also evaluated the

hypothesis of a shared genetic basis between mammographic density and sex-hormone levels [121] and discarded it, suggesting to consider the two features as two independent sets of traits. The same study also showed that dense and nondense areas of the breasts were partly influenced by the same genetic factors, although in opposite directions [63]. There is uncertainty as to whether subjects with known BRCA mutations have higher density than non-carriers or the general population [137–141].

1.2.4.7 Benign breast disease

It is well established that women who develop benign breast disease (BBD) are at increased risk of developing breast cancer, with the magnitude of the increased risk depending on the degree of epithelial atypia [142, 143] and being particularly high for women age 45 or less with atypical hyperplasia [144, 145]. Furthermore, Boyd *et al.* [146] showed that having high density ($> 75\%$) increased the risk of having hyperplasia without atypia (RR=13.85), and the risk of having atypical hyperplasia or lobular carcinoma in situ (RR=9.23). Thus, the two features may not be independent factors. Further investigations by Byrne *et al.* [142], however, suggest that benign breast disease histology and mammographic density are independent breast cancer risk factors. Tice and colleagues addressed this hypothesis on a cohort of 42,818 women with at least one benign breast biopsy. They found that women with atypical hyperplasia and very high breast density had the highest risk for breast cancer, however women with almost entirely fatty breasts were at low risk for future breast cancer regardless of the histology of their breast biopsy [147]. This suggests that there may be synergy between atypia and density in their effect on subsequent breast cancer risk [147].

1.2.5 Breast density and screening

Mammographic density, appearing as opaque areas on a mammogram, can mask some radiological features of the breast and thus render it more difficult to detect tumours [70]. As noted in 1.2.3 above, because density influences the detection of cancer, estimates of the risk of breast cancer associated with mammographic density may be distorted [26]. It has been found that interval cancers occurring within a short period after screening are more common in women with dense breasts, in particular Boyd and colleagues [26] observed an increased OR of 17.8 in women with over 75% of breast density, compared to subjects with primarily fatty (less than 10% dense) breasts in the first 12 months

after a negative screening, therefore, implying that breast density is associated with a shorter potential lead time in mammography screening [148]. Nevertheless, in the same study [26] the association between density and breast cancer risk remained significant, though weaker, with an OR of 5.7 (2.1-15.5), for cancers detected more than a year after a negative screening, suggesting that this issue of masking can only partially explain such an association.

The radiation dose required to penetrate fibroglandular tissue and to compensate for less effective compression (i.e. greater thickness) is greater in dense breasts, so that cumulative lifetime exposure to radiation due to mammography screening, is also greater in women with dense breasts [149]. Although this may suggest that mammographic screening could have a less favourable benefit-harm ratio for women with dense breasts, particularly younger women, it has been estimated that this exposure has a lifetime attributable risk of fatal breast cancer of 1.3 per 100,000 women aged 40 at exposure and <1 case per 1,000,000 women aged 80 at exposure, whereas 292 lives would be saved as a result of annual screening [150]. Recent results from a major study of adjunct ultrasound screening in women with mammographically dense breasts suggest that with this additional screening modality the interval breast cancer rate in women with at least 50% density could be brought down to a similar level as that currently seen in women with less than 50% density [151]. Similarly, early results from a non-randomised study of automated whole breast ultrasound and mammography in asymptomatic women with BIRADS density 3 or 4 suggest that the breast cancer detection rate could be doubled for women in these categories with the use of this technology.

As noted previously, MRI is considerably more effective than mammography for the surveillance of women with high-risk BRCA mutations [152–154], even in BRCA carriers with low mammographic density [155], thus mammographic density does not entirely explain the difference in performance between these two screening modalities. It should be remembered that surveillance in such women tends to begin relatively early in adult life when the breast tissue is denser, which may explain why there is conflicting evidence as to whether mammographic density is generally greater in BRCA carriers than in non carriers or the general population [137–140].

MRI is resource intensive, and demands a greater commitment from the screenee than does mammography. Hand-held ultrasound is labour-intensive and subjective, but there is now an automated whole-breast ultrasound system (AWBUS) [156–158] that promises to improve sensitivity, although increasing the numbers of false positives [150, 157, 159]. Hence screening modalities which may mitigate the effect of breast density on sensitivity and which are more appropriate for high-throughput population screening are digital mammography and automated ultrasound [150, 156, 158, 160]. Personalised screening

strategies, according to the level of breast density and other risk factors such as family-history, have the potential to improve cost-effectiveness of the breast cancer screening programme.

1.2.6 Breast density and models for predicting breast cancer risk

Currently, there are a number of risk evaluation models available for assessing a woman's breast cancer risk based on her risk factors but none have yet been validated in the general population or have yet incorporated mammographic density into their risk assessment. The majority have been developed and validated in high-risk subgroups of women. In a study comparing the performance of the Gail [161, 162], Tyrer-Cuzick [1], Claus [163–165] and Ford [166] models at predicting breast cancer risk in women with a family history of breast cancer who were attending the Family History Evaluation and Screening Programme in Manchester, UK, Amir *et al.* [167] found that the Gail, Claus and Ford models all significantly underestimated risk and that the Tyrer-Cuzick model was consistently the better performer, with the ratio of expected to observed numbers of breast cancers .48 for the Gail model, .56 for Claus, .49 for Ford and .81 for Tyrer-Cuzick. As regards ability to predict individual cases there was far less difference among the models, with the area under the ROC curve being .735 for Gail, .716 for Claus, .737 for Ford and .762 for Tyrer-Cuzick. Thus, it seems that the Tyrer-Cuzick model, incorporating extensive family history information, endogenous estrogen exposure and benign breast disease, is the best at predicting the average level of risk of a population, and therefore the overall number of cases, but has only a slight advantage as regards individual discrimination and prediction. Other studies have also found that the Gail model [161] has low discriminatory power (c-statistic = .58) [13] but to some extent this is due to the fact that the age-specific breast cancer rates on which it is based require updating [168]. The Tyrer-Cuzick model significantly overpredicts breast cancer risk in women with atypical hyperplasia [169] whereas the Gail model significantly underpredicts [170]. Both models could therefore be improved by addressing known weaknesses and adding information on new risk factors, such as genotype and mammographic density.

Previous attempts [2, 3, 78, 171] to introduce mammographic density into risk prediction models have led to only modest improvements in predictive ability. In 2005, Tice and colleagues[3, 171] built a model containing only breast density adjusted for age and ethnicity, which performed as well as the Gail model. Similarly, in 2006 Barlow et al. [2] used a cohort of 1,007,600 women to develop two models, one for pre-menopausal and

one for post-menopausal women, that include both traditional risk factors and breast density. In both cases, breast density, as a single predictor, appeared to be as powerful as age, however only a moderate further predictive ability was obtained by adding this feature to the other factors. During the same year Chen *et al.* [78] proposed a model based on age at first live birth, number of biopsies, number of affected relatives, mammographic density and body weight, as an alternative to the Gail model but the improvement in discriminatory power, assessed with the c-statistic, was modest (Gail model: .602 vs Chen model: .664) [162]. It should be noted, however, that in a multivariate risk model for breast cancer, most individual risk factors typically account for only a modest proportion of the predictive power.

These results need validation on independent datasets, and may lead to the conclusion that major determinants of breast cancer risk remain to be discovered [2]. Nevertheless measures of mammographic density are likely to make at least a modest contribution. Given also that all these models used BIRADS [2, 3, 78, 171] as the measure of density (i.e. a measure with 4 or 5 categories), and it is possible that a more detailed measure may be more suitable for this context. It may be that continuous measures would lead to a better prediction. It is also worth noting that there are limitations regarding the c-statistic, or area under the ROC curve (AUC), as a measure for capturing discrimination and, therefore, for assessing the usefulness of markers [172, 173].

These issues concerning the predictive power of risk evaluation models are important to both clinical practice and the current controversy regarding whether or not a woman should be informed of her mammographic density. Current recommendations in the US from the American Society of Clinical Oncologists are that if a woman has a 5 year breast cancer risk of greater than 1.66% by the Gail model then she should be considered high-risk and offered a risk-reducing intervention such as tamoxifen or raloxifene [174]. However action advocacy groups, in the USA, are pushing for legislation, which would mandate that women attending screening should be informed regarding their breast density; thirteen states have already adopted this policy [63]. Similarly, in the UK, the NICE guidelines [175] recommend a threshold of 8% for the 10 year breast cancer risk calculated from the Tyrer-Cuzick or a similar model when determining whether a risk-reducing intervention should be offered. Because mammographic density has not yet been incorporated into these models, it is still possible for a woman to have high mammographic density and not satisfy the above criteria. Thus, a woman who is informed of her high mammographic density might have her fears raised but be denied appropriate counselling and risk-reducing interventions. It is a matter of some urgency, therefore, that major studies to investigate the complex interrelationship between mammographic density, established breast cancer risk factors and risk of breast cancer, are

carried out, as the results of these will inform the existing risk evaluation models and enable them to be further developed. It is of particular importance that we learn how best to incorporate mammographic density into existing or new models so that the process of assessing risk becomes more comprehensive and coherent. Work on this is already underway. The PROCAS (Predicting Risk Of breast Cancer At Screening) study [176] in the Greater Manchester area of the UK is currently collecting comprehensive information on women's breast cancer risk factors (including mammographic density and SNPs) through the NHS Breast Screening Programme and has already enrolled about 52,000 women (http://www.uhsm.nhs.uk/research/Pages/PROCASstudy.aspx). The data collected, together with subsequent follow-up, will provide a valuable resource for investigating the feasibility and acceptability of assessing breast cancer risk within a population-based mass screening programme and for the further development and testing of the statistical models used in the evaluation of individual breast cancer risk. Specifically, it is intended that the Tyrer-Cuzick model will be updated on the basis of this study to include two new risk factors, mammographic density and a genetic risk score based on a number of SNPs. Including information from mammographic density and SNPs in risk prediction models is the most topical challenge for breast cancer risk estimation [79]. A similar study, KARMA (http://www.karmastudien.se/) in Sweden was designed to update the Gail model. So far Darabi *et al.* recorded a limited but significant improvement in terms of AUC, including adding percent density, BMI and a collection of SNPs to the Gail model. It has been suggested that AUC may not be the most appropriate tool to assess the predictive ability gained adding new information to the model [173]. However, the current lack of standardisation in breast density assessment is the more important issue to be addressed.

1.2.7 Future applications

The key drawbacks of the established area-based methods of measuring breast density are that they are subjective (visual assessment, and to a lesser extent computer assisted) and labour intensive (computer assisted). The intra- and inter-reader error (respectively, the difference between repeat readings of a single mammogram by the same reader and the difference between repeat readings of a single mammogram by different readers) can be as much as 30% and 35%, respectively, for visual assessment [27]. The labour-intensive nature of density assessment, particularly if measured using interactive threshold methods, and the subjective decision-making involved in visual assessment, render it difficult to incorporate into mass screening. For example, it would clearly add an unsustainable

burden to the UK National Breast Screening programme which screens approximately two million women per year [177]. However, the advent of digital mammography means that both these issues may soon be resolved as there are a number of fully-automated methods of assessing breast density from digital mammography currently being developed, ranging from bespoke image analysis software designed specifically to work with particular mammography machines (Quantra [51, 178], Volpara [52, 179]) to stand-alone generic packages which accept any digital images (Image J), which will assess and record breast density as the mammogram is being taken.

The potential this technological progress will release is considerable, as major population-based studies will now become more feasible, and new volumetric and absolute measures of mammographic density will be readily available, adding little or no labour or time to population screening. If the automated methods predict risk as effectively as visual assessment and interactive thresholding, then their use in risk-management could be incorporated into population screening. Further work is required, however, to validate these new methods of density assessment for it is not yet known whether they will work as well as the current methods at determining breast cancer risk.

It should also be noted that the modifiability of breast density may be a major factor in future applications. As noted above, density has been shown to be modifiable by hormonal drugs and by changes in body habitus. The fact that in high-risk women treated with tamoxifen, only those whose density was reduced after administration showed a subsequent reduction in breast cancer risk suggests that density can be used not only to select subjects at risk as candidates for preventive interventions, but also to monitor the effectiveness of such interventions on an individual basis. [17]

1.2.8 Conclusions

Technological advances suggest that the breast density research community is on the brink of major progress. Nevertheless, there is currently no fully validated objective means of assessing mammographic density (see section on Measurement of mammographic density 1.2.2) and continuing uncertainty as to how to utilise mammographic density in clinical practice (see sections on interpretation of density and risk 1.2.3 and future applications 1.2.7). Without this knowledge the current demand in the US for the introduction of Breast Density Inform Law (for example, Public Act No. 09-41 passed in the State of Connecticut in 2009) may be premature. Whilst breast density has clearly been shown to be a powerful factor for predicting the risk of developing breast cancer, its potential role in assessing hormonal preventive regimes and helping to tailor screening

algorithms cannot be fully realised until we have density measures which both predict risk with accuracy and can be used in a population screening context.

1.3 Aims and objectives

Research plan

1.3.1 Part I: Mammographic density as a biomarker for breast cancer risk

Mammographic density is recognised as one of the strongest independent risk predictors for breast cancer risk, but the reasons accounting for this association are not yet fully understood. Thus it appears to have great potential as a biomarker for risk of this disease, and great interest lies in investigating its properties.

In this part we analyse studies that have recorded information on density and other risk factors for breast cancer. It should be underlined that the datasets presented in these chapters were already existing as part of previous larger studies, and no further data was collected during the PhD project. However these datasets had not been fully analysed and exploited for what concerns density and its association to breast cancer risk. In order to clarify the novelty element of this thesis, Table 1.1 compares the published results from these studies and the issues that will be addressed in this work. Hence these data will allow us to investigate the relationship of mammographic density with breast cancer risk, and with other risk factors. This in turn will clarify the potential of density as a tool in breast cancer risk management.

The studies involved in the analyses are here described in brief.

TABLE 1.1: Main published findings of the studies analysed in part I and new issues addressed in this thesis

Study	Published Results	Unreported issues for analyses in this thesis	Data
IBIS-I	Prophylactic tamoxifen reduces the risk of breast cancer by about a third. [180] The risk-reducing effect of tamoxifen persisted for at least 10 years, but most side effects of tamoxifen do not continue after the 5-year treatment period. [181] The 12- to 18-month change in mammographic breast density is an excellent predictor of response to tamoxifen in the preventive setting. [17]	Alternative quantification of the extent of surrogacy	Nested case-control study N=1065 (123 cases)
ATAC	Anastrozole showed both short-term and long term superior efficacy than Tamoxifen in both improving survival and lessening severe side effects. [182, 183]	Analyses of the relationship of density and its changes with risk of recurrence and first tumour characteristics	Nested dataset N=601 (132 cases)
FH01	Annual mammography in women aged 40-49 years at enhanced familial risk is both clinically effective in reducing breast cancer mortality and cost-effective. [184]	Assessment of the strength of the association between density and risk in this high-risk population, with and without adjustment for other risk factors and Tyrer-Cuzick risk estimate	Nested case-control study N=298 (103 cases)

1.3.1.1 Potential of visually assessed density as a marker of chemopreventive effect: a case-control study within a breast cancer chemoprevention trial (IBIS-I), Chapter 2

Data The International Breast cancer Intervention Study (IBIS-I)[17] is 10-year-long prevention trial, April 1992 to March 2001, that involved 7,152 high-risk subjects from the UK (60%), Australia or New Zealand (37%) and the rest of Europe (3%), and investigated the effectiveness of therapy with Tamoxifen, an oestrogen antagonist, in preventing breast cancer. Women, aged 30-70 years, were considered at high risk if they had risk factors for breast cancer indicating at least a two-fold relative risk at ages 45-70 years, a four-fold relative risk at ages 40-44 years or a roughly ten-fold relative risk at ages 35-39 years. After recruitment one half of the study population was randomly assigned to a 5-year treatment with Tamoxifen and the other to placebo.
After the main trial results had been reported [181], 123 women who had been diagnosed with breast cancer and 942 controls were selected and their data analysed to examine the relationship between changes in breast cancer risk and mammographic density due to Tamoxifen [17]. The study found that those women treated with tamoxifen where density decreased most were also the population whose wisk was most reduced. However, traditional measurement of the extent to which the Prentice criterion for surrogacy was satisfied were modest [17, 185]. It has been noted in the past that it would be difficult to find surrogate endpoint which satisfies the Prentice criterion perfectly or nearly perfectly [186].
This nested case-control data will therefore be used for the following analysis.

Proposed analysis

- To further examine the recent published result [17] that those whose density decreases as a result of treatment, also have corresponding reduced risk.
- To assess an alternative criterion for surrogate endpoints [185] in order to further clarify the extent to which reductions in density can be considered as a valid indicator of the success of the treatment.

1.3.1.2 Correlates of density with prognostic markers and outcome in a trial of treatment of breast cancer (ATAC), Chapter 3

Data The Arimidex (anastrozole), Tamoxifen, Alone or in Combination (ATAC) study [187] was an international double-blind clinical trial aimed to find a therapy successful in preventing breast cancer recurrence in post-menopausal women with hormone receptor-positive tumours. Between July 1996 and 2000 the study enrolled 9,366 postmenopausal women with localized breast cancer, and randomised them to receive either 5 years of adjuvant treatment with anastrozole alone, tamoxifen alone or a combination of the two. The dataset for analysis will comprise 601 women from ATAC, 208 treated with anastrozole, 198 with Tamoxifen, and 195 with a combination of the two, including 22% of cases of recurrence, in each arm respectively 42, 47 and 43.

Proposed analysis

- To analyse the relationship between mammographic density and risk of recurrence of breast cancer.
- To investigate the interactions of density, treatment and other risk factors.
- To study how density correlates with biomarkers and tumour histochemistry (e.g. her2 etc.).
- To replicate findings from IBIS-I, that size of density reduction is related to treatment efficacy, here measured by time to recurrence, not prevention like in IBIS-I.

1.3.1.3 Additional risk information from breast density in a population at enhanced familial risk of breast cancer (FH01), Chapter 4

Data The Family History (FH01) study [188] was designed to evaluate the effectiveness of annual mammographic surveillance in young women (50 years old or less) with a family history of breast cancer. In the UK 6,710 subjects were enrolled because of their "moderate risk" of developing breast cancer, i.e. their family history was such that they had a substantially increased of risk of developing breast cancer, but generally not sufficiently strong to suspect a BRCA mutation. These women received annual invitation to screening and were observed for a minimum of 5 years.
Although mammographic density is well established as a risk factor, its interplay with other risk factors is less clear. In particular, it is not known whether it adds further information on risk in a population already known to be at high risk.
At the end of the intervention, a nested case-control study was performed, including 103 cancer cases and 195 matched controls, which investigated how mammographic density affected risk of developing breast cancer and interacted with familial factors.
The case-control study will be used for the following analysis.

Proposed analysis

- To estimate the additional effect of density on risk, in a population already known to be at enhanced risk of breast cancer.
- To investigate the association of density and breast cancer risk, adjusted and unadjusted for other factors in this higher risk population.
- To compare percent and absolute measures of density.
- To study how the relationship between density and breast cancer risk varies according to menopausal status.

1.3.2 Part II: Automated volumetric assessments of mammographic density (Quantra and Volpara)

Fully-automated techniques are expected to be the future of measuring mammographic density, since they will hopefully provide predictive ability as regards breast cancer risk and quick assessments, would be possible to implement on a population scale and would

ensure reproducibility of the results. In addition to this, while repeatability and reproducibility are important in measurement of density, as with measurement of any other attribute, the crucial feature is the ability to predict breast cancer risk. Therefore a number of technologies are currently being developed or undergoing validation.
Quantra [51, 178] and Volpara [52, 179] are two of such methods in need of evaluation, before being recommended in routine screening. In this part of the thesis data from Princess Grace Hospital, a private facility in central London, are analysed to investigate the properties of the two new techniques, comparing their predictive ability. In order to clarify the novelty element of this thesis, Table 1.2 compares the published results from these studies and the issues that will be addressed in this work.
The studies involved in the analyses are here described in brief.

TABLE 1.2: Main published findings of the studies analysed in part II and new issues addressed in this thesis

Chapter	Published Results	Unreported issues for analyses in this thesis	Data
Chapter 5	-	Analyses of the association between breast cancer risk and Quantra absolute and relative density measures, also in comparison to BIRADS visual assessments. Evaluation of the impact of age and area of residence on such association.	Case-control study N=400 (200 cases)
Chapter 6	-	Assessment of changes in breast composition as estimated by Quantra, according to age and area of residence.	Longitudinal dataset N=332 (no cases)
Chapter 7	-	Analyses of the association between breast cancer risk and Volpara absolute and relative density measures, also in comparison to Quantra and BIRADS. Evaluation of the impact of age and area of residence on such association.	Case-control study N=366 (182 cases)
Chapter 8: 1. IBIS-I	*see Table 1.1*	Evaluation of intra- and inter-reader agreement in Cumulus and visual (21 agreement.	*see Table 1.1*
2. Chapters 5 and 6	-	Assessment of potential measurement error in Quantra density estimates and its impact on their association with breast cancer risk. Evaluating the relationship between volumetric and area density assessments.	*see above*

1.3.2.1 A case-control evaluation of a fully automated volumetric density measure as a predictor of breast cancer risk (1) (Quantra), Chapter 5

Data As previously noted, to be practically useful in the context of population screening and risk management, automatic measurement of density, with little or no call on manpower, is desirable. In this study details were recorded of 200 cases, female patients with histopathologically verified breast cancer, and 200 matched controls. For every subject, mammographic density was assessed both with Quantra volumetric measures of absolute and relative density and visually with the BIRADS classification. The available covariates also included age, area of residence and the Quantra measure of total breast volume.

Proposed analysis

- To compare Quantra measures of absolute and relative volumetric density as regards their ability to discriminate between cases and controls.
- To examine the relationship between mammographic density, as assessed with Quantra, and the established BIRADS method.
- To assess how the association between density estimates and breast cancer risk varies according to age and area of residence.

1.3.2.2 Serial volumetric density measures using Quantra, Chapter 6

Data This dataset recorded details of 332 women, (231 younger than 50 years and 101 aged 50 or more years) undergoing two mammographic examinations, typically 18 months apart. Density was assessed with Quantra and we also had information on age and area of residence.

Proposed analysis

- To assess changes in breast composition (i.e. amount of dense and fat tissue) over time, for the combined study group and each age cohort separately.

1.3.2.3 A case-control evaluation of a fully automated volumetric density measure as a predictor of breast cancer risk (2) (Volpara), Chapter 7

Data The mammograms of 366 women (182 cases and 184 controls) from the Quantra case-control study, were reassessed using Volpara[52, 179], another fully-automated volumetric measure.

Proposed analysis

- To examine the relationship between the mammographic density measures produced with Volpara and the established BIRADS method.
- To compare Volpara measures of absolute and relative volumetric density as regards their ability to discriminate between cases and controls.
- To assess how the association between density estimates and breast cancer risk varies according to age and area of residence.
- To compare directly Volpara and Quantra assessments.

1.3.2.4 Measuring mammographic density: results, issues and potential implications, Chapter 8

This chapter includes analyses on data described in Chapters 2, 4, 5 and 6 and it focuses on variability from measurement error and population variation in breast composition.

Proposed analysis

- To evaluate inter- and intra-reader agreement in Cumulus [28] and visual (21-categories) assessments using data from the IBIS-I study [17], described in Chapter 2.

- To investigate how measurement error in density assessed with a fully-automated volumetric method (Quantra) may influence the association between density and breast cancer risk, observed in Chapter 5.
- To adjust the association between risk and Quantra absolute density estimates (Chapter 5) for the potential measurement error, using data introduced in Chapter 6 to estimate variability components.
- To study the relationship between two- and three-dimensional density assessments, and their difference in variability, using data from the longitudinal study presented in Chapter 6 and CADET1 study [4].
- To compare the propensity for risk prediction of most of the density measures available for this project (visual 21-category assessment from Chapter 2, Cumulus from Chapters 2 and 4, Quantra from Chapter 5 and Volpara from Chapter 7) using standardised odds ratios and the areas under the ROC curve (AUCs).

1.3.3 Implications for practice and for future research

As noted above, the aim of this project is to clarify the role of breast density as a marker of risk, and to indicate its scope in:

- identification of high-risk individuals for preventive or surveillance interventions,
- monitoring of interventionss to reduce risk at individual level,
- risk and surveillance: management in the context of population screening.

The concluding chapter draws together implications put in terms of what can be implemented in health care now, and which issues remains to be researched.

Part I

Mammographic density as a biomarker for breast cancer risk

Chapter 2

Potential of visually assessed density as a marker of chemopreventive effect: a case-control study within a breast cancer chemoprevention trial (IBIS-I study)

2.1 Introduction

Studies focusing on cancer incidence as an endpoint within cancer prevention and treatment trials require large numbers of subjects, follow-up over many years, and, consequently, imply great expense. Surrogate markers are measures that can be used as early indicators of treatment effect because of their ability to predict the outcome of the therapy on this endpoint. They are therefore very attractive for study designers because of the opportunity they provide to make studies smaller, faster and less expensive. A surrogate (S) is expected to mediate the effect of the therapy or exposure (Z) on the true endpoint (T), in other words S should lie on the causal pathway connecting Z to T. Given the complexity of carcinogenesis, finding an appropriate surrogate marker is not an easy task [185, 186, 189].

The appearance of breast tissue on a mammogram varies between subjects due to differences in breast composition, leading to different levels of risk of developing breast cancer. In particular, a larger amount of white (dense) areas, i.e. the projection of stroma and epithelium or fibroglandular tissue, corresponds to higher breast cancer risk [6]. Decades of research on this topic indicated that this association holds regardless of the tumour-type [81]. Nevertheless the understanding of the biology underlying the association between breast cancer risk and mammographic density is still limited [190]. Current hypotheses suggest increased stromal collagen, delayed involution and increased expression of aromatase [63]. One of the most interesting properties of breast density as a risk factor for breast cancer is that it can be altered and is particularly responsive to hormonal manipulation. Exposure to hormone replacement therapy, normally progesterone and oestrogen combined, leads to increases in both breast cancer risk and mammographic density, although possibly through independent pathways [101, 103, 191]. Tamoxifen therapy, on the other hand, is associated with a decline in mammographic density as well as in risk [17–19]. Tamoxifen influences breast cancer risk by blocking oestrogen stimulation on breast cancer cells, inhibiting both translocation and nuclear binding of the oestrogen receptor [192]. Recent publications [17–19] showed that subjects who experienced a considerable reduction ($\geq$10%) in mammographic density, were more responsive to tamoxifen treatment. This was observed both in a high-risk population in a cancer-prevention study (IBIS-I) [17] and in the context of reducing risk of recurrence or death [18, 19]. These findings strengthen the hypotheses that changes in mammographic density may be a good indicator for the outcome of therapy with tamoxifen.

Previously published findings from the IBIS -I study [17, 181] suggested that: (1) tamoxifen (Z) is associated with reduction in breast cancer risk (T), (2) tamoxifen is associated with reduction in mammographic density (S), and (3) the reduction in mammographic density is associated with reduction in breast cancer risk. Hence reduction in mammographic density satisfies three of the four conditions necessary for a surrogate end-point [185, 193]. The fourth requires that

$$f(T|S,Z) = f(T|S) \tag{2.1}$$

where $f(T|S,Z)$ and $f(T|S)$ are the conditional probability distribution of T. In simpler terms this equation stipulated that the effect of tamoxifen on density should be sufficient information to infer its effect on breast cancer risk.

The analyses reported in this chapter are based on data from the case-control study nested in IBIS-I [17] and are aimed at verifying whether this condition holds too, i.e. to investigate the extent to which mammographic density mediates the effect of tamoxifen

therapy on breast cancer risk, and thus evaluate the potential of mammographic density changes as a predictor of the outcome of the treatment.

2.2 Materials and Methods

2.2.1 Study Setting and Population

The IBIS-I chemoprevention study is an international trial designed to evaluate the effect of tamoxifen on breast cancer risk in high-risk women. It recruited a total of 7,152 women aged 30-70 [180]. A nested case-control study was conducted subsequently to investigate relationships between treatment, change in mammographic density and known breast cancer risk factors. It comprised 123 cases from the UK and Finland and 942 British controls, and its results were published in 2011 by Cuzick and colleagues [17].
Detailed breast cancer risk factors were collected at study entry. Of the women in our subset 507 were treated with tamoxifen and 558 received a placebo treatment, with random allocation in the treatment arms [180, 181]. Mammographic density was assessed visually by one radiologist (R.M.L. Warren) who viewed the film mammograms and visually estimated the proportion of total breast area that comprised dense tissue (to the nearest 5%). Baseline films were taken at or up to a year before randomisation, and follow-up films were taken at least 12 months after randomization, with a median time between randomization and first follow-up mammogram of 18 months. The assessment of mammographic density for both cases and controls was based on a composite assessment of both the left and right mediolateral oblique views, with the exception of 13 case subjects who were diagnosed at the first follow-up mammogram. For these women the density assessment was made using only the film from the contralateral breast [17].
In this study we only focused on age, body mass index and percent breast density at baseline as risk factors for breast cancer. Change in breast density between the two mammograms was classified in four categories as follows: “Increase”, “No change” “Reduction 5-10%” and “Reduction $\geq$10%”.
Further details on the data collection are provided in Cuzick and colleagues’ paper [17].

2.2.2 Prentice's criteria and how to quantify the effect of a surrogate marker

Prentice's necessary and sufficient condition for a surrogate endpoint [185] is as follows: a test of the null hypothesis of no relationship of the surrogate to treatment is also a valid test of no relationship of the true endpoint to treatment.
If T is the clinical outcome or true endpoint, S is the surrogate endpoint, Z is the treatment indicator and $f(S|Z)$, $f(S)$, $f(T|Z)$ and $f(T)$ are conditional and unconditional distribution of S and T, this can be expressed as: $f(T|S,Z) = f(T|S)$ where T and Z are associated to each other and to the surrogate S. Further, Qu distinguished between surrogate marker and surrogate endpoint, as the former is drug-specific, whereas the latter is a surrogate for any efficacious treatment [193]. Hence in our case it is more appropriate to use the term "surrogate marker", as we tested it only in relation to tamoxifen treatment.
Freedman and colleagues [189] proposed the *proportion of treatment effect* (PTE) explained by a surrogate marker to quantify the extent to which Prentice's criterion is satisfied. PTE is a function of the ratio of the estimated treatment effects, i.e. regression coefficients γ_Z and β_Z, respectively from a model including both treatment (Z) and surrogate (S) and one including only treatment (Z) as predictor variable.

$$PTE = 1 - \frac{\gamma_Z}{\beta_Z}. \tag{2.2}$$

In more familiar terms, γ_Z is the regression coefficient of the treatment on the true endpoint adjusted for the surrogate, and β_Z is the unadjusted coefficient. The main drawbacks of this index are that it is unbounded and that results, in particular the variability of γ_Z, can easily be affected by a strong collinearity between treatment and surrogate marker. Because the treatment is generally effective in improving the surrogate marker, a strong correlation between treatment and the surrogate marker might be expected [194]. In this situation the coefficient estimates of the regression model including both S and Z may change erratically in response to small changes in the model or the data [195]; thus making PTE a less reliable estimate of the validity of S as a surrogate marker, as intended by Prentice [185].
More recently Qu and Case proposed another method, the *proportion of information gain* (PIG), to evaluate a surrogate marker [193, 194]. PIG is based on the likelihood function and is defined as

$$PIG = LRT(S:1)/LRT(S,Z:1) \tag{2.3}$$

where LRT(S:1) is the likelihood ratio test statistic comparing the model including intercept and marker S as explanatory variables for outcome and the model only including the intercept. Likewise LRT(S,Z:1) is the statistic referring to the model including both surrogate (S) and treatment (Z), along with the intercept, in comparison to the model based on the intercept only. Simulations showed that PIG overcomes some of the disadvantages of PTE [193, 194]. The major conceptual difference between PTE and PIG is that the former compares the effect of treatment on the true endpoint, adjusted and unadjusted for the surrogate, whereas the latter compares the effect of the surrogate on the true endpoint, adjusted and unadjusted for the treatment.
Alonso *et al.* [196] also developed a method to assess the validity of a potential surrogate marker according to the Prentice's criteria, the *likelihood reduction factor* (LRF). This method uses the same models necessary to compute PTE and LRF is defined as

$$LRF(Z, S : Z) = 1 - exp\{-LRT(Z, S : Z)/n\} \tag{2.4}$$

where LRT$(Z, S : Z)$ is the likelihood ratio test (LRT) based on the models were the true endpoint is explained by the treatment effect adjusted and unadjusted by the surrogate, and n is the number of observations. This measure, like PIG, is based on likelihood function and, as Qu and Case pointed out [194], there is a close relationship between LRF$_a$, i.e. the adjusted LRF, bounded by [0,1], and PIG,that can be approximated as follows: $1 - LRF_A(S, Z : S) \approx PIG$.

2.2.3 Statistical Analyses

The distributions of demographic and other variables at baseline are summarised as percentages or mean and SDs, as appropriate, both overall and stratified by treatment arm.
Some results on this dataset have been previously published [17]. These suggested change in percent density after tamoxifen treatment as a potential predictor of the outcome of the therapy. Subjects experiencing a reduction in density of at least 10% appeared more responsive to the therapy in terms of risk reduction.
As noted above, results already published from IBI-I suggest that changes in density reflect changes in risk. Our aim in this chapter was therefore to assess the extent of density's potential as a surrogate for the outcome ofbreast cancer, using alternate formal measures of this extent. In particular, we assessed the validity of reduction in density as a surrogate marker for risk reduction using both PTE and PIG methods. Alonso's LRF

was considered redundant and not computed because of its close relationship to the PIG measure. The results of this work apply to an enhanced risk population rather than the general population. This is appropriate since it is specifically in such a enhanced risk population that tamoxifen can be used as primary prevention.

Logistic regression models were used to assess the relationship between risk of breast cancer and both treatment with tamoxifen and reduction in mammographic density. We used weights to reflect the sampling fraction, since, by study design, the two treatment arms have different sample size. In particular we weighted the subjects with the inverse of the probability that the observation is included because of the sampling design, therefore subjects in the placebo arm were weighted 532.5/558 and those in the tamoxifen arm 532.5/507.

All analyses were performed using Stata, version 12.1.

2.3 Results

Baseline characteristics of the 1,065 subjects, overall and by treatment group, are summarised in Table 2.1. More than half of the subjects in both arms had on average received treatment for at least 5 years. Other characteristics such as age and BMI at entry in the study were well balanced between the two arms.

The proportion of cases in the two arms are similar, because of the study design, since the placebo arm here included 558 cases and controls, whereas the tamoxifen arm included only 507 subjects, i.e. larger numbers of unaffected subjects were sampled for the placebo arm. Hence sampling weights were used in the logistic regression. However the protective effect of tamoxifen against breast cancer has been widely recognised and it was especially evident in the long-term prospective results from the whole IBIS-I study [180, 181].

Baseline density levels were similar but differed post-treatment as the two therapies affected mammographic density in different ways. Almost half of the women treated with tamoxifen (46%) experienced a reduction in the percent density of at least 10%, whereas only a quarter of the subjects in the placebo arm (25%) showed such a change.

TABLE 2.1: Characteristics of the study sample

	Overall N=1065		Tamoxifen N=507		Placebo N=558	
Number of cases (%)	123	(11.6)	51	(10.1)	72	(12.9)
Time on treatment						
Mean (SD)	4.9	(0.7)	4.9	(0.7)	4.9	(0.7)
Age in years at baseline						
Mean (SD)	50.2	(6.2)	50.2	(6.3)	50.1	(6.1)
Body Mass Index at baseline						
Mean (SD)	26.7	(4.8)	26.7	(4.9)	26.7	(4.8)
% breast density at baseline, No.(%)						
0 %	114	(10.7)	58	(11.4)	56	(10.0)
1-10 %	108	(10.1)	47	(9.3)	61	(10.9)
11-25 %	132	(12.4)	62	(12.2)	70	(12.5)
26-50 %	252	(23.7)	129	(25.4)	123	(22.0)
51-75 %	240	(22.5)	109	(21.5)	131	(23.5)
76-100 %	219	(20.6)	102	(20.1)	117	(21.0)
Mean(SD)	44.5	(30.2)	44.1	(30.0)	44.8	(30.4)
Change in breast density categories after 12 months, No. (%)						
Increase	86	(8.1)	20	(3.9)	66	(11.8)
No change	394	(37.0)	161	(31.8)	233	(41.8)
Reduction 5-10%	213	(20.0)	94	(18.5)	119	(21.3)
Reduction $\geq$10%	372	(34.9)	232	(45.8)	140	(25.1)

It is clear from Table 2.2 that the numbers and the percentages of events varied according to the change in density that the patients experienced, therefore this is a strong risk factor. Likewise tamoxifen reduced the number of breast cancer events. Table 2.3 shows the estimates required for computing both Freedman's and Qu's methods [189, 193, 194] for surrogate markers.

We fitted the models with and without treatment (Z) and our potential surrogate marker (S), Table 2.3. Freedman's PTE was then obtained as:

$$PTE = 1 - \frac{-0.18}{-0.28} = .37. \tag{2.5}$$

TABLE 2.2: Change in density and breast cancer incidence according to randomised treatment group (P=placebo, T=tamoxifen)

Change in density after 12-month treatment	Nr. patients at risk/events P	Nr. patients at risk/events T	Percent of events P	Percent of events T
Increase	66/9	20/4	13.6	20.0
no change	233/27	161/20	11.6	12.4
Reduction >5%	119/21	94/12	17.6	12.8
Reduction ≥10%	140/15	232/15	10.7	6.5

TABLE 2.3: Intermediate endpoint validation analysis

Model	β_z	SE(β_Z)	LR (χ^2)	df
$ln(p/1-p) = \beta_0 + \beta_Z Z + \beta_S S$	-.18	.21	9.85	4
$ln(p/1-p) = \beta_0 + \beta_Z Z$	-.28	.19	2.09	1
$ln(p/1-p) = \beta_0 + \beta_S S$	-	-	8.77	3
$ln(p/1-p) = \beta_0$	-	-	0	0

Qu's PIG instead is computed as:

$$PIG = \frac{LRT(S:1)}{LRT(S,Z:1)} * 100 = \frac{8.77}{9.85} * 100 = 89\%. \tag{2.6}$$

This low PTE may be due to collinearity between S and Z. This hypothesis is supported by the significant association between treatment arm and density reduction observed in Table 2.2 (χ^2 test: p<.01). Qu and Case observed similar disagreement between PTE and PIG in their example based on the MORE (Multiple Outcome of Raloxifene Evaluation) study [194]. This issue will be further discussed in the next section.

2.4 Discussion

Reduction in mammographic density is a potential surrogate marker in trials with tamoxifen, since there is a very high *proportion of information gain* (PIG) [193, 194], 89%. Conversely, Freedman's *proportion of treatment effect* (PTE) explained by a surrogate marker [189] appeared modest, (.37). A similar disagreement between the two methods was observed also by Qu and Case [194] in their example using data from the MORE study [197]. They felt this was an illustration of the limitations of PTE: this measure is unbounded and susceptible to collinearity between treatment and surrogate

marker. As a matter of fact the standard error for PTE in their simulation was quite high, especially in comparison with the other measures. This is supporting evidence of the collinearity existent between the surrogate and the treatment, as it would make parameter estimates, in the regression model including both factors, more susceptible to small changes in the data [195]. Another disadvantage of Freedman's PTE is that it does not consider whether the coefficient of the treatment remain significant after the surrogate marker is introduced to the model. It would be difficult to observe a PTE equal to 1 as it would require that the adjustment for the surrogate marker nullifies the coefficient of the treatment. This rarely happens but the significance of the effect of the treatment may be more strongly influenced, in other words: even if the introduction of the surrogate marker decreases only slightly the coefficient of the treatment, the latter may no longer be significant.

PIG on the other hand addresses significance , and therefore consistency, of effects rather than their size, being based on likelihood ratio test statistics of the models, which also makes it robust to collinearity. This method is based on the concept that the treatment will not contribute much information to the clinical outcome, once the surrogate markers are available, if these are truly effective surrogates [194]. The disadvantage of this method is that there is no established cut-off point for PIG above which the surrogate marker is deemed to be valid. However 89% appears high enough to consider reduction of at least 10% in mammographic density as a good predictor of the efficacy of treatment with tamoxifen.

There is also the difference noted that PTE and PIG are not measuring exactly the same phenomenon. PTE measures the attenuation of the magnitude of the treatment effect by adjusting for the surrogate; PIG measures the additional information provided by the treatment effect in addition to that provided by the surrogate. Also, as noted above, the PIG is more related to consistency rather than magnitude of effect.

Density as a surrogate may not apply to treatment with other therapies though, as different treatments may act through different causal pathways. Risk-reduction strategies using raloxifene, aspirin or aromatase inhibitors have not led to larger changes in mammographic density compared to the placebo arms [107, 111, 112, 198–200], despite being more successful in reducing risk of breast cancer. The fibroglandular tissue may be related to potential oestrogen binding sites in the breast but not to the amount of bioavailable estradiol, so that mammographic density is not an appropriate surrogate marker for the effect of aromatase inhibitors in reducing the risk of breast cancer [200]. It is also possible that mammographic appearance is not the right tool for assessing changes in breast composition due to aromatase inhibitors and thus alternative techniques, such as MRI or ultrasound tomography may need to be evaluated [200]. To

support this theory, Eng-Wong and colleagues [107] assessed the change in breast density in high-risk premenopausal women from both mammograms and breast magnetic resonance imaging (MRI) and observed that the breast MRI volume (MRIV) declined significantly after use of raloxifene, whereas mammographic density did not. It should be observed that because of the limited size of the sample (N=27) this result would require validation on a larger dataset. Cigler and colleagues [111] found that letrozole does not significantly affect breast density, however it should be noted that their subjects were previously treated with tamoxifen and therefore this may have already achieved maximal density-lowering effects. Vachon *et al.* deduced similar conclusions from their study on the impact of letrozole in women completing 5 years of tamoxifen [110]. A more recent study [112] enrolled 369 early-stage breast-cancer postmenopausal patients and 269 matched healthy controls to study the efficacy of aromatase inhibitors, anastrazole and exemestane. In this case as well, no therapy induced reduction in density was reported, other than the natural decline observed also in the controls. Raloxifene is a selective oestrogen receptor modular (SERM) with similar mechanisms to tamoxifen, but a review on raloxifene trials [198] found no treatment-induced change in density, in contrast to the IBIS-I study [17]. This could be due to a difference in the population: IBIS-I included healthy pre- and post-menopausal women at high risk of developing breast cancer whereas the studies on raloxifene included in the review were performed mostly on healthy post-menopausal women not at high risk. Hence subjects in these studies had lower baseline breast density, which may have made reductions in breast density less evident [198].

Even though change in mammographic density appears promising as a surrogate for tamoxifen treatment, it is not ready to be used as such, because the assessment of mammographic density has yet to be standardised [28, 201]. Currently the most popular density classification method is BIRADS, which uses four wide categories and therefore may not be sensitive enough to reflect change in density at the required level of detail. The IBIS-I case-control study [17] used visually assessed density in 21 categories, and found that a reduction in density of at least 10% was a good indicator for a successful outcome of the treatment. Kim and colleagues [18] observed similar results in preventing recurrence using interactive thresholding to estimate density. In Sweden Li *et al.* assessed mammographic density, at baseline and follow-up, using an automated thresholding method [55], and in their data a reduction in dense area of at least 20% corresponded to a decrease in mortality of 50%, in comparison to little or no change. Finally in a recent Korean study [20] mammographic density reduction was defined dichotomously comparing baseline and follow-up mammograms classified using BIRADS categories. In this study patients who showed a reduction in breast density according to

BI-RADS classification after an average of 19 months of adjuvant tamoxifen treatment had a 65% lower risk of recurrence than patients who did not show such reduction. Other density assessment methods, e.g. volumetric methods such as Quantra or Volpara, may need different cut-off points in density reduction. A standardisation of breast density assessment with general accepatibility to patients and practitioners is necessary for its use in risk-prediction of breast cancer in routine mammography screening. Likewise the use of breast density reduction as a surrogate marker is limited without standardised measurements.

In conclusion, this study found that mammographic density reduction could be a valid surrogate for a positive outcome with tamoxifen treatment in high-risk women. However, it is not clear whether it could be also used for other therapies for reducing risk of breast cancer. Alternative imaging methods, for example MRI or tomosonography, may be better than mammography in detecting the relevant changes in breast composition. In any case, mammographic density reduction requires an established standardised method of measurement in order to be used as a surrogate. The potential for automated methods to fill this role is further investigated in the second part of this thesis.

Chapter 3

Mammographic density, risk of breast cancer recurrence and characteristics of the tumour (ATAC study)

3.1 Introduction

Despite advances in therapy a substantial number of breast cancer patients remain at risk for late recurrences. In particular almost one in five breast cancer survivors after 5 years of adjuvant therapy suffers a recurrence within 10 years after treatment [202]. Being able to identify subjects more likely to develop a recurrence would be crucial to design focused and more effective disease control strategies. Mammographic density, i.e. the amount of radiodense tissue in a mammogram, is a strong and independent risk-predictor for breast cancer, but it is unclear if it could be helpful in a prognostic setting. A British study, including 759 screened women with an invasive breast cancer [203], reported that mammographic density and prognosis, subsequent to diagnosis and treatment, were unrelated. Later results on a Danish cohort [84], by contrast, indicated that breast tumours arising within mixed or dense breasts were less aggressive than tumours detected in primarily fatty breasts. Finally, in 2009 a study comprising 335 women, who underwent breast-conserving surgery for invasive breast cancer [204], suggested that mammographic density could provide a good indicator of risk of local breast cancer recurrence. However, these results were limited to subjects who did not receive

breast irradiation.

The ATAC trial [182, 183] was designed to evaluate anastrazole and tamoxifen as adjuvant treatments for prevention of breast cancer recurrence. Therapy with anastrazole, and other aromatase inhibitors, has not been observed to have a significant effect on density [112, 205], whereas tamoxifen is known to affect the breast composition of the subjects, substantially decreasing density [17–20]. Recent evidence suggests that reduction in density of at least 10%, following tamoxifen therapy, in unaffected women at moderate to high-risk of breast cancer, might lead to a risk reduction of 63% [17]. Likewise Kim and colleagues [18] suggested that patients, with ER-positive breast cancer, who did not experience a reduction of at least 10% in density after a year of tamoxifen therapy, were twice as likely to have a recurrence than those who did experience such a reduction. Similar results were reported on a cohort of Swedish post-menopausal breast cancer patients [19] and in Korea on a group of premenopausal women with ER-positive breast cancers [20]. Hence, reduction in the amount of fibroglandular tissue could provide an indicator of the effectiveness of the treatment [17–20]. This background suggests two hypotheses: that density might be employed as a risk predictor also for recurrence; and that targeting density reduction could be a successful strategy for dynamic therapy decisions to prevent breast cancer recurrences.

Breast cancer is a heterogenous disease; not only have different pathological subtypes of breast cancer distinct clinico-morphological features and require different treatment strategies, but they are also associated with distinct risk factors, suggesting an etiologic heterogeneity [80, 206]. For this reason the relationship between density and breast tumour subtypes has been subject of several investigations that have reported contradicting results [80, 207–211]. A recent meta-analysis [81] concluded that mammographic density should be recognised as an important and strong marker of overall and of subtype specific risk.

The analyses reported here are on a subset of subjects from the ATAC trial, thus postmenopausal women with histologically proven operable invasive breast cancer and randomly assigned to treatment with anastrazole, tamozifen or a combination of the two, after completing primary surgery, chemotherapy and, often, radiotherapy. The data included details on tumour characteristics, as well as information on the cell proliferation marker Ki-67. Mammographic density was assessed retrospectively on baseline and subsequent yearly mammograms.

This gave us the opportunity to investigate: the role of mammographic density in association to risk of recurrence, both as an indicator of higher risk and as a target for more effective treatment; and how mammographic density and its changes interact with the tumour characteristics, including (Ki-67), both overall and within the treatment arms.

3.2 Materials and Methods

3.2.1 Study setting and population

The Arimidex (anastrazole), Tamoxifen, Alone or in Combination (ATAC) trial has been described in detail previously [182, 183]. In brief it was designed to compare the efficacy and safety of anastrazole and tamoxifen to prevent breast cancer recurrence, during a 5-year treatment and a 10-year follow-up. The trial enrolled 9,366 postmenopausal women with localised breast cancer, and randomly assigned (1:1:1) them to one of the three treatment arms, anastrazole, tamoxifen and a combination of the two.

This chapter presents analyses on a subset of these women, that comprises 601 subjects from centres in England and Wales, including 132 that experienced breast cancer recurrences (20-24% in each arm). The dataset recorded information about age and Body Mass Index (BMI) at entry in the study. Characteristics of the first tumour, such as histological grade, tumour size, nodal and ER status, were collected through CRF pathological reports from each centre. Moreover, information on HER2 (human epidermal growth factor receptor 2) status and Ki-67, a cellular marker for proliferation, was available for a subset of the subjects. HER2 status was determined using the Dako HercepTest (k5207; Dako Cytomation, Carpinteria, CA) followed by the Vysis fluorescence in situ hybridization (FISH; Downers Grove, IL) test for tumors [212]. Ki67 was scored as the percentage of positively stained cells among 1,000 malignant cells [213].

Data included details regarding time to recurrence, to death or exit from the study. Please note that in these analyses "recurrence events" refers to local and distant recurrences, but also death before recurrence if deemed a consequence of the primary breast tumour.

Mammographic density was assessed visually (to the nearest 5%) using the MLO view on the contralateral breast from scanned film mammograms taken at baseline and ideally yearly thereafter. Our analyses focused on density at baseline and its changes after one, two and five years of treatment.

After the initial analysis at 33 months of follow-up [182], the combination group was stopped because no benefit compared with tamoxifen alone was seen, in terms of either efficacy or tolerability, and follow-up data were not subsequently collected. However, the mammograms were collected retrospectively, after randomising on the sample firstly enrolled in ATAC, and blinded to the treatment arm. Thus we had details on density

also for women in the combination arm even after its termination.

3.2.2 Statistical Analyses

The distributions of demographic and other variables at baseline are summarised as percentages or mean and SDs, as appropriate, both overall and stratified by treatment arm.

Percent density distribution was presented using both categories, following Boyd's classification ("0 %", "1-10 %", "11 - 25%", "26-50%", "50-75%" and ">75%") [28], and mean and SDs, again both overall and stratified by treatment arm. Likewise changes in density after 12, 24 and 60 months were summarised with mean and SDs, highlighting the distribution of changes after two years of treatment categorised as follow: "$> -5\%$", "$-5 - -10\%$", "$\leq -10\%$" and "Unknown".

Associations between risk of breast cancer recurrence and risk factors, such as age and body mass index (BMI) at baseline and the first breast tumour characteristics were evaluated using univariate logistic regression models. Univariate and multivariable logistic regression models were run to analyse the association between mammographic density at baseline and its changes after 12, 24 and 60 months and risk of breast cancer recurrence. These analyses were repeated after adjusting for the available risk factors (age and BMI at baseline and first breast tumour characteristics). Multivariable Cox regression models evaluated how mammographic density and its changes over time affected the time free from recurrences.

We illustrated the relationship between the characteristics of the first breast tumour and mammographic density at baseline and its changes over time, both overall and according to treatment arm, using means, SDs, and box-plots. Significance testing was done using non-parametric test for trend [214] or Wilcoxon rank-sum test, for dichotomous variables. Repeated measures of density at baseline and after 1, 2 and 5 years were also analysed graphically and with linear regression models over time, adjusting for age and BMI, in order to investigate potential different trends according to tumour characteristics. The regression models took into account the potential correlation of repeated measures on the same subjects.

Results from the two tests on HER2 status were defined as "Positive" or "Negative". This was available only for a limited number of subjects (N=268), none of whom were in the combination group.

The cell proliferation marker Ki-67 was first log-transformed, because of the approximate

lognormal distribution of the data [215], then scatter plots and Pearson's correlation coefficients allowed an evaluation of the relationship between the quantitative Ki-67 level and mammographic density and its changes over time, both overall and according to treatment arm. Note that, similarly to HER2, there were no data on Ki-67 for the subjects in the combination group so only the anastrazole and tamoxifen arm are compared.

3.3 Results

As noted above, this study comprised 601 post-menopausal women, diagnosed with breast cancer and enrolled in the ATAC trial from centres around England and Wales. Median follow-up for this analysis was 102 months (range 3-127), including a total of 4650 women-years (1640 women-years for anastrazole, 1557 for tamoxifen and 1453 for the combination arm).
A similar number of recurrences (20-24 %) occurred in these subsets of the three study arms though they differed in time to recurrence (Table 3.1). As reported previously [183] disease-free survival was significantly better in the anastrazole group compared to the tamoxifen arm. The distributions of baseline age and BMI appeared similar across the groups (Kruscal-Wallis non-parametric ANOVA p-values: .69 and .31 respectively), whereas mammographic density was similar in the tamoxifen and anastrazole arms, but appeared higher in the combined group. However this difference was not statistically significant, according to ANOVA analysis (p=.27). Likewise for the change in density, the combination group, two and five years in the study, showed a higher, although not significantly higher, reduction, probably due to a regression to the mean phenomenon related to the higher initial level of density.

The analyses of the association between the risk of breast cancer recurrence and the available risk factors and first tumour characteristics revealed that age, histological grade, nodal status and tumour size were strongly and positively associated with risk of recurrence (Table 3.2). However in the multivariable logistic regression model, only age at diagnosis and tumour size remained significant. In other words women diagnosed with a larger tumour (>2 cm) and who entered the ATAC study at an older age were especially prone to recurrence.

TABLE 3.1: Characteristics of the study samples in the three treatment arm

	Treatment Arm							
	Anastrazole (N=208)		Tamoxifen (N=198)		Combined (N=195)		Total (N=601)	
Number of recurrences (%)	42	(20.2)	47	(23.7)	43	(22.1)	132	(22.0)
*Time at recurrences, No.(%)**								
0-12 months	0	(0.0)	3	(6.4)	1	(2.3)	4	(3.0)
12-24 months	3	(7.1)	7	(14.9)	3	(7.0)	13	(9.8)
24-60 months	17	(40.5)	12	(25.5)	18	(41.9)	47	(35.6)
>60 months	22	(52.4)	25	(53.2)	21	(48.8)	68	(51.5)
Age in years at baseline								
Mean (SD)	63.2	(7.7)	63.3	(8.4)	63.9	(8.7)	63.5	(8.2)
Body Mass Index in kg/m^2 at baseline								
Mean (SD)	27.3	(5.0)	27.1	(4.6)	26.7	(4.8)	27.0	(4.8)
% breast density at baseline, No.(%)								
0%	17	(8.2)	15	(7.6)	8	(4.1)	40	(6.7)
1-10%	37	(17.8)	26	(13.1)	31	(15.9)	94	(15.6)
11-25%	44	(21.2)	53	(26.8)	44	(22.6)	141	(23.5)
26-50%	59	(28.4)	52	(26.3)	48	(24.6)	159	(26.5)
51-75%	39	(18.8)	46	(23.2)	53	(27.2)	138	(23.0)
76-100%	12	(5.8)	6	(3.0)	11	(5.6)	29	(4.8)
Mean(SD)	33.6	(24.9)	34.2	(24.6)	37.4	(25.3)	35.0	(25.0)
Median (IQR)	30	(40)	30	(40)	35	(45)	30	(40)
Change in breast density, mean (SD)								
after 12 months (N=464)	-3.8	(10.0)	-3.3	(10.7)	-3.6	(11.6)	-3.6	(10.7)
after 24 months (N=525)	-4.9	(10.0)	-4.3	(11.3)	-5.5	(12.4)	-4.9	(11.2)
after 60 months (N=506)	-6.8	(11.6)	-7.1	(12.9)	-8.5	(14.2)	-7.4	(12.8)
Change in breast density categories after 24 months, No. (%)								
$> -5\%$	88	(42.3)	85	(42.9)	75	(38.5)	248	(41.3)
$-5 - -10\%$	48	(23.1)	34	(17.2)	38	(19.5)	120	(20.0)
$\leq -10\%$	49	(23.6)	52	(26.3)	56	(28.7)	157	(26.1)
Unknown	23	(11.1)	27	(13.6)	26	(13.3)	76	(12.6)

Note: (*) percent values refer to the amount of recurrences

TABLE 3.2: Odds ratios (ORs) for the risk of recurrence of breast cancer from univariate logistic models

Variable	case subjects (N=132)		controls subjects (N=469)		OR	Univariate [95% Conf.	Interval]
Age at baseline in y, No. (%)							
≤59	33	(36.3)	185	(39.5)	1.00	*(referent)*	
60-69	51	(4.1)	190	(4.5)	1.50	.93	2.44
≥70	48	(23.6)	94	(2.0)	2.86	1.72	4.76
P_{trend}						<.01	
Body mass index in kg/m^2, No. (%)							
≤23	23	(17.4)	83	(17.7)	1.00	*(referent)*	
24-25	19	(14.4)	83	(17.7)	.83	.42	1.63
26-30	52	(39.4)	178	(38.0)	1.05	.60	1.84
>30	26	(19.7)	101	(21.5)	.93	.49	1.75
Unknown	12	(9.1)	24	(5.1)		-	
P_{trend}						.93	
Histological grade, No. (%)							
(1) Well	24	(18.2)	113	(24.1)	1.00	*(referent)*	
(2) Moderate	66	(5.0)	233	(49.7)	1.33	.79	2.24
(3) Poor/Undiff	36	(27.3)	93	(19.8)	1.82	1.02	3.27
Unknown	6	(4.5)	30	(6.4)		-	
P_{trend}						.04	
Nodal status, No. (%)							
Negative	67	(5.8)	324	(69.1)	1.00	*(referent)*	
Positive	58	(43.9)	125	(26.7)	2.24	1.49	3.37
Unknown	7	(5.3)	20	(4.3)		-	
P_{trend}						<.01	
Tumour size, No. (%)							
≤2cm	71	(53.8)	328	(69.9)	1.00	*(referent)*	
>2cm	61	(46.2)	140	(29.9)	2.01	1.36	2.99
Unknown	0	(0.0)	1	(0.2)		-	
P_{trend}						<.01	
ER status, No. (%)							
Positive	94	(71.2)	361	(77.0)	1.00	*(referent)*	
Negative	21	(15.9)	48	(1.2)	1.68	.96	2.94
Unknown	17	(12.9)	60	(12.8)	1.09	.61	1.95
P_{trend}						.40	

Note: Subjects classified as "Unknown" were excluded when testing for trend.

TABLE 3.3: ORs for risk of breast cancer recurrence from univariate and multivariable logistic regression models using mammographic density and its change over time

Variable	category	obs	OR	[95% Conf.	Interval]	P>z
% breast density at baseline	per 10%	601	.96	.89	1.04	.349
% breast density at baseline	per 10%	464	.95	.87	1.05	.314
12-month change in density	per 10% reduction		.94	.76	1.17	.599
% breast density at baseline	per 10%	525	.97	.89	1.07	.576
24-month change in density	per 10% reduction		.91	.73	1.12	.370
% breast density at baseline	per 10%	506	.96	.87	1.07	.490
60-month change in density	per 10% reduction		.97	.78	1.20	.754

In univariate and multivariate analyses (Table 3.3), mammographic density and its changes were not significantly related to the risk of breast cancer recurrence nor to the probability of remain recurrence free for the whole follow-up.

TABLE 3.4: ORs for risk of breast cancer recurrence from multivariable logistic regression models using mammographic density and its change over time adjusted for other risk-factors*

Variable	category	obs	OR	[95% Conf.	Interval]	P>z
% breast density at baseline	per 10%	531	.98	.89	1.08	.627
% breast density at baseline	per 10%	398	.96	.85	1.07	.443
12-month change in density	per 10% reduction		.95	.73	1.23	.689
% breast density at baseline	per 10%	466	.96	.86	1.08	.542
24-month change in density	per 10% reduction		1.05	.82	1.34	.714
% breast density at baseline	per 10%	447	.94	.82	1.08	.374
60-month change in density	per 10% reduction		1.11	.87	1.42	.408

Note: (*) age and BMI at baseline, tumour size, hystological grade, nodal and ER status

After adjusting for age, BMI and first breast tumour characteristics (Table 3.4), the results were not materially different.

We repeated the multivariable logistic regression analyses, adjusted for the other risk factors and tumour characteristics, in the three treatment arms separately (Table 3.5). Once again mammographic density and its changes over time appeared ineffective in

TABLE 3.5: ORs for risk of breast cancer recurrence from multivariable logistic regression models using mammographic density and its change over time stratified by treatment arms adjusted for other risk-factors*

Variable	category	obs	OR	[95% Conf.	Interval]	P>z
Anastrazole						
% breast density at baseline	per 10%	183	.88	.73	1.05	.163
% breast density at baseline	per 10%	134	.87	.68	1.12	.286
12-month change in density	per 10% reduction		1.08	.58	1.98	.814
% breast density at baseline	per 10%	164	.84	.66	1.05	.126
24-month change in density	per 10% reduction		1.12	.66	1.91	.623
% breast density at baseline	per 10%	167	.83	.64	1.07	.149
60-month change in density	per 10% reduction		1.11	.67	1.85	.675
Tamoxifen						
% breast density at baseline	per 10%	175	1.05	.89	1.25	.558
% breast density at baseline	per 10%	130	.94	.77	1.15	.564
12-month change in density	per 10% reduction		1.07	.69	1.65	.774
% breast density at baseline	per 10%	151	1.14	.93	1.39	.224
24-month change in density	per 10% reduction		.97	.62	1.50	.881
% breast density at baseline	per 10%	140	1.04	.82	1.33	.738
60-month change in density	per 10% reduction		1.20	.79	1.84	.394
Combination						
% breast density at baseline	per 10%	173	.97	.82	1.15	.734
% breast density at baseline	per 10%	134	.97	.80	1.17	.727
12-month change in density	per 10% reduction		.75	.48	1.20	.230
% breast density at baseline	per 10%	151	.91	.74	1.11	.340
24-month change in density	per 10% reduction		1.08	.74	1.58	.685
% breast density at baseline	per 10%	140	.94	.75	1.18	.609
60-month change in density	per 10% reduction		1.00	.66	1.51	.994

Note: (*) age and BMI at baseline, tumour size, hystological grade, nodal and ER status

discriminating subjects who experienced a second tumour event and those who did not, in every treatment arm.

TABLE 3.6: HRs for risk of breast cancer recurrence from multivariable Cox survival regression models using mammographic density and its change over time adjusted for other risk-factors*

Variable	obs	HR	[95% Conf.	Interval]	P>z
% breast density at baseline	531	1.00	.99	1.01	.545
% breast density at baseline	398	1.00	.99	1.01	.419
12-month change in density		1.00	.98	1.03	.732
% breast density at baseline	466	1.00	.99	1.01	.553
24-month change in density		1.00	.98	1.02	.825
% breast density at baseline	447	1.00	.98	1.01	.437
60-month change in density		.99	.97	1.01	.488

Note: (*) age and BMI at baseline, tumour size, hystological grade, nodal and ER status

Results from the Cox regression analyses (Table 3.6) confirmed those observed with logistic regression models, and the hazard ratios related to mammographic density and how it changed after 12, 24 and 60 months were all not significant.
We repeated the analyses reported in Tables 3.3, 3.4, 3.5 and 3.6 without adjusting for baseline density, and results were not affected: odds ratios and hazard ratios related to changes in density remained not significant (results not shown).

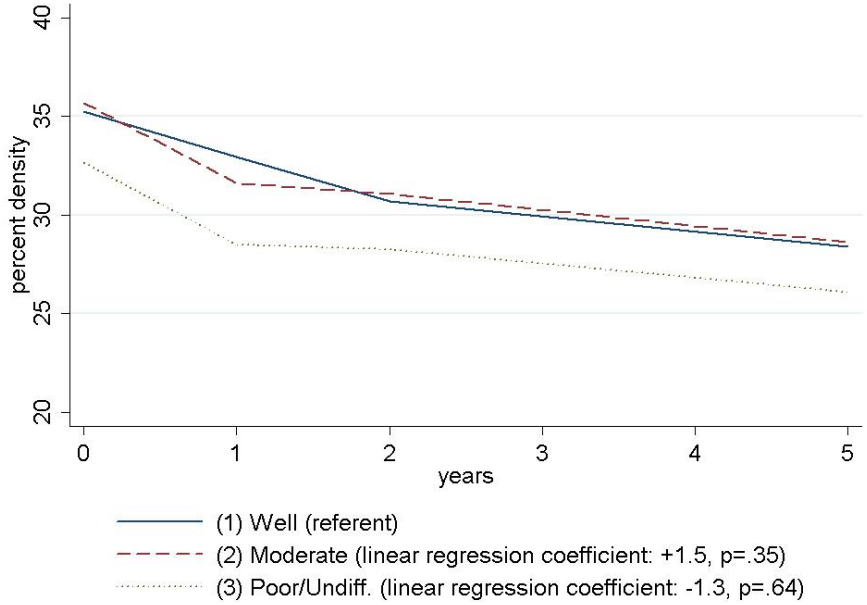
In Table 3.7 the distributions of mammographic density and its changes over time are summarised with means and SDs according to histological grade ("(1) well differentiated", "(2) moderately differentiated" and "(3) poorly differentiated"). Overall mammographic density at baseline appeared lower for subjects who were classified as "(3) Poor/Undiff.". These subjects appeared also the least responsive to the treatment in terms of breast density reduction. In particular after one year of treatment we registered a trend across the three histological categories, however after 2 and 5 years the women classified as "(2) Moderate" experienced a reduction similar to the subjects whose cells were differentiating "(1) Well", and even the differences between these two groups and the "(3) Poor/Undiff." group are no longer significant when assessed with Wilcoxon rank-sum test (p=.10 and p=.40 for 24 and 60 months respectively). These are further illustrated in Figure 3.1, that also suggests a similar decreasing trend for each of the histological grade, but a lower baseline density for subjects with poorly differentiated

tumours, although not significantly lower, as confirmed in the linear regression analyses adjusted for age and BMI (Linear regression coefficient for "(3) poorly differentiated": -1.3, p=.64). These results did not differ according to treatment arm (results not shown).

TABLE 3.7: Distribution of mammographic density at baseline and its change after 12, 24 and 60 months according to histological grade

	(1) Well N=137		(2) Moderate N=299		(3) Poor/Undiff N=129		
Variable	mean	(SD)	mean	(SD)	mean	(SD)	P_{trend}
% density at baseline	35.2	(25.6)	35.7	(25.2)	32.6	(23.2)	.490
12-month change in density	-4.8	(10.3)	-3.8	(11.1)	-1.4	(8.9)	.029
24-month change in density	-5.5	(10.4)	-5.3	(12.4)	-3.8	(9.1)	.069
60-month change in density	-8.6	(13.6)	-7.6	(13.5)	-6.0	(9.4)	.325

FIGURE 3.1: Mammographic density over time according to histological grade



Overall results regarding the relationship between nodal status and mammographic density (Table 3.8) reported nothing significant. Mammographic density and its changes at the different time-points were similar in subjects whose nodal status was classified as "Negative" or "Positive/Unknown", suggesting similar decreasing patterns in density for both groups. There was a baseline significant difference at 24 months, with the node negative group showing greater changes. This appears evident in Figure 3.2, where, after the second year of treatment, density in the two groups decline in similar magnitude although values in the node negative group are consistently lower. Linear

regression analysis of density over time, adjusted for age and BMI, confirmed this result. No substantial difference according to nodal status, instead, was observed in the analyses stratified by treatment arm (results not shown).

TABLE 3.8: Distribution of mammographic density at baseline and its change after 12, 24 and 60 months according to nodal status

	Positive or Unknown N=210		Negative N=391		
Variable	mean	(SD)	mean	(SD)	$P_{Wilcoxon}$
% density at baseline	35.6	(25.5)	34.7	(24.7)	.697
12-month change in density	-3.9	(11.0)	-3.4	(10.6)	.808
24-month change in density	-3.7	(12.0)	-5.4	(10.8)	.086
60-month change in density	-6.6	(13.3)	-7.9	(12.6)	.126

FIGURE 3.2: Mammographic density over time according to nodal status

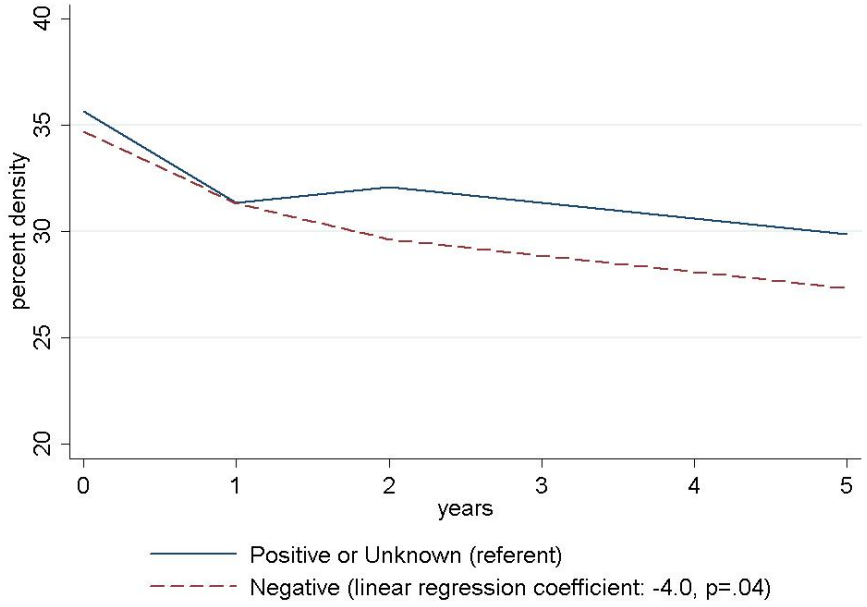


Table 3.9 and Figures 3.3-3.4 illustrate how tumour size and mammographic density and its changes over time interact. Overall no significant difference was recorded in density or density reductions between subjects who had a first tumour larger than 2 cm and those who had a smaller tumour. Only after 2 years of treatment the difference is borderline significant, indicating that women who had a larger first tumour were less responsive to treatment in terms of density reduction. This result is further supported by Figure 3.3 and by the linear regression analysis of density over time. Average density levels in

TABLE 3.9: Distribution of mammographic density at baseline and its change after 12, 24 and 60 months according to tumour size

Variable	≤ 2cm N=399 mean	(SD)	>2cm N=201 mean	(SD)	$P_{Wilcoxon}$
% density at baseline	34.7	(25.1)	35.6	(24.7)	.605
12-month change in density	-3.7	(11.1)	-3.3	(10.0)	.551
24-month change in density	-5.5	(11.5)	-3.7	(10.5)	.064
60-month change in density	-8.0	(13.3)	-6.2	(11.8)	.173

subjects who had larger primary tumours were consinstently higher (+6%, p<.01). This difference between the two tumour size groups appeared stronger in the tamoxifen arm (Figure 3.4.b). However after 5 years of treatment this difference was less notable, even in the tamoxifen arm (Wilcoxon rank-sum test: p=.10).

FIGURE 3.3: Mammographic density over time according to tumour size

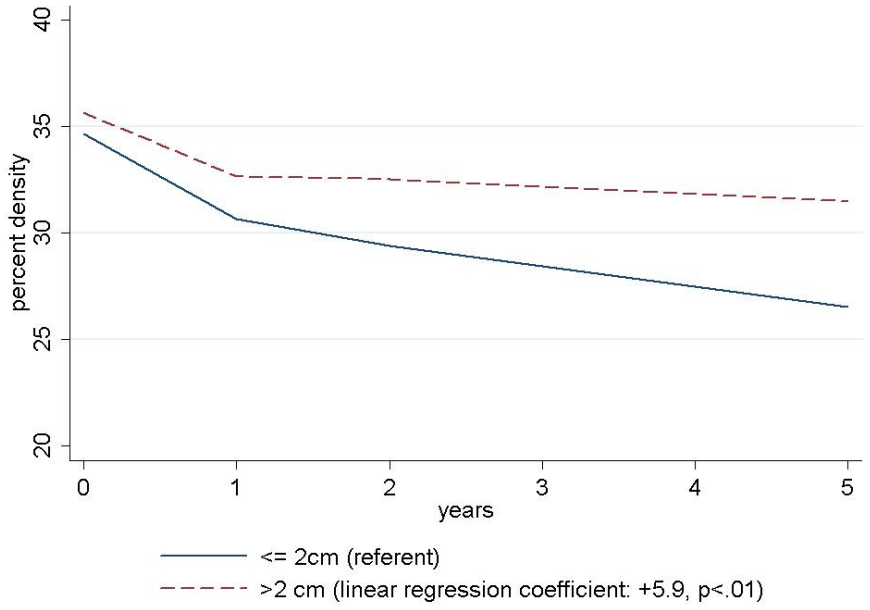


FIGURE 3.4: Distribution of changes in mammographic density after 12 and 24 months according to tumour size and treatment arm

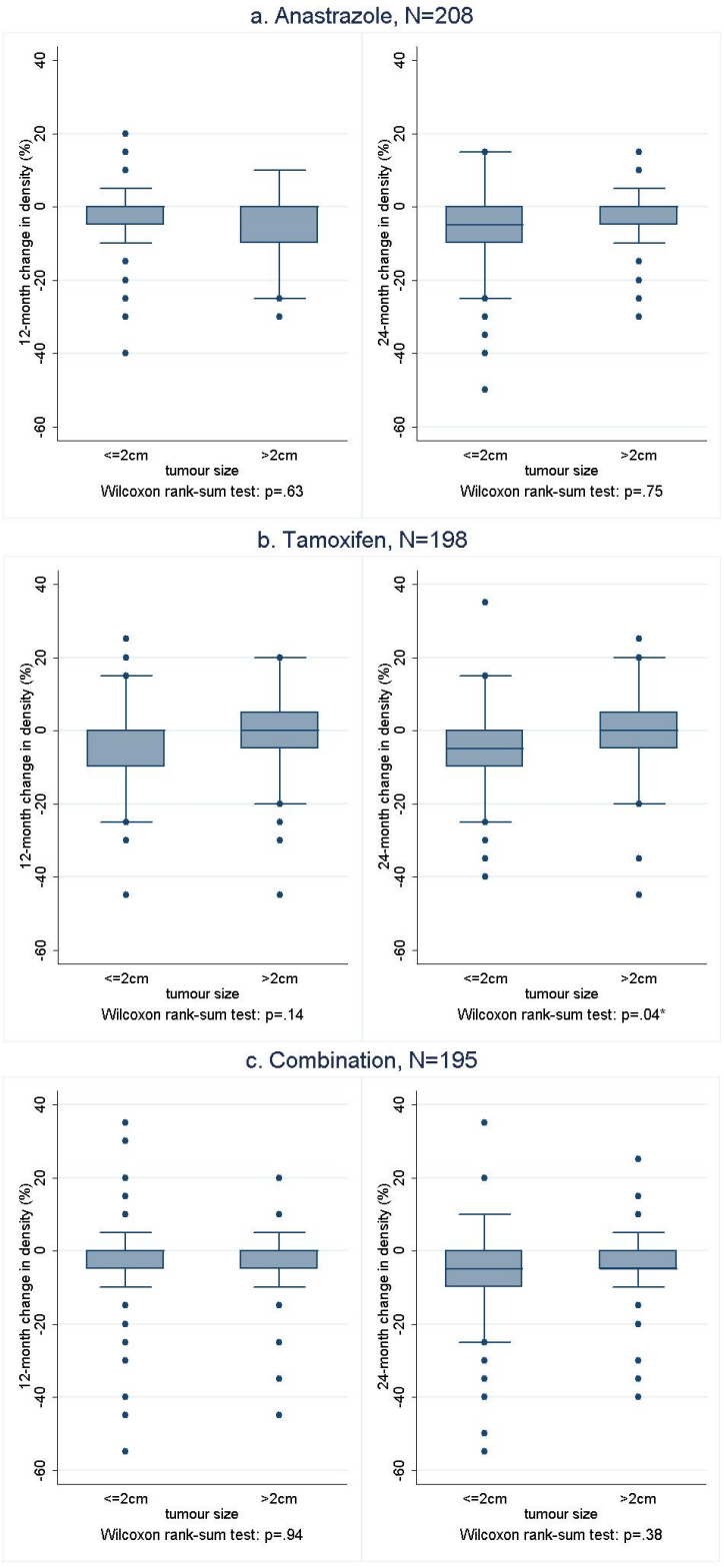


TABLE 3.10: Distribution of mammographic density at baseline and its change after 12, 24 and 60 months according to ER status

	Positive N=455		Negative N=69		
Variable	mean	(SD)	mean	(SD)	$P_{Wilcoxon}$
% density at baseline	35.3	(25.0)	36.2	(25.0)	.785
12-month change in density	-3.9	(10.5)	-5.8	(13.6)	.806
24-month change in density	-5.0	(11.2)	-6.8	(12.9)	.884
60-month change in density	-7.7	(12.9)	-9.0	(14.2)	.999

Most of the subjects in our study had a first breast tumour which was oestrogen receptor positive (455/524). Table 3.10 and Figures 3.5-3.6 display the results of the analyses of the relationship between ER status and density, both overall and according to treatment arm. Overall, contrary to expectations, the mean reductions at 12, 24 and 60 months were consistently larger in the "Negative" group, however these differences were not significant. The linear regression analysis of density over time confirmed these results. The stratified analyses on changes after 24 months highlighted that in the anastrazole and in the tamoxifen arms the reduction were larger, though not significantly, in the subjects who had ER positive tumours, as one would expect. In the combination group, instead, we observe a significant trend in the opposite direction. This is likely due to the limited number of subjects in the combination arm who had an ER negative tumour (N=22).

FIGURE 3.5: Mmmographic density over time according to ER status

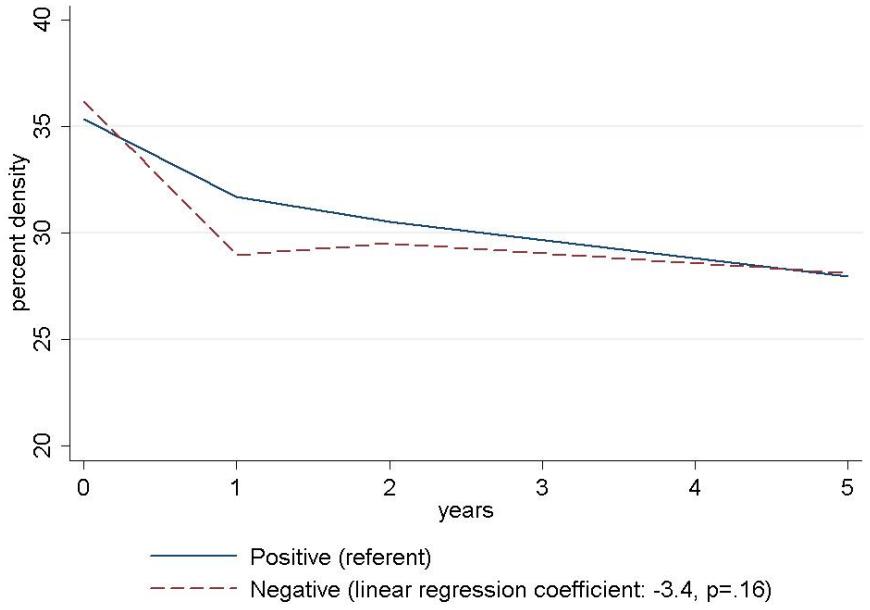


FIGURE 3.6: Distribution of changes in mammographic density after 12 and 24 months according to ER status and treatment arm

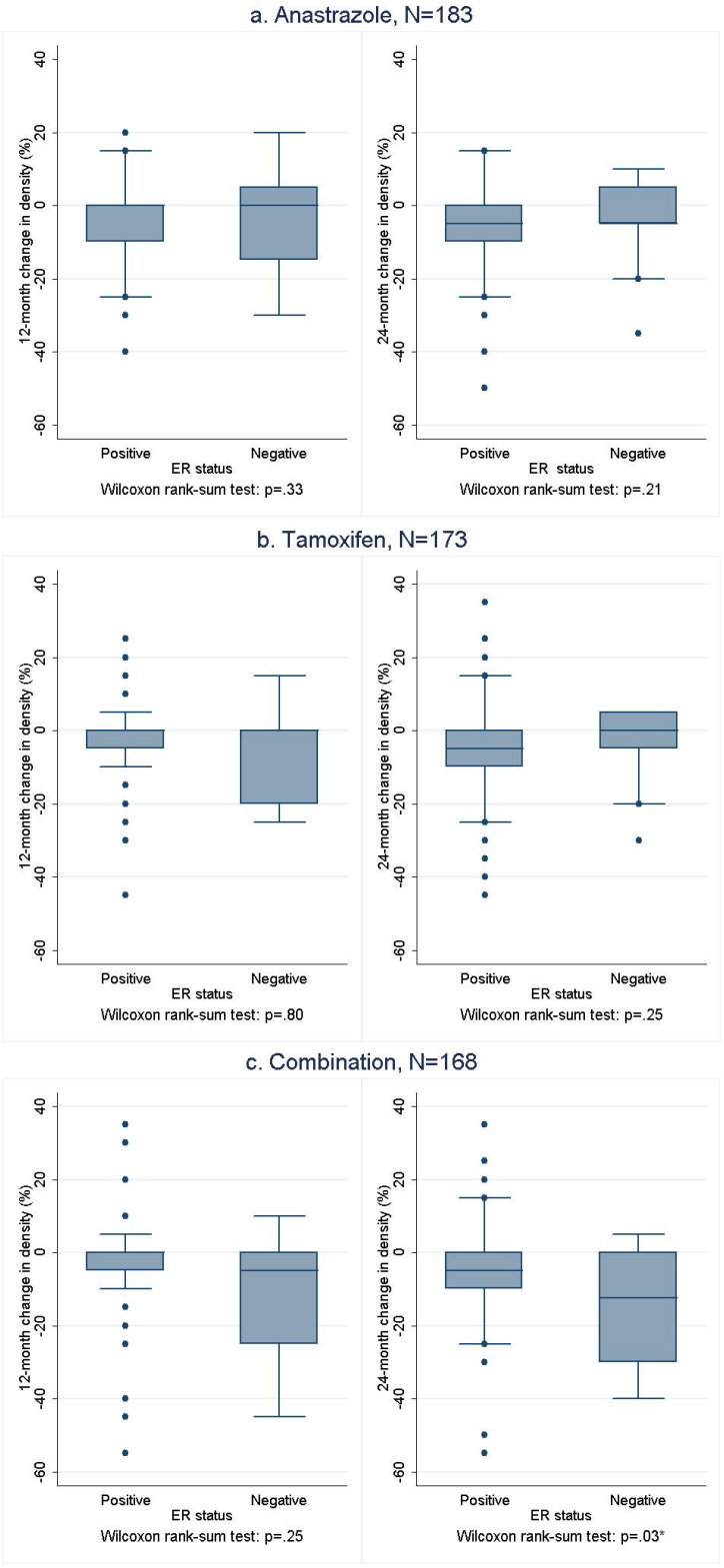


TABLE 3.11: Distribution of mammographic density at baseline and its change after 12, 24 and 60 months according to HER2 status

	Positive N=26		Negative N=242		
Variable	mean	(SD)	mean	(SD)	$P_{Wilcoxon}$
% density at baseline	33.9	(24.5)	36.7	(27.1)	.666
12-month change in density	-3.8	(10.1)	-2.5	(8.0)	.617
24-month change in density	-4.8	(10.8)	-5.2	(9.8)	.889
60-month change in density	-7.6	(12.8)	-6.6	(10.8)	.661

Analyses of the interaction between mammographic density and HER2 status were limited by the small number of HER2 positive subjects (N=26). Results showed no substantial difference in density and its changes between the two groups (Table 3.11). However, this observation is qualified by the small proportion of HER2 positive subjects.

FIGURE 3.7: Association between the amount of protein ki67 and (a) mammographic density at baseline and its change after (b) 12, (c) 24 and (d) 60 months

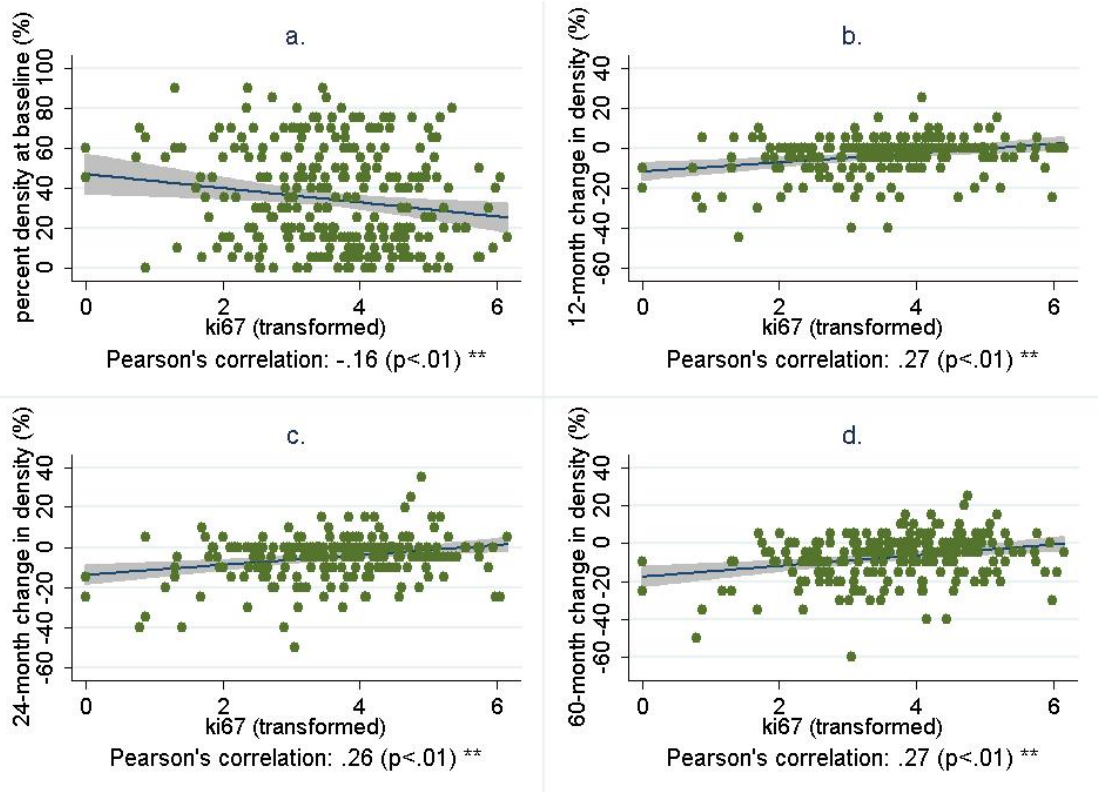
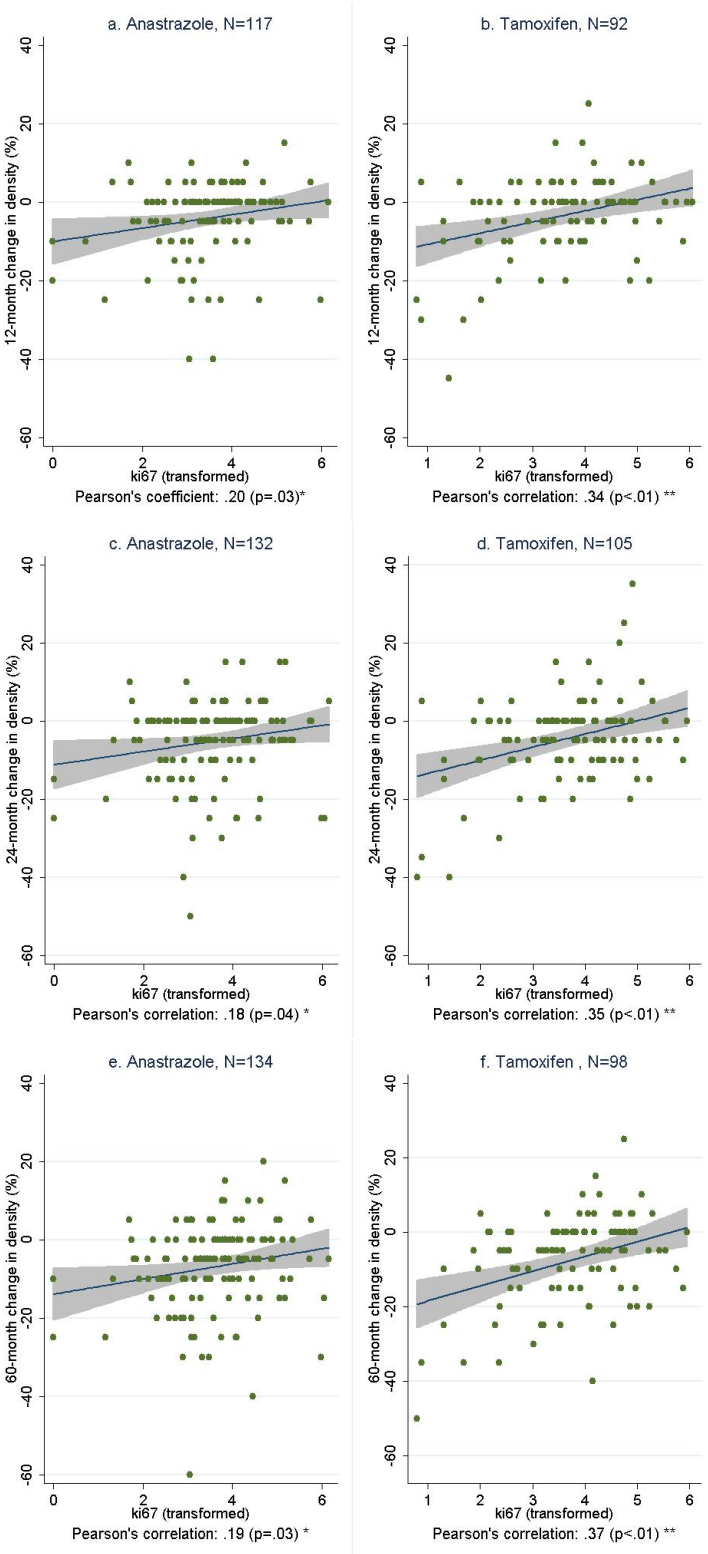


FIGURE 3.8: Association between the amount of protein ki67 and the changes in mammographic density after (a,b)12, (c,d) 24 and (e,f) 60 months according to treatment arm



Results on density and cellular proliferation, indicated by levels of the protein Ki-67, are reported in Figures 3.7 and 3.8. They revealed a significant negative relationship between density and Ki-67, which means that the higher the density the lower the levels of Ki-67 and vice versa. However, the lower the level of KI-67 at baseline, the larger the reduction in density for each of the time-points. In other words breasts with a lower level of cellular proliferation were more likely to experience a higher reduction in the amount of fibroglandular tissue. This was evident also in both treatment groups, although the relationship between the level of Ki-67 and the changes in density after 12, 24 and 60 months was stronger in the tamoxifen arm (Figure 3.8).

3.4 Discussion

Mammographic density at baseline appeared unrelated to risk of recurrence, and larger reductions in density did not necessarily lead to a lower risk of recurrence. Our findings support Porter and colleagues' previous study [203], reporting a lack of association between primarily dense breast composition and breast cancer prognosis, whereas they contradict Cil *et al.*'s [204] conclusions, that suggested that denser breasts were more likely to experience a recurrence. In comparison with both these studies, ours had the advantage of considering density assessed in greater quantitative detail (21 visual categories), whereas Porter and colleagues had it classified according to the four BIRADS categories and Cil's study used a variation of Wolfe's parenchymal patterns [89]. On the other hand, in both Porter's and Cil's data, mammographic density was assessed on the contralateral breast on a diagnostic pretreatment mammogram, whereas our study used as baseline mammograms those taken at enrollment in the ATAC trial. Since entry in the trial was after the subjects underwent chemotherapy and, often (65% of subjects in our subset), radiotherapy treatment, density may be affected by these. Chemotherapy is known to affect mammographic appearance [216], whereas the relationship between density and radiotherapy has not been established.

Similar changes in density were reported across the treatment arms (Table 3.1), and unlike previous findings [18–20] a reduction in dense tissue of 10% or higher did not suggest better outcomes.

Despite being a strong predictor of primary breast cancer, higher levels of density do not increase risk of death or recurrence. This result is consistent with Gierach and colleagues' findings on a cohort of 9,232 women diagnosed with primary invasive breast carcinoma [85], that reported that BIRADS density was not related to breast cancer

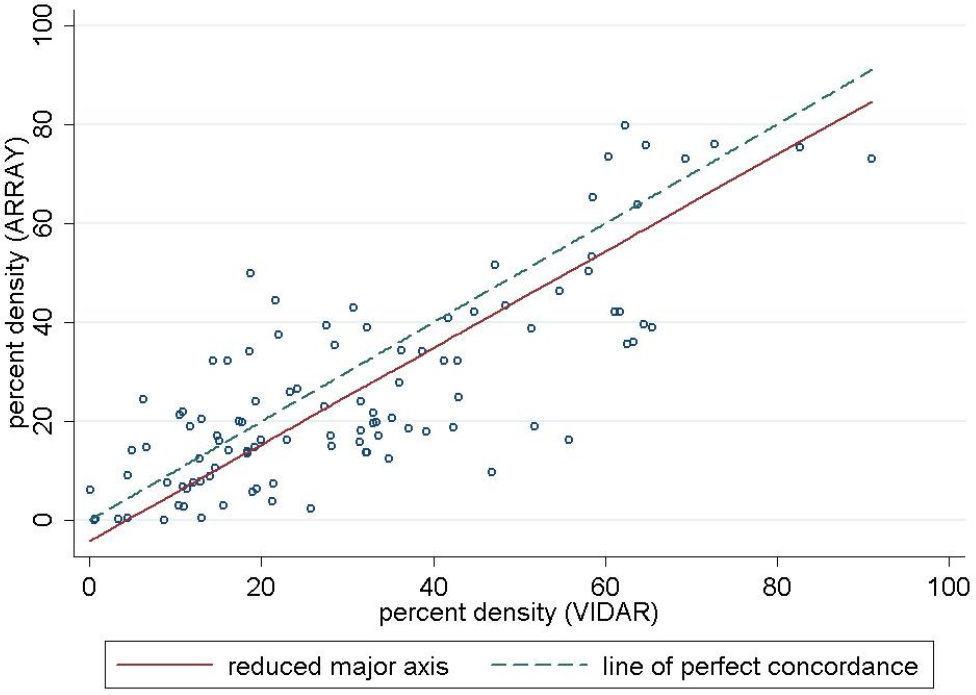
death, especially after adjusting for tumour characteristics and risk factors. Likewise, results from a Danish study [84] indicated that subjects with denser breasts had a higher incidence of breast tumours, but the proportion of breast cancer deaths among them was lower than among women with fattier breasts. Thus, risk factors for the development of breast cancer may not necessarily be the same as factors influencing the risk of death after developing a breast cancer [84, 85]. Our findings suggest that the same may be applied to risk of recurrence.

Our analyses did not suggest an association between mammographic density and risk of more aggressive tumours, as previous studies suggested [80, 211]. However it is known that dense breasts decrease mammographic sensitivity, resulting in delayed detection and corresponding to larger and more advanced tumours [207]; and this could explain why those studies observed such a relationship. A recent meta-analysis on mammographic density and tumour subtypes [81] suggested that denser breast tissue increases the risk of developing a breast tumour overall, no matter what the sub-type is.
Our results suggest a modest association of histological grade with changes in density, consistent with the effect of grade on response to treatment. Good or moderate differentiation in the tumour was associated with a significantly larger reduction in density, at least for the first two years of treatment. Density appeared affected by nodal status after the second year of treatment, and the "Negative-node" group experienced higher reductions in the amount of fibroglandular tissue after 24 months. Our analyses suggested similar findings regarding ER and HER2 status, however the small number of ER Negative (N=69) and HER2 Positive (N=26) subjects prevented a proper comparison.
In our data, tumour size appeared not influenced by mammographic density, although, after two years, subjects with smaller tumours experienced significantly greater reductions in breast dense tissue, especially in the tamoxifen arm (p=.04). This effect was temporary and after 5 years the difference in reduction between the two groups ($\leq$ 2 cm and $>$ 2cm) was no longer significant. However studying the trend in density over time according to tumour size, it was more evident that density was consistently higher and less likely to reduce in women with larger primary tumours.
It is worth reminding also that for all histological grade, nodal status and tumour size, there was a borderline significant difference 24 months into the study, always showing the larger reduction in the category with the better prognosis, i.e. histological grade 1 ("well differentiated"), negative nodal status and smaller primary tumour.
Levels of the proliferation marker Ki-67 were significantly lower in those with higher density at baseline in this study. In contrast to our findings, Harvey and colleagues

[217] observed that increased breast density was significantly associated with Ki-67 activity in the ducts. However their analyses considered only 56 women and focused on the effect of the use of hormone replacement therapy and the histologic changes it caused, whereas our sample comprised 268 subjects. Another study of 55 tamoxifen users [218] reported no correlation between breast density and cell proliferation. These findings may also be due to the limited sample size. Furthermore, in our data, a lower baseline level of Ki-67 corresponded to a larger reduction in fibroglandular tissue. This may be partly due to the initial higher levels of density in those with low Ki-67, however we observed also a stronger association between Ki-67 and reduction in density in the tamoxifen arm. Since this phenomenon is affected by treatment given, this suggestes that it does indeed reflect a biological association. The higher the levels of proliferation at baseline, the higher the chance that subjects would experience no change or even an increase in density over the years of treatment, particularly in the tamoxifen arm. There is extensive clinical evidence that tamoxifen decreases cell-proliferation in breast tissue [218]. In 2005 Dowsett *et al.* compared the effect on levels of Ki-67 of treatment with anastrazole and tamoxifen, alone or combined [215]. Results showed significantly greater suppression of the proliferation marker within the anastrazole arm, which could explain the weaker correlation between Ki-67 and density changes we observed in that treatment group. Unfortunately we only had data on Ki-67 levels at baseline, so we could not compare the changes in both suppression of proliferation and breast density during treatment. Nevertheless, we can hypothesize that increased cell proliferation would limit reductions in density, and therefore the higher the level of Ki-67, the lesser the resulting decrease of breast density. If anastrazole is more effective than tamoxifen in suppressing proliferation, but not in decreasing density, than the relationship between changes in fibroglandular tissue and baseline level of Ki-67 would be weaker in this arm, as we observed. Further investigations on this relationship are needed, as levels of Ki-67 could be crucial to understanding how aromatase inhibitors and tamoxifen affect breast cancer risk and the tissue composition of the breast.

This study had limitations. Firstly, because therapy with aromatase inhibitors was a major focus, it comprised only post-menopausal women, thus our findings may not apply to premenopausal breast cancer patients. Secondly, as we noted above, mammographic density was assessed after chemotherapy and, in most subjects (65%), radiotherapy, which may have affected not only the breast composition but also how this tissue responds to treatment, as well as the relationship between density and Ki-67. Moreover, our sample included relatively few cases of ER negative or HER2 positive tumours.

FIGURE 3.9: Comparison of percent density assessments using images scanned with VIDAR (old scanner) and ARRAY (new scanner)



Hence our results regarding the relationship between density and these two tumour categories, as well as their response to treatment, are not conclusive. Larger studies focusing on these two tumour subtypes and non-hormonal treatments may be needed. Finally, it may be hypothesised that results might have been influenced by a possible sub-optimal quality of the images, because of the scanner used (VIDAR). To verify if this was the case, a batch of mammograms (N=103) were re-scanned with a new machine (Array 2905HD Laser Film Digizer) and we compared density assessments using both images for each mammogram (Figure 3.9). Results showed a good agreement between the two scanners (Pearson's correlation coefficient: 86.4%) and reassured usabout the reliability of the findings we reported in this chapter (Intraclass correlation coefficient: 76%, 95% CI (67.8%, 84.2%)).

In conclusion, mammographic density at baseline did not predict risk of recurrence and greater reductions in the proportion of dense tissue were not necessarily suggesting better outcomes of the treatment. However, subjects with primary invasive tumour of poor histological grade experienced smaller reductions in breast density. After two years in the study, the differences in density were more evident also between subjects with different nodal status and size of the primary tumour, with larger reductions in the negative

node and smaller tumour groups. Further, women with a higher baseline level of cell proliferation were less likely to experience changes in breast composition. Thus, we cannot yet assign a role for density in management of such patients. Further studies are needed to determine the role mammographic density plays in the pathways of adjuvant treatments, histologic factors and improved survival, as well as the effect of radiotherapy on breast tissue composition.

Chapter 4

Mammographic density and breast cancer risk in a young population at enhanced familial risk (FH01 study)

4.1 Introduction

In UK the NHS Breast Screening Programme (NHSBSP) invites women aged 50 to 70 for mammographic screening every 3 years, and the age range is currently being extended from 47 to 73. The majority of women in their 40s are, therefore, not included because it has been observed that screening in women aged 40 to 49 has lower sensitivity and, due to the lower incidence in this age group, fewer lives are saved in return for the financial and human costs [219]. However, there is evidence that targeted screening in this age group could lower breast cancer mortality in women at higher than average risk by detecting tumours at an earlier stage [13, 73, 188, 220]. Since family history of breast cancer is a well-established indicator of risk of developing the disease [1, 221], in the UK, many women under 50, i.e. not yet eligible for the national screening programme, are offered early mammographic screening when found to be at moderate or high familial risk.

High breast density reduces screening sensitivity as well as increasing risk of developing a breast tumour [2, 74]. As noted in the previous chapters, mammographic density is a reflection of the amount of connective and epithelial tissue in the breast [222] and

declines with age [88]. More precisely the proportion of epithelial tissue increases until age 30-40 years, then declines progressively [88], and between the ages of 40 and 49 years the majority of women have high- or very high- density breasts, respectively 57% and 17% of the subjects, according to BIRADS classification [223]. However, because of its strong independent association to breast cancer risk, it has been suggested that density should also be considered when deciding the appropriate surveillance regimen for a high-risk patient [73].

An important question to be clarified before such a strategy could be adopted as policy is whether density adds further information on risk in a population already defined as at enhanced risk, even after adjustment for other risk factors. FH01 is a study recruiting precisely such a population. It is a longitudinal, single arm study of five years of mammographic surveillance in 6,710 women aged 40-49, at moderate familial risk of breast cancer [188, 224].

The present investigations were run on a case-control study nested within FH01 [188, 224]. Our primary objective was to evaluate mammographic density as a risk factor for breast cancer, investigating its ability to discriminate between cases and controls in this population, that have not only been classified as high risk because of their family history [188, 224], but also have a predominant proportion of high density subjects because of the age range [223]. Secondly, we examined the interactions between density and other breast cancer risk factors, e.g. age at menarche and parity, and how these affected the association between density and risk of breast cancer. Finally, we studied the additional effect of density adjusted for a 10-year risk of breast cancer established from multiple familial and non familial risk factors [1]. The results of the work in this chapter necessarily pertain only yo a younger population at enhanced familial risk. However, there is a need to clarify the role of mammographic density in further risk delineation in this risk group. The effect of density in such risk groups is considerably less well researched than in the general population.

4.2 Materials and Methods

4.2.1 Study setting and population

The FH01 study [188, 224] was designed to evaluate the effectiveness of annual mammographic screening in women younger than 50 years who have a clinically significant family history of breast cancer. Across the UK, a total of 6,710 women, aged 40-49

years, were enrolled between 2003 and 2007 because of their moderate familial risk of developing breast cancer. The intervention offered every subject annual mammography over 5 years and the study was concluded in 2009 [188, 224].

Here we report on a case-control study nested within FH01, with 103 breast cancer cases, each matched by age and time in the study, until diagnosis, to either one or two controls free from breast cancer. There were a total of 195 controls.

Mammographic density was assessed on the last mammogram prior to diagnosis for the cases, and for the controls density was assessed on the image closest to this in time. Density measures were estimated, by trained radiologists (Professor R.M.L. Warren, I. Warsi, Dr. M. Kataoka), using the semi-automatic computerised interactive thresholding programme (Cumulus) [28]. This method provides estimates of total breast area and absolute density (cm^2), allowing percent density and non-dense breast area to be computed.

The dataset included other risk factors for breast cancer: age at menarche, parity, age at first live birth, use of hormone replacement therapy and menopausal status. We also had absolute risk estimates using the Tyrer-Cuzick model. The Tyrer-Cuzick model provides 10 year risk estimates for developing breast cancer based on extensive family history information, multiple other breast cancer risk factors mainly representing exposure to endogenous oestrogen, and benign breast disease [1].

4.2.2 Statistical Analyses

The distributions of demographic and other variables at the first examination are summarised as percentages for discrete variables, or means and SDs for continuous variables. Percent density was categorised following Boyd's classification ("$< 5\%$", "$5 - 10\%$", "$10 - 25\%$", "$25 - 50\%$", "$50 - 75\%$" and "$> 75\%$") [28] and then reduced to five and four categories, collating first "$< 5\%$" with "$5 - 10\%$" and then "$50 - 75\%$" with "$> 75\%$", in order to have more than five subjects for each level of density, both in cases and controls.

The relationship between breast cancer risk and density, both percent and absolute, was assessed using odds ratios obtained from conditional logistic regression, taking into account the matching of the study design. The relationship between density estimates and other variables was investigated both graphically and using the Pearson's correlation coefficient, the Wilcoxon rank-sum test and the non-parametric test for trend [214], as appropriate. Multivariable conditional logistic regression was used to adjust for other risk factors. In order to include nulliparous subjects in the conditional logistic

regression model with age at first pregnancy, we created a new discrete variable *age at first pregnancy* * with the three following categories "Under 30", "30 or more" and "Nulliparous". The relationship between breast cancer risk and density was also studied adjusted for the Tyrer-Cuzick estimate of 10-year risk of breast cancer. However, since these estimates are based on age, besides age at menarche, parity, menopausal status and family history, conditional logistic would have led to adjusting for age twice (both in the model and in the matching). Thus standard unconditional logistic was performed instead for this analysis.
Secondary analyses, including only premenopausal subjects, the majority of this population (75%), were also performed.
All analyses were performed with STATA software version 11.

4.3 Results

The characteristics of the 103 case subjects and 195 control subjects are shown in Table 4.1. On average the subjects underwent the mammographic examination used for the density assessment 3 months after their 44th birthday and had about 2-years follow-up in the study at that time. They experienced their first period aged 12 years and 8 months and gave birth for the first time when they were 26. The modal number of children was two (41%), only 14 % had been prescribed hormone replacement therapy and the majority (75%) were premenopausal.
Percent density and non-dense area had similar mean values and standard deviations in the two groups. Conversely, absolute density and, consequently, total breast area had substantial differences in distribution, both mean and SD, between case and control subjects.

TABLE 4.1: Characteristics of the study sample

	cases N=103		controls N=195	
Age at mammogram in years				
mean (SD)	44.38	(2.33)	44.21	(2.18)
Years of follow-up				
mean (SD)	2.32	(1.63)	2.09	(1.49)
Menopausal status (%)				
Pre-menopausal	86	(83.50)	148	(75.90)
Post-menopausal	5	(4.85)	17	(8.72)
Unknown	12	(11.65)	30	(15.38)
Age at menarche in years				
mean (SD)	12.84	(1.60)	12.63	(1.43)
Number of live births (%)				
0	19	(19.79)	34	(18.48)
1	16	(16.67)	28	(15.22)
2	39	(40.63)	83	(45.11)
3	20	(20.83)	26	(14.13)
$\geq$4	2	(2.08)	13	(7.07)
Unknown	7	-	11	-
Age at first live birth in years				
mean (SD)	26.70	(5.79)	25.58	(5.23)
Hormone replacement therapy or HRT (%)				
Users	13	(13.54)	28	(15.82)
Tyrer-Cuzick 10-years estimate of risk, *mean (SD)*				
Absolute risk	.068	(.041)	.057	(.039)
Relative risk	3.71	(2.48)	3.06	(2.23)
Breast composition,*mean (SD)*				
Absolute density (cm^2)	56.86	(41.38)	47.80	(31.74)
Percent density (%)	36.71	(19.29)	34.55	(19.21)
Non-dense area (cm^2)	105.38	(65.61)	100.54	(62.48)
Total breast area (cm^2)	162.24	(73.77)	148.34	(61.88)

FIGURE 4.1: Distribution of (a) absolute and (b) percent density among cases and controls

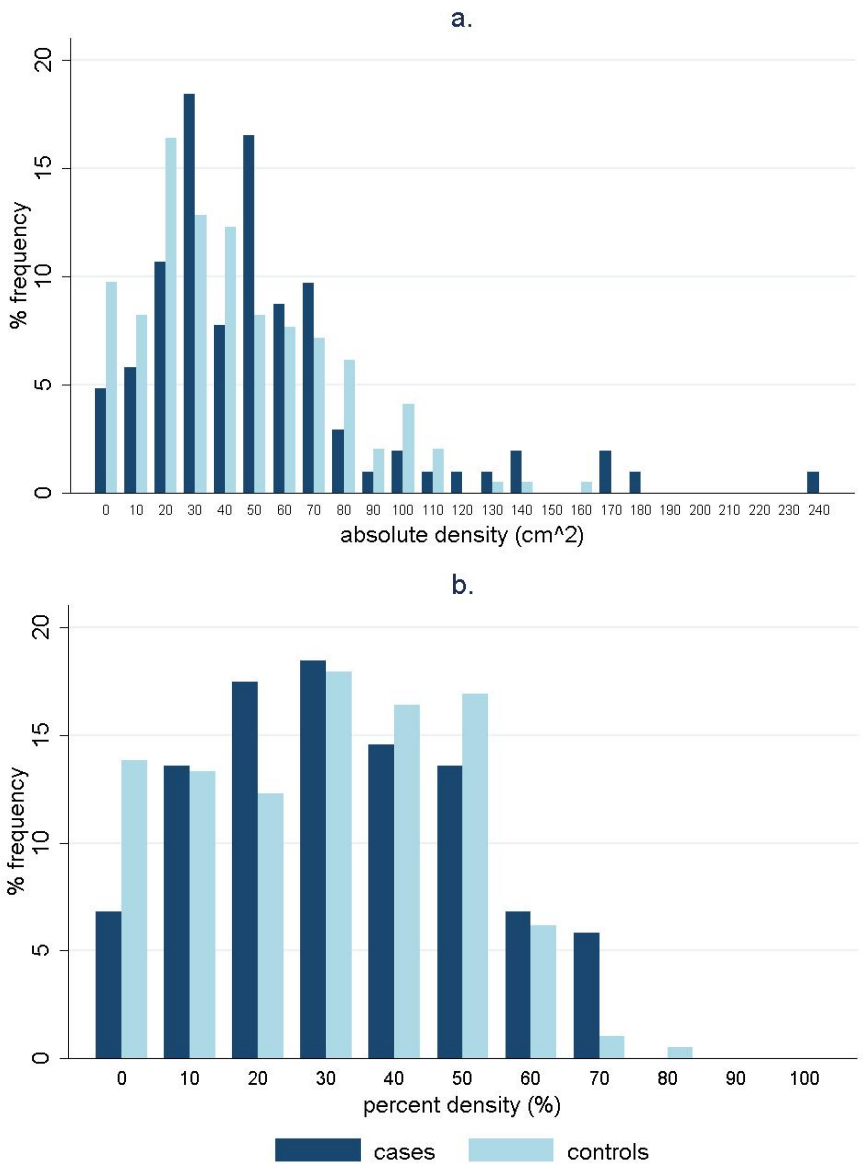


Figure 4.1 and Table 4.2 show in more detail how the distribution of absolute and percent density varies between cases and controls. In both it is seen that the median for absolute density in the cases is higher than that for the controls, although the Wilcoxon rank-sum test found this difference to be weakly significant (p=.09). On the other hand percent density appeared to be distributed similarly in the two groups.

TABLE 4.2: Distribution of cases and controls in density categories based on quintiles of the controls distribution

	cases (%)		controls (%)		Total (%)	
Absolute density						
0-20.77 cm^2	13	(12.62)	38	(19.49)	51	(17.11)
20.77-32.41 cm^2	13	(12.62)	39	(20.00)	52	(17.45)
32.41-49.93 cm^2	24	(23.30)	39	(20.00)	63	(21.14)
49.93-77.02 cm^2	32	(31.07)	39	(20.00)	71	(23.83)
$\geq$77.02 cm^2	21	(20.39)	40	(20.51)	61	(20.47)
Total	103	(100.00)	195	(100.00)	298	(100.00)
Percent density						
0-14.58 %	14	(13.59)	38	(19.49)	52	(17.45)
14.58-29.93 %	26	(25.24)	39	(20.00)	65	(21.81)
29.93-40.86 %	20	(19.42)	39	(20.00)	59	(19.80)
40.86-52.44 %	18	(17.48)	39	(20.00)	57	(19.13)
52.44-100 %	25	(24.27)	40	(20.51)	65	(21.81)
Total	103	(100.00)	195	(100.00)	298	(100.00)

4.3.1 Mammographic density and breast cancer risk

Table 4.3 shows how mammographic density, assessed in percent and absolute terms, relates to breast cancer risk, using both control quintiles and a continuous effect expressed per 10-unit increase, and adjusting for non-dense area, which has been suggested as a proxy for BMI [33, 225, 226]. Categorisation by quintiles of the distribution among controls did not seem able to discriminate between cases and controls, as was also the case in Table 4.2. However, we observed a significant continuous effect of absolute density, in discriminating cases and controls, especially after adjusting for non-dense area. This may be related to the fact that among cases, absolute density has a skewed distribution, with a longer tail for extreme values (Figure 4.1.a).

We also categorised percent density using Boyd's cut-off points (Table 4.4), and conditional logistic regression led to significantly and steadily increasing ORs (p=.034), after adjusting for non-dense area (Table 4.5). To be more specific, subjects with percent density between 10 and 25 % were three (3.1) times more likely to have breast cancer than women with a breast density of less than 10%, when they have similar non-dense area. This difference in risk increases with the categories with up to a 12.6 (95% CI (1.40,114.06)) OR between the two extreme density groups, i.e. "< 10% dense" and "> 75% dense" (Table 4.5).

TABLE 4.3: ORs for risk of developing breast cancer from conditional logistic regression models using mammographic density

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
absolute density	0- 20.77 cm^2	292	1.00	-	-	.099
	20.77-32.41 cm^2		1.11	0.44	2.79	
	32.41-49.93 cm^2		2.13	0.87	5.20	
	49.93-77.02 cm^2		2.75	1.18	6.42	
	> 77.02 cm^2		1.57	0.67	3.72	
absolute density	per 10cm^2	292	1.06	0.99	1.14	.064
percent density	0- 14.58%	292	1.00	-	-	.614
	14.58-29.93%		1.84	0.82	4.11	
	29.93-40.86%		1.41	0.63	3.15	
	40.86-52.44%		1.19	0.53	2.67	
	52.44-100%		1.63	0.74	3.59	
percent density	per 10%	292	1.05	0.93	1.19	.449
absolute density	per 10cm^2	292	1.07	1.00	1.15	.040
non-dense area	per 10cm^2		1.02	0.98	1.07	.257
percent density	per 10%	292	1.15	0.97	1.37	.105
non-dense area	per 10cm^2		1.04	0.99	1.10	.111

These results could have been influenced by the scarcity of subjects in the highest category (Table 4.4), so we repeated the conditional logistic analyses combining the two highest categories in "> 50% dense" (Table 4.6). Here again we observed a significant and steady increase in the ORs across Boyd's categories (p=.048), and the odds ratio between the two extreme density groups was 4.43 (1.26,15.58).

TABLE 4.4: Distribution of percent density, categorised using Boyd's cut-off points (5 categories)

Categories	controls (%)		cases (%)		Total (%)	
<10%	27	(6.93)	7	(13.99)	34	(11.56)
10-25%	38	(21.78)	22	(19.69)	60	(20.41)
25-50%	79	(44.55)	45	(40.93)	124	(42.18)
50-75%	47	(23.76)	24	(24.35)	71	(24.15)
>75%	2	(2.97)	3	(1.04)	5	(1.70)
Total	193	(100.00)	101	(100.00)	294	(100.00)

TABLE 4.5: ORs for risk of developing breast cancer from conditional logistic regression models using mammographic percent density classified in Boyd's density categories (5) and adjusted for amount of non-dense tissue, n=292

Factor	category	OR	[95% Conf.	Interval]	P>z
percent density	<10% dense	1.00	-	-	.231
	10-25% dense	2.24	.83	6.05	
	25-50% dense	2.10	.84	5.23	
	50-75% dense	2.00	.76	5.23	
	>75% dense	5.19	.71	37.70	
percent density	<10% dense	1.00	-	-	.034
	10-25% dense	3.09	1.04	9.23	
	25-50% dense	3.81	1.22	11.86	
	>50% dense	4.25	1.20	15.07	
	>75% dense	12.62	1.40	114.06	
non-dense area	per 10 cm^2	1.06	1.00	1.12	.046

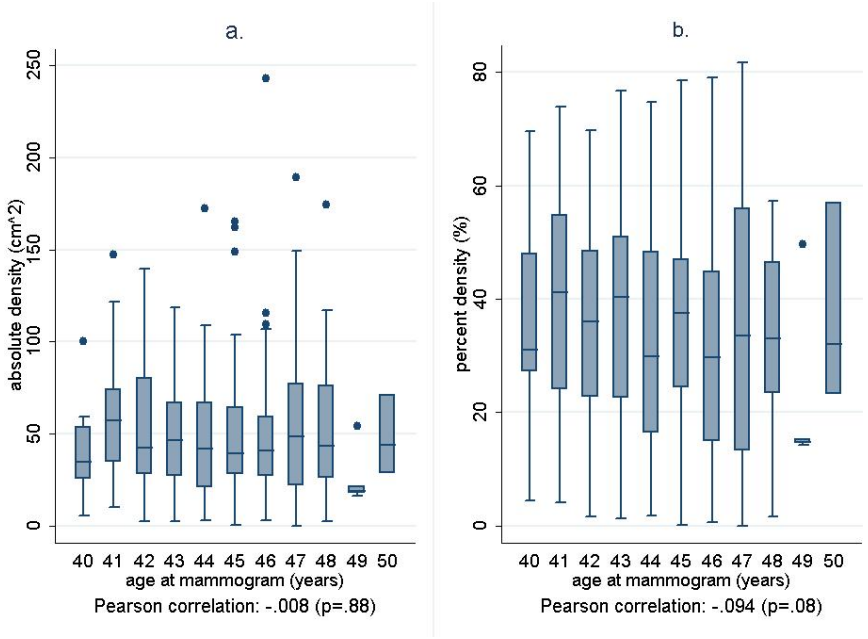
TABLE 4.6: ORs for risk of developing breast cancer from conditional logistic regression models using mammographic percent density classified in Boyd's density categories and adjusted for amount of non-dense tissue, n=292

Factor	category	OR	[95% Conf.	Interval]	P>z
percent density	<10% dense	1.00	-	-	.279
	10-25% dense	2.26	.84	6.09	
	25-50% dense	2.08	.84	5.17	
	>50% dense	2.13	.82	5.52	
percent density	<10% dense	1.00	-	-	.048
	10-25% dense	3.08	1.04	9.17	
	25-50% dense	3.69	1.19	11.42	
	>50% dense	4.43	1.26	15.58	
non-dense area	per 10 cm^2	1.06	1.00	1.11	.070

4.3.2 Mammographic density and other risk factors

Figure 4.2 shows the relationship between age and mammographic density, both in absolute and percent terms. Absolute density did not appear to be correlated with age (p=.88) whereas percent density did, but only weakly (p=.08). However, it is reasonable to believe that we observed this because of the design of the study, that included a limited age-range.

FIGURE 4.2: Association between age at mammogram and mammographic density measures, (a) absolute and (b) percent



The associations between density measures and risk factors for breast cancer (use of HRT, age at menarche, age at first live pregnancy and number of live pregnancies) are shown in Figures 4.3 and 4.4. The unexpected significantly negative association between the use of HRT and both percent and absolute density (Figures 4.3.a and 4.4.a) was possibly due to the strong confounding effect of menopause. Apart from this, absolute density appeared unrelated to the other risk factors for breast cancer (Figure 4.3). On the other hand, percent density was significantly negatively related to parity and positively related to age at first birth (Figures 4.4.c and .d).

We repeated these analyses using only controls (results not shown), and the results were similar to those using the whole dataset.

FIGURE 4.3: Association between absolute density (cm^2) and risk factors for breast cancer

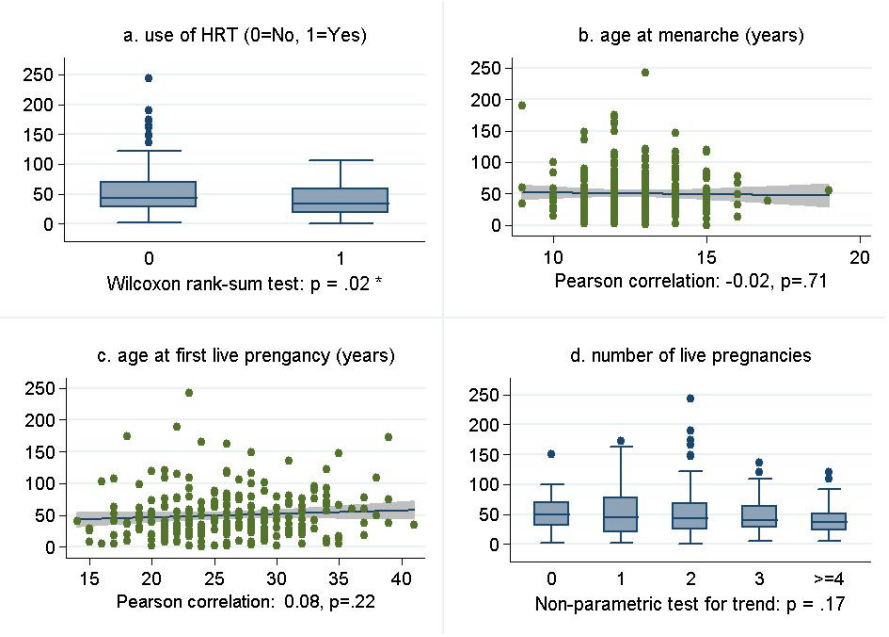
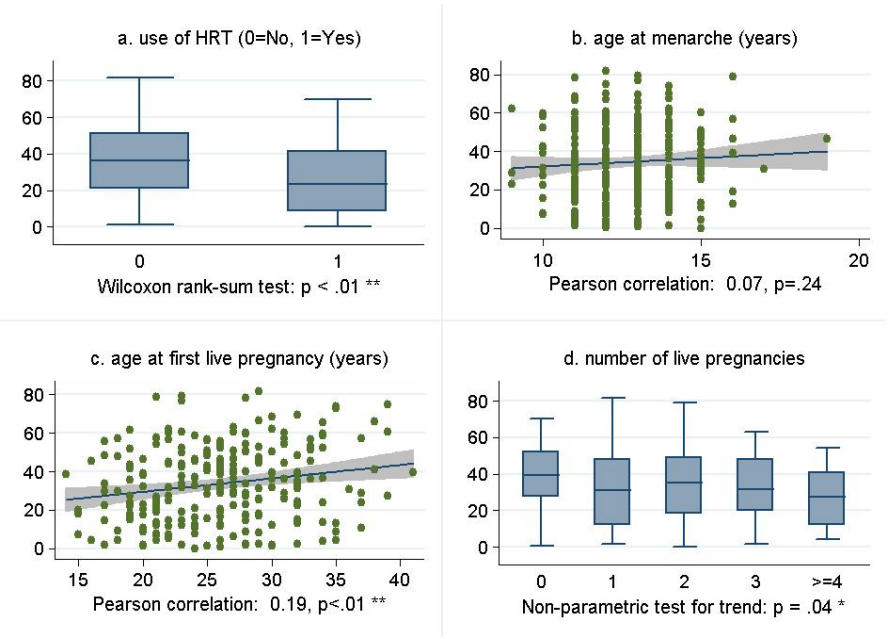


FIGURE 4.4: Association between percent density (%) and risk factors for breast cancer



The risk factors available were included in multivariable conditional logistic regression models, along with density measures, to study their impact on the association between density and breast cancer risk. Results are shown in Tables 4.7, 4.8 and 4.9. Absolute density was more strongly associated with breast cancer risk than percent density, after adjustment for other risk factors for breast cancer. Moreover, the association between absolute density and breast cancer risk appeared independent of other risk factors. The ORs were consistently significant at about 1.08 per 10 cm^2, increasing slightly to 1.10 per 10 cm^2 after adjusting for age at first pregnancy (Table 4.9), possibly related to the exclusion of nulliparous subjects in the analysis (N=174). This hypothesis is supported by the fact that when nulliparous women were included in the model (Table 4.10) the OR per 10 cm^2 of dense tissue was again 1.07 (1.00,1.15), after adjusting for non-dense area, consistent with the results obtained from the analyses with the other risk factors.

TABLE 4.7: ORs for risk of developing breast cancer from conditional logistic regression models using mammographic density measures adjusted for use of hormone replacement therapy and age at menarche

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
absolute density	per 10cm^2	261	1.07	1.00	1.15	.044
hrt	Never		1.00	-	-	.977
	Currently or previous		0.99	0.44	2.21	
percent density	per 10%	261	1.11	0.96	1.27	.155
hrt	Never		1.00	-	-	.989
	Currently or previous		1.01	0.44	2.28	
absolute density	per 10cm^2	261	1.08	1.00	1.16	.039
non-dense area	per 10cm^2		1.01	0.97	1.06	.648
hrt	Never		1.00	-	-	.910
	Currently or previous		0.95	0.42	2.17	
percent density	per 10%	261	1.19	0.99	1.43	.063
non-dense area	per 10cm^2		1.04	0.98	1.10	.223
hrt	Never		1.00	-	-	.957
	Currently or previous		0.98	0.43	2.22	
absolute density	per 10cm^2	240	1.08	1.01	1.16	.033
age at menarche	per year		1.10	0.92	1.31	.295
percent density	per 10%	240	1.11	0.97	1.28	.122
age at menarche	per year		1.08	0.91	1.28	.400
absolute density	per 10cm^2	240	1.09	1.01	1.17	.026
non-dense area	per 10cm^2		1.02	0.97	1.06	.509
age at menarche	per year		1.11	0.93	1.33	.247
percent density	per 10%	240	1.23	1.02	1.50	.032
non-dense area	per 10cm^2		1.05	0.99	1.11	.125
age at menarche	per year		1.11	0.93	1.32	.264

TABLE 4.8: ORs for risk of developing breast cancer from conditional logistic regression models using mammographic density measures adjusted for parity

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
absolute density	per 10cm^2	259	1.07	1.00	1.15	.050
parity	0		1.00	-	-	.677
	1		0.98	0.39	2.47	
	2		0.80	0.41	1.55	
	3		1.36	0.61	3.03	
	≥4		0.33	0.07	1.70	
percent density	per 10%	259	1.08	0.94	1.24	.286
parity	0		1.00	-	-	.666
	1		0.99	0.40	2.48	
	2		0.84	0.43	1.63	
	3		1.31	0.59	2.90	
	≥4		0.32	0.06	1.63	
absolute density	per 10cm^2	259	1.08	1.00	1.16	.040
non-dense area	per 10cm^2		1.01	0.97	1.06	.534
parity	0		1.00	-	-	.701
	1		0.96	0.38	2.42	
	2		0.79	0.41	1.54	
	3		1.38	0.62	3.08	
	≥4		0.34	0.07	1.73	
percent density	per 10%	259	1.16	0.96	1.40	.128
non-dense area	per 10cm^2		1.03	0.97	1.10	.269
parity	0		1.00	-	-	.773
	1		0.97	0.39	2.45	
	2		0.84	0.43	1.63	
	3		1.37	0.61	3.07	
	≥4		0.36	0.07	1.84	

TABLE 4.9: ORs for risk of developing breast cancer from conditional logistic regression models using mammographic density measures adjusted for age at first live pregnancy

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
absolute density	per 10cm^2	174	1.10	1.01	1.19	.025
age at 1st pregnancy	per year		1.03	0.97	1.09	.308
percent density	per 10%	174	1.14	0.97	1.34	.104
age at 1st pregnancy	per year		1.03	0.97	1.09	.397
absolute density	per 10cm^2	174	1.10	1.01	1.19	.029
non-dense area	per 10cm^2		1.00	0.95	1.06	.974
age at 1st pregnancy	per year		1.03	0.97	1.09	.310
percent density	per 10%	174	1.20	0.97	1.49	.094
non-dense area	per 10cm^2		1.03	0.96	1.10	.479
age at 1st pregnancy	per year		1.03	0.97	1.09	.412

TABLE 4.10: ORs for risk of developing breast cancer from conditional logistic regression models using mammographic density measures adjusted for age at first live pregnancy, including nulliparous women

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
absolute density	per 10cm^2	292	1.06	.99	1.13	.076
age at 1st pregnancy*	Under 30		1.00	-	-	.414
	30 or more		1.60	.84	3.08	
	Nulliparous		1.22	.69	2.16	
percent density	per 10%	292	1.04	.92	1.18	.534
age at 1st pregnancy*	Under 30		1.00	-	-	.415
	30 or more		1.62	.85	3.10	
	Nulliparous		1.22	.68	2.17	
absolute density	per 10cm^2	292	1.07	1.00	1.15	.048
non-dense area	per 10cm^2		1.02	.98	1.07	.271
age at 1st pregnancy*	Under 30		1.00	-	-	.370
	30 or more		1.58	.82	3.03	
	Nulliparous		1.25	.70	2.22	
percent density	per 10%	292	1.14	.96	1.35	.149
non-dense area	per 10cm^2		1.04	.99	1.10	.143
age at 1st pregnancy*	Under 30		1.00	-	-	.432
	30 or more		1.55	.81	2.98	
	Nulliparous		1.21	.68	2.16	

TABLE 4.11: ORs for risk of developing breast cancer from conditional logistic regression models using mammographic density measures adjusted for the other available risk factors

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
absolute density	per 10 cm^2	143	1.11	1.01	1.21	.025
non-dense area	per 10 cm^2		1.00	.94	1.07	.982
hrt	Never		1.00	-	-	.301
	Current or previous		.47	.12	1.95	
age at menarche	per year		1.06	.82	1.36	.665
age at 1st pregnancy	per year		1.04	.96	1.12	.304
parity	per pregnancy		1.06	.64	1.76	.823

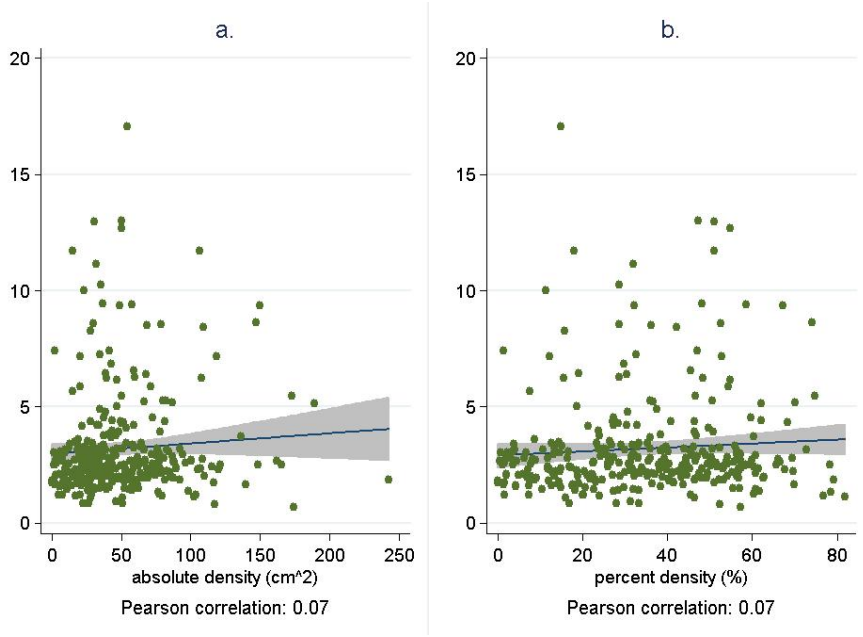
TABLE 4.12: ORs for risk of developing breast cancer from conditional logistic regression models using mammographic density measures adjusted for the other available risk factors, including nulliparous women

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
absolute density	per 10 cm^2	216	1.08	1.00	1.16	.056
non-dense area	per 10 cm^2		1.01	.96	1.06	.660
hrt	Never		1.00	-	-	.180
	Current or previous		.47	.16	1.42	
age at menarche	per year		1.09	.91	1.31	.360
age at 1st pregnancy*	Under 30		1.00	-	-	.958
	30 or more		1.39	.65	2.95	
	Nulliparous		.74	.23	2.41	
parity	per pregnancy		.77	.50	1.20	.252

The inclusion of all risk factors in a conditional logistic regression model (Table 4.11) reduced the number of observations to 143, mostly because of missing values, but also because 83 subjects were dropped by the conditional analyses due to all positive or all negative outcomes. Nevertheless absolute density appeared strongly and significantly associated with breast cancer risk after adjusting for the other available risk factors (non-dense area, use of HRT, age at menarche, age at first pregnancy and number of live births), with an OR of 1.11 (1.01,1.21) per 10 cm^2 of dense tissue. This analysis was repeated including the nulliparous women and results are displayed in Table 4.12. The number of observations increased to 216 and the OR per 10 cm^2 of dense tissue was 1.08 (1.00,1.17), in agreement with the previous results of independence between absolute density and the other risk factors, except for non-dense area.

4.3.3 Mammographic density and Tyrer-Cuzick 10-year risk estimate

FIGURE 4.5: Association between mammographic density measures, (a) absolute and (b) percent, and the Tyrer-Cuzick 10-year personal risk [1]



The relationship between mammographic density measures and Tyrer-Cuzick 10-year personal risk estimates is displayed in Figure 4.5. In both cases there is a negligible and non-significant association.

After adjusting for the Tyrer-Cuzick 10-year relative risk of developing breast cancer (Tables 4.13 and 4.14), the association between absolute density and breast cancer risk was not altered, in that the OR per 10 cm^2 of dense tissue was 1.08 (1.00,1.16) without adjusting for non-dense area and 1.09 (1.01,1.17) after adjustment. This indicates that absolute density added risk information to the Tyrer-Cuzick risk estimates and could potentially improve risk stratification and delineation of high-risk groups.

In the analysis using percent density classified in Boyd density categories, Table 4.14, the association between density and breast cancer risk was slightly weakened by adjustment for Tyrer-Cuzick risk: the ORs ranged between 2.61 and 3.90, and, overall, lost significance.

TABLE 4.13: ORs for risk of developing breast cancer from unconditional logistic regression models using mammographic density measures adjusted for non dense area and Tyrer-Cuzick estimates for 10-year personal risk [1]

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
absolute density	per 10cm^2	292	1.08	1.00	1.16	.038
personal risk	<.06 risk		1.00	-	-	<.001
	≥.06 risk		2.95	1.75	4.98	
percent density	per 10%	292	1.06	.93	1.21	.375
personal risk	<.06 risk		1.00	-	-	<.001
	≥.06 risk		2.93	1.74	4.92	
absolute density	per 10cm^2	292	1.09	1.01	1.17	.024
non-dense area	per 10cm^2		1.02	.98	1.07	.240
personal risk	<.06 risk		1.00	-	-	<.001
	≥.06 risk		3.01	1.78	5.10	
percent density	per 10%	292	1.19	.99	1.42	.062
non-dense area	per 10cm^2		1.05	1.00	1.11	.068
personal risk	<.06 risk		1.00	-	-	<.001
	≥.06 risk		3.01	1.78	5.08	

TABLE 4.14: ORs for risk of developing breast cancer from unconditional logistic regression models using mammographic percent density classified in Boyd's density categories and adjusted for amount of non-dense tissue and Tyrer-Cuzick estimates for 10-year personal risk [1], N=294

Factor	category	OR	[95% Conf.	Interval]	P>z
percent density	<10% dense	1.00	-	-	.079
	10-25% dense	2.61	.87	7.86	
	25-50% dense	3.27	1.01	10.56	
	>50% dense	3.90	1.05	14.52	
non-dense area	per 10 cm^2	1.05	1.00	1.11	.068
personal risk	< .06 risk	1.00	-	-	<.001
	≥ .06 risk	2.79	1.65	4.71	

4.3.4 Results for premenopausal women

As noted above, the majority of women were premenopausal, but there was a minority of postmenopausal women, which appeared to introduce confounding to the relationship between HRT and density, and may have other unforeseen confounding effects.
Logistic regression analyses were therefore repeated on the premenopausal subset (78.5%) of the sample. In the univariate analyses, every additional 10 cm^2 of dense tissue was associated with a 12% increase in odds of breast cancer (OR: 1.12 (1.03,1.22)), Table 4.15. Even after adjusting for other risk factors (Table 4.16), results remained strongly significant and the OR ranged between 1.12 and 1.14.
Among premenopausal women, percent density also had a significant association with breast cancer risk, after adjustment for non-dense area or other risk factors (Tables 4.15 and 4.16). However, it appears that adjusting percent density for non-dense breast area gives similar information as we would obtain from absolute density.

TABLE 4.15: ORs for risk of developing breast cancer from conditional logistic results for density in premenopausal women

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
absolute density	0- 20.77 cm^2	200	1.00	-	-	.026
	20.77-32.41 cm^2		1.59	.51	5.00	
	32.41-49.93 cm^2		2.62	.85	8.10	
	49.93-77.02 cm^2		3.54	1.21	10.36	
	> 77.02 cm^2		2.47	.85	7.18	
absolute density	per 10cm^2	200	1.12	1.03	1.22	.008
percent density	0- 14.58%	200	1.00	-	-	.194
	14.58-29.93%		1.97	.72	5.41	
	29.93-40.86%		1.72	.63	4.67	
	40.86-52.44%		1.16	.44	3.04	
	52.44-100%		2.53	.99	6.49	
percent density	per 10%	200	1.13	0.97	1.31	.108
absolute density	per 10cm^2	200	1.14	1.04	1.24	.005
non-dense area	per 10cm^2		1.03	0.98	1.08	.228
percent density	per 10%	200	1.34	1.07	1.67	.009
non-dense area	per 10cm^2		1.08	1.01	1.15	.033

TABLE 4.16: ORs for risk of developing breast cancer from conditional logistic regression models using mammographic density measures adjusted for other risk factors for breast cancer, in premenopausal women

Density measure	adjusted for	obs	OR	[95% Conf.	Interval]	P>z
Absolute density *per 10 cm^2*	hrt	183	1.13	1.03	1.23	.007
	hrt non-dense area	183	1.14	1.04	1.25	.005
	age at menarche	180	1.12	1.03	1.22	.009
	age at menarche non-dense area	180	1.14	1.04	1.25	.005
	parity	185	1.12	1.03	1.22	.008
	parity non-dense area	185	1.14	1.04	1.25	.005
	age at 1st pregnancy	119	1.16	1.04	1.28	.007
	age at 1st pregnancy non-dense area	119	1.15	1.03	1.29	.011
Percent density *per 10 %*	hrt	183	1.20	1.02	1.41	.030
	hrt non-dense area	183	1.42	1.12	1.79	.003
	age at menarche	180	1.15	0.98	1.34	.078
	age at menarche non-dense area	180	1.37	1.09	1.73	.008
	parity	185	1.13	0.96	1.32	.137
	parity non-dense area	185	1.32	1.05	1.66	.018
	age at 1st pregnancy	119	1.26	1.04	1.53	.020
	age at 1st pregnancy non-dense area	119	1.37	1.04	1.79	.024

On this premenopausal subset, absolute density also added significant risk information to the Tyrer-Cuzick personal risk estimates, with or without adjustment for non-dense area, Table 4.17. The ORs were smaller than those obtained in the previous conditional logistic regression analyses, possibly because of use of standard unconditional analyses. After adjustment for Tyrer-Cuzick risk estimates, percent density had a significant effect only after adjusting for non-dense area which was also significantly associated with risk in this model. This is not surprising, since percent density and non-dense breast area together give information similar to absolute density on its own.

TABLE 4.17: ORs for risk of developing breast cancer from unconditional logistic results for density adjusted for Tyrer-Cuzick estimates for risk [1], in premenopausal women

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
absolute density	per 10cm^2	232	1.09	1.01	1.18	.027
personal risk	<.06 risk		1.00	-	-	.002
	≥.06 risk		2.50	1.40	4.47	
percent density	per 10%	232	1.10	.95	1.28	.189
personal risk	<.06 risk		1.00	-	-	.002
	≥.06 risk		2.51	1.41	4.46	
absolute density	per 10cm^2	232	1.10	1.02	1.19	.019
non-dense area	per 10cm^2		1.02	.98	1.07	.288
personal risk	<.06 risk		1.00	-	-	.002
	≥.06 risk		2.55	1.42	4.58	
percent density	per 10%	232	1.29	1.05	1.59	.017
non-dense area	per 10cm^2		1.07	1.01	1.13	.032
personal risk	<.06 risk		1.00	-	-	.001
	≥.06 risk		2.61	1.45	4.68	

4.4 Discussion

In these analyses with 298 subjects, including 103 breast cancer cases, at enhanced familial risk of breast cancer [188], an association between mammographic density and breast cancer risk was evident, especially when expressing density continuously and in absolute terms (cm^2). This result is consistent with other studies, on the general population, where it has been observed that an absolute measure of density is more predictive of risk than percent density [32, 33, 225–227] [Chapters 5 and 7], possibly because it is

a closer estimate of the amount of stromal and fibroglandular tissue and is less affected by confounding with BMI.

Although adjustment of percent density estimates for non-dense area has been considered a reasonable substitute for adjusting for BMI, interpretation is problematic. Firstly, the structural negative collinearity of percent density and non-dense area may lead to instability of estimates. Secondly, empirically we observed that risk prediction using absolute density unadjusted was similar to that using percent density with adjustment for non-dense areas, both in terms of effect size and information as measured by log-likelihood differences.

A 10 cm^2 increase in dense tissue corresponded to an increase in the odds of about 8%, when having a similar amount of non-dense tissue, and this result was not altered by adjusting for other risk factors. This independence of the association of density and breast cancer risk from other risk factors is consistent with observations in other studies [13, 16, 92].

Another of our results was that, in our high or moderate familial risk population, the risk association was more evident in the premenopausal majority. A possible explanation is the fact that some high risk women experienced a premature menopause because of a hysterectomy and/or oophorectomy, and therefore the menopausal status for these subjects was unclear and thus sometimes classified as "unknown". To support this hypothesis we repeated the conditional logistic regression analyses excluding only the subjects with "unknown" menopausal status. On the remaining 232 subjects, we observed an OR of 1.11 (1.02,1.20) per 10 cm^2 of dense tissue after adjusting for non-dense breast area. The reduction of the sample, from 298 to 232, could help explain this change. Nevertheless, further investigations on larger samples of these under 50, high-risk women with postmenopausal or unclear menopausal status, could provide a clearer explanation.

In our study, age at mammogram did not appear to have an effect on either absolute or percent density, unlike most of the previous literature [91, 228]. This is most likely to be due to the limited age range available in our study, rather than to the enhanced familial risk that characterised the subjects in FH01. However, in another study with volumetric assessment of density [Chapter 6], we observed that absolute density did not vary with increasing age, whereas the percent density did, because of a steady increase with age in non-dense breast volume.

Despite HRT being known to increase mammographic density [101, 103–105], we observed a significant negative association between these two features. This is most likely to be explained by a confounding effect of menopausal status, as HRT is normally prescribed to postmenopausal women and menopause is known to lead to a significant

decrease in density [90, 91]. Also, the age range of the women in this study meant that they were mainly premenopausal.

In this data, age at menarche appeared unrelated either to absolute or percent density. These results are consistent with the weak or null association between density and age at menarche observed in other studies [12, 226, 229, 230]. De Stavola and colleagues [229] observed in a cohort of 5,104 women a significant tendency for later age of menarche to be associated with a denser Wolfe pattern, however this association was only significant among postmenopausal women once adjusted for other risk factors, such as BMI. Likewise, Vachon *et al.* [12] in 2000 conducted a study with 1,900 subjects reaching a similar conclusion, that late age at menarche was associated with increased density, but the magnitude and statistical significance were attenuated in multivariate analyses. If this assumption is real, it is an interesting example of negative confounding. However, a case-control study, comprising 607 cases and 667 controls, from a multi-ethnic population [230] concluded that mammographic density was not related to age at menarche, in agreement with our results. More recently, Haars and colleagues [226] investigated the relationship between density expressed in both absolute and percent terms and other risk factors on a population of 418 postmenopausal women, and observed, as in in our study, no significant tendency of increasing density at increasing age at menarche.

Reproductive factors (age at first birth and parity) were strongly related to percent breast density. This is in agreement with previous literature and consistent with biological mechanisms, since parity leads to changes in breast morphology and biochemistry and it is known that some of these changes are reflected by a reduction in mammographic density [231]. Hence, it is generally observed that late age of first pregnancy is associated with higher breast density, as well as the fact that breast density decreases with increasing number of births [12, 89, 226, 232, 233]. However, mammographic density expressed in absolute rather than percent terms did not appear to be related to these reproductive factors in our data. It is interesting to compare this result with the study from Haars *et al.* [226], as in both cases there is a comparison between proportional and absolute amount of density in association with other factors. In their study, both measures showed an association between density and parity, and both were unrelated to age at first childbirth, unlike in our data. The populations of the two studies differed considerably: theirs comprised only postmenopausal subjects, aged 49.2-65.8 years, whereas ours were under 50, mostly premenopausal women at enhanced familial risk. More recently, Woolcott and colleagues obtained data from four case-control studies (1,699 cases and 2,422 controls) in order to investigate if the association between density, assessed with Cumulus in both absolute and percent terms, and breast cancer risk was modified by reproductive factors [231]. As in Haars's study, analyses showed that percent density

was lower among women with greater parity and did not vary by age at first birth, and results for absolute dense area followed a similar pattern.

One of the goals of this study was to evaluate the ability of breast density to further discriminate cases from controls, after adjusting for Tyrer-Cuzick 10-year risk estimates [1]. Both density measures were found independent from Tyrer-Cuzick 10-year risk-estimates of developing breast tumours. Absolute density, in particular, appeared to add discriminatory ability to the model, both on its own and after adjusting for non-dense breast area. These results further highlight the potential of density in risk management and surveillance policy, and support the suggestion of using it to identify high-risk subjects and to design personalised screening programmes [73].

The cases included in the study were primarily screen detected (80%). Thus, the effect of density on risk may be artificially reduced in this study due to masking, particularly as this is a relatively young population with denser breast tissue than is typical in the National Screening Programme. However, the effect of density is consistently observed in this dataset. When data were restricted to screen-detected cases, the OR we obtained (1.06 (.98-1.14) per 10 cm^2 of dense tissue) did not vary substantially from the original results but it was no longer significant, possibly because of the loss of statistical power due to the sample reduction, from 292 to 239 subjects.

The main weakness of this study is the small number of cases. However, it has the strength of allowing the analyses to focus on this selected high- and moderate-risk population, that, because of their relatively young age, could benefit the most from individually tailored surveillance.

In summary, absolute density aided discrimination between cases and controls even in this population at enhanced familial risk for breast cancer. Moreover, our study provides further evidence that absolute density should be preferred over percent density. Not only was absolute dense area more predictive than percent density, it was not associated with most of the risk factors that percent dense area was, thus making it a simpler risk-predictor and less affected by confounding factors. In addition, absolute density appeared able to improve significantly the risk-prediction provided by Tyrer-Cuzick 10-year risk estimates. Further and larger studies are needed to validate these results and to have a better understanding of how density should be used to provide more effective screening programmes.

Part II

Automated volumetric assessments of mammographic density (Quantra and Volpara)

Chapter 5

A case-control evaluation of a fully automated volumetric density measure (Quantra) as a predictor of breast cancer risk

5.1 Introduction

As noted in Chapter 1, area based breast density assessments measure the amount of breast stromal and epithelial, or dense, tissue that appears white in a mammogram. There are currently two main approaches for measuring density: computer-assisted, e.g. interactive thresholding (*Cumulus*) [30], and visual, e.g. Boyd's six-category classification (SCC) [25, 26] and Breast Imaging Reporting and Data System (*BIRADS*)[234, 235]. The latter was developed in the 1990s by the American College of Radiology, and is a standardized lexicon that classifies density into four categories according to the proportion of dense tissue in the mammogram. It is widely employed because of its ease of use in routine screening, and its reasonably high inter- and intra-observer agreement rates [27]. In addition to this, it has been shown to be a successful risk predictor [2, 3]. Tice and colleagues [3] observed, within a cohort of 81,777 women (5.1 average years of follow-up, 955 cases), that a model based on BIRADS alone, adjusted for age and ethnicity, could predict breast cancer risk as accurately as the Gail model [161, 162], a multivariable statistical model commonly implemented for this purpose. Similarly, the results of Barlow *et al.* [2] suggested that, as a single predictor,

BIRADS is almost as powerful as age. However, the dense area on the mammogram is a two-dimensional representation of a three-dimensional feature, the volume of fibroglandular tissue in the breast; hence, in principle, a volumetric measure is preferable. Dense volume is expected to be more directly related to risk and to capture more accurately the underlying biological processes that lead to breast cancer, since it is believed to be a more precise estimate of the amount of fibroglandular tissue in the breast [30].

It has long been thought that the introduction of a calibration object, beside the breast during imaging, would allow the projected dense area to be directly related to the thickness and the composition of the breast tissue, and thereby provide an estimate of the fibroglandular volume [30, 236]. Recent fully-automated volumetric approaches have used X-ray attenuation at each pixel to estimate the amount of fibroglandular tissue it represents, and hence estimate both the overall and the dense volume [38]. One example is Quantra, a software package that analyses raw images from full-field digital mammograms (FFDM) to quantify dense tissue. This method appears promising according to the results presented at conferences [38, 48, 237–244], and recent publications [51, 178], partially in terms of reproducibility, but there is no certainty that this will be the case in a risk-prediction context.

Disappointingly, other automated volumetric approaches proposed so far [8, 36, 245] seem to provide similar or weaker risk prediction than area-based measures. Ding and colleagues [245] compared the standard mammography form (SMF), a software tool for volumetric assessment, to Wolfe's parenchymal patterns, visual percent density (21-point scale) and interactive thresholding. Their data suggested that the volumetric measurements had a lesser association with breast cancer risk than interactive thresholding, the current gold standard [28]. The predictive ability of SMF was also investigated by Aitken *et al.* [36] within a case-control study comprising 969 subjects (355 cases), that led to similar discouraging results. Likewise, findings from Boyd and colleagues's comparison of area-based Cumulus with the new volumetric version (*Cumulus V*) [8] indicated that the volumetric measure provided similar risk information to that of the area method. These results highlighted the need to develop new techniques or to refine current algorithms, and to validate the novel approaches using larger datasets. The pursuit of a fully automated density measurement continues, as density could only be used for risk management at a population level if it could be measured automatically.

In this study, we examine the relationship between Quantra and BIRADS and evaluate their performance in predicting risk of breast cancer, and also compare Quantra absolute and percent measurements of dense volume.

5.2 Material and Methods

5.2.1 Data collection

We conducted a case-control study, whose participants were women attending the Princess Grace Hospital, an independent hospital in central London, who had full-field digital mammographic examinations in 2005-2009. Two hundred symptomatic cases were randomly selected from patients with histopathologically verified breast cancer, and each case was matched with a healthy control by date of birth, age at examination and laterality of mammogram used for density determination.

All mammograms were high quality two-view images, soft copy reported on high resolution monitors. Image quality was not a factor in case selection before randomization, and because all images were adequate, none were rejected during the density estimation. Among cases, density was assessed on the last mammogram of the contralateral breast before cancer detection; for controls, density was estimated on the same side as the matched case, using mammograms dating as close as possible to the matched case diagnosis. Both views were used for assessing density, as it is known that the craniocaudal view tends to give higher density estimates than the mediolateral oblique view [4]. The area-based density was estimated by two experienced radiologists (Dr. Nick Perry and Dr. Katja Pinker-Domenig) using the standard density grades according to the American College of Radiology Breast Imaging Reporting and Data System (BIRADS) [234]. This system classifies density into four categories: (A) $< 25\%$ *dense (almost entirely fatty)*; (B) $25 - 50\%$ *dense (scattered fibroglandular density)*; (C) $51 - 75\%$ *dense (hetereogeneously dense)*; (D) $> 75\%$ *dense (extremely dense)* [246].

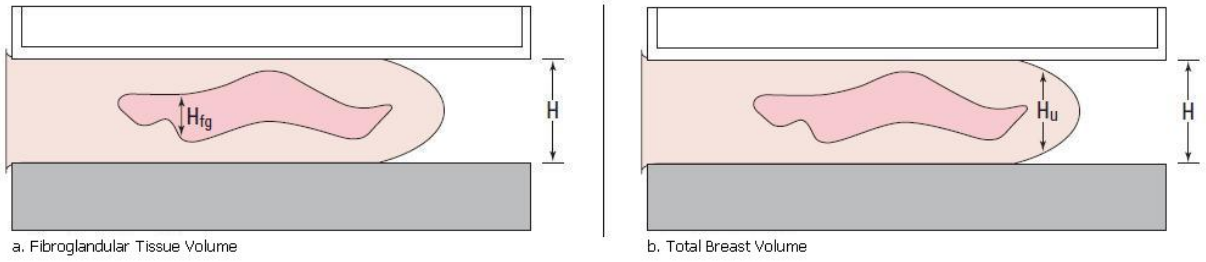
Breast density was also assessed using an integral automated volumetric breast density measurement system (Hologic, Quantra). Further details of Quantra are provided in the next section.

As previous results suggested higher density in women resident in London [246], we also recorded data on area of residence. Area of residence was classified as London, any other region in the UK (based on the post-code of residence), or outside the UK, referred to as "abroad" for brevity. In addition to age, area of residence, and density assessed with both Quantra and BIRADS, no other data on breast cancer risk factors was available. Complete information was collected for every variable in the study (breast density and total volume estimates, age and area of residence) thus there were no missing values.

5.2.2 Quantra

Quantra is a software application that quantifies the total volume of the breast, its fibroglandular, *dense*, component (*absolute density*) in cm^3, and proportional breast density as a ratio of the two (*percent density*).
Quantra accepts and evaluates raw image data, produced by Full-Field Digital Mammography (FFDM). It evaluates density from even just a single view and can use either cranio-caudal or mediolateral oblique images. If both views are available the software estimates density as an average from both images. In our study the two views of the contralateral breast were taken into account for each case, and the same side was chosen for the matched control.
The algorithm is based on a physical model of the X-ray imaging chain that relates breast tissue X-ray attenuation to the digital mammography images provided. It estimates the amount of fibroglandular tissue that an X-ray must have penetrated in order to deposit a measured amount of energy at the detector. Thus Quantra provides a result in centimetres of fibroglandular tissue penetrated (H_{fg}) at each pixel of the image (Figure 5.1.a). After completing the analysis of every pixel, all the H_fg are aggregated into the volume of fibroglandular tissue (absolute density), given in cm^3.

FIGURE 5.1: Quantra volumetric assessment of fibroglandular tissue (a) and total breast volume (b)



Source: Understanding R2 Quantra 1.3, Hologic

Likewise, for the total volume, Quantra considers the entire outline of the imaged breast, compensating for the decrease in thickness in the uncompressed portions (H_u), Figure 5.1.b. The percent density is then computed as a ratio.
For further information on Quantra algorithm:
http://www.hologic.com/en/breast-screening/volumetric-assessment/

5.2.3 Statistical analyses

Demographic and other variables are summarised using percentages for categorical variables or medians and IQRs for continuous variables. We had data on age, area of residence classified as "London", "Other UK regions" and "abroad", total breast volume, dense volume, non-dense volume and BIRADS classification. Total breast volume and non-dense volume were used as potential indirect indicators of BMI, which was not available, and age under 50 versus greater than or equal to 50 was used as a surrogate for menopausal status.

Conditional logistic regression was used to calculate odds ratios (OR) for risk of breast cancer after adjustment for potential confounding factors. This analysis was repeated after log-transformation of Quantra estimates. Secondary analyses were also performed stratified by area of residence (London vs Other UK), and age.

The relationship between BIRADS and Quantra was investigated graphically and non-parametric tests [214] were used to test for differences for trend across ordered groups. All analyses were performed using STATA, version 12.1.

5.3 Results

Characteristics of the 200 cases and 200 controls are summarised in Table 5.1. The matching algorithm, based on date of birth, age at examination and laterality of mammogram used for density determination, was successful as the distribution of these factors was the same for each group. There was no difference between cases and controls in the proportions of women from London, from other British regions, and from abroad. The distribution of total breast volume, dense volume (absolute density) and percent density were compared between the case and control groups using the Wilcoxon rank-sum test. Results indicated that there was no significant difference between cases and controls in both total breast volume and percent density, $p = .15$ and $p = .40$ respectively, whereas cases seemed to have systematically greater dense volume in terms of cm^3 ($p = .01$). Figure 5.2 supports this result, as controls tended to group at lower levels than cases for absolute density, whereas this trend was less evident for percent density.

TABLE 5.1: Characteristics of the study samples

	Cases N=200		Controls N=200		Total N=400	
Age at mammogram, No.(%)						
< 45	35	(17.5)	35	(17.5)	70	(17.5)
45-54	63	(31.5)	63	(31.5)	126	(31.5)
55-64	53	(26.5)	52	(26.0)	105	(26.2)
≥ 65	49	(24.5)	50	(25.0)	99	(24.8)
Area of residence, No. (%)						
London	102	(51.0)	107	(53.5)	209	(52.2)
Other UK	81	(40.5)	78	(39.0)	159	(39.8)
Abroad	15	(7.5)	17	(8.5)	32	(8.0)
Measures of mammographic breast density assessed by Quantra, Median (IQR)						
Percent breast density, in %	20	(14.0)	19.5	(13.0)	20	(13.5)
Absolute breast density, in cm^3	91	(76.5)	75.5	(60.5)	82	(73.5)
Total breast volume, in cm^3	441.5	(399.5)	407.5	(347.0)	422.5	(392.5)

FIGURE 5.2: Distribution of (a) percent and (b) absolute breast density in cases and controls

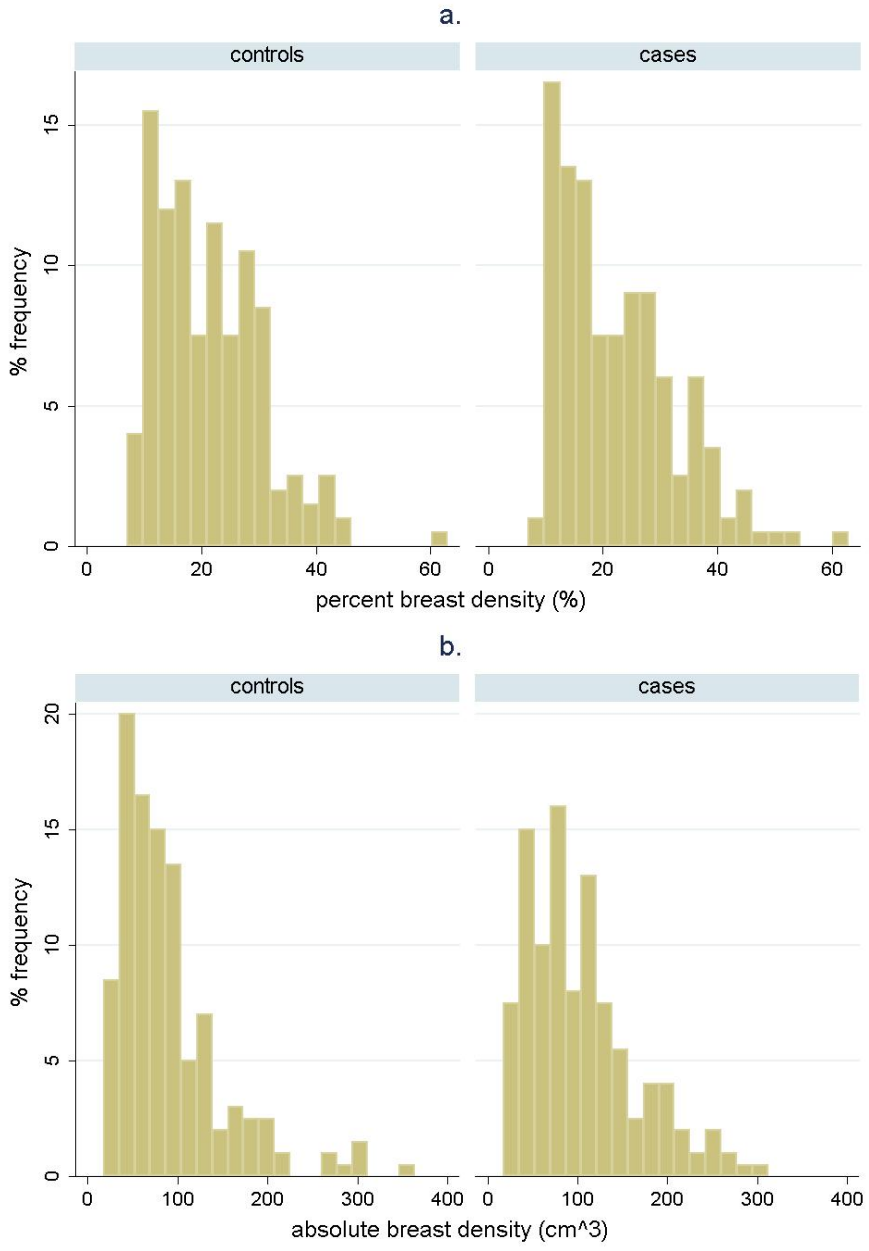


FIGURE 5.3: Distribution of BIRADS dense categories in cases and controls, (a) overall and (b) stratified by age

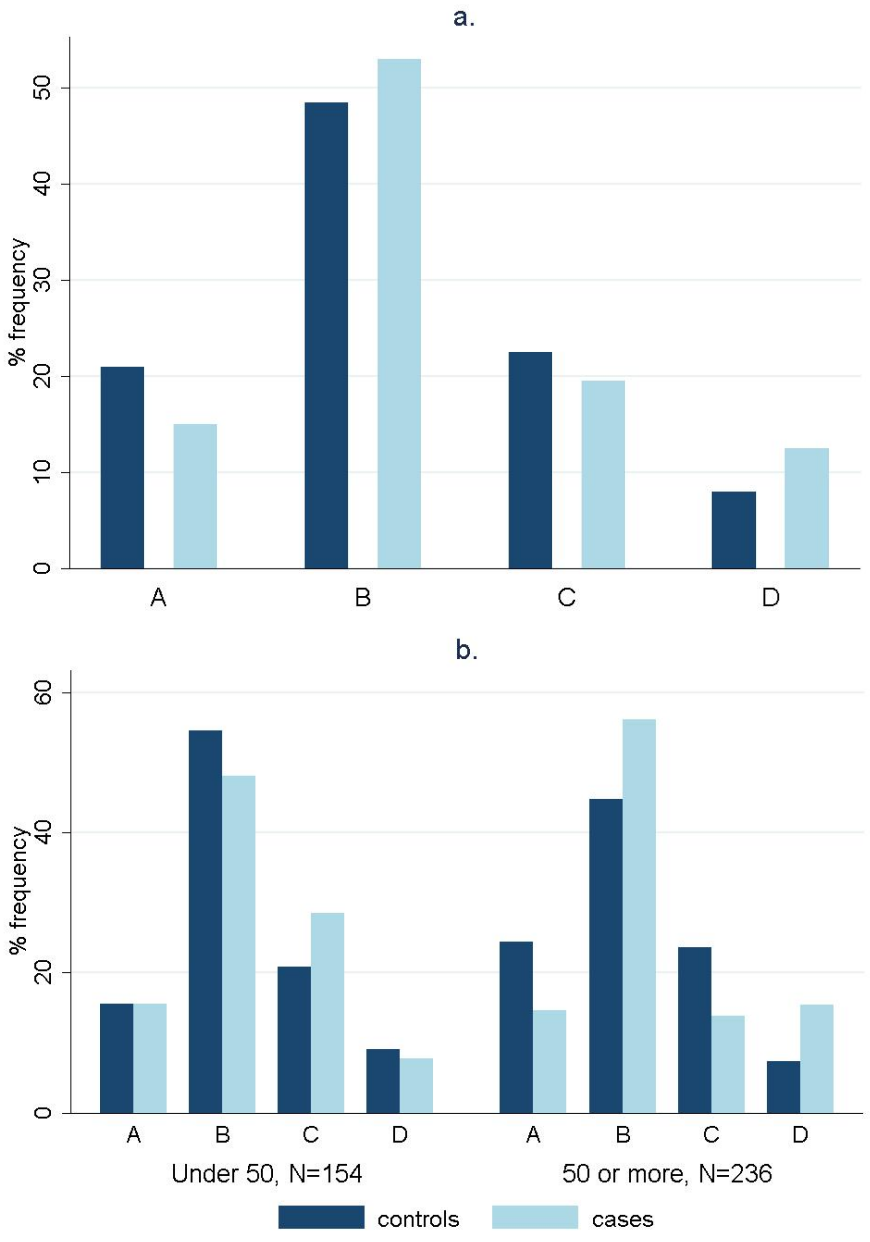


Figure 5.3 depicts the distribution of BIRADS classification between cases and controls. It shows a pattern of a smaller proportion of cases in the lowest density category, and a greater proportion in the highest. This was more pronounced in women aged 50 or more (Figure 5.3.b).

We also investigated the association of percent and absolute density, assessed using Quantra, with the risk of breast cancer (Tables 5.2 and 5.3). The odds ratios related to

absolute density and non-dense volume are expressed per 10 cubic centimetre increase, whereas those regarding percent density are given per 10% increase. Comparing the performance of risk prediction of the two variables derived by the Quantra system (Tables 5.2 and 5.3), the outcomes suggest that only absolute breast density had a significant univariate main effect (p=.03), and that an increase of 10 cm^3 [1] in fibroglandular volume would lead to a 4% increase in risk of breast cancer. The introduction of the other risk factors (non-dense volume and area of residence) to the model did not substantially alter these results, whereas the observed effect of Quantra percent density was strengthened by adjustment for these factors, OR 1.47 (95% CI (1.09, 1.98)) per 10% density (Table 5.2). In both cases, likelihood ratio tests indicated that the fit of the model did not improve significantly from the inclusion of other risk factors.

The introduction of area of residence in the model caused a reduction in sample size for two reasons: (a) 32 women from abroad were excluded from the analysis in order to study the urban effect of London against other areas in the UK; (b) a further 28 observation were dropped by the analyses because of all positive or all negative outcomes.

TABLE 5.2: ORs for developing breast cancer from conditional logistic regression models using Quantra percent density

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
univariate						
% density	per 10%	400	1.15	.92	1.44	.211
adjusted for non-dense volume						
% density	per 10%	400	1.29	1.00	1.66	.050
non-dense volume	per 10 cm^3		1.01	1.00	1.02	.055
adjusted for London residence						
% density	per 10%	340	1.31	1.01	1.70	.040
residence	Other UK		1.00	-	-	.660
	London		.90	.56	1.44	
adjusted for non-dense volume and London residence						
% density	per 10%	340	1.47	1.09	1.98	.011
non-dense volume	per 10 cm^3		1.01	1.00	1.02	.081
residence	Other UK		1.00	-	-	.864
	London		.96	.60	1.54	

[1] The average absolute density is 97.31 cm^3.

TABLE 5.3: ORs for developing breast cancer from conditional logistic regression models using Quantra absolute density

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
univariate						
absolute density	per 10 cm^3	400	1.04	1.00	1.08	.030
adjusted for non-dense volume						
absolute density	per 10 cm^3	400	1.04	1.00	1.09	.063
non-dense volume	per 10 cm^3		.99	.99	1.01	.719
adjusted for London residence						
absolute density	per 10 cm^3	340	1.04	1.00	1.08	.062
residence	Other UK		1.00	-	-	.956
	London		.99	.62	1.58	
adjusted for non-dense volume and London residence						
absolute density	per 10 cm^3	340	1.05	1.00	1.10	.063
non-dense volume	per 10 cm^3		1.00	.99	1.01	.480
residence	Other UK		1.00	-	-	.869
	London		.96	.60	1.54	

Table 5.4 shows the ORs associated with quintiles of absolute dense volume, based on the control distribution, in order to further understand the linear association between Quantra absolute density and risk of developing breast cancer. According to these results, the presence of at least 125 cm^3 of dense tissue in the breast would correspond to a 77% increase in breast cancer relative risk, compared to subjects with less than 47 cm^3 of dense tissue. The effect on breast cancer risk of absolute density categorised in this way was significant (p=.02).

TABLE 5.4: ORs for developing breast cancer from a conditional logistic regression model using the quintiles of Quantra absolute density among controls

Quintile	OR	[95% Conf.	Interval]	P>z
1st quintile (< 47 cm^3)	1.00	-	-	.020
2nd quintile ($47 - 65$ cm^3)	.78	.39	1.55	
3rd quintile ($66 - 87$ cm^3)	1.08	.56	2.06	
4th quintile ($88 - 124$ cm^3)	1.26	.68	2.35	
5th quintile (≥ 125 cm^3)	1.77	.95	3.29	

Results of the analyses using Quantra measures after log-transformation are summarised in Table 5.5. The evidence sustaining the association between breast cancer risk and absolute density appears stronger in comparison with the results obtained using the original measures. However, they led to similar conclusions to those of Tables 5.2 and 5.3, and we therefore returned the original measures for the following analyses, this also facilitates the interpretation of the results. Nevertheless, in the analysis based on the log-transformed densities, the chi-square statistics demonstrate that adjustment of the percent density effect for total breast volume is returning the same information as absolute density but at the cost of adding a degree of freedom.

TABLE 5.5: ORs for developing breast cancer risk from conditional logistic regression models using Quantra log-transformed measures

Factor	obs	OR	[95% Conf.	Interval]	P>z	modelχ^2	df
univariate							
log-absolute density	400	1.50	1.07	2.11	.020	5.62	1
adjusted for log-non-dense volume							
log-absolute density	400	1.76	1.08	2.85	.022	6.47	2
log-non-dense volume		.83	.56	1.24	.358		
adjusted for log-total volume							
log-absolute density	400	1.83	1.03	3.25	.038	6.38	2
log-total volume		.79	.47	1.34	.386		
univariate							
log-% density	400	1.35	.82	2.24	.236	1.42	1
adjusted for log-non-dense volume							
log-% density	400	1.86	1.02	3.40	.044	5.20	2
log-non-dense volume		1.39	.99	1.94	.055		
adjusted for log-total volume							
log-% density	400	1.73	.98	3.03	.057	5.66	2
log- total volume		1.43	1.01	2.01	.042		

Table 5.6 reports the results from the analyses on the association between risk of developing a tumour in the breast and percent breast density classified according to BIRADS categories. The relationship between BIRADS density and risk is a non-linear one, nevertheless the risk of developing breast cancer in subjects classified as "extremely dense" (D) is more than two-fold higher than women defined "almost entirely fatty" (A). The

univariate OR is estimated to be 2.26 (1.01, 5.05), and the effect is stronger when adjusting for area of residence (OR 2.72 (1.12, 6.62)).

BIRADS and absolute Quantra measures assess breast density in different ways: the first gives a subjective evaluation of the dense tissue area in terms of 25% intervals, whereas the second is an automatic system that estimates the volume of fibroglandular tissue in cm^3. Bearing this in mind, a comparison between the risk-predictive performances of the two variables must be tentative (Tables 5.3 and 5.6).

TABLE 5.6: ORs for developing breast cancer from conditional logistic regression models using BIRADS classification

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
univariate						
BIRADS	(A) <25% dense	400	1.00	-	-	.161
	(B) 25-50% dense		1.57	.90	2.73	
	(C) 50-75% dense		1.24	.66	2.35	
	(D) ≥75% dense		2.26	1.01	5.05	
adjusted for non-dense volume						
BIRADS	(A) <25% dense	400	1.00	-	-	.188
	(B) 25-50% dense		1.57	.90	2.73	
	(C) 50-75% dense		1.24	.65	2.34	
	(D) ≥75% dense		2.18	.97	4.88	
non-dense volume	per 10 cm^3		1.00	1.00	1.01	.276
adjusted for London residence						
BIRADS	(A) <25% dense	340	1.00	-	-	.057
	(B) 25-50% dense		1.56	.86	2.84	
	(C) 50-75% dense		1.42	.71	2.84	
	(D) ≥75% dense		2.78	1.15	6.74	
residence	Other UK		1.00	-	-	.825
	London		.96	.59	1.54	
adjusted for non-dense volume and London residence						
BIRADS	(A) <25% dense	340	1.00	-	-	.065
	(B) 25-50% dense		1.57	.86	2.85	
	(C) 50-75% dense		1.42	.71	2.84	
	(D) ≥75% dense		2.72	1.12	6.62	
non-dense volume	per 10 cm^3		1.00	.99	1.01	.594
residence	Other UK		1.00	-	-	.902
	London		.98	.60	1.60	

Conditional logistic regression analyses were repeated in women aged under 50 years and 50 or more separately (Tables 5.7,5.8 and 5.9). Quantra percent density was significant in discriminating cases from controls in the younger group, particularly when the model took non-dense volume into account, with an OR per 10% density of 1.61 (1.05, 2.48) (Table 5.7). Conversely, Quantra absolute density was not significant in the model based only on women aged under 50 (OR per 10 cm^3: 1.02 (.97, 1.07)), but there was a considerate loss of power as the sample size reduced from 400 to 154 (Table 5.8). This result is partly explained by the dependency of percent density on non-dense volume, as Figure 5.4 illustrates. At younger ages, cases tend to have a higher percent density not because of a greater absolute volume of dense tissue, but due to a generally smaller amount of non-dense tissue than controls. Furthermore, among women aged under 50, there was no substantial difference in breast cancer risk in subjects belonging to the two extreme categories of BIRADS classification, i.e. "almost entirely fatty" (A) and "extremely dense" (D). At older ages, on the other hand, breasts classified as "extremely dense" have 4.12 (1.41, 12.07) times higher odds of cancer than those classified as "almost entirely fatty" (Table 5.9).

TABLE 5.7: ORs for developing breast cancer from conditional logistic regression models using Quantra percent density, stratified by age

Factor	category	obs	OR	[95 Conf.	Interval]	P>z
univariate						
Under 50						
% density	per 10%	154	1.46	1.01	2.10	.045
50 or more						
% density	per 10%	246	.97	.72	1.31	.850
adjusted for non-dense volume						
Under 50						
% density	per 10%	154	1.61	1.05	2.48	.029
non-dense volume	per 10 cm^3		1.01	.99	1.02	.336
50 or more						
% density	per 10%	246	1.09	.78	1.52	.607
non-dense volume	per 10 cm^3		1.01	1.00	1.02	.107

FIGURE 5.4: Comparison of medians and inter-quartile ranges of Quantra (a) percent density, (b) non-dense volume and (c) absolute density by age group, between cases and controls

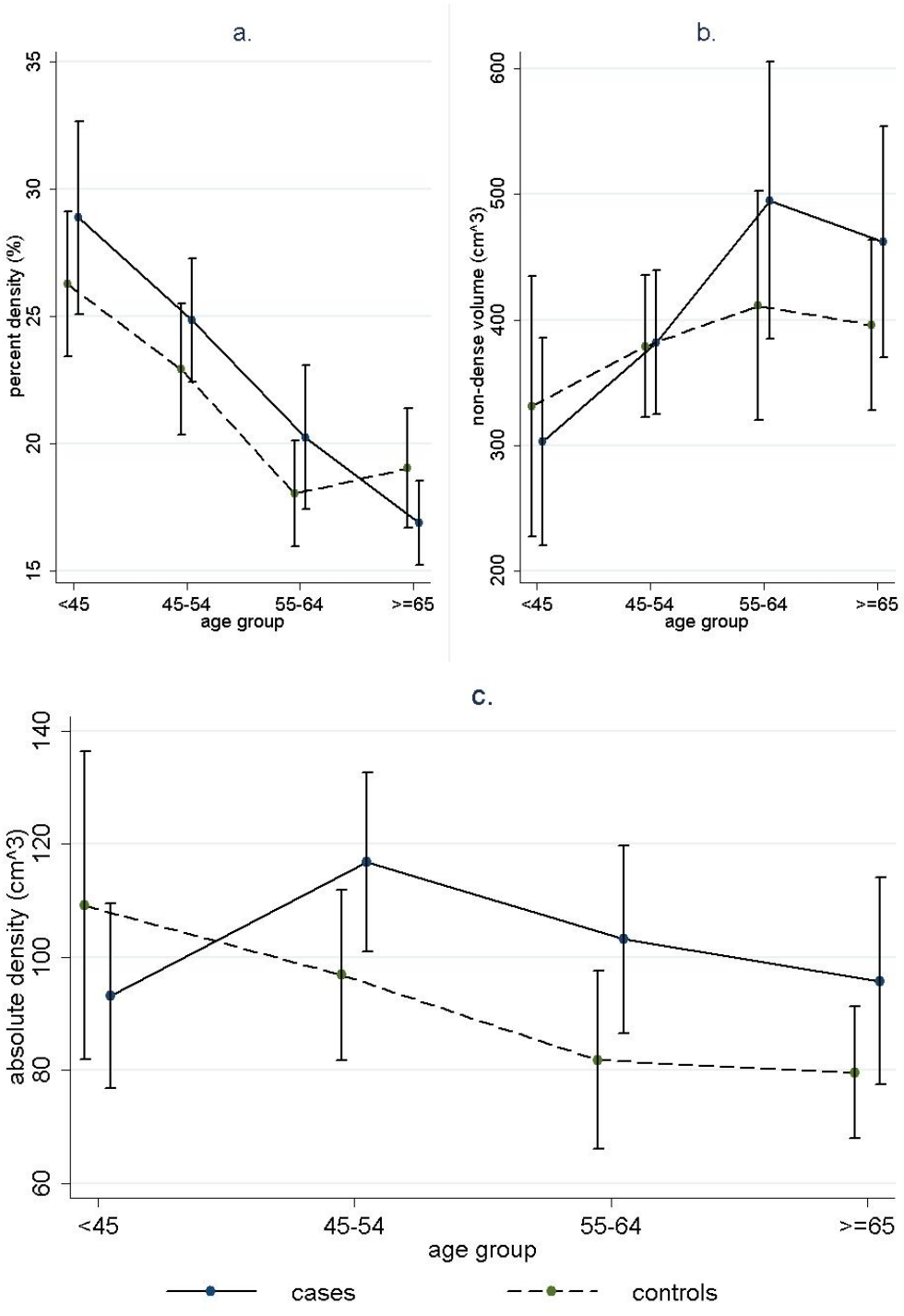


TABLE 5.8: ORs for developing breast cancer from conditional logistic regression models using Quantra absolute density, stratified by age

Factor	category	obs	OR	[95 Conf.	Interval]	P>z
univariate						
Under 50						
absolute density	per 10 cm^3	154	1.02	.97	1.07	.493
50 or more						
absolute density	per 10 cm^3	246	1.06	1.01	1.11	.022
adjusted for non-dense volume						
Under 50						
absolute density	per 10 cm^3	154	1.04	.97	1.11	.279
non-dense volume	per 10 cm^3		.99	.98	1.01	.384
50 or more						
absolute density	per 10 cm^3	246	1.06	.99	1.13	.083
non-dense volume	per 10 cm^3		1.00	.99	1.01	.984

TABLE 5.9: ORs for developing breast cancer from conditional logistic regression models using BIRADS classification, stratified by age

Factor	category	obs	OR	[95 Conf.	Interval]	P>z
univariate						
Under 50						
BIRADS	(A) <25% dense	154	1.00	-	-	.692
	(B) 25-50% dense		.89	.34	2.36	
	(C) 50-75% dense		1.51	.48	4.74	
	(D) ≥75% dense		.88	.24	3.29	
50 or more						
BIRADS	(A) <25% dense	246	1.00	-	-	.148
	(B) 25-50% dense		2.24	1.11	4.54	
	(C) 50-75% dense		1.10	.50	2.44	
	(D) ≥75% dense		4.12	1.41	12.07	
adjusted for non-dense volume						
Under 50						
BIRADS	(A) <25% dense	154	1.00	-	-	.707
	(B) 25-50% dense		.87	.32	2.35	
	(C) 50-75% dense		1.48	.47	4.70	
	(D) ≥75% dense		.86	.23	3.27	
non-dense volume	per 10 cm^3		1.00	.99	1.01	.891
50 or more						
BIRADS	(A) <25% dense	246	1.00	-	-	.233
	(B) 25-50% dense		2.16	1.06	4.39	
	(C) 50-75% dense		1.05	.47	2.34	
	(D) ≥75% dense		3.67	1.24	10.89	
non-dense volume	per 10 cm^3		1.01	1.00	1.02	.188

TABLE 5.10: ORs for developing breast cancer from conditional logistic regression models using Quantra percent density, stratified by UK location of residence

Factor	category	obs	OR	[95 Conf.	Interval]	P>z
univariate						
London						
% density	per 10%	124	1.25	.83	1.86	.283
Other UK						
% density	per 10%	72	1.49	.91	2.44	.112
adjusted for non-dense volume						
London						
% density	per 10%	124	1.41	.89	2.23	.146
non-dense volume	per 10 cm^3		1.01	.99	1.02	.236
Other UK						
% density	per 10%	72	2.02	1.07	3.82	.030
non-dense volume	per 10 cm^3		1.03	1.00	1.06	.046

TABLE 5.11: ORs for developing breast cancer from conditional logistic regression models using Quantra absolute density, stratified by UK location of residence

Factor	category	obs	OR	[95 Conf.	Interval]	P>z
univariate						
London						
absolute density	per 10 cm^3	124	1.03	.98	1.09	.231
Other UK						
absolute density	per 10 cm^3	72	1.11	1.00	1.22	.050
adjusted for non-dense volume						
London						
absolute density	per 10 cm^3	124	1.04	.97	1.12	.291
non-dense volume	per 10 cm^3		1.00	.98	1.02	.824
Other UK						
absolute density	per 10 cm^3	72	1.10	.98	1.24	.121
non-dense volume	per 10 cm^3		1.00	.98	1.03	.890

In the analyses stratified by area of residence (Tables 5.10, 5.11 and 5.12), 32 women living abroad were excluded in order to focus on within-UK differences. A further 172 observations, 85 in the "London" group and 87 in the "Other UK" group, did not contribute to the matched statistical analysis due to all positive or all negative outcomes. Therefore the reduced sample sizes affect the precision of the results. However, Quantra

TABLE 5.12: ORs for developing breast cancer from conditional logistic regression models using BIRADS classification, stratified by UK location of residence

Factor	category	obs	OR	[95 Conf.	Interval]	P>z
univariate						
London						
BIRADS	(A) <25% dense	124	1.00	-	-	.025
	(B) 25-50% dense		4.18	1.54	11.34	
	(C) 50-75% dense		1.61	.50	5.25	
	(D) ≥75% dense		12.14	2.06	71.38	
Other UK						
BIRADS	(A) <25% dense	72	1.00	-	-	.896
	(B) 25-50% dense		.40	.07	2.20	
	(C) 50-75% dense		.40	.07	2.37	
	(D) ≥75% dense		.70	.09	5.46	
adjusted for non-dense volume						
London						
BIRADS	(A) <25% dense	124	1.00	-	-	.023
	(B) 25-50% dense		4.21	1.54	11.50	
	(C) 50-75% dense		1.64	.50	5.35	
	(D) ≥75% dense		12.19	2.07	71.72	
non-dense volume	per 10 cm^3		1.00	.99	1.02	.484
Other UK						
BIRADS	(A) <25% dense	72	1.00	-	-	.915
	(B) 25-50% dense		.35	.06	2.13	
	(C) 50-75% dense		.29	.04	1.90	
	(D) ≥75% dense		.57	.07	4.92	
non-dense volume	per 10 cm^3		1.01	.99	1.04	.175

densities, both absolute and percent (the latter after adjustment for non-dense breast volume) seemed to be able to discriminate better cases from controls in British women not living in London (Tables 5.10 and 5.11), whereas BIRADS density registered strongly increasing ORs in London residents (Table 5.12).

We repeated these stratified analyses using unmatched logistic models adjusted for age, in order to preserve a larger portion of the sample (Tables 5.13, 5.14 and 5.15). These unconditional analyses relating to Quantra density estimates confirmed the previous results. In the models with BIRADS, on the other hand, the larger sample, 368 subjects instead of 196, led to less extreme results, especially among London residents, i.e. the OR for the most dense breasts in comparison to the primarily fatty ones ("<25% dense") was

3.5 (1.3, 9.6) instead of 12.1 (2.1, 71.4), as observed in the conditional logistic analyses.

TABLE 5.13: ORs for developing breast cancer from unconditional logistic regression models using Quantra percent density adjusted for age, stratified by UK area of residence

Factor	category	obs	OR	[95 Conf.	Interval]	P>z
adjusted for age						
London						
% density	per 10%	209	1.07	.80	1.44	.632
age	per year		1.01	.99	1.04	.386
Other UK						
% density	per 10%	159	1.46	1.00	2.14	.050
age	per year		1.00	.97	1.03	.792
adjusted for non-dense volume and age						
London						
% density	per 10%	209	1.17	.84	1.63	.366
non-dense volume	per 10 cm^3		1.01	.99	1.02	.301
age	per year		1.01	.99	1.04	.343
Other UK						
% density	per 10%	159	1.59	1.05	2.40	.027
non-dense volume	per 10 cm^3		1.01	1.00	1.02	.270
age	per year		1.00	.96	1.03	.763

TABLE 5.14: ORs for developing breast cancer from unconditional logistic regression models using Quantra absolute density adjusted for age, stratified by UK area of residence

Factor	category	obs	OR	[95 Conf.	Interval]	P>z
adjusted for age						
London						
absolute density	per 10 cm^3	209	1.09	.99	1.08	.177
age	per year		1.01	.99	1.04	.361
Other UK						
absolute density	per 10 cm^3	159	1.04	.98	1.10	.181
age	per year		.99	.96	1.01	.318
adjusted for non-dense volume and age						
London						
absolute density	per 10 cm^3	209	1.04	.98	1.10	.236
non-dense volume	per 10 cm^3		1.00	.98	1.01	.803
age	per year		1.01	.99	1.04	.346
Other UK						
absolute density	per 10 cm^3	159	1.06	.98	1.15	.117
non-dense volume	per 10 cm^3		.99	.98	1.01	.370
age	per year		.99	.96	1.02	.511

TABLE 5.15: ORs for developing breast cancer from unconditional logistic regression models using BIRADS classification adjusted for age, stratified by UK area of residence

Factor	category	obs	OR	[95 Conf.	Interval]	P>z
adjusted for age						
London						
BIRADS	(A) <25% dense	209	1.00	-	-	.056
	(B) 25-50% dense		2.43	1.16	5.10	
	(C) 50-75% dense		1.67	.70	4.00	
	(D) >75% dense		3.46	1.25	9.56	
age	per year		1.01	.99	1.04	.378
Other UK						
BIRADS	(A) <25% dense	159	1.00	-	-	.726
	(B) 25-50% dense		.80	.30	2.13	
	(C) 50-75% dense		.95	.31	2.85	
	(D) >75% dense		1.13	.28	4.51	
age	per year		.98	.96	1.01	.264
adjusted for non-dense volume and age						
London						
BIRADS	(A) <25% dense	209	1.00	-	-	.059
	(B) 25-50% dense		2.40	1.14	5.05	
	(C) 50-75% dense		1.66	.69	3.99	
	(D) >75% dense		3.41	1.23	9.46	
non-dense volume	per 10 cm^3		1.00	.99	1.01	.524
age	per year		1.01	.99	1.04	.402
Other UK						
BIRADS	(A) <25% dense	159	1.00	-	-	.748
	(B) 25-50% dense		.80	.30	2.13	
	(C) 50-75% dense		.94	.31	2.84	
	(D) >75% dense		1.11	.28	4.45	
non-dense volume	per 10 cm^3		1.00	.99	1.01	.764
age	per year		.98	.95	1.01	.249

Figure 5.5 compares medians of Quantra measures by case-control status and age group, stratified by area of residence and may help to explain such result. In London residents, case and control lines are relatively close, especially for median percent density, implying poor discrimination. Moreover, comparing the controls from London and other UK regions, subjects from the capital appear to have a higher percent density, even presenting similar absolute density estimates, because they have lower non-dense volume. The Wilcoxon rank-sum test comparing non-dense breast volume among women living in London or elsewhere in the UK supports this result (p=.03). This may mean that London women have lower percent density because they are thinner than women resident in other UK regions.

FIGURE 5.5: Comparison of medians and inter-quartile ranges of Quantra (a) percent density, (b) non-dense volume and (c) absolute density by age group, between UK residents from London or other British regions, stratified by case-control status

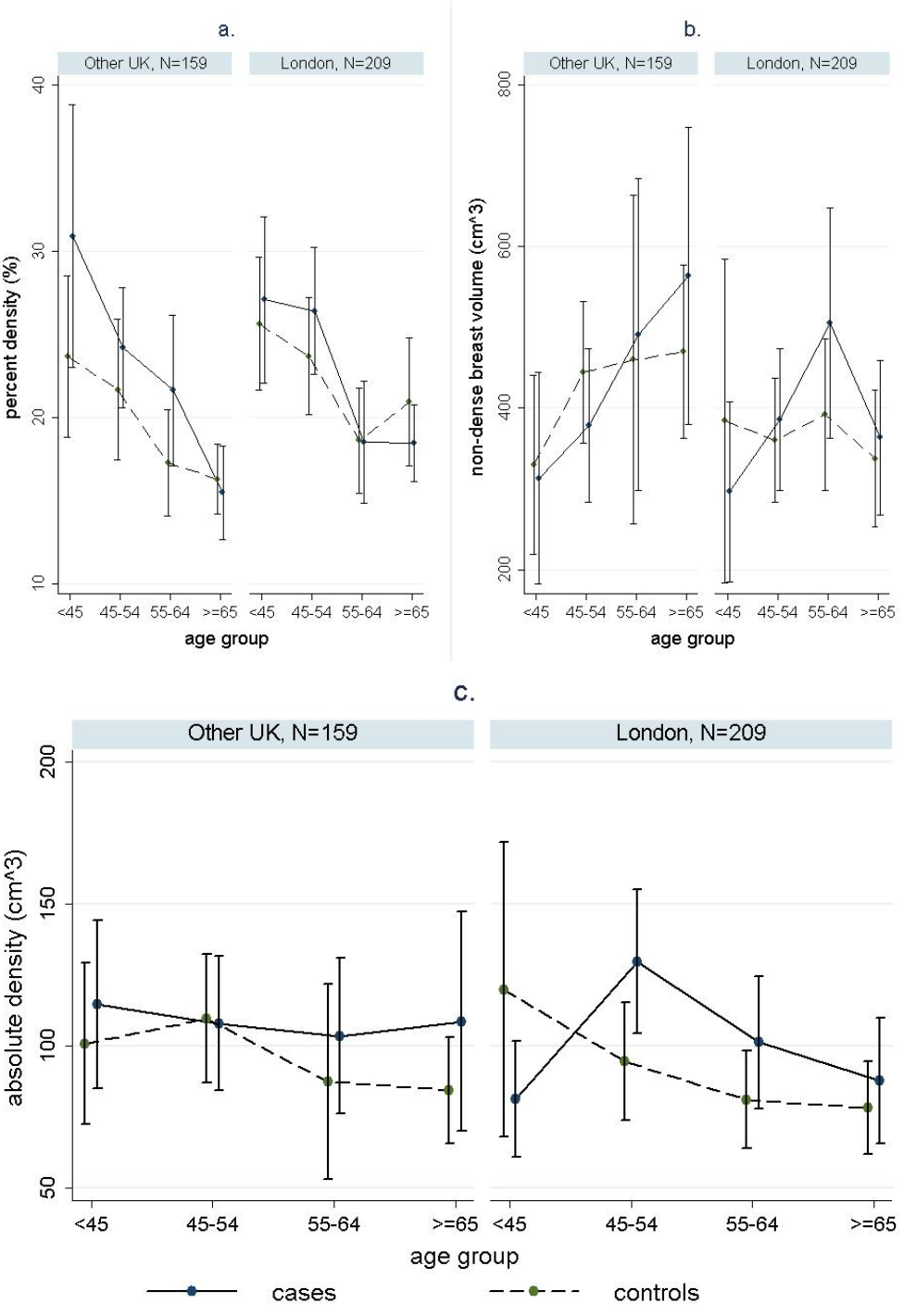
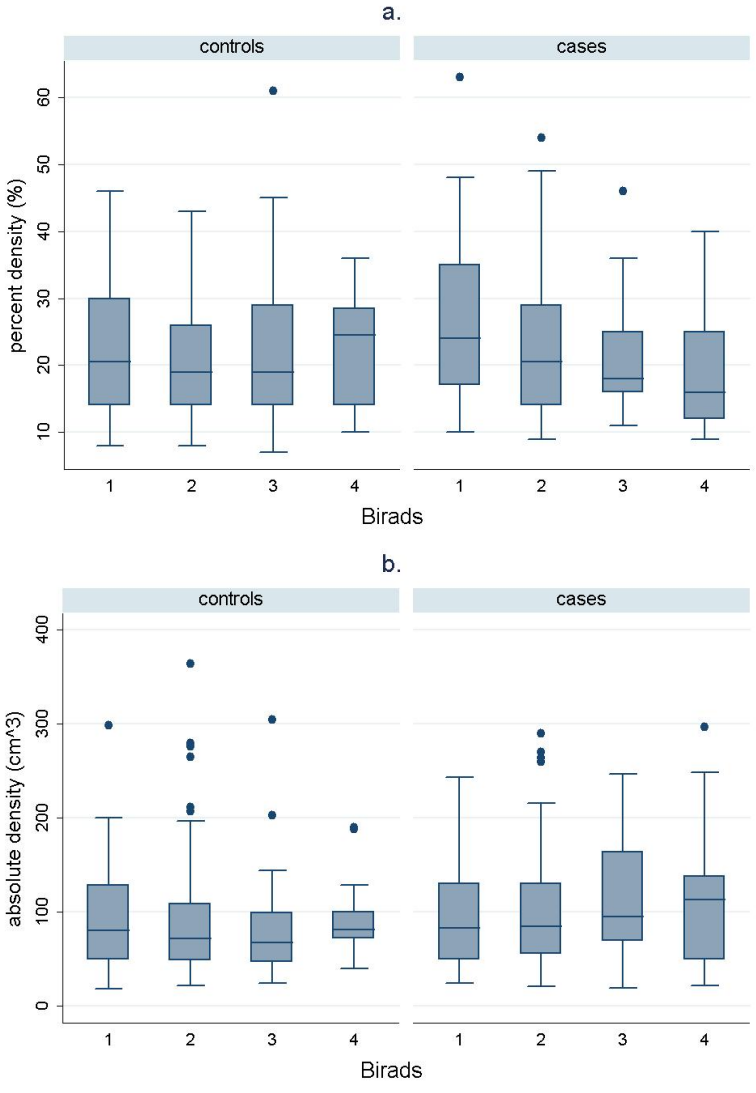


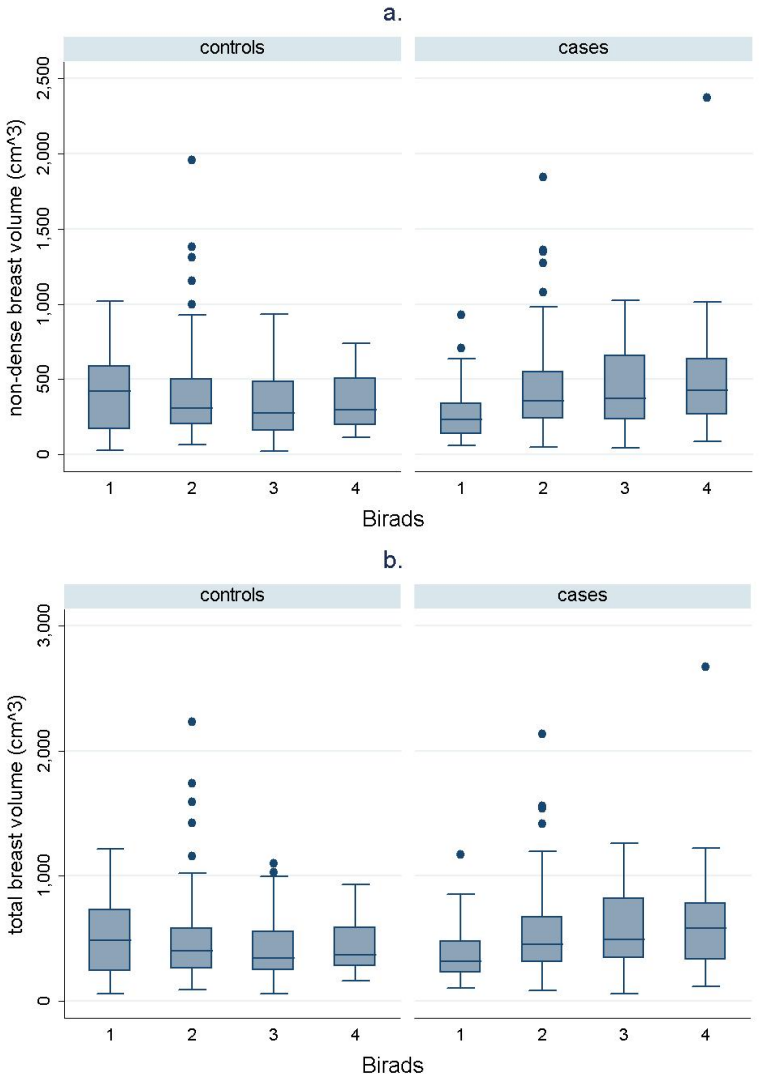
Figure 5.6 shows the relationship between BIRADS and Quantra (a) percent and (b) absolute density. The boxplots represent the distribution of percent and absolute breast density in each BIRADS category. Ideally, with increasing BIRADS category, one would expect to observe an increasing trend among the boxes, as evidence of a positive association between density measured by the two methods. Overall results from tests for trend suggest the absence of association between BIRADS and Quantra percent density among controls (p=.35). However, focusing only on cases, the two measures appear significantly but inversely related (p<.01). There appears to be no association between BIRADS classification and Quantra absolute density, and tests for trend are consistent with this (cases: p=.38; controls: p=.76).

FIGURE 5.6: Boxplots of (a) percent and (b) absolute breast density stratified by BIRADS categories and case-control status



Analyses of the other measures estimated by Quantra, i.e. (a) non-dense and (b) total breast volume (Figure 5.7), yielded similar results. Among controls, tests for trend showed no significant association between BIRADS and Quantra non-dense and total breast volume (respectively p=.35 and p=.38). In contrast, among cases, tests for trend showed a significant and positive association with BIRADS categories for both non-dense (p<.01) and total breast volume (p=.01). This indicates that the result for percent density in cases is due mainly to the non-dense volume.

FIGURE 5.7: Boxplots of (a) non-dense and (b) total breast volume stratified by BI-RADS categories and case-control status



Because of the lack of association observed between the two measures, we performed a conditional logistic regression adjusting Quantra absolute density for BIRADS, with results shown in Table 5.16. These results provide further evidence that there is no

TABLE 5.16: ORs for developing breast cancer risk from conditional logistic regression models using Quantra absolute density and adjusting for BIRADS categories

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
absolute density	per 10 cm^3	400	1.04	1.00	1.07	.035
BIRADS	(A) <25% dense		1.00	-	-	.191
	(B) 25-50% dense		1.58	.90	2.78	
	(C) 50-75% dense		1.23	.65	2.35	
	(D) ≥75% dense		2.19	.97	4.93	

relationship between BIRADS classification and Quantra absolute density, as the values do not substantially vary from those reported in previous Tables 5.3 and 5.6. The two measures appeared to have independent effects on risk. This suggests that they may be measuring different aspects of breast density.

5.4 Discussion

Our data seem to support the hypothesis that absolute volume of dense tissue, is a better predictor of breast cancer risk than percent density, possibly because it gives a closer estimate of the number of cells at risk of suffering a malignant transformation. However it should be noted that the OR for the highest quintile compared to the lowest is only 1.77 (Table 5.4); in other words, the observed difference in the number of cases and controls between high and low amounts of dense volume is modest. In order to validate and give further support to our findings, it would be interesting to compare the performances of Quantra dense volume and interactive thresholding, the current gold standard. Previously, other volumetric estimates [8, 36, 245] failed to be more informative in similar comparisons. It should be noted that our study population is on average younger than the target population for screening in the UK, and is likely yo be of higher socioeconomic status than the general population, since it is made up of users of a private healthcare facility. Also, the cases are symptomatic. It is feasible that the levels of breast density may differ from the general population. However, the comparison between cases and controls would be expected to be valid.

In our data, the lack of association between Quantra and BIRADS is surprising and raises questions about the validity of both measures. Moreover the association between breast cancer risk and BIRADS classification here appears weaker than one would expect. As McCormack and dos Santos Silva's meta-analysis [6] reported BIRADS category

"extremely dense ($\geq$ 75% dense)" typically presents a risk of 4.08 (2.96, 5.63) relative to the lowest category, "almost entirely fat ($\leq$ 25% dense)", whereas in this study it was 2.26 (1.01, 5.05). A first suggestion could be to repeat these analyses on further and larger datasets for validation. If these results were confirmed we should start investigating what Quantra is actually measuring, since it seems to be strongly related to breast cancer risk. In particular, the significant inverse positive association between BIRADS and Quantra non-dense breast volume, among cases is surprising. It could indicate that part of the visually assessed dense tissue is detected as fat by Quantra. It should be mentioned that, recently, Ciatto and colleagues compared Quantra estimates and BIRADS readings by 11 expert radiologists on 418 digital mammograms [51] and their findings suggested that, despite Quantra estimates being consistently lower than visual classification, there was a good agreement between automated and visual density assessments. This is in opposition to what we observed, thus further analyses on a larger dataset are needed to clarify this issue.

FIGURE 5.8: Percent distribution of BIRADS in the present study compared to Barlow 2006 [2]

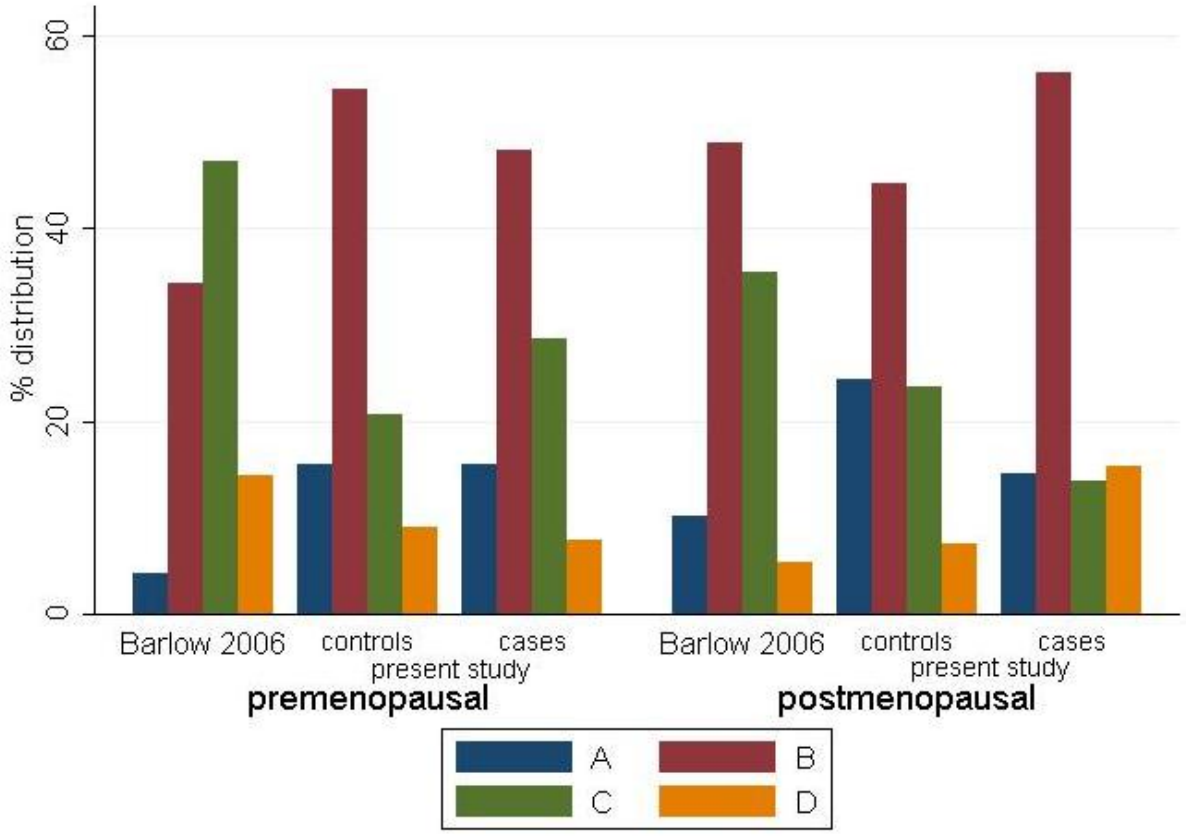


TABLE 5.17: Distribution (%) of BIRADS in Tice *et al.* 2005 [3], Barlow *et al.* 2006 [2] and the present study

BIRADS dense categories	Tice *et al.* 2005		Barlow *et al.* 2006		present study *controls*		*cases*	
	(%)		(%)		(%)		(%)	
(A) <25% dense	7,890	(9.7)	142,660	(8.6)	42	(21.0)	30	(15.0)
(B) 25-50% dense	36,543	(44.7)	744,080	(45.2)	97	(48.5)	106	(53.0)
(C) 50-75% dense	31,282	(38.2)	633,954	(38.5)	45	(22.5)	39	(19.5)
(D) ≥75% dense	6,062	(7.4)	126,680	(7.7)	16	(8.0)	25	(12.5)
Total	81,777	(100.0)	1,647,374	(100.0)	200	(100.0)	200	(100.0)

Supplementary investigation of the distribution of BIRADS, stratified by menopausal status (Figure 5.8), revealed that our data differ significantly from those observed in larger studies [2, 3]. Barlow's dataset was chosen for the comparison because the high number of women (N=1,647,374) assessed with BIRADS and the similarity of proportions in the categories found in Tice's analysis on 81,777 subjects [3] make it a good estimate of the general population. Table 5.17 shows the distributions of BIRADS categories in the cases and controls in the present study and in the larger studies of Tice *et al.* [3] and Barlow *et al.* [2]. In our sample women, both cases and controls, seem to be more likely to have less dense breasts (BIRADS categories A or B) than in the the other two studies, with a corresponding deficit in category C. The cases in our study do have a higher proportion in category D, as one would expect. The differences between the distributions in the controls here and in the Barlow population were significant overall, and in pre- and post-menopausal women separately ($p<.01$ in every comparison). Thus it may be that our sample is not representative of the general population; and the fact that the data were collected in a private medical facility could have introduced selection bias. Also, of course, the controls are not a population sample, but are age-related to cases.

Another possibility is that the visual classification of density on processed, full-field digital images is not reliable. The image-digitization process is not designed to assess breast density but to help detect tumours. The processing algorithm is developed to remove image errors and to attenuate areas of whiteness (dense tissue) within the raw images, so that tumours are more visible. Consequently, breasts are classified as less dense as they would be if the assessment were based on film mammograms. This fact might explain not only the differences in distribution between our sample and Barlow's but also the poor performance and small ORs of a well-established method, (BIRADS) in the present case-control study. Based on our findings, one would have to conclude that visual estimation from digital mammograms is a less reliable measure. However,

a previous study [34] compared density measures, using both Wolfe's patterns and the Boyd's six category classification (SCC), assessed from film and digital mammograms, and found substantial agreement. More recently Harvey *et al.* [247] found no difference in reported BIRADS breast density categories according to acquisition method, i.e. film or digital images. Thus, it is by no means clear that visual assessment by BIRADS or other systems is not reliable with digital mammography.

There have been studies [48, 50, 248] that compared Quantra volumetric assessments to MRI breast density [47] that should help evaluate what Quantra is actually measuring. Although the total breast volume appeared to be estimated with a good agreement, the fibroglandular tissue volume, or absolute density, was poorer (squared Pearson's correlation coefficient: .36), leading also to low correlation regarding percent density (squared Pearson's correlation coefficient: .51) [50, 248]. Nevertheless Wang and colleagues [50] concluded that the observed lack of agreement between mammographic density and MRI density was most likely driven by differences in total breast volume. Kontos et al. [48], instead, observed that MRI estimates of absolute density were significantly lower than Quantra estimates for women with very low density breasts. This finding suggests that Quantra may consider as dense some fatty tissue, one may speculate that it could be due to the compression applied during the mammographic exam. Nevertheless Kontos and colleagues were using a limited sample (N=32), thus their results require validation from a larger dataset.

Non-dense breast volume is utilised as a proxy measure for BMI, though it also allows absolute density to be adjusted to become percent density and *vice versa*. Interestingly, the association between breast cancer risk and percent density increases in strength after adjusting for non-dense volume, so it seems that there is independent information in each measure. It is also probable that this finding is due to the fact that percent density and non-dense breast volume together allow the computation of the amount of dense breast volume in terms of cm^3 and thus are just as informative as absolute density on its own. The structural collinearity of non-dense volume and percent density, however, make the adjusted estimates difficult to interpret. This may be distinct from the issue of confounding with BMI, which could be effectively answered in a study including BMI data on the subjects. Clearly, non-dense volume is at best a partial substitute for BMI.

Case subjects included in this study were symptomatic, and this may have affected our findings due to masking, i.e. symptomatic subjects are likely to have denser breasts that can hide tumours and lead to negative screening. Thus, masking should reduce observed associations of density with risk of screen-detected cases and exaggerate associations with post-screening, symptomatic cancers. However, if this was the case, we would have expected a much stronger association between density and risk.

The results with respect to area of residence suggest that the higher percent density in London women is largely due to lower non-dense breast volume. This is consistent with our previous finding, and with the observation that women residing in London are thinner than women living outside the capital [246, 249]. The differing results by "menopausal" groups (age below and above 50) may be chance findings and need confirmating by further studies, with explicit data on menopausal status.

These findings need to be validated using a larger dataset with richer data on other risk factors in order to be confirmed and to help explain the apparent lack of association between Quantra density measures and BIRADS categories. Our main finding that, overall, absolute dense volume is a better predictor of risk than percent, is consistent with our own previous findings [Chapter 4]. Quantra volumetric method requires further validation as well, but it suggests that this is a hopeful direction for research and for risk management in practice.

Chapter 6

Serial volumetric density measures using Quantra: changes with time and age

6.1 Introduction

Mammographic density has been consistently found to be lower in older women than in younger women [92], and the relationship between mammographic density and time has been investigated in several studies [16, 88, 91]. Better understanding of how breast composition varies with age, how density increases with age up to age 40 and then decreases [88], could give us a wider insight into the association of density with risk, and improve breast screening and breast cancer prevention programmes.

As a woman ages the mammographic appearance of her breasts changes because of an involution process, i.e. there is a reduction in glandular tissue, and a simultaneous increase in fat and connective tissue [16]. In particular Hutson and colleagues [88] analysed breast samples, collected at necropsy, from subjects aged 10 to 80, and concluded that the proportion of the breast comprised of epithelial tissue increases in the first decades after menarche (approx. age 10 to 40), and then declines progressively until menopause. In post-menopausal subjects this proportion remained fairly constant [88].

In 2010 McCormack *et al.* focused on the relationship between breast composition and age in 645 women aged 50 to 65, studying the screening mammograms available in a 3- to 12-year interval [91]. Their findings suggest a non-linear relationship between age and mammographic density, both percent and absolute dense area, with a greater drop

in the first years after the 50th birthday, intuitively due to the beginning of menopause, and then a gradual stabilization by age 65. The proportion of non-dense tissue in the mammograms correspondingly increased but the rate of change was almost double that of dense area. However, it is hoped that volumetric methods of assessment for mammographic density should provide more precise measures of the changes in breast composition [91].

The current study comprised 331 women without breast cancer, whose density was assessed on two occasions using Quantra, a fully-automated volumetric method [51]. These data allow investigations of the change in breast composition over time within subjects, and how the rate of change varies according to age. In addition, we performed a cross-sectional analysis on the association between age and the amount of dense and non-dense breast tissue at baseline. The potential presence of an "urban effect" [246] was also a subject of investigation.

6.2 Materials and methods

6.2.1 Data collection

Between 2002 and 2008, 413 women attending for mammography screening at the Princess Grace Hospital, a private facility in central London, were randomly selected to enter a cohort study of volumetric changes in breast composition. Cohort members were aged 34 to 81 and had no diagnosis of breast cancer before or during the study. These women underwent two digital mammographic examinations, on average 18 months apart. Due to misrecording and missing values, the final dataset included complete details on 332 women (231 younger than 50 years and 101 aged 50 or more years at entry). The younger group included one subject whose exact age was not known, but was known to be under 50.

Density was assessed with Quantra, an automated volumetric method. This technique is described in the previous chapter [Chapter 5]. It provides estimates of the total and dense volumes in cm^3, and the percent density, which is a ratio of the two. We also computed the volume of fatty tissue (non-dense volume), as the difference between total and dense volumes.

As previous results suggested higher density in women resident in London, data on area of residence was also recorded. This was classified as London, any other region in the UK (based on the post-code of residence), or overseas.

Aside from residence, age, breast composition assessed by Quantra and screening history, no other personal data was available.

6.2.2 Statistical analysis

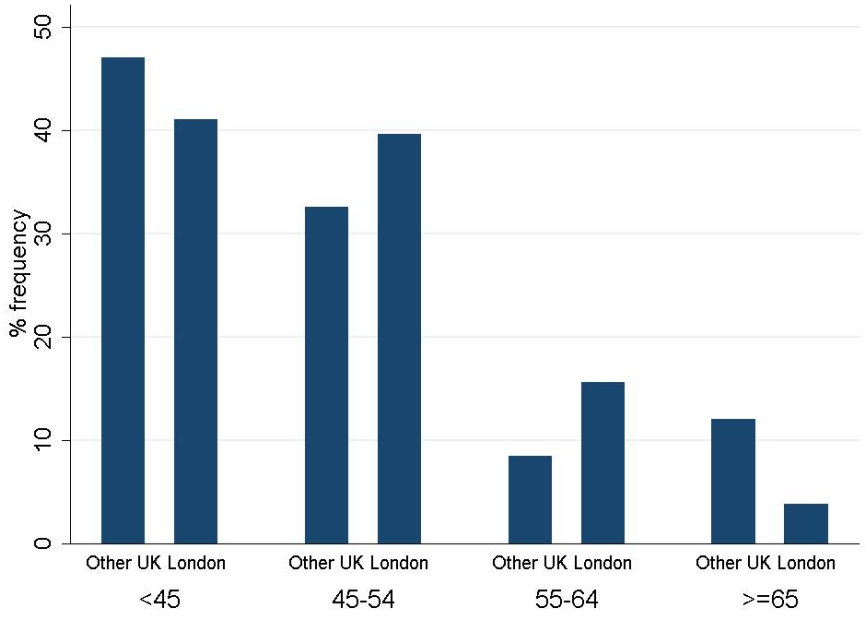
The distributions of demographic and other variables at the first examination were summarised as percentages for discrete variables, or means and SDs for continuous variables. Total breast volume and non-dense volume were used as a proxy for BMI, and the subdivision between women aged under 50 and those aged 50 or over, at the first screening, provided an approximate indicator of menopausal status.
Changes in breast composition between the two mammogaphic examinations were the main subject of investigation, and were first summarised with means and SDs. The null hypothesis of "no change" was assessed using the non-parametric Wilcoxon matched-pairs signed-ranks test. The relationship between age at baseline and the changes in absolute and percent density, non-dense and total breast volume, was studied graphically and through linear regression analyses. The hypothesis that the rate of change in breast composition is time/age-dependent was investigated using boxplots, non-parametric tests for trend across ordered groups and linear regression. Note that the differences in mammographic measures between screens were computed as follows: $screen2 - screen1$, so that the change is positive if the measure has increased over time, and negative if it has decreased.
The relationship between age and breast composition at baseline was investigated using linear regression and Pearson's correlation coefficients.
Secondary analyses, stratified by age at baseline and by area of residence (London or elsewhere in the UK), were also performed.
All analyses were performed with STATA software version 11.

6.3 Results

The sample recruited in the study comprised 332 women, resident mainly in London (64%) and on average aged 47 (Table 6.1). The age distribution varied significantly according to area of residence (χ^2 test: p=.02), because women living outside the capital tended to be more often in the extreme age classes, "under 45" and "65 or more" (Fig.

6.1).

FIGURE 6.1: Distribution of age at baseline and area of residence (London: N=213, Other UK: N=83)

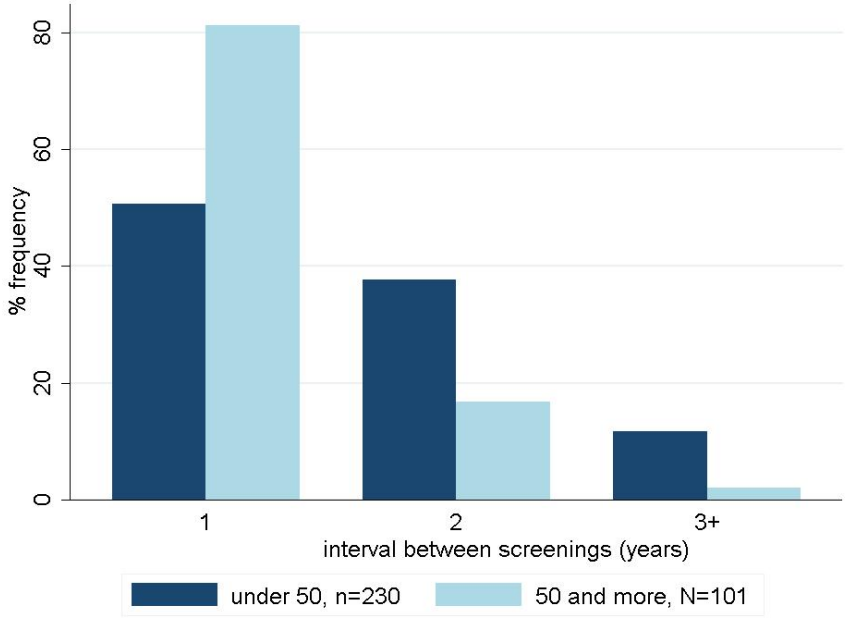


Overall, breast composition at baseline, shown in Table 6.1, was as one would expect: at older ages women had, on average, significantly lower percent density (Wilcoxon rank-sum test: $p<.01$) and higher amounts of fatty and total volume ($p<.01$ for both), whereas absolute breast density did not, on average, change significantly with age ($p=.78$).

TABLE 6.1: Characteristics of the study samples at baseline

	Age groups under 50 N=231		50 or more N=101		Total N=332	
Age at baseline						
mean (sd)	42.1	(3.4)	58.4	(6.8)	47.1	(8.9)
Area of residence (%)						
London	154	(66.7)	59	(58.4)	213	(64.2)
Other UK	58	(25.1)	25	(24.8)	83	(25.0)
Abroad	19	(8.2)	17	(16.8)	36	(10.8)
Measures of mammographic breast density assessed by Quantra at baseline						
mean (sd)						
Percent breast density (%)	26.9	(11.4)	20.0	(9.4)	24.8	(11.3)
Absolute breast density (cm^3)	102.9	(72.1)	96.5	(52.0)	101.0	(66.6)
Non-dense volume (cm^3)	321.0	(250.8)	446.9	(277.5)	359.3	(265.2)
Total breast volume (cm^3)	423.9	(303.3)	543.4	(309.8)	460.2	(309.7)

FIGURE 6.2: Distribution of the time interval (in years) between screenings and age at baseline



For the group aged 50 or more mammography screening was repeated after one year in 81% of cases, most other cases after 2 years, whereas 49% of younger women (49.4%) had their second mammogram 2 to 5 years after (Figure 6.2). In all women considered, we observed a significant decrease of percent density between the first and second mammogram (-.3%, p=.03), a stronger effect among older women (-.8%, p<.01). This was due to the fact that both groups experienced a substantial increase in volume of fatty tissue (under 50: +8.6 cm^3, p<.01; 50 or more: +10 cm^3, p<.05). Absolute density did not change substantially between screens (Table 6.2).

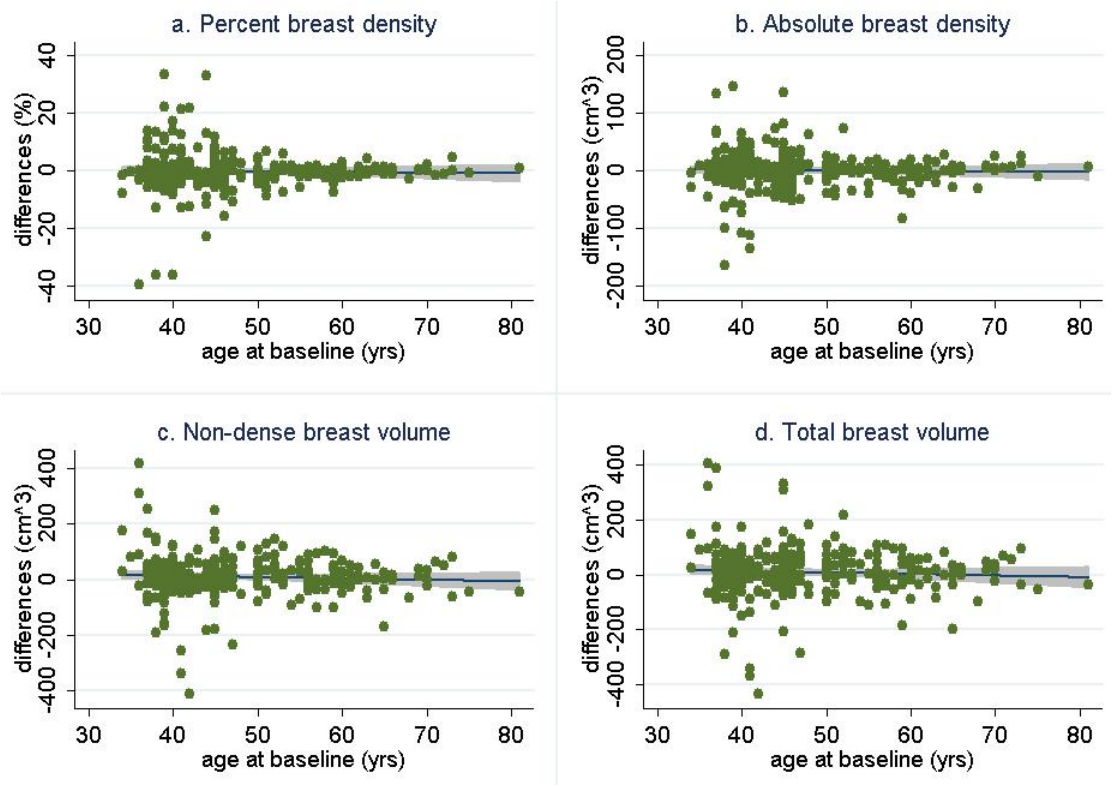
TABLE 6.2: Breast composition measures at first and second screen

	screen 1		screen 2		
Variable	Mean	(SD)	Mean	(SD)	p
Under 50, N=231					
Percent breast density (%)	26.9	(11.4)	26.9	(12.0)	.39
Absolute breast density (cm^3)	102.9	(72.1)	103.6	(74.7)	.31
Non-dense breast volume (cm^3)	321.0	(250.8)	329.5	(262.3)	<.01
Total breast volume (cm^3)	423.9	(303.3)	433.1	(313.4)	<.01
50 or more, N=101					
Percent breast density (%)	20.0	(9.4)	19.3	(9.1)	<.01
Absolute breast density (cm^3)	96.5	(52.0)	94.5	(48.9)	.80
Non-dense breast volume (cm^3)	446.9	(277.5)	456.9	(268.8)	.05
Total breast volume (cm^3)	543.4	(309.8)	551.3	(296.6)	.11
Total, N=332					
Percent breast density (%)	24.8	(11.3)	24.6	(11.7)	.03
Absolute breast density (cm^3)	101.0	(66.6)	100.8	(68.0)	.46
Non-dense breast volume (cm^3)	359.3	(265.2)	368.3	(270.3)	<.01
Total breast volume (cm^3)	460.2	(309.7)	469.1	(312.7)	<.01

Figure 6.3 depicts the changes in (a) percent density, (b) absolute density, (c) non-dense volume and (d) total volume between the two screens according to age of the subjects; it shows also the fitted linear slope with related 95% confidence interval (area in grey). In every case the distribution was a funnel-shaped pattern, with a broader range of values for change in density between screens at earlier ages and a narrower distribution at later. The intervals around the fitted lines are narrower at younger ages because of a higher number of subjects. The linear trend described by these slopes was not significantly different from zero, and in particular in the charts regarding the change in density, both percent and absolute, the lines are almost exactly parallel to the X-axis (coefficient for percent density: -.02, p=.6; coefficient for absolute density: -.08, p=.7). Interestingly,

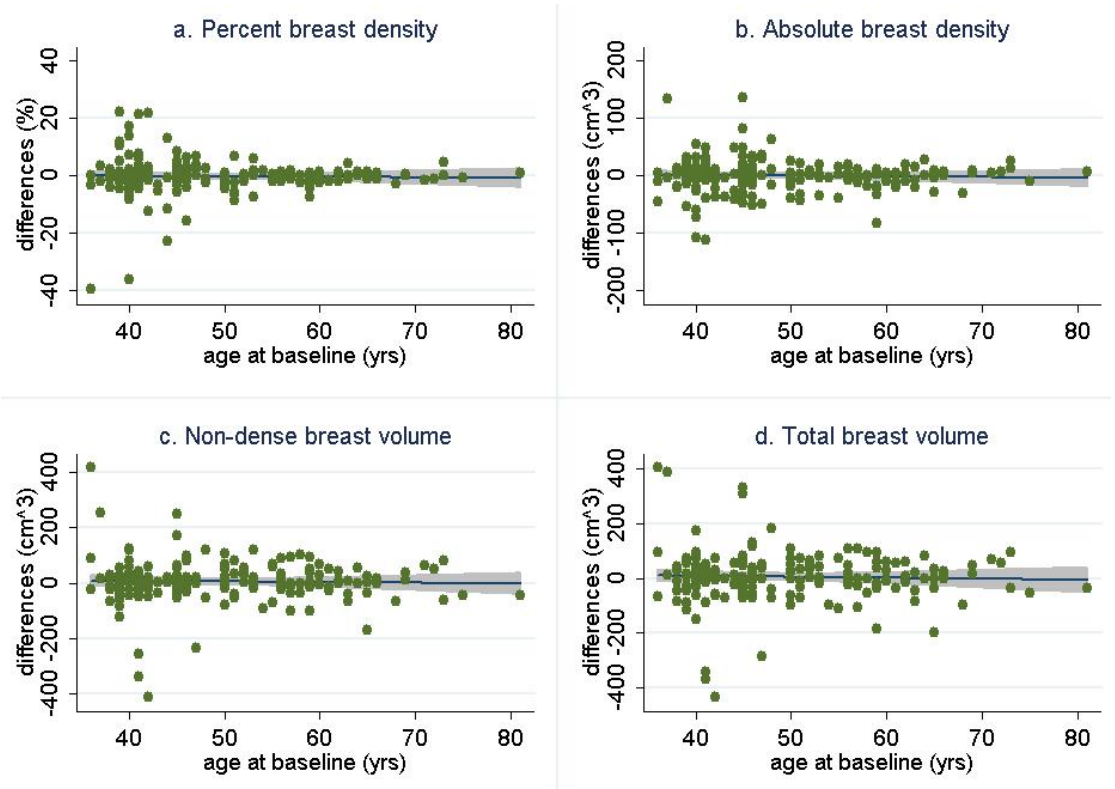
between 40% and 50% of young women experienced an increase in breast density (both percent and absolute) between the first and the second mammogram.

FIGURE 6.3: Changes in breast composition between subsequent mammograms according to age at baseline (years)



The fact that at younger ages the interval between screens was wider and therefore subject to a higher variability could also explain the funnel shape. Thus, we repeated this analysis considering only the subjects screened a year or less apart (Figure 6.4). However results did not differ substantially from the previous ones.

FIGURE 6.4: Changes in breast composition between subsequent mammograms according to age at baseline (years), with maximum one-year interval between screens



Investigating if and how changes in breast composition depend on time between measurements (Figures 6.5 and 6.6), the results suggested that changes in non-dense and total breast volume increased with time between mammograms. This is consistent with the findings above that in the younger women, the percent and absolute density did not change significantly between screens, and that the significant change in percent density in all ages combined was driven by an increase in non-dense volume (Table 6.2).

The relationship between change in density and time between screens was investigated only overall and in the young group, because women aged 50 or more underwent their second screen mostly (98%) only one or two years after the first (Figure 6.2). Thus, there was insufficient variation in time interval to assess its effect in the older group.

FIGURE 6.5: Dependence on time of changes in breast composition between two subsequent mammograms (N=332)

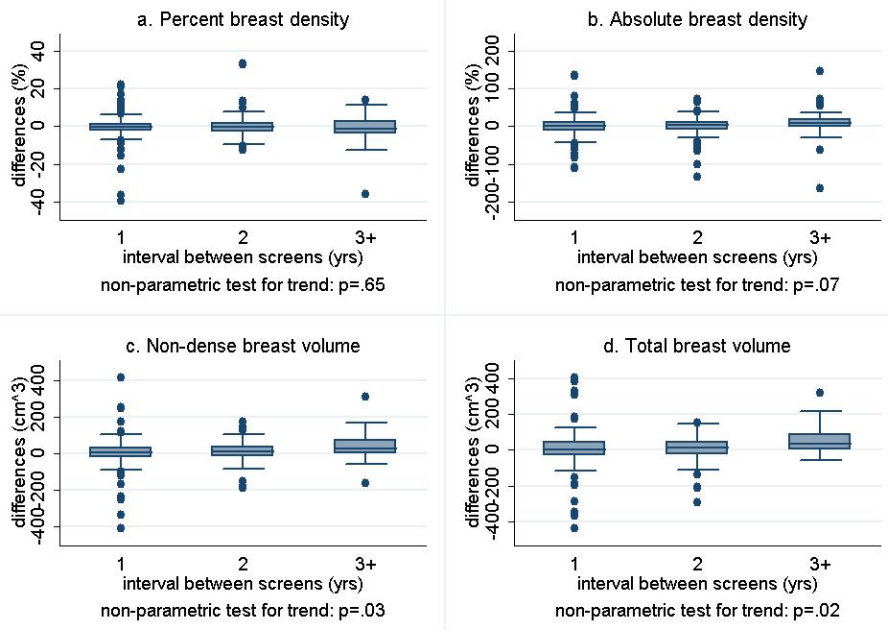


FIGURE 6.6: Dependence on time of changes in breast composition between two subsequent mammograms, in women aged under 50 (N=231)

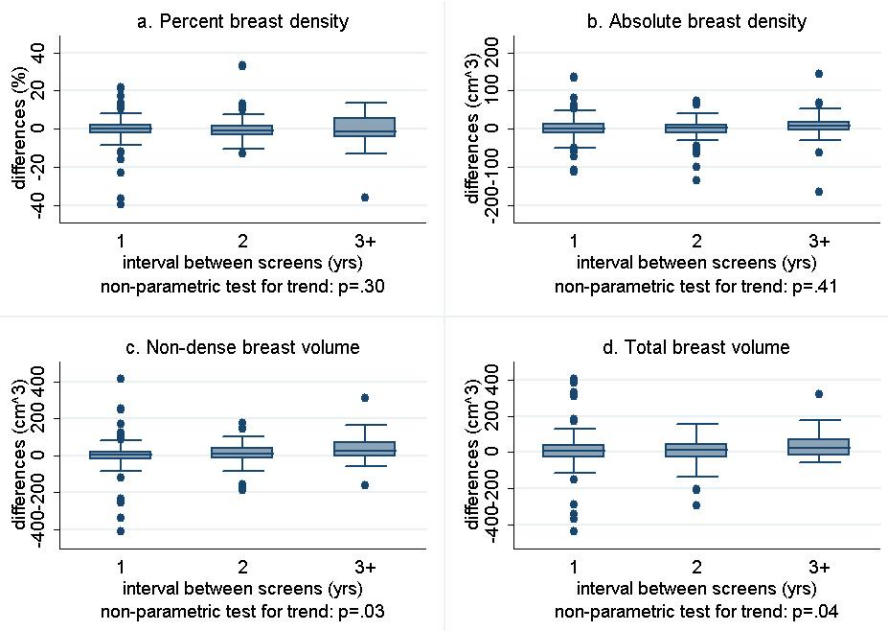


Table 6.3 shows the results of regression analyses for the change in breast composition factors by time between screens, both adjusted and unadjusted for age at baseline. Significant associations were only observed for non-dense breast volume (p=.03 unadjusted; p=.05 adjusted for age) and total breast volume (p=.02 unadjusted; p=.04 adjusted for age). Age at baseline, on the other hand, seemed to have no effect on the changes, indicating that the rates of change in breast composition were similar at all ages. This can also be seen in Table 6.2.

TABLE 6.3: Linear dependance of changes in mammographic volumetric measures on time between screens

Factor	Coef.	[95% Conf.	Interval]	p-value
Response: Difference in percent density				
Interval between screenings	.02	-1.00	1.04	.97
Interval between screenings	-.08	-1.15	.99	.88
Age at baseline	-.02	-.11	.06	.58
Response: Difference in absolute density				
Interval between screenings	2.95	-1.74	7.63	.22
Interval between screenings	2.74	-2.16	7.63	.27
Age at baseline	-.02	-.41	.38	.93
Response: Difference in non-dense volume				
Interval between screenings	12.13	1.35	22.91	.03
Interval between screenings	11.28	-.02	22.58	.05
Age at baseline	-.22	-1.13	.70	.64
Response: Difference in total volume				
Interval between screenings	15.07	2.31	27.83	.02
Interval between screenings	14.01	.65	27.38	.04
Age at baseline	-.23	-1.31	.85	.67

Scatter plots in Figure 6.7 illustrate the relationship between age and breast composition, i.e. (a) percent and (b) absolute density, (c) non-dense and (d) total breast volume, at baseline. They suggest a steady increase of non-dense and total breast volume with age, respectively +6.8 cm^3/year (p<.01) and +6.3 cm^3/year (p=<.01) and a consequent constant decrease in percent density (-.41%/year, p<.01). The absolute amount of dense

tissue did not change significantly (-.47 cm^3/year, p=.26).

FIGURE 6.7: Relationship between breast composition and age at baseline

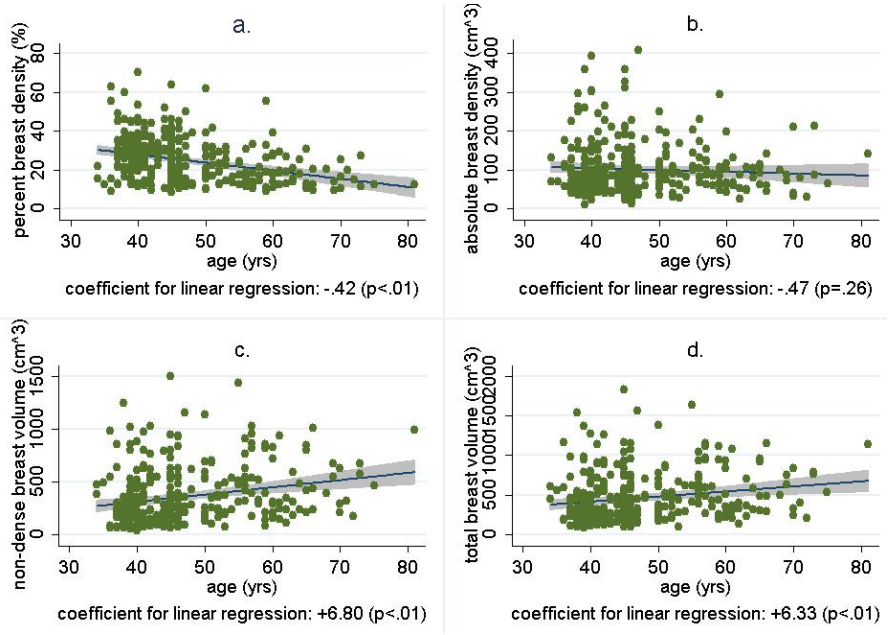


TABLE 6.4: Breast composition measures at first and second screen, stratified by residence

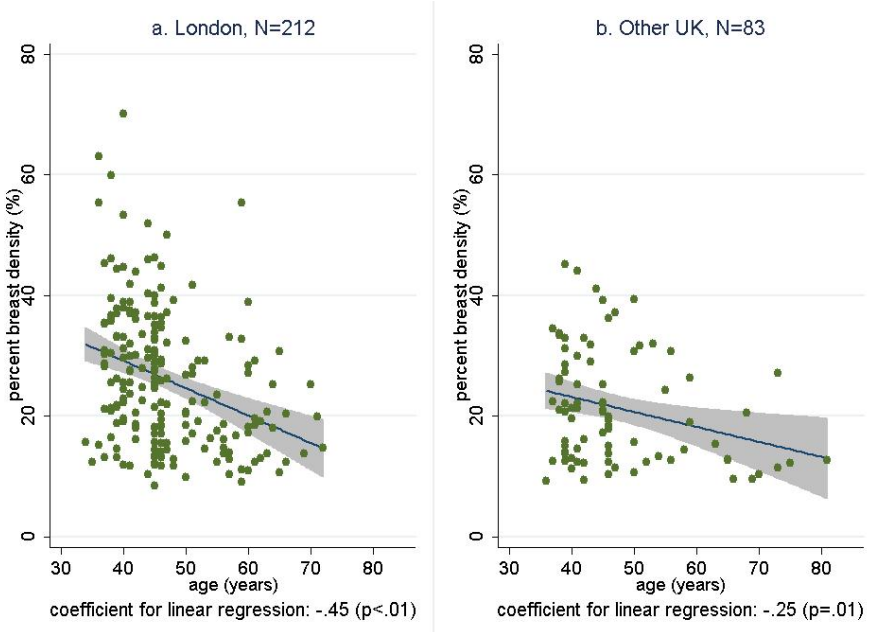
	screen 1		screen 2		
Variable	Mean	(SD)	Mean	(SD)	*p*
London, N=213					
Percent breast density (%)	26.1	(11.3)	25.3	(11.5)	*.01*
Absolute breast density (cm^3)	101.1	(67.9)	100.3	(70.7)	*.96*
Non-dense breast volume (cm^3)	341.4	(271.1)	351.3	(279.5)	< .01
Total breast volume (cm^3)	442.5	(319.2)	451.5	(328.0)	*.03*
Other UK, N=83					
Percent breast density (%)	21.3	(9.3)	22.2	(10.8)	*.51*
Absolute breast density (cm^3)	96.8	(60.1)	100.0	(65.4)	*.15*
Non-dense breast volume (cm^3)	401.9	(239.9)	402.3	(228.2)	*.30*
Total breast volume (cm^3)	498.7	(277.3)	502.3	(260.7)	*.21*

Table 6.4 examines change in breast composition between successive screens according to area of residence. UK residents living outside London seemed not to have experienced any significant changes from screen 1 to screen 2, whereas in London residents, on average, there was a significant decrease in percent density (-.71%, p=.01) because of a substantial increase in non-dense tissue (+9.84 cm^3, p<.01). It is possible that the

absence of observed differences in the non-London residents is due to the smaller number of subjects; however there are not even any suggestive differences in this group. Another hypothesis to explain these differences between the two groups is a different age composition (Figure 6.1), but London residents are, on average, only slightly younger (46.7 years) than women living elsewhere in the UK (47.5 years). Furthermore a Wilcoxon rank-sum test, comparing age distributions in the two groups, was unequivocally non-significant (p=.81).

Studying Table 6.4 vertically between London and other UK locations, it appears that at baseline all the subjects presented with similar volumes of fibroglandular tissue in volumetric terms (Wilcoxon rank-sum test: p=.88), but that London residents had lesser fatty (non-dense) tissue and smaller total breast size (p<.01 for non-dense breast volume, p=.02 for total breast volume) leading to a higher average percent density (p<.01).

FIGURE 6.8: Relationship between percent breast density and age at baseline, stratified by residence



Figures 6.8, 6.9, 6.10 and 6.11 illustrate the correlations between breast composition and age at baseline. The increase in non-dense volume per year of age was less steep in London (+5.61 cm^3/year, p=.01) than elsewhere in the UK (+8.51 cm^3/year, p<.01), likewise for the total breast volume (London: +4.69 cm^3/year, p=.08; Other UK: +9.27 cm^3/year, p<.01), arguably consistent with the smaller breast size and lower fatty tissue volume in London.

FIGURE 6.9: Relationship between absolute breast density and age at baseline, stratified by residence

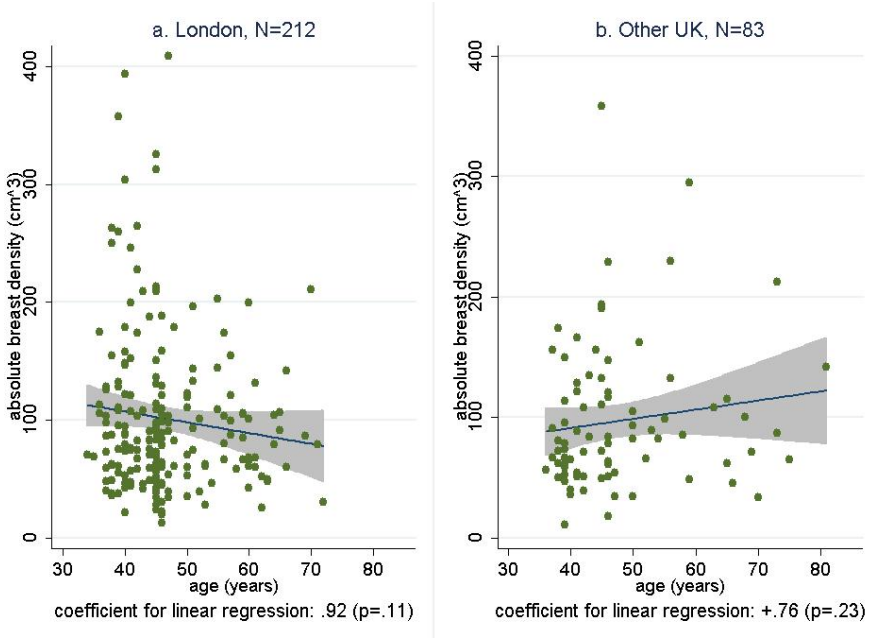


FIGURE 6.10: Relationship between non-dense breast volume and age at baseline, stratified by residence

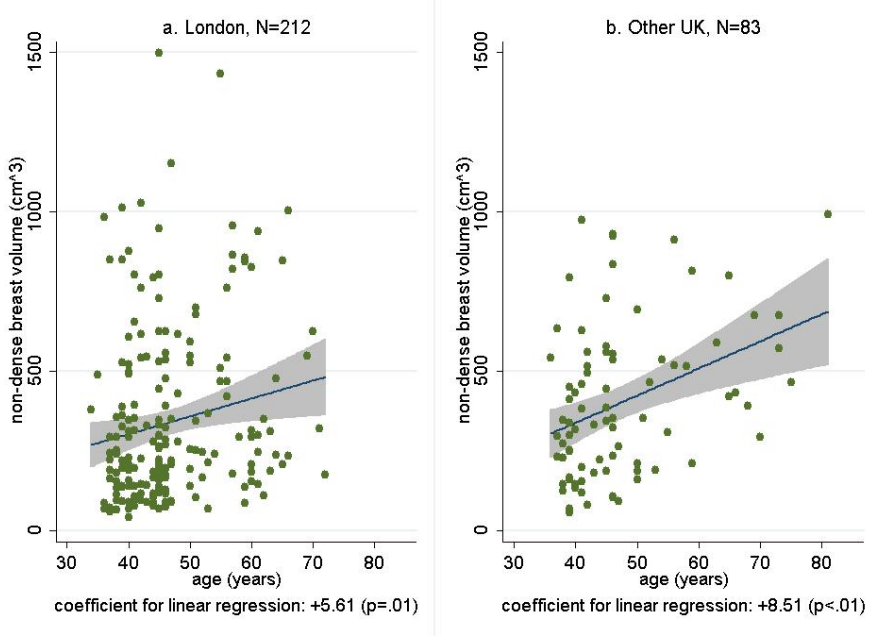
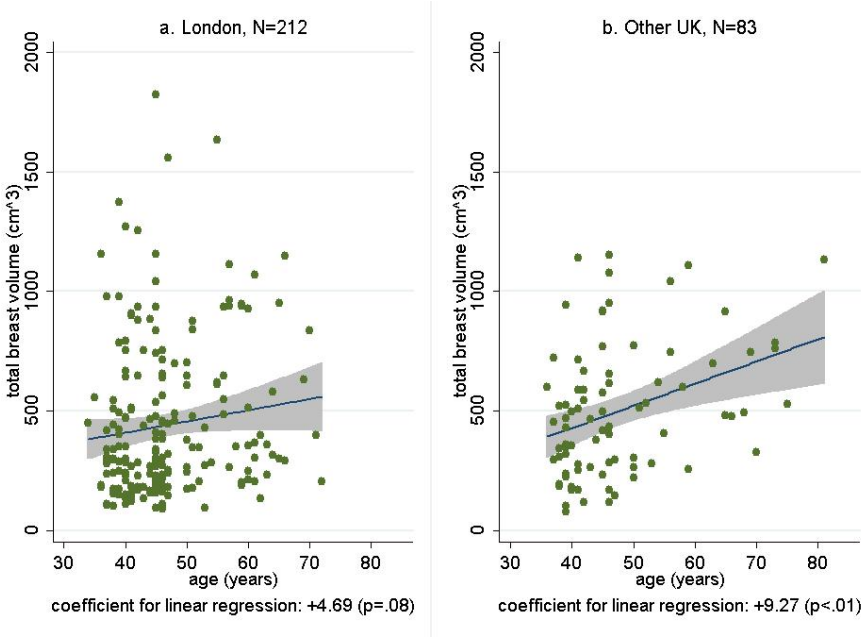


FIGURE 6.11: Relationship between total breast volume and age at baseline, stratified by residence



6.4 Discussion

In the present study we witnessed no significant change in absolute density with age at baseline or between successive screens, indicating that the observed variations in breast composition, and in particular the decrease in proportion of dense tissue, are to be attributed to the growth in volume of non-dense tissue. Between screens percent density had, overall, a limited, but statistically significant, reduction of .27% because the non-dense tissue was volumetrically increasing at a higher rate (+11 cm^3/year, Table 6.3) than the dense tissue. This result is in agreement in McCormack's and colleagues's study [91] where the growth in non-dense area was much larger in magnitude than the decline in dense area.

However unlike in the previous literature, including McCormack *et al.*'s findings [91], in our case, the dense tissue did not appear to vary in absolute size. This could be due to lack of power investigating changes in fibroglandular tissue, but also the different methods used for assessing density could have contributed: McCormack and colleagues used interactive-thresholding, a computer-assisted technique that estimates breast components in terms of areas [30], whereas the current data were assessed with Quantra, a volumetric fully-automated system [51]. The estimates from Quantra were anticipated to be more reliable because they are not affected by the rater subjectivity, unlike those from interactive thresholding that depend on human input; nevertheless the Quantra

method is still undergoing validation and currently there is no conclusive evidence of its accuracy in measuring density. Another possible explanation for the disagreement between the current findings and those of McCormack *et al.* is the different age composition of the samples: in the present study age at baseline ranged from 34 to 81, and almost the 70% of subjects were aged under 50, whereas McCormack and colleagues included in their study only women aged 50 to 65. In premenopausal women breast tissue is known to be more susceptible to both endogenous and exogenous factors [12], and this is reflected in a wider variability in breast composition at younger ages which was also seen in our data. This variability could have been enhanced by the longer intervals between screens, i.e. up to 5 years, in women aged under 50, but when we repeated the analyses including only subjects screened one year apart the results were not affected. Furthermore, analyses run separately in two age groups, "under 50" and "50 or more", did not report significant differences. Such findings could also indicate that longitudinal changes in breast composition occur at similar rates at different ages. However this could be due to lack of power, since these results are derived from relatively small subsets of data.

Menopause affects the characteristics of mammographic appearance [21, 89, 90], promoting the involution process and leading to an average further reduction in breast percent density of 3.26%, over 5 years, due to a reduction in glandular, i.e. dense, tissue, and an increase in connective tissue and fat, i.e. non-dense, tissue [90]. McCormack and colleagues [91] observed this drop and consequently described as non-linear the trajectory of reduction in absolute and percent dense area over time. In the current study, by contrast, the rate of change in breast tissue components seemed to be unaffected by age as a substitute for menopausal status. Moreover pre- and post-menopausal (age 50) mammographic appearances were significantly different in terms of percent density because of non-dense volume. The volume of dense tissue appeared approximately constant.

Since details on area of residence were available it was possible to verify the existence of an "urban effect". A previous study by Perry and colleagues [246] classified 962 women (318 from London, 654 resident outside of London) according to BIRADS density categories and then compared the incidence of each category between subjects from London or other UK areas. It was observed that women resident in London had significantly higher Birads categories than women resident outside the capital, after adjusting for age. Similarly, in the current data, percent density in London residents was substantially higher (Table 6.4). Once more, we found that it was more attributable to larger volumes of fatty tissue in women resident outside London, than to a difference in absolute density. In the previous publication reporting higher percent density in London

women [246], it was speculated that this was due to lower body mass in London women, as evident from survey data [246, 249]. The results here support this speculation.
A limitation of this study is the fact that it only comprised two screens within a relatively short interval. It is possible that with a longer interval between screens we might have observed a reduction in absolute density, due to involution, within the subjects. However the volume of dense tissue stayed constant across age even when we performed a cross-sectional analysis of the data at baseline, supporting our longitudinal results.
Measurement error in Quantra estimates could be partially responsible for some of the unexpected results. As noted above, the Quantra method is currently undergoing validation to assess the extent to which it is correctly measuring mammographic density. This issue will be further developed in Chapter 8.
The availability of details on only age and residence, besides volumetric estimates of breast components, represents another limitation. Information on other factors, e.g. exogenous hormone use, potential pregnancies between the screens and menopausal status, could have helped us explain more clearly the variations in breast composition.
In the current study, the volume of dense tissue seemed not to be susceptible to age and time. It has been suggested that an absolute measure of density should be chosen over a percent one in breast cancer risk prediction contexts [33, 226], since it is less affected by BMI. Our results suggest that further study of the non-dense component might yield further understanding of this.
In conclusions the main implications of this analysis are:

1. Absolute non-dense volume, as measured by Quantra, appears to be more sensitive to temporal and age-related changes than percent or absolute dense volume.

2. Research in the future should focus on whether the previous implication is due to the measurement qualities of Quantra or to a greater biological role for non-dense tissue. If the latter, there may be major implications for prevention and for monitoring of preventive interventions.

3. Assuming that the results are not primarily due to measurement properties of the Quantra instrument, the results strongly support previous findings that percent density is higher in London women, probably due to their lower body mass.

4. These results broadly support those of others using automated or semi-automated methods, that absolute measures of breast composition are at least as important as percentages (and probably more so).

Chapter 7

A case control evaluation of a fully automated volumetric density measure (Volpara) as a predictor of breast cancer risk

7.1 Introduction

In Chapter 5, we examined the association of a fully-automated volumetric density measure, Quantra, with breast cancer risk in a case-control study. In that chapter we noted that, whilst there is good evidence for density as a risk factor [6], there are practical difficulties in its use at population level unless a reliable and fully-automated method is available. In the USA and in other countries, the visually assessed BIRADS categorisation is in general use, and in parts of the USA it is mandated by statute [2, 3, 234, 235, 250–252]. Nevertheless, being based on human input, it is susceptible to subjectivity [28, 250] and its reliability on digital mammograms still requires proper validation [32] [Chapter 5], although Harvey and colleagues recently observed good agreement in reported BI-RADS breast density categories according to acquisition method, i.e. film or digital [247]. As already remarked, it is thought that a continuous density assessment could provide a more informative and accurate measure [28].

In the late 1990s, technology advances led to the development of a computer-assisted method for assessing density, based on interactive thresholding [30]. This technique, available as the computer programme *Cumulus*, is considered by many in the field to

be the gold-standard. It provides continuous information regarding both absolute and percent dense area observed in mammograms. Not only does Cumulus record more complete and comprehensive information regarding breast density, it has also been observed to be very effective in breast cancer risk predictions [6, 26]. The drawbacks of this technique are that it requires trained observers and digitized images, along with being time- and labour-intensive. For these reasons, Cumulus is difficult to use in the context of high-volume, population mammographic screening.

Automated approaches are expected to minimize or solve these limitations, by providing quick and non-subjective assessments with little or no call on human resources. In particular, the newly developed methods have been focusing on volumetric assessments [36–42, 51, 52, 178, 236], because it is anticipated that volumetric estimates of density may provide more precise information about the amount of fibroglandular tissue [30, 201]. These methods are currently undergoing validation, however the results presented so far have suggested that volumetric estimates were no more predictive of risk than bi-dimensional measures, assessed with interactive thresholding [8, 36, 201, 245]. Volpara [52] is one of these novel fully-automated volumetric methods. It is a software package that analyses raw images from full-field digital mammograms (FFDM) to quantify dense tissue. Despite the current scarcity of evidence in the public domain supporting Volpara, its implementation has already spread internationally, e.g. in UK, Netherlands, Korea and US [253], and made its evaluation and validation urgent and necessary.

Here we report an evaluation of the association of Volpara with breast cancer risk in a case-control dataset, using, largely but not entirely, the same cases and controls as in the Quantra study reported in Chapter 5.

7.2 Materials and Methods

7.2.1 Data collection

For this study, we used mostly the same digital mammograms used in the Quantra case-control study [Chapter 5], from women attending the Princess Grace Hospital, London, who had full-field digital mammographic examinations between 2005-2009 . Originally, two hundred cases were randomly selected from patients with histopathologically verified breast cancer, along with two hundred controls. Due to missing raw digital information,

124 (62 cases) subjects could not be assessed with Volpara. Consequently 90 new subjects (43 cases) were recruited.

All mammograms were high quality two-view images, soft copy reported on high resolution monitors. Image quality was not a factor in case selection and none were rejected on quality grounds during density estimation. Among cases, density was assessed on the last mammogram of the contralateral breast before diagnosis; for controls, density was estimated on the same side as the matched case, using mammograms dating as close as possible to the matched case diagnosis. Estimation of density was based on an average of the two-views, as it is known that the craniocaudal view tends to give higher density estimates than the mediolateral oblique. The area-based density was estimated by one of two experienced radiologists (Dr Nick Perry and Dr Katja Pinker Domenig), using the standard density grades according to the American College of Radiology Breast Imaging Reporting and Data System (BIRADS) [234], as described in Chapter 5.

Results regarding Quantra density estimates are given in Chapter 5. In this chapter, we report the corresponding results using the Volpara software [52], acquired more recently. As previous results suggested higher density in women residing in London [246], we also recorded data on area of residence. Area of residence was classified as London, any other region in the UK (based on the post-code of residence), or outside the UK, referred to as "abroad" for brevity. In addition to age at mammogram (henceforth referred to simply as age), area of residence, and density assessed with Volpara, Quantra and BIRADS, no information on other breast cancer risk factors was available.

Volpara assessment was only available for 366 subjects (182 cases and 184 controls). Table 7.1 shows the numbers of cases and controls used in both chapters.

TABLE 7.1: Data available for both measures

Density measure	cases	controls
Volpara only	43	47
Quantra only	61	63
Both	139	137

7.2.2 Statistical analyses

Demographic and other variables are summarised using percentages for categorical variables or medians and IQRs for continuous variables. As noted above, we had data on age,

area of residence classified as "London", "Other UK regions" and "Abroad", Volpara and Quantra estimates for total breast volume, dense volume and non-dense volume and BIRADS classification. Total breast volume and non-dense volume were used as a proxy for BMI, which was not available, and age under 50 versus greater than or equal to 50 was used as a surrogate for menopausal status.

The distributions of percent and absolute density, as well as non-dense and total breast volume, were compared between cases and controls using the Wilcoxon rank-sum test.

Unconditional logistic regression was used to calculate odds ratios (OR) for risk of breast cancer after adjustment for age and other potential confounding factors. Loss of information at merging made conditional logistic regression analyses, used in Chapter 5, no longer applicable to the data as this would have entailed loss of large numbers of matched cases and controls. This analysis was repeated after log-transformation of Volpara estimates. Secondary analyses were also performed stratified by age and area of residence (London vs other UK).

The relationship between BIRADS and Volpara was investigated graphically and Cuzick's nonparametric test was used to evaluate differences for trend across ordered groups [214].

The agreement between the two automated density measures was assessed using Pearson's correlation coefficents and contingency tables of categorised measures. Finally the associations of Volpara and Quantra density estimates with breast cancer risk were compared using standardised odds ratios.

Quantra and BIRADS analyses are presented in Chapter 5. Henceforth, in this chapter, Volpara density estimates are referred to simply as density, unless otherwise specified.

All analyses were performed using STATA, version 12.1.

7.3 Results

Characteristics of the 182 case subjects and 184 control subjects are presented in Table 7.2. Age at mammogram and areas of residence appeared equally distributed between cases and controls. The distribution of percent and absolute density, as well as non-dense and total breast volume, were compared between cases and controls using the Wilcoxon-rank-sum test. There was no significant differences in these measures between the two groups, but absolute density was borderline significantly higher among cases than controls (p=.07).

TABLE 7.2: Characteristics of the study samples

	Cases N=182		Controls N=184		Total N=366	
Age at mammogram, No. (%)						
<45	22	(15.8)	21	(15.3)	43	(15.6)
45-54	46	(33.1)	47	(34.3)	93	(33.7)
55-64	37	(26.6)	36	(26.3)	73	(26.4)
≥ 65	34	(24.5)	33	(24.1)	67	(26.3)
Area of residence, No. (%)						
London	71	(51.1)	81	(59.1)	152	(55.1)
Other UK	53	(38.1)	46	(33.6)	99	(35.9)
Abroad	15	(10.8)	10	(7.3)	25	(9.0)
Measures of mammographic breast density assessed by Volpara, Median (IQR)						
Percent breast density, in %	11.9	(7.2)	11.4	(7.0)	11.7	(7.1)
Absolute breast density, in cm^3	66.4	(38.3)	58.7	(31.4)	62.5	(35.2)
Non-dense breast volume, in cm^3	634.4	(416.3)	588.5	(372.5)	611.3	(395.0)
Total breast volume, in cm^3	700.7	(428.2)	647.3	(382.3)	673.9	(406.1)

Table 7.3 summarises the distribution of BIRADS classification between cases and controls. Overall, there was a smaller proportion of cases in the lowest density category, and a greater proportion in the highest. These differences appeared stronger in the older age group. Further results regarding BIRADS on this dataset are presented in Chapter 5.

TABLE 7.3: Distribution of BIRADS in cases and controls (percentages)

BIRADS	Cases (%) N=139		Controls (%) N=137	
Under 50				
(A) <25% dense "almost entirely fatty"	11	(19.0)	11	(20.4)
(B) 25-50% dense "scattered fibroglandular density"	28	(48.3)	26	(48.1)
(C) 50-75% dense "hetereogeneously dense"	15	(25.8)	12	(22.2)
(D) >75% dense "extremely dense"	4	(6.9)	5	(9.3)
50 or more				
(A) <25% dense "almost entirely fatty"	11	(13.6)	16	(19.3)
(B) 25-50% dense "scattered fibroglandular density"	50	(61.7)	40	(48.2)
(C) 50-75% dense "hetereogeneously dense"	6	(7.4)	19	(22.9)
(D) >75% dense "extremely dense"	14	(17.3)	8	(9.6)
Overall				
(A) <25% dense "almost entirely fatty"	22	(15.8)	27	(19.7)
(B) 25-50% dense "scattered fibroglandular density"	78	(56.1)	66	(48.2)
(C) 50-75% dense "hetereogeneously dense"	21	(15.1)	31	(22.6)
(D) >75% dense "extremely dense"	18	(13.0)	13	(9.5)

FIGURE 7.1: Distribution of (a) percent and (b) absolute breast density in cases and controls

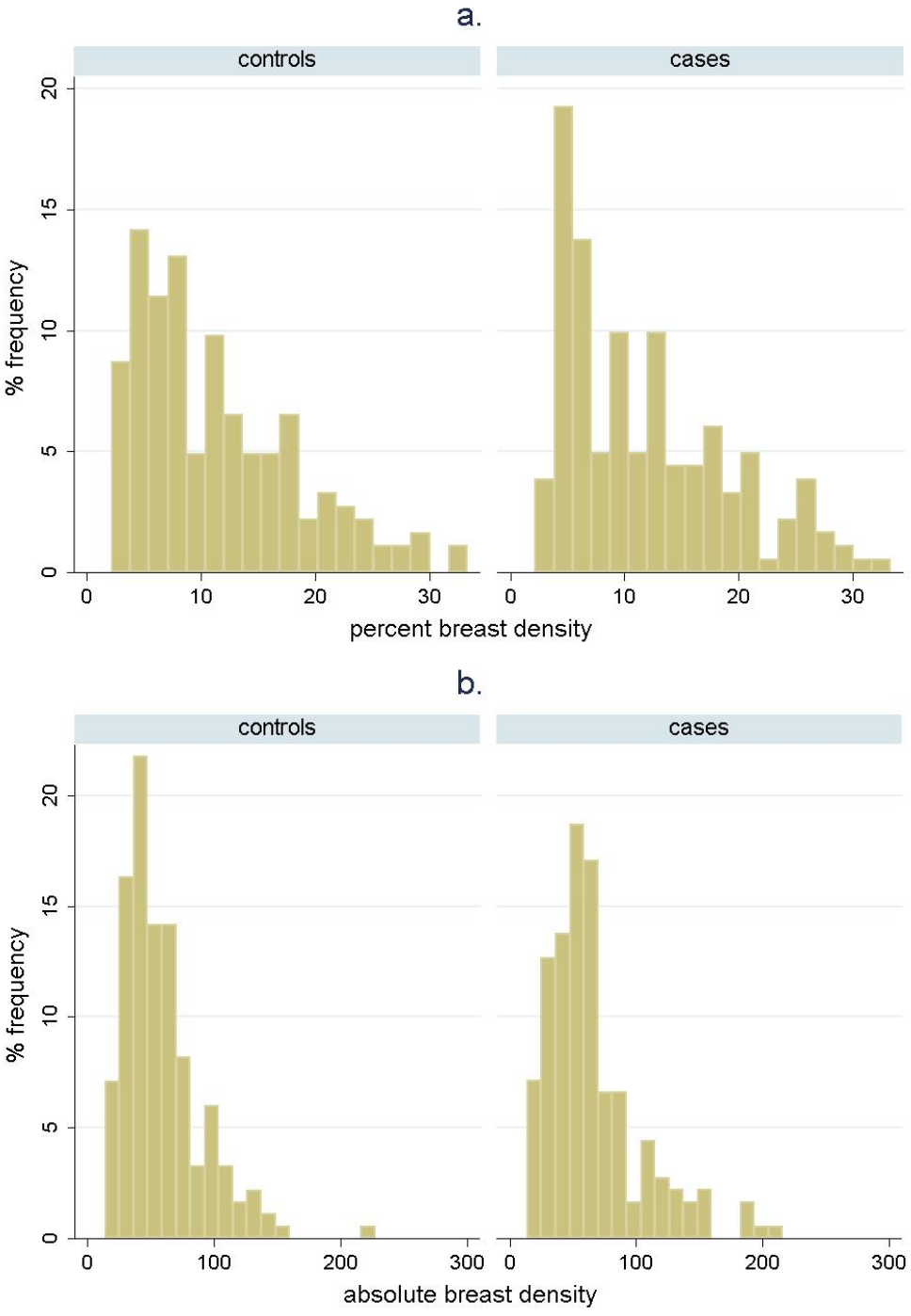


TABLE 7.4: ORs for developing breast cancer from unconditional logistic regression models using Volpara absolute density

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
adjusted for age						
absolute density	per 10 cm^3	366	1.07	1.00	1.13	.038
age	per year		1.00	.98	1.02	.809
adjusted for age and non-dense volume						
absolute density	per 10 cm^3	366	1.06	1.00	1.13	.069
age	per year		1.00	.98	1.02	.914
non-dense volume	per 10 cm^3		1.00	1.00	1.01	.602
adjusted for age and London residence						
absolute density	per 10 cm^3	251	1.06	.98	1.14	.123
age	per year		1.00	.97	1.02	.754
area of residence	Other UK		1.00		*(referent)*	.309
	London		.77	.46	1.28	
adjusted for age, non-dense volume and London residence						
absolute density	per 10 cm^3	251	1.06	.98	1.15	.129
age	per year		1.00	.97	1.02	.800
non-dense volume	per 10 cm^3		1.00	1.00	1.01	.841
area of residence	Other UK		1.00		*(referent)*	.302
	London		.76	.46	1.28	

We also investigated the association of percent and absolute density assessed using Volpara with breast cancer risk (Figure 7.1). The odds ratios relating to absolute density and non-dense volume are expressed per 10 cubic centimetre increase, whereas those relating to percent density are per 10% increase. Comparing the performance in risk prediction of the two density measures derived by the Volpara system (Tables 7.4 and 7.5), the results suggested that only absolute density had a significant main effect (p=.04), after age-adjustment. An increase of 10 cm^3 [1] in fibroglandular volume would lead to a 7% increase in the odds of breast cancer. The inclusion of non-dense volume to the model did not substantially alter the result, whereas adjusting for area of residence weakened it, possibly by causing a reduction in the number of observations (Table 7.4). Volpara percent density, on the other hand, had a non-significant association with breast cancer risk, even after adjustment for the other available risk factors (Table 7.5).

The introduction of the area of residence in the model (Table 7.4-7.5) caused a reduction

[1] the average absolue density is 62.5 cm^3

TABLE 7.5: ORs for developing breast cancer from unconditional logistic regression models using Volpara percent density

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
adjusted for age						
% density	per 10 %	366	1.11	.82	1.50	.513
age	per year		1.00	.98	1.02	.886
adjusted for age and non-dense volume						
% density	per 10 %	366	1.37	.94	2.00	.105
age	per year		1.00	.98	1.02	.872
non-dense volume	per 10 cm^3		1.01	1.00	1.01	.064
adjusted for age and London residence						
% density	per 10 %	251	1.20	.83	1.75	.335
age	per year		1.00	.97	1.02	.823
area of residence	Other UK		1.00		*(referent)*	.247
	London		.74	.44	1.23	
adjusted for age, non-dense volume and London residence n=251						
% density	per 10 %	251	1.40	.88	2.22	.154
age	per year		1.00	.97	1.02	.865
non-dense volume	per 10 cm^3		1.00	1.00	1.01	.267
area of residence	Other UK		1.00		*(referent)*	.273
	London		.75	.45	1.25	

in sample size for two reasons: (a) 25 foreign women were excluded from the analysis, in order to study the urban effect of London against other areas in the UK; (b) further 90 observations were dropped because of missing values. While it increases the standard error and reduces significance, adjustment for area of residence does not notably change the ORs for density, nor is area of residence significant. Therefore, adjustment for area of residence is probably counter-productive, lowering precision of estimation with no gain in validity.

Table 7.6 shows the ORs associated with quintiles of absolute dense volume, based on the control distribution, in order to facilitate the understanding of the shape of the association between Volpara absolute density and risk of developing breast cancer. There was no significant difference in the odds across the quintiles (p=.12). This may suggest

TABLE 7.6: ORs for developing breast cancer from an unconditional age-adjusted logistic regression model using the quintiles of Quantra absolute density among controls, N=366

	OR	[95% Conf.	Interval]	P>z
1st quintile ($0 - 34.75$ cm^3)	1.00	*(referent)*		.119
2nd quintile ($34.75 - 43.58$ cm^3)	.97	.49	1.92	
3rd quintile ($43.58 - 59.15$ cm^3)	1.23	.63	2.41	
4th quintile ($59.15 - 80.78$ cm^3)	1.43	.75	2.75	
5th quintile (> 80.78 cm^3)	1.35	.77	2.37	
age	1.00	.98	1.02	

that the association between breast cancer risk and density which we observed previously (Table 7.4) is substantially influenced by small numbers of extreme values among cases, as Figure 7.1 also suggests. It may also suggest a threshold effect as the main difference in Table 7.6 is between quintiles 1-2 combined and 3-5 combined.

Results of the analyses using Volpara measures after log-transformation are summarised in Table 7.7. Evidence of an association between breast cancer risk and absolute density appears weaker, i.e. only borderline significant (p=.06), in comparison with the results obtained using the original measures. However, overall they led to similar conclusions to those of Tables 7.4 and 7.5. We therefore returned to the original measures for subsequent analyses, to facilitate interpretation of results. In addition to this, in the analysis based on the log-transformed densities, the χ^2 statistics demonstrate that adjustment of the percent density variable for non-dense or total breast volume is returning the same information as absolute density but at the cost of adding a degree of freedom.

TABLE 7.7: ORs for developing breast cancer risk from unconditional age-adjusted logistic regression models using Volpara log-transformed measures

Factor	obs	OR	[95% Conf.	Interval]	P>z	model χ^2	df
adjusted for age							
log-absolute density	366	1.47	.98	2.19	.061	3.55	2
age		1.00	.98	1.02	.842		
adjusted for age and log-non-dense volume							
log-absolute density	366	1.45	.94	2.22	.091	3.59	3
age		1.00	.98	1.02	.890		
log-non-dense volume		1.03	.74	1.45	.846		
adjusted for age and log-total volume							
log-absolute density	366	1.44	.92	2.24	.109	3.59	3
age		1.00	.98	1.02	.890		
log-total volume		1.04	.71	1.53	.842		
adjusted for age							
log-% density	366	1.12	.79	1.58	.525	0.41	2
age		1.00	.98	1.02	.902		
adjusted for age and log-non-dense volume							
log-% density	366	1.51	.93	2.43	.094	3.55	3
age		1.00	.98	1.02	.904		
log-non-dense volume		1.49	.96	2.32	.079		
adjusted for age and log-total volume							
log-% density	366	1.44	.92	2.24	.109	3.59	3
age		1.00	.98	1.02	.892		
log- total volume		1.49	.96	2.33	.077		

The logistic regression analyses were repeated in women aged under 50 and 50 or more separately (Tables 7.8 and 7.9). Volpara absolute density was not significantly related to risk in both groups, but there was a considerable loss of power as the sample size was reduced from 366 to 130 and 236 respectively (Table 7.8). Also, the size of OR was the same as the significant results for all ages combined. Similar results were obtained from the analyses using percent density (Table 7.9).

TABLE 7.8: ORs for developing breast cancer from unconditional age-adjusted logistic regression models using Volpara absolute density, stratified by age

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
adjusted for age						
under 50						
absolute density	per 10 cm^3	130	1.07	.98	1.17	.151
age	per year		.98	.93	1.04	.577
50 or more						
absolute density	per 10 cm^3	236	1.07	.98	1.16	.123
age	per year		1.01	.97	1.04	.735
adjusted for age and non-dense volume						
under 50						
absolute density	per 10 cm^3	130	1.08	.98	1.19	.132
age	per year		.99	.93	1.04	.604
non-dense volume	per 10 cm^3		1.00	.99	1.01	.643
50 or more						
absolute density	per 10 cm^3	236	1.06	.97	1.15	.228
age	per year		1.00	.97	1.04	.798
non-dense volume	per 10 cm^3		1.00	1.00	1.01	.393

TABLE 7.9: ORs for developing breast cancer from unconditional age-adjusted logistic regression models using Volpara percent density, stratified by age

Factor	category	obs	OR	[95 Conf.	Interval]	P>z
adjusted for age						
under 50						
% density	per 10 %	130	1.16	.74	1.82	.505
age	per year		.99	.94	1.04	.686
50 or more						
% density	per 10 %	236	1.05	.68	1.62	.815
age	per year		1.01	.97	1.04	.741
adjusted for age and non-dense volume						
under 50						
% density	per 10 %	130	1.32	.73	2.39	.351
age	per year		.99	.94	1.04	.679
non-dense volume	per 10 cm^3		1.00	.99	1.02	.511
50 or more						
% density	per 10 %	236	1.38	.81	2.35	.235
age	per year		1.00	.97	1.04	.806
non-dense volume	per 10 cm^3		1.01	.99	1.01	.087

FIGURE 7.2: Comparison of medians and inter-quartile ranges of Volpara (a) percent density, (b) non-dense volume and (c) absolute density by age group, between cases and controls

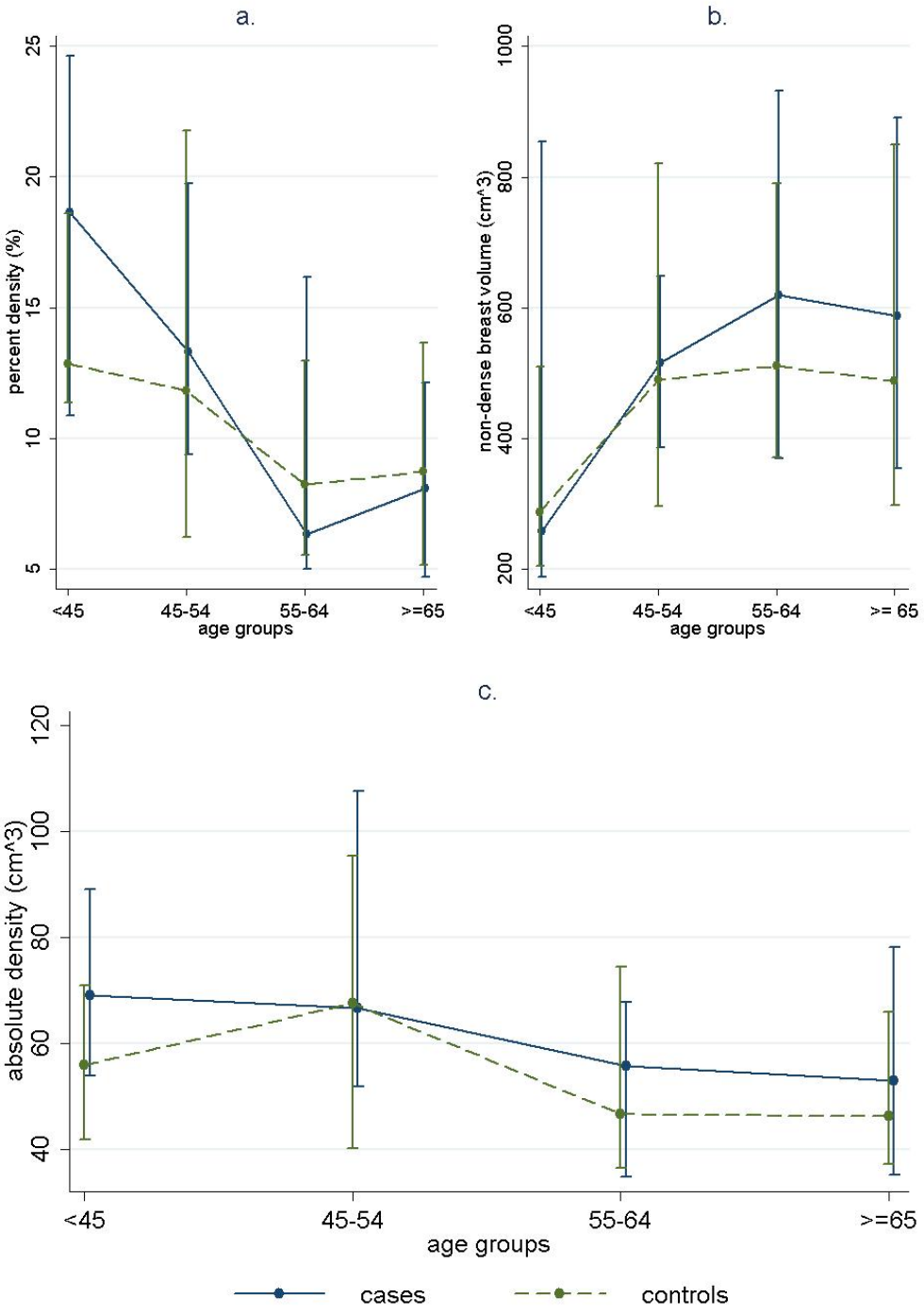


Figure 7.2 depicts the differences in breast composition between cases and controls according to age group. Results showed no substantial difference at any age and in any of the breast characteristics here presented, (a) percent density, (b) non-dense breast

volume and (c) absolute density.

TABLE 7.10: ORs for developing breast cancer from unconditional age-adjusted logistic regression models using Volpara absolute density, stratified by UK location of residence

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
adjusted for age						
London						
absolute density	per 10 cm^3	152	1.10	.99	1.21	.069
age	per year		1.01	.98	1.05	.374
other UK						
absolute density	per 10 cm^3	99	1.02	.91	1.13	.768
age	per year		.98	.94	1.01	.162
adjusted for age and non-dense volume						
London						
absolute density	per 10 cm^3	152	1.08	.97	1.19	.163
age	per year		1.01	.98	1.04	.546
non-dense volume	per 10 cm^3		1.01	1.00	1.02	.206
other UK						
absolute density	per 10 cm^3	99	1.06	.94	1.20	.355
age	per year		.98	.95	1.02	.364
non-dense volume	per 10 cm^3		.99	.98	1.00	.121

In the analyses stratified by area of residence (Tables 7.10, 7.11 and 7.12) 25 women living abroad were excluded to focus on the potential influence of urban and provincial environment within the UK on breast density and its risk-predictive ability. A further 90 observations were dropped because of missing values resulting in a reduced sample size which may affect the reliability of the results. Volpara absolute density seemed to better discriminate cases from controls among London residents, although this was only only borderline significant (p=.07), whereas Volpara percent density was borderline significant (p=.08) in discriminating cases from controls among British women not living in the capital.

According to Figure 7.3, cases residing in London or elsewhere in the UK have a very similar pattern in absolute density (c) across ages, and have no significant differences either in (a) percent density or (b) non-dense breast volume. Likewise, among controls (Figure 7.4) we observed no difference among (a) percent density, (b) non-dense breast

TABLE 7.11: ORs for developing breast cancer from unconditional age-adjusted logistic regression models using Volpara percent density, stratified by UK location of residence

factor	category	obs	OR	[95% Conf.	Interval]	P>z
adjusted for age						
London						
% density	per 10 %	152	.97	.61	1.55	.906
age	per year		1.01	.98	1.04	.658
other UK						
% density	per 10 %	99	1.85	.93	3.67	.078
age	per year		.99	.95	1.03	.510
adjusted for age and non-dense volume						
London						
% density	per 10 %	152	1.49	.81	2.76	.199
age	per year		1.01	.98	1.04	.524
non-dense volume	per 10 cm^3		1.01	1.00	1.02	.035
other UK, n=99						
% density	per 10 %	99	1.68	.77	3.69	.196
age	per year		.99	.95	1.03	.506
non-dense volume	per 10 cm^3		1.00	.99	1.01	.639

TABLE 7.12: Distribution of cases and controls according to area of residence and age at mammogram

	controls (%)		cases (%)		Total (%)	
London, n=152						
Under 50	33	(40.7)	29	(40.8)	62	(40.8)
50 or more	48	(59.3)	42	(59.2)	90	(59.2)
Other UK, n=99						
Under 50	15	(32.6)	23	(43.4)	38	(38.4)
50 or more	31	(67.4)	30	(56.6)	61	(61.6)

FIGURE 7.3: Comparison of medians and inter-quartile ranges of Volpara (a) percent density, (b) non-dense volume and (c) absolute density by age group, between UK residents from London or other British regions, in cases

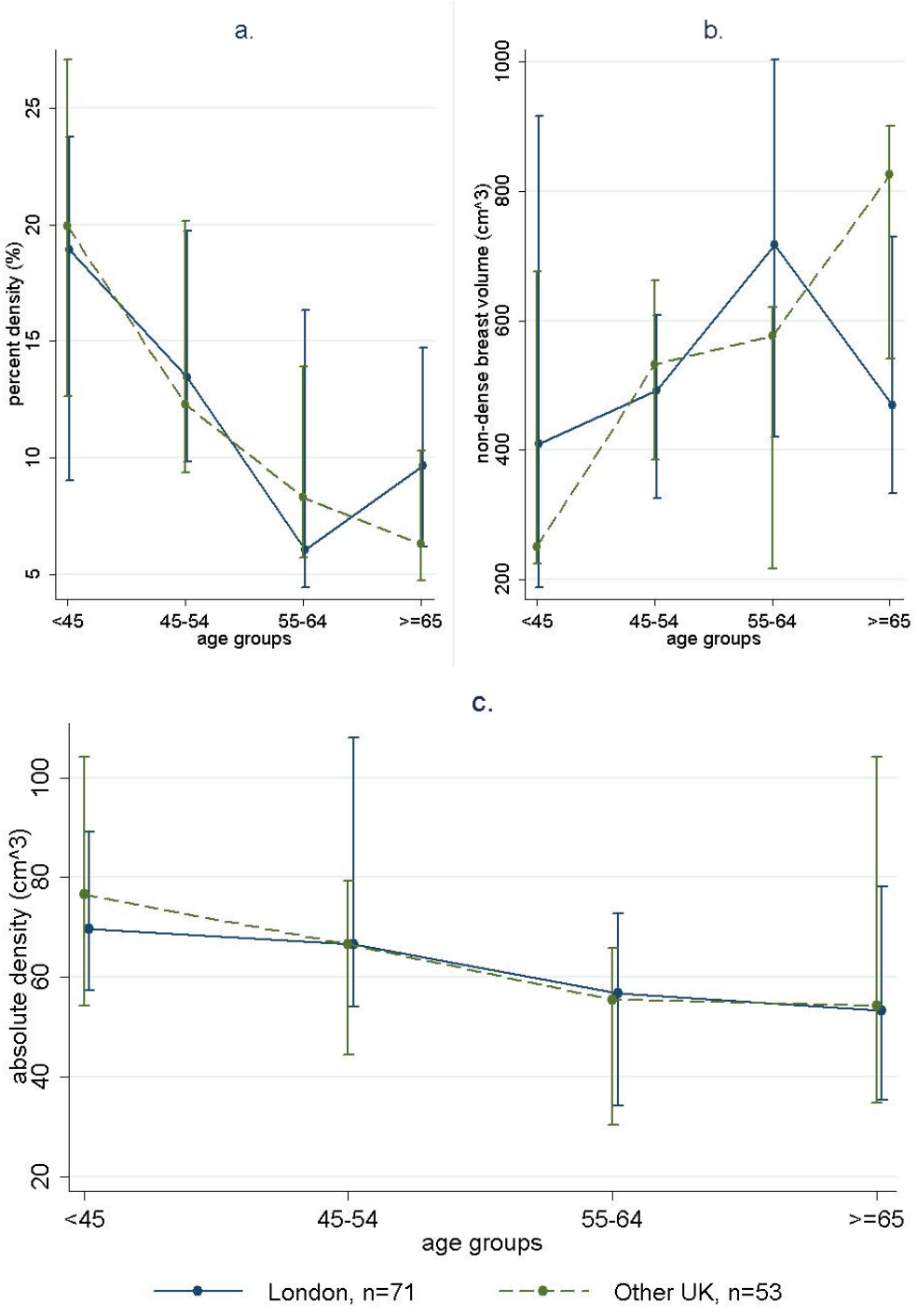
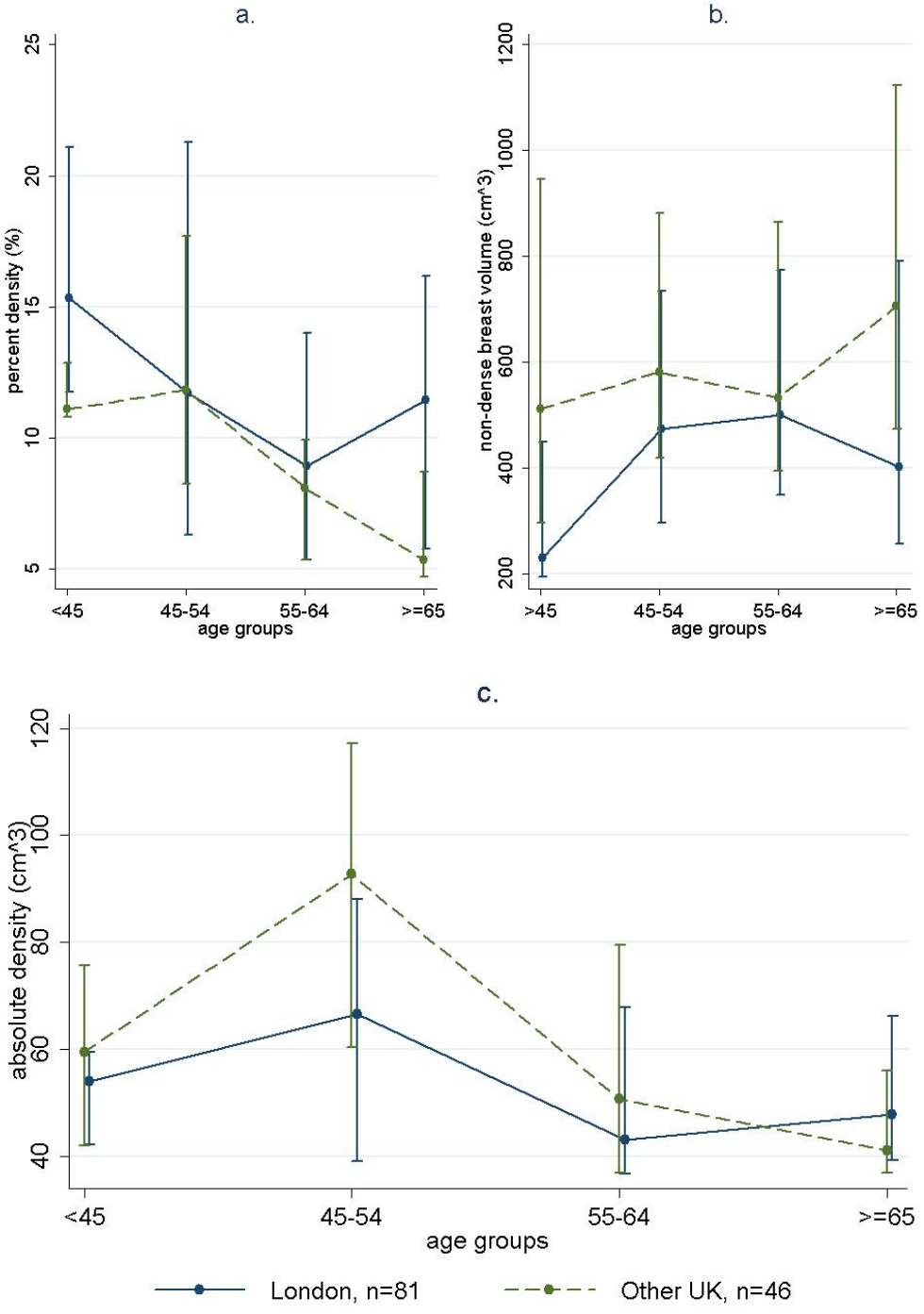


FIGURE 7.4: Comparison of medians and inter-quartile ranges of Volpara (a) percent density, (b) non-dense volume and (c) absolute density by age group, between UK residents from London or other British regions, in controls



volume and (c) absolute density comparing subjects living in London or in another UK area. Nevertheless it can be observed that London control residents have consistently higher median percent density than women resident elsewhere in the UK, despite the fact that women with a different area of residence have mostly higher absolute dense volume. This is observed because London residents had consistently lower non-dense volume, according to Volpara estimates.

FIGURE 7.5: Boxplots of Volpara (a) percent and (b) absolute breast density stratified by BIRADS categories and case-control status)

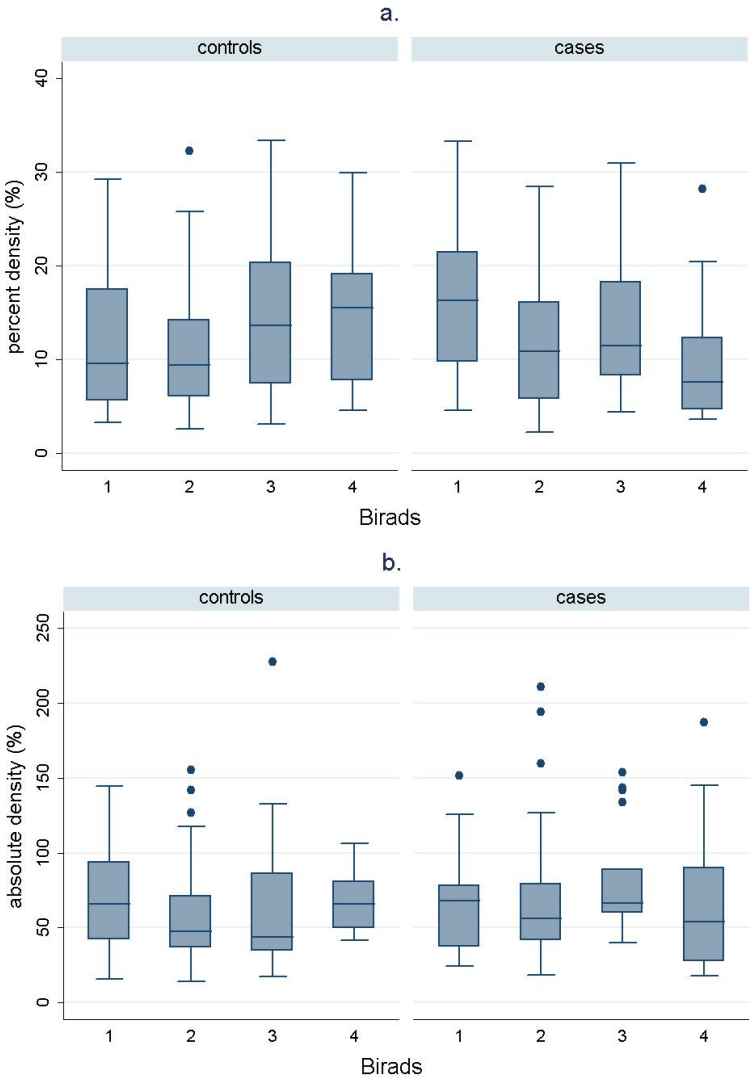
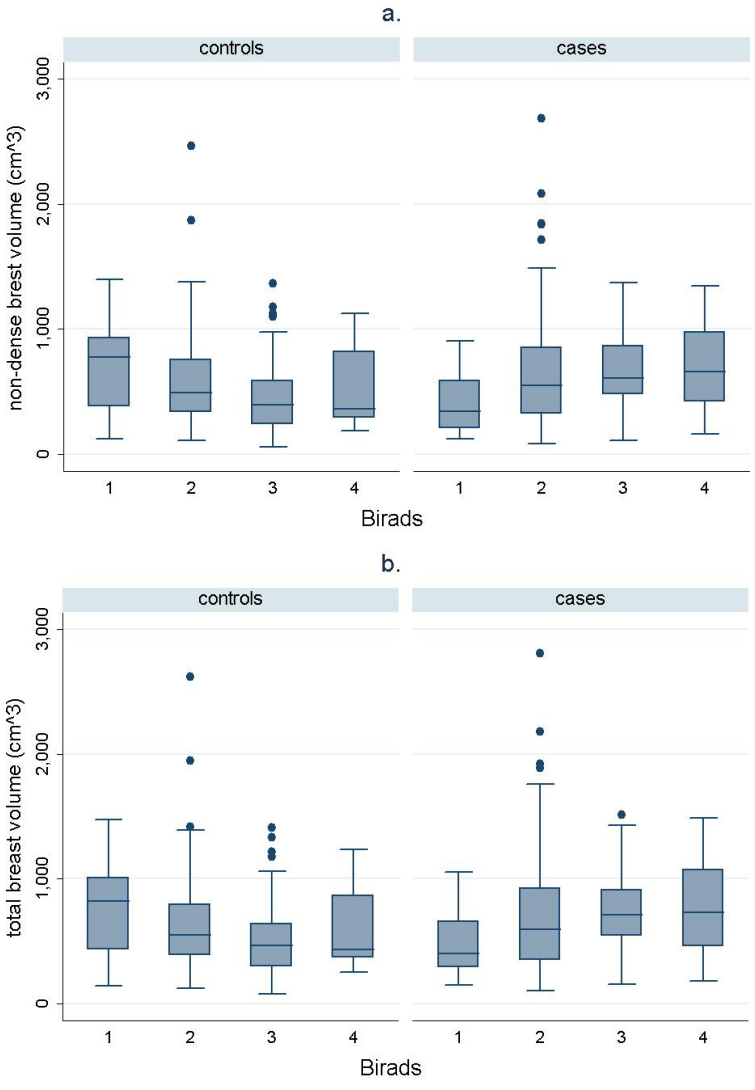


Figure 7.5 shows the relationship between BIRADS and Volpara density measures. These boxplots represent the distribution of percent and absolute density in each BIRADS category. Ideally, with increasing BIRADS category, we would expect to observe an increasing trend, at least of Volpara percent density, as evidence of a positive association

between the two methods. Results from tests for trend in Volpara percent density with increasing BIRADS category suggest the absence of an overall association (p=.70), but a positive association among controls (p=.05) and a negative trend among cases (p=.01). However, there appears to be no association between BIRADS and Volpara absolute density, and tests for trend confirm this lack of association between the two measures (overall: p=.92, controls: p=.95 and cases: p=.99).

FIGURE 7.6: Boxplots of Volpara (a) non-dense and (b) total breast volume stratified by BIRADS categories and case-control status)



Analyses of the other measures estimated by Volpara, non-dense and total breast volume (Figure 7.6), help to explain these results. Overall no significant trend was observed for either measure (non-dense breast volume: p=.59, total breast volume: p=.58), whereas a significant declining trend for both was observed among controls (non-dense breast

volume: p=.04, total breast volume: p=.04), and a significant increasing trend among cases (non-dense breast volume: p<.01, total breast volume: p<.01).

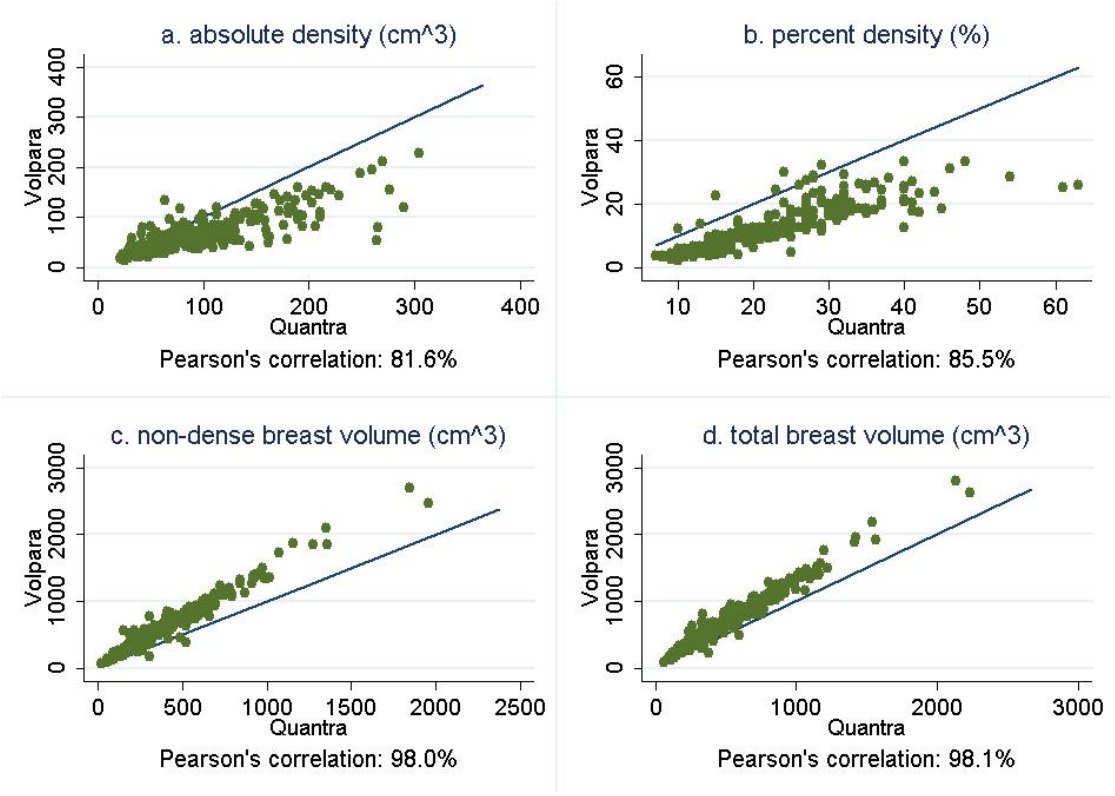
TABLE 7.13: ORs for developing breast cancer risk from unconditional age-adjusted logistic regression models using Volpara absolute density and adjusting for BIRADS categories

Factor	category	obs	OR	[95% Conf.	Interval]	P>z
absolute density	per 10 cm^3	276	1.06	.99	1.14	.086
BIRADS	0-25% dense		1.00	*(referent)*		.794
	25-50% dense		1.51	.78	2.91	
	50-75% dense		.80	.36	1.79	
	75-100% dense		1.73	.69	4.33	
age	per year		1.00	.97	1.02	.619

Given the lack of association between BIRADS and Volpara absolute density, we performed an age-adjusted unconditional logistic regression including both measures, Table 7.13. These results provide further evidence of the lack of impact that BIRADS classification has on Volpara absolute density and *vice versa*, as the values do not vary greatly from those reported previously in Table 7.4. However, the association between Volpara density and breast cancer risk was slightly weakened.

Further results regarding BIRADS on this dataset are presented in Chapter 5.

FIGURE 7.7: Association between Volpara and Quantra estimates



Analyses of the association between Quantra and Volpara methods showed high and significant correlations between all the breast composition characteristics estimated, (a) percent and (b) absolute density, (c) non-dense and (d) total breast volume (Figure 7.7). Volpara density measures were consistently lower than those estimated by Quantra, perhaps because the former method does not treat skin as dense tissue. These plots also highlighted a worse agreement between the two methods for higher values of Quantra. As regards non-dense and total breast volumes, Volpara estimates were consistently higher. Consequently standard deviations were higher for Volpara measures of non-dense and total breast volume than those obtained by Quantra (395.0 vs 291.8 for non-dense breast volume and 406.1 vs 334.3 for total breast volume). Conversely, Volpara density estimates had a consistently lower standard deviations compared to Quantra density measures (7.1 vs 9.7 for percent density and 35.2 vs 60.2 for absolute density).

The association between Quantra and Volpara estimates for absolute density was also investigated by comparing their quintile distributions in a contingency table (Table 7.14). The cut-off points vary between the two methods, but because of the differences in the two algorithms, standardised categories would lead to a poorer agreement, thus it was preferred to compare Quantra and Volpara estimates on their own scales. Results indicated an overall agreement of 54.7 %, with 26.5 % of subjects classified in higher quintiles

by Volpara, and *vice versa* for the remaining 18.8%. The absence of observations in the two most extreme cells (top right and bottom left) and low percent frequency in the adjacent cells are also indicators of good agreement between the two measures.

TABLE 7.14: Absolute and relative (%) cell frequencies comparing Volpara and Quantra quintile distributions for absolute density

	Quantra					
Volpara	0-46.6 cm^3	46.6-68 cm^3	68-91 cm^3	91-125.4 cm^3	>125.4 cm^3	Total
0-34.8 cm^3	32(11.6)	16(5.8)	1(.4)	0(.0)	0(.0)	49(17.8)
34.8-43.6 cm^3	10(3.6)	18(6.5)	7(2.5)	5 (1.8)	1(.4)	41(14.9)
43.6-59.2 cm^3	4(1.5)	8(2.9)	24(8.7)	9 (3.3)	7(2.5)	52(18.8)
59.2-80.8 cm^3	2(.7)	5(1.8)	25(9.1)	29(10.5)	6(2.2)	67(24.3)
>80.8 cm^3	0(.0)	4(1.5)	4(1.5)	11(4.0)	48(17.4)	67(24.3)
Total	48(17.4)	51(18.5)	61(22.1)	54(19.6)	62(22.4)	276(100.0)

Finally, we compared Volpara and Quantra with respect to their associations with breast cancer risk (Table 7.15), in terms of the effect per 10 cm^3 for absolute density and per 10% for percent density. We then report the standardised effects (i.e. OR per standard deviation) to enable a scale-free comparison of the two methods. ORs and p-values were very similar between the two methods for both absolute and percent density, and this result is confirmed by the standardised ORs.

Further results assessing Quantra using this dataset are presented in Chapter 5.

TABLE 7.15: ORs for developing breast cancer risk from age-adjusted unconditional logistic regression using Volpara and Quantra density estimates

Factor		obs	OR	[95% Conf.	Interval]	P>z
Absolute density						
Volpara	per 10 cm^3	366	1.07	1.00	1.13	.038
age	per year		1.00	.98	1.02	.809
Quantra	per 10 cm^3	400	1.04	1.00	1.07	.030
age	per year		1.00	.98	1.02	.848
Percent density						
Volpara	per 10 %	366	1.11	.81	1.50	.513
age	per year		1.00	.98	1.02	.886
Quantra	per 10 %	400	1.15	.92	1.42	.221
age	per year		1.00	.99	1.02	.679
Standardised OR						
Volpara	per 1 standard deviation	366	1.26	1.01	1.56	.038
age	per year		1.00	.98	1.02	.809
Quantra	per 1 standard deviation	400	1.25	1.02	1.53	.030
age	per year		1.00	.98	1.02	.848

7.4 Discussion

The data support the hypothesis that absolute volume of dense tissue could be a better predictor of breast cancer risk than percent density. This may be because absolute density is a closer measure of the quantity of tissue at risk of malignant transformation. However, despite being statistically significant, the ORs do not indicate a particularly strong relationship, i.e. increases in absolute density do not reflect substantial increase in risk. In fact, when comparing the extreme quintiles (Table 7.6), the odds of expected cases and controls are not substantially different (OR: 1.35, 95% CI (.77-2.37)).

As noted in Chapter 5, the study population here is younger and likely to be of higher socioeconomic status than generla population women in the UK screening programme. However, one would expect case-control differences in breast density to prevail in our population as in the general population.

The overall lack of association between Volpara and BIRADS is surprising and raises questions about the validity of both measures. In particular, stratified analyses revealed a positive association between BIRADS and Volpara percent density among controls and

a negative one among cases. This was observed because Volpara non-dense and total breast volume decreased with increasing BIRADS category among controls and increased along with BIRADS categories among cases. Volpara absolute volume and BIRADS categories, on the other hand, appeared independent in both groups. Likewise, Quantra measures of non-dense and absolute breast volume increased steadily with increasing BIRADS categories [Chapter 5]. This leads to doubts about what the two volumetric methods are actually assessing. Volpara percent density was non-significantly related to breast cancer risk even after adjusting for age and non-dense volume. Also, the association between BIRADS and risk of developing breast tumours was weaker than expected [6] [Chapter 5]. This may be due to an inherent difficulty in visual estimation of density from processed digital images. The use of processed images for visual assessment, and raw data for automated assessment, may have implications for the association between the two.

In 2013 Wang and colleagues [50] ran a study on 99 women comparing mammographic measures of volumetric breast density, including Volpara, to MRI density [47]. They found that average total breast volume was higher for Volpara than for MRI, probably due to the fact that Volpara includes the edge in their estimates, a moderate correlation and agreement were observed between the two methods measuring percent density (squared Pearson's correlation coefficient: .73, and kappa statistics: .68). These results are encouraging, suggesting that Volpara is capturing the right amount of dense tissue, nevertheless larger datasets are needed for validation.

Investigation of the relationship between the two volumetric methods suggest a strong association and similarity. Pearson's correlation coefficients showed an almost perfect correlation (98%) for non-dense and total volume, and were lower, although still very high (over 80%), in the density measures. This lower correlation is likely to be due to the exclusion of skin tissue from Volpara dense estimates. Nevertheless, for the association with breast cancer risk (Table 7.15), Volpara and Quantra density estimates were very similar, both for ORs and p-values. The standardised ORs provide further support of the hypothesis of equivalence between the two techniques, in terms of breast cancer risk prediction.

In terms of costs, both Quantra and Volpara have multiple packages with diminishing additional costs per additional unit, but both would cost approximately US $ 50,000 for the software and hardware. On cost alone, it would be difficult to decide between the two.

As in the Quantra study (Chapter 5), this dataset was comprised of symptomatic cases, which may have affected our findings due to masking. Again, if this were the case, as in Chapter 5, we would have expected a much stronger association between density and risk.

Interestingly, for analysis of area of residence, the higher percent density among London residents without breast cancer (Figure 7.4) was driven by a lower non-dense volume rather than a higher dense volume. This is in agreement with the Quantra results [Chapter 5] and is consistent with the hypothesis that a higher percent density in London residents is largely due to lower BMI in London women [246, 249].
The main weakness of this study lies in the limited sample size, which suffered a further loss of information during merging. The analyses may have also benefited from additional details on other risk factors, which could have helped to explain and clarify the unexpected results we observed. Thus these findings require further validation on a richer and larger dataset.

In conclusion, Volpara absolute density is a breast cancer risk predictor of similar strength to Quantra, and the two measures are strongly related. Validation in a larger dataset with information on potential confounders would be useful.

Chapter 8

Measuring mammographic density: results, issues and potential implications

8.1 Introduction

In the review in Chapter 1, we described how the methods of measurement of mammographic density have developed since 1976 [5] and discussed some of the outstanding issues delaying a more extensive use of density in breast cancer screening.
Mammographic density is a strong risk factor for breast cancer. Evidence of this association has been published in the past decades [15, 16, 254, 255] and a recent meta-analysis [6] confirmed that a dense breast indicates a four- to six-fold increase of risk of developing a breast cancer compared to a non-dense. Nevertheless mammographic density is underused as a means of assessing breast cancer risk. One explanation for this is that, despite the wide availability of technology and the high attendance in breast screening programs, how mammographic density should be assessed effectively and feasibly in routine screening has not been determined yet.
To date, the most commonly applied methods for measuring density are based on visual assessment (usually BIRADS), and interactive thresholding (Cumulus, a computer-assisted technique), both of which have drawbacks. Firstly, visual assessments depend on human judgement, and this introduces subjectivity in the form of inter- and intra-rater variability, raising questions of reproducibility [27]. Advancements in technology led to the introduction of interactive thresholding [256], arguably the current gold standard,

which is less subjective, although it still relies on reader input, and has the advantage of providing more comprehensive information about total breast area, dense area and non-dense area [28, 30, 256]. However, the main barrier to density being routinely measured with this method lies in its time-consuming nature that makes it impractical for density assessment in routine population screening. In addition, both methods reduce a three-dimensional feature, the breast, to its bi-dimensional projection obtained on a mammogram, and may lose accuracy by not taking into account the thickness of the dense tissue of the breast. It is therefore anticipated that volumetric measures of dense tissue will provide more precise estimates of the amount of fibroglandular tissue within the breast and thus will be a better predictor of breast cancer risk [201]. For this reason, recent years witnessed the development of several fully-automated volumetric methods [36–42, 51, 52, 178, 236], that are undergoing validation.

In this chapter, we focus on some of these issues relating to variability, from both measurement error and genuine population variation in breast composition, in measurement of mammographic density and suggest possible solutions. Firstly, we analyse intra- and inter-reader agreement in visual (21 categories) and Cumulus [256] assessments, using data from the IBIS-I study [17]. Secondly, we investigate the potential effect of measuring error on observed risk associations, in density assessed with the novel fully-automated volumetric method Quantra [51, 178], using datasets introduced in chapters 5-6, and applying a modification of Rosner's and colleagues's method [257–259] for correcting for the measurement error. This method is based on the method of moments and requires the estimates of the variability components from an external dataset with serial assessments of the factor prone to measurement error, in this case Quantra absolute density estimates. Then, we study the relationship between two- and three-dimensional density assessments, focusing on their difference in variability, albeit in data from two different population, with data from CADET1 [4] and the Quantra longitudinal study [Chapter 6]. Finally, we compare the association with breast cancer risk of most of the density measures available for this project: visual (Chapter 2), Cumulus (Chapters 2 and 4), Quantra (Chapter 5) and Volpara (Chapter 7). In this analysis, we use the exponential of the standard logistic regression coefficient as the measure of association for each density method. Thus each measure of association is the odds ratio per standard deviation of the density measure, so that all methods are assessed on a comparable scale.

8.2 Inter- and intra-reader agreement (IBIS-I)

Cumulus is said to be preferable to visual assessment because it provides more complete information regarding breast composition, thus leading to more accurate estimates of breast cancer risk, and it is to some extent less subjective, because it is partially automated. However there is a scarcity of literature regarding reader reproducibility.
In this section we focus on quantifying variability in Cumulus estimates from the same mammograms, and compare them with visual density assessment (to the nearest 5%). We also evaluate both inter- and intra-reader agreement in visual assessment.

8.2.1 Materials and Methods

8.2.1.1 Data Collection

The IBIS-I chemoprevention study was an international trial designed to evaluate the role of tamoxifen for breast cancer risk reduction in high-risk women. The study population comprised a total of 7,152 women aged 30-70 years. A nested case-control study was conducted to investigate relationships between treatment, change in mammographic density and known breast cancer risk factors. It comprised data from 123 cases and 942 controls and some results were published in 2011 [17].
Mammographic density was assessed on these 1,065 subjects both at baseline and after 12-month follow-up. Three trained readers (J. Stone, J. Warwick and P. Allgood) used Cumulus software to measure both absolute and percent dense area. The same readers received training by one experienced radiologist (R.M.L. Warren) and classified the proportion of dense tissue in the subjects visually on a 21-point scale (to the nearest 5%). Prof. Warren also performed visual density assessment on the whole sample, on two separate occasions in order to allow analyses of intra-reader agreement. One reader (PA) performed her assessment on only 56% of the sample. However most of the subjects (98%) had density assessed with Cumulus by 2 readers (JS and JW) and visually by 3 (JS, JW and RW). Note that there is a difference in the number of mammograms read visually and using Cumulus by each reader.
Since Prof. Warren provided the training and was by far the most experienced reader, we considered her as the referrent reader when comparing visual assessments.

8.2.1.2 Statistical Analyses

The focus of the analyses reported here was the inter-reader agreement and not the change in density between baseline and follow-up, hence each mammogram was counted separately, allowing the size of the sample for the analyses to double. To verify that the inclusion of two mammograms for each subject would not influence the results, we repeated the analyses using only baseline mammograms. As we obtained very similar outcomes (results not shown), we decided to keep all the mammograms available. Likewise the exclusion of the cancer cases gave similar results.

For an exploratory analysis, we compared the distributions in the readers' assessments using mean, SD and percentiles for Cumulus estimates, and histograms for visual.

We compared the readers in pairs using Bland-Altman limits of agreement plots [260] and concordance correlation coefficients [261, 262] for Cumulus continuous estimates and Cohen's kappas for visual discrete measures, as well as average absolute differences. Bland and Altman are sceptical about the intraclass correlation coefficient (ICC), as it models measurements by different raters as replicates, whereas they may be systematically different due to individual rater tendencies. It has been noted, however, that the ICC is a useful summary of the relative size of the between-subject variation (reader variation of the attribute from person to person) and the within-subject variation (variation between measures on the same subject due to measurement error) [263]. We therefore calculated ICCs as a secondary measure.

Prof. Warren repeated her readings in two occasions 6-month apart, and intrareader agreement was also studied.

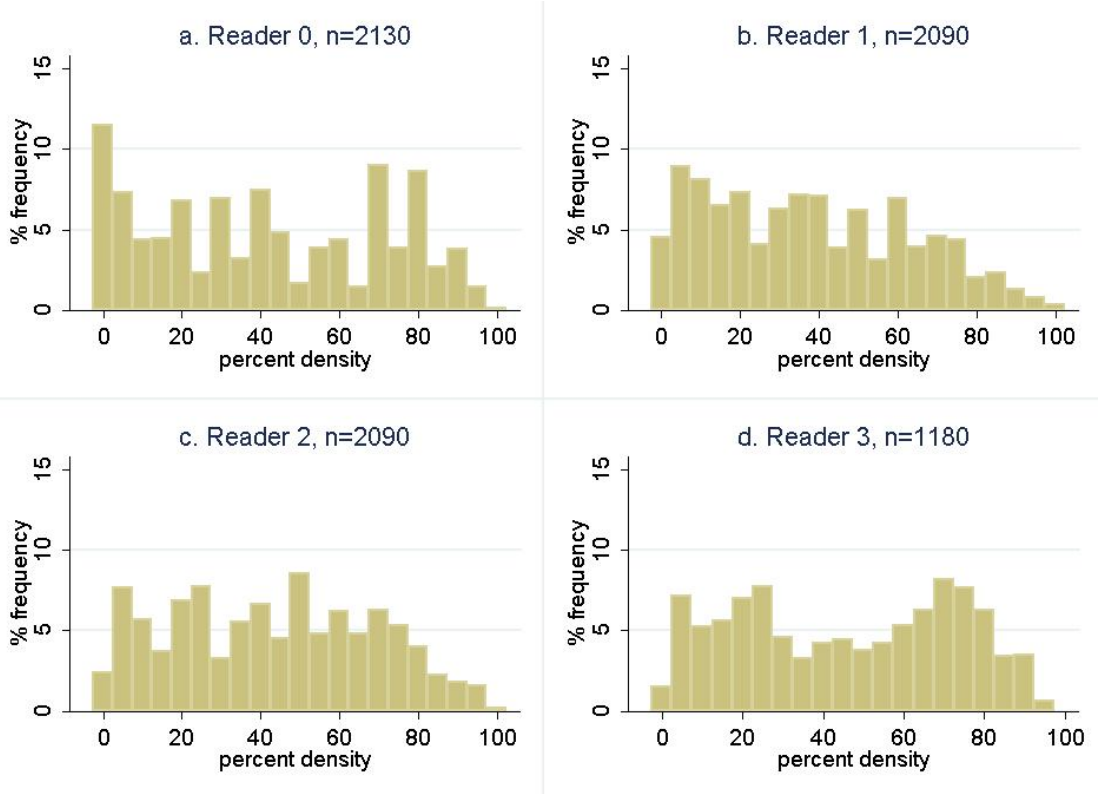
8.2.2 Results

The distributions of all the density measures considered are summarised in Table 8.1 and Figure 8.1. Concerning Cumulus estimates, the two main readers (1 and 2) had similar distributions (Wilcoxon rank-sum test: p=.12 for absolute density and p=.15 for percent density), whereas assessments by the third reader appeared consistently lower, both for absolute and percent density (Wilcoxon rank-sum test: p<.01 in each comparison). For visual assessments (Figure 8.1), Prof. Warren is identified as "reader 0", having trained the other three. The distributions, in this case, appeared uneven but, overall, similar.

TABLE 8.1: Summary statistics of Cumulus measures by reader

Reader	obs	Mean	SD	1st quartile	median	3rd quartile
Absolute density (cm^2)						
1	2084	24.62	15.54	13.75	23.29	33.33
2	2099	23.93	15.20	13.34	22.16	32.50
3	1192	21.08	14.64	10.33	18.77	28.85
Percent density (%)						
1	2084	36.24	21.06	20.21	36.43	51.85
2	2099	37.25	21.37	20.71	36.83	53.16
3	1192	32.78	20.23	16.10	31.72	47.83

FIGURE 8.1: Distribution of visual density assessments according to reader

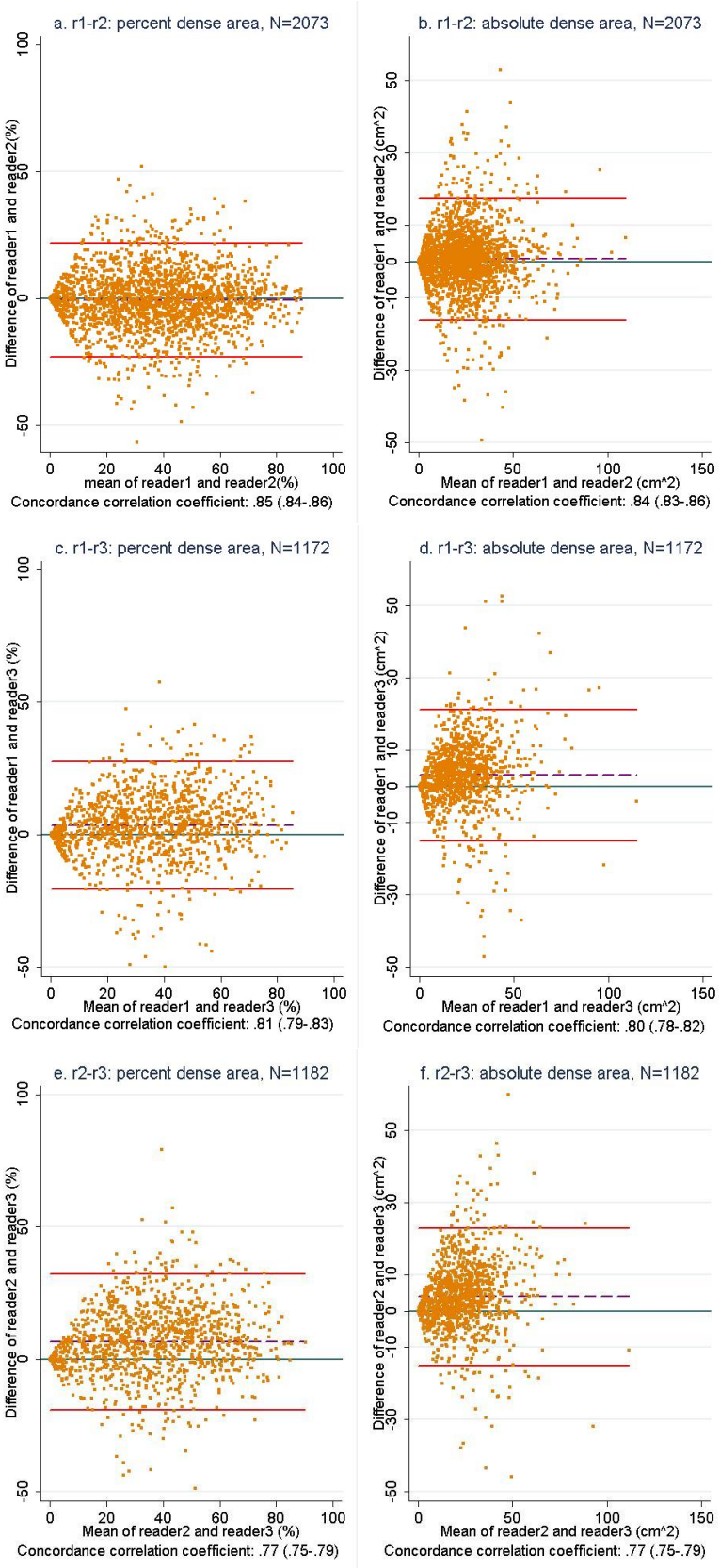


Overall concordance coefficients (Figure 8.2) suggested about 80% concordance between readers for both absolute and percent density estimated with Cumulus, with a peak of 85% between the two main readers. Bland-Altman limits of agreement plots allow to evaluate how the agreement was affected by the level of density, in other words if the mammograms were overall darker or whiter. The shapes of the scatter plots suggest a better agreement for extremely low values, i.e. darker mammograms, then the variability increased in an approximately linear fashion with increasing density estimates with a peak at about 40% for percent dense area and between 30 and 50 cm^2 for absolute measures. On the other hand, the width between the 95% limits of agreement line indicated that the estimates were quite variable between readers, and also support the hypothesis that readers 1 and 2 have the better agreement. Further evidence for this hypothesis is the fact that the lines of observed average agreement between readers 1 and 2 are close to the one of perfect agreement both for percent and absolute density estimates (Figures 8.2.a and .b). Comparing these two readers with reader 3 (Figures 8.2.c-.f), it appeared that this third reader would estimate density with on average lower values, as suggested by the scatter plots and especially by the observed average agreement line. It can be speculated that the better agreement between readers 1 and 2 is due to the larger samples of mammograms used for the analyses. However repeating the analyses on a reduced number of mammograms, including only those read also by reader 3, the concordance correlation agreement between the first two readers remained the highest (.86 (.85-.88) for both percent and absolute dense area, compared with .81 (.79-.83) for percent and .80 (.78-.82) for absolute between readers 1 and 3, and .77 (.75-.79) for percent and .77 (.75-.79) for absolute between readers 2 and 3).

In this comparison there was no reader known to be the referent as they all received the same training. However, comparing the visual readings with reader 0 (our referent), reader 1 had the highest concordance correlation coefficient (.89 (.88-.89)), and was therefore considered as referent for the Cumulus readings.

Note that readers 1-3 were abbreviated as r1-3.

FIGURE 8.2: Bland-Altman's limits of agreement plots and concordance correlation coefficients between readers



y=0 is the line of perfect average agreement, the dotted line is the observed average agreement and the two external lines are the 95% limits of agreement.

Absolute differences between readers assessments are summarised in Table 8.2. Cumulus density estimates were on average less than 8 cm^2 apart, but for almost 120 mammograms were over 15 cm^2 when comparing reader 3 with the two other readers. Differences were lower among the readers 1 and 2, and differences between assessments of at least 15 cm^2 were recorded only on 8% of the mammograms. Cumulus percent dense estimates were on average approximately 10 units apart between readers, with differences of over 20% for about 1 mammogram in 10. Again, readers 1 and 2 had the better agreement, whereas reader 2 and reader 3 disagreed more often.

TABLE 8.2: Absolute differences in Cumulus density estimates between readers

Readers	Obs.	Mean	SD	Min	Max	1^{st} decile	9^{th} decile
Absolute density (cm^2)							
r1-r2	2073	5.86	6.33	.00	53.09	.651	13.65
r1-r3	1172	6.90	6.90	.00	52.63	.744	15.65
r2-r3	1181	7.51	7.34	.00	59.97	.831	17.03
Percent density (%)							
r1-r2	2073	8.39	7.83	.00	56.94	1.11	18.48
r1-r3	1172	9.45	8.61	.00	57.48	1.01	20.91
r2-r3	1181	11.05	9.38	.00	57.20	1.34	23.62

The ICC was 82% (95% CI (81%, 83%)) for both absolute and percent dense area, indicating considerably more variance between subjects than between measures on the same subjects, and suggesting reproducibility of density assessed with Cumulus by trained readers, both in absolute and percent terms.

TABLE 8.3: Weighted Cohen's kappas for visual density estimates between readers

Readers	Agreement	Expected Agreement	Kappa	Std. Err.	Z	Prob>Z
r0-r1	90.19%	67.95%	.69	.014	50.77	<.001
r0-r2	86.14%	67.87%	.57	.014	41.51	<.001
r0-r3	87.15%	66.86%	.61	.019	32.8	<.001
r1-r2	88.74%	70.30%	.62	.013	46.32	<.001
r1-r3	88.76%	68.88%	.64	.018	35.5	<.001
r2-r3	87.15%	69.87%	.57	.018	31.22	<.001

TABLE 8.4: Absolute differences in visual percent (to the nearest 5%) density estimates between readers

Readers	Obs.	Mean	SD	Min	Max	1^{st} decile	9^th decile
r0-r1	2090	9.81	9.06	0	80	0	20
r0-r2	2090	13.86	12.89	0	95	0	30
r0-r3	1180	12.85	12.44	0	75	0	30
r1-r2	2090	11.26	12.26	0	85	0	25
r1-r3	1180	11.24	11.06	0	65	0	25
r2-r3	1180	12.85	12.14	0	80	0	25

Visual classification (21 categories) showed excellent inter-rater agreement between readers, between 86% and 90% (Table 8.3). There was on average a distance of 2 categories (over 10%) for every pair of readers (Table 8.4), except for the pair composed by reader0 and reader 1 who had an average smaller distance (9.8%), that suggests a better agreement. In around 10% of cases, two readers would classify the mammograms at least four units apart (a difference of 20% or more). In particular, reader 2's assessments were the furthest from reader 0's, who was considered as the gold standard because of her experience.

The intraclass correlation coefficient, calculated over all four readers, was 81% (80%, 82%), supporting the reliability of visual density assessments.

FIGURE 8.3: Correlation in visually assessed percent density between first and second reading by reader 0

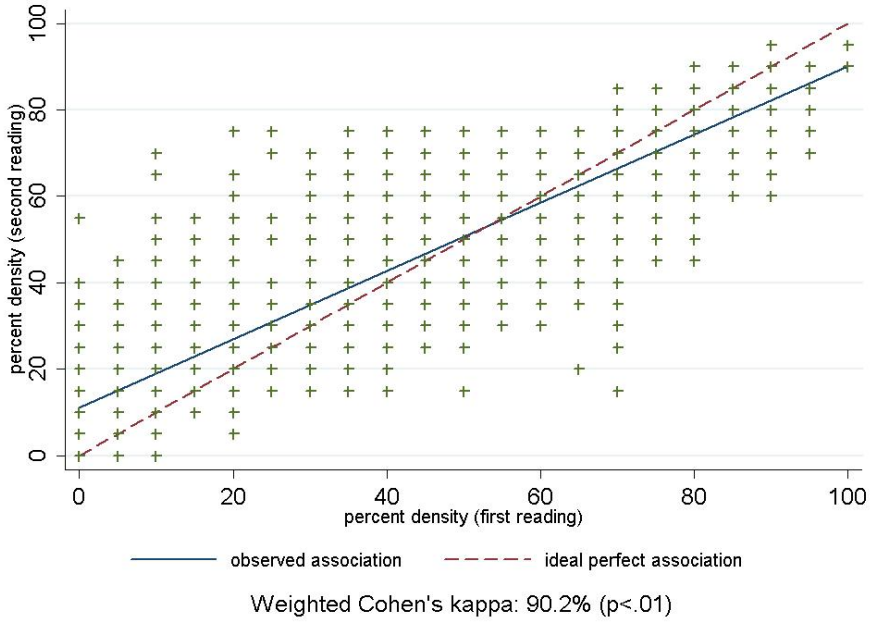
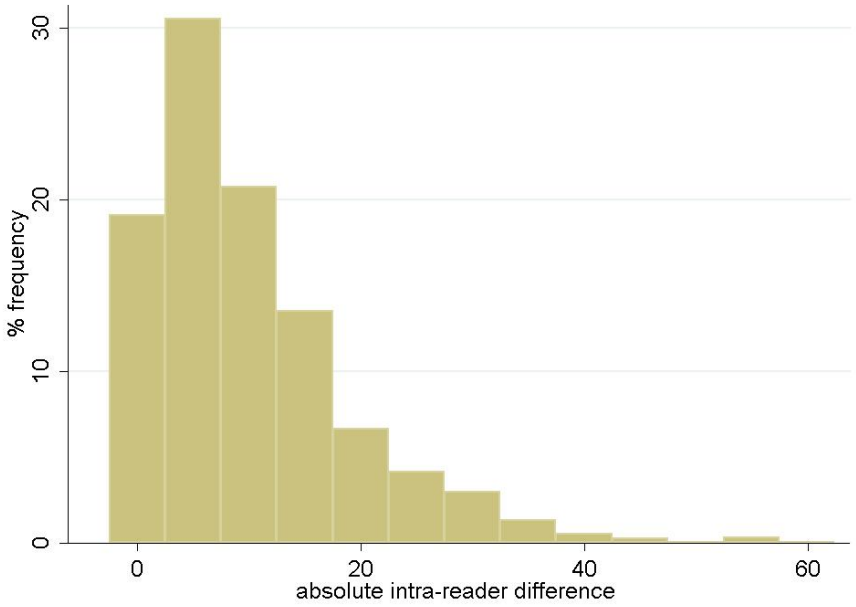


FIGURE 8.4: Absolute differences in visually assessed percent density between first and second reading by reader 0



Intra-reader agreement was assessed on reader 0 repeated assessments (Figure 8.3), and appeared especially strong (weighted Cohen's kappa: 90.2%, p<.01). In particular first and second assessments differed on average by less than 10%, and in 70% of cases the mammograms would be reclassified either in the same category or in 1 or 2 categories apart (Figure 8.4). The intraclass correlation coefficient also suggests the high reliability in intra-reader visual assessment, being 89% (88%, 90%).

8.2.3 Discussion

There are differences in Cumulus estimates according to reader, however the overall reproducibility of this method was high for both absolute and percent estimates. Likewise visual assessments appeared reliable, with good agreement both inter- and intra-reader. Although we had good agreement on average, the fact that there were a few mammograms for which the readings differed of up to 60 cm^2 or 60% in Cumulus absolute and percent density respectively, and even 80% or more in visual assessments, suggests that these methods are still subjective and have room for improvement in reproducibility, if they are to be used for individual risk prediction and management.

8.3 Adjusting for measurement error in estimating odds ratios associated with volumetric density measures from Quantra

Fully-automated volumetric methods are anticipated to resolve issues regarding reproducibility of measurements and to provide a more precise estimate of the amount of fibroglandular tissue; for this reason several have been or are being developed [36–42, 236]. However the results so far suggest that volumetric methods provide similar or weaker risk prediction than the area-based measures [201]. Recent evidence suggests that this problem in volumetric breast density estimation may be due to inaccuracy in assessing the thickness of the breast, as an effect of compression when performing a mammogram [264, 265]. A particular challenge, "paddle tilt", which occurs when compression is applied with a flexible compression paddle, allowing the upper plate to tilt, results in variation in breast thickness from the chest wall to the outer side of the breast [265].

Our case-control study of the fully automated Quantra volumetric density measures [Chapter 5] found only moderate association with risk. Since fully automated density analysis is still very much in development, it is reasonable to hypothesise that measurement error is at least partially responsible for the weakness of Quantra's risk predictions. In this section we attempt to correct the risk estimates for random measurement error using established statistical methods in our case-control study [Chapter 5]. More specifically we apply a modification of Rosner's and colleagues's method [257–259] to the case-control data, estimating the necessary components of the variance from our longitudinal dataset of repeated Quantra measurements on a similar population [Chapter 6]. In addition, since for Quantra we observed a weaker association with risk than was shown in previous studies using Cumulus, we estimated how substantial the measurement error in Quantra estimates would need to be in order to have a strength in the association with risk similar to the that of Cumulus in Stone *et al.*'s study [266].

8.3.1 Materials and Methods

8.3.1.1 Data Collection

We used the two datasets [Chapters 5 and 6] collected from women attending breast cancer screening at Princess Grace Hospital. All the subjects underwent digital mammography and their breast density was assessed using Quantra software, as described in [Chapter 5], providing estimates in cubic centimetres for total breast volume and dense breast tissue volume (*absolute density*). Percent density (%) and non-dense volume (cm^3) were computed by division and subtraction respectively.
The first dataset is described in Chapter 5 and was used to evaluate the association between density estimates and risk. The second dataset had a longitudinal structure with repeated measurement [Chapter 6] and thus allowed us to estimate the variance components. In the case-control study [Chapter 5] details were recorded of 200 cases, female patients with histopathologically verified breast cancer, and 200 matched controls. The longitudinal dataset [Chapter 6] comprised information on 332 women, (231 younger than 50 years and 101 aged 50 or more years) undergoing two mammographic examinations, typically 12-24 months apart. See Chapter 6 for further details.

8.3.1.2 Statistical Analyses

We used the method of Rosner and colleagues, based on the method of moments, to correct for measurement error [257–259].
Suppose the model relating our true amount of fibroglandular tissue (*mammographic density*) X, the probability (risk) of developing breast cancer B and the probability of not developing breast cancer $\bar{B}$ is of logistic form, whereby

$$ln\left(\frac{Pr(B|X)}{Pr(\bar{B}|X)}\right) = \alpha + \beta X \tag{8.1}$$

where $ln(\cdot)$ is the natural logarithmic function, α and β are unknown parameters. From now on we refer to $ln\left(\frac{Pr(B|X)}{Pr(\bar{B}|X)}\right)$ simply as Y.
Furthermore, suppose that a linear relationship is assumed to exist between true density X and observed, error-prone, density Z that is of the form

$$Z = X + \epsilon \tag{8.2}$$

where ϵ is the random error and has a standardised Normal distribution ($N(0, \sigma^2)$). Note that in equation (8.2) there is no fixed intercept or coefficient for X because that would indicate the presence of a systematic error whereas in this case we suppose that the measurement error is random (unbiased) and distributed as normal with mean 0.
In the empirical application, the assumption of random error is an approximation, but it is a reasonable one. While it is known that density reduces with age, in the longitudinal dataset, the average values of the two consecutive density measures did not differ significantly. It is thus reasonable to assume that the differences between the two consecutive measures on the same subjects are mainly due to random variation and that the time between masurement is not sufficient for changes in density with age to contribute materially to the differences. We refer to this as measurement error, although it may come from two sources: (1) random error of determination of density by the measurement tool (including position of the subject); and (2) fluctuation of the breast composition over time, around a central "average" or typical composition.
Since the true density X is mistakenly observed as Z, the relationship between density and breast cancer risk which we estimate is not equation (8.1), but as follows

$$Y = \alpha + \tilde{\beta} Z. \tag{8.3}$$

Hence we focus on the relationship between β and $\tilde{\beta}$.
In a linear regression equation (8.1) with true density X, β is commonly estimated as

$$\beta = \frac{\sigma_{xy}}{\sigma_x^2} \tag{8.4}$$

where σ_{xy} is the covariance between X and Y, and σ_x^2 the variance of X. However, in our observed data (8.3) Z is replacing X, therefore

$$\tilde{\beta} = \frac{\sigma_{zy}}{\sigma_z^2}. \tag{8.5}$$

where σ_{zy} is the covariance between Z and Y and σ_z^2 the variance of Z. Since the measurement error here is assumed independent from Y and X we can conclude that equation (8.5) resolves as follows

$$\tilde{\beta} = \frac{\sigma_{xy}}{\sigma_x^2 + \sigma_\epsilon^2}. \tag{8.6}$$

Consequently, a corrected estimate β^* is

$$\beta^* = \frac{\sigma_x^2 + \sigma_\epsilon^2}{\sigma_x^2}\tilde{\beta} = \frac{\tilde{\beta}}{\lambda} \qquad \text{where } \lambda = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\epsilon^2}. \tag{8.7}$$

In a longitudinal dataset, where no significant change in density is observed over time, we could estimate σ_ϵ^2 as the variance of measurements within the same subject (*variance WITHIN*), whereas σ_x^2 represents the variance between the subjects (*variance BETWEEN*). Therefore λ represents an estimate of the reliability of Z and is also called the intraclass correlation coefficient (ICC).
Since we have a logistic link, the exact correction would not be in closed form. However, Rosner *et al.* [257] show that the above formula (8.7) is a second order Taylor approximation to the true value β^*.
The variance of the new adjusted β^* is then computed and the confidence interval varies accordingly, following the next formulas [257]:

$$\text{var}(\beta^*) = \text{var}(\frac{\tilde{\beta}}{\lambda}) = \frac{1}{\lambda^2}\text{var}(\tilde{\beta}) + \frac{\tilde{\beta}^2}{\lambda^4}\text{var}(\lambda) \tag{8.8}$$

$$\text{CI}_{95\%}(\beta^*) = [\beta^* \mp 1.96\text{SE}(\beta^*)] = \left[\frac{\tilde{\beta}}{\lambda} \mp 1.96\text{SE}(\frac{\tilde{\beta}}{\lambda})\right]. \tag{8.9}$$

The confidence limits for the associated odds ratio $\exp(\beta^*)$ are given by

$$\text{CI}_{95\%}(OR) = exp[\beta^* \mp 1.96\text{SE}(\beta^*)]. \tag{8.10}$$

After verifying and justifying the comparability of the population in the two datasets, i.e. that it is reasonable to use the estimates of ICC from the longitudinal study to correct estimates in the case-control study, we applied this method empirically. The case-control dataset was used to estimate the association of the observed density with risk ($\tilde{\beta}$), and then the longitudinal study provided the data necessary to compute λ, that is the ICC. The variance of $\tilde{\beta}$ was calculated by logistic regression, and the variance of λ was estimated using Smith's method [267].
Menopause affects the characteristics of mammographic appearance [21, 89, 90], and therefore the variance of density. For this reason the different age composition in the two datasets may lead to different variances. Thus analyses were run separating the subjects into those aged less than 50 and those aged 50 or more, considering baseline age in the longitudinal study.

The focus of the analyses here is on the absolute density, because not only was its association with breast cancer risk in our case-control study stronger, it is also less affected by confounding with BMI [33, 226] [Chapter 5].
For the second part of our investigation, we compared the odds ratios estimating breast cancer risk associated with increasing quintiles from our case-control data, with those of Stone and colleagues' [266] which used the Cumulus interactive thresholding method to assess breast dense area. To promote a more direct comparison between the two assessing techniques, we considered whether to transform the volumetric measures into areas, by the approximating assumption that volumes are spherical, using the following formulas:

$$V = \frac{4}{3}\pi r^3 \Rightarrow r = \left(\frac{3V}{4\pi}\right)^{\frac{1}{3}} \quad \text{Therefore:} \quad A = \pi\left(\frac{3V}{4\pi}\right)^{\frac{2}{3}}$$

where V is volume, A is area and r stands for radius. However, Stone et al. presented the ORs per quintile and, since the transformation of volumes into areas is monotonic, it would not affect the ranking of the measures and, consequently, the ORs per quintile. In view of this, we retained the volumetric estimates.
We also used equation (8.7), to obtain the λ, or ICC, which would be required to give a corrected association of the same strength as that of Stone and colleagues. In this case there was no age-stratification, since only overall results were available [266].
All analyses were performed with STATA software version 12.1.

8.3.2 Results

Previous Chapters 5 and 6 reported the analysis of these two datasets from the Princess Grace Hospital in more detail. The longitudinal data [Chapter 6] was used to investigate both the change in breast composition by time and age, and the relationship between breast components, i.e. density and non-dense volume, and age. In the case-control study [Chapter 5], we examined the relationship between risk of breast cancer and mammographic density measures. According to the results, absolute density had a stronger association with breast cancer risk than percent density and the final OR per 10 extra cm^3 of dense tissue was 1.04 (95% CI(1.00 , 1.08)).

Descriptive statistics for absolute density in the two studies are summarised in Table 8.5. The means and standard deviations obtained among controls and in the longitudinal study were similar in the younger group (Wilcoxon rank-sum test: p=.85). Women

TABLE 8.5: Characteristics of the study samples in the two datasets

	Case-control study				Longitudinal study	
	cases		controls		*at baseline*	
	N=200		N=200		N=332	
	mean	(SD)	mean	(SD)	mean	(SD)
Under 50						
Absolute density (cm^2)	111.38	(59.76)	104.17	(71.49)	102.91	(72.11)
50 or more						
Absolute density (cm^2)	99.18	(60.54)	82.33	(49.11)	96.48	(52.02)
Overall						
Absolute density (cm^2)	103.88	(60.39)	90.74	(59.53)	100.96	(66.63)

in the older groups of the two studies had significantly different means (p=.02), possibly due to a different age composition (mean age: 62 years in the case-control study and 58 in the longitudinal study). Consequently we performed a linear regression of absolute density on a categorical variable identifying the two studies, in order to explain absolute density differences due to the two studies after age-adjustment. Results would suggest that subjects in the longitudinal study would have higher absolute density (linear regression coefficient β=+ 12.6 cm^3), yet not significant (p=.07). In any case, we require the two studies to be comparable mainly in terms of variability rather than absolute average, as the longitudinal study is used to estimate the variance components. Standard deviations were similar in both groups (controls from the case-control study and subjects in the longitudinal study), suggesting that the two samples can be used together. This is as we would expect since the subjects were both selected from the same population, the Princess Grace Hospital breast surveillance service.

Step 1: Estimate β. Coefficients for association between breast cancer risk and the observed absolute density were obtained from the case-control study (Table 8.6) using conditional logistic regression.

TABLE 8.6: Coefficients from the conditional logistic regression models using Quantra absolute density

Coefficient	SE	[95% Confidence	Interval]	P> |z|
Under 50, N=154				
.0017	.0025	-.0032	.0066	.493
50 or more, N=246				
.0056	.0024	.0008	.0104	.022

Step 2: Compute the intraclass correlation coefficient. Table 8.7 displays the ICCs obtained in the longitudinal study. These values are the estimates of λ in the two age groups.

TABLE 8.7: Intraclass correlation coefficients, for absolute density, estimated in the longitudinal study

ICC	SE	[95% Confidence	Interval]
Under 50, N=231			
.887	.014	.859	.914
50 or more, N=101			
.930	.013	.904	.956

Step 3: Compute the adjusted β^*. Following equations (8.7), (8.8) and (8.9) we obtained the adjusted coefficient for absolute density and relative variance and confidence intervals.

For women aged under 50:

$$\beta^* = \frac{\tilde{\beta}}{\lambda} = \frac{.0017}{.887} = .0019$$

$$\text{var}(\beta^*) = \text{var}(\frac{\tilde{\beta}}{\lambda}) = \frac{1}{.887^2}\left(.0025 * \sqrt{154}\right)^2 + \frac{.0017^2}{.887^4}\left(.0141 * \sqrt{231}\right)^2 = .0012$$

$$CI_{95\%}(\beta^*) = [\beta^* \mp 1.96 SE(\beta^*)] = \left[.0019 \mp \sqrt{\frac{.0012}{154}}\right] = [-.0036, .0075]$$

Similarly for the subjects *50 or more*:
$\beta^* = .0060$, $\text{var}(\beta^*) = .0017$ $\text{CI}_{95\%}(\beta^*) = [.0009, .0112]$.

Step 4: Obtain the ORs and 95% CIs. Finally, from equation (8.10), among younger women (Under 50) an extra 10 cm^3 of fibroglandular tissue would lead to an OR of 1.020 (.965, 1.078), whereas in the 50 or more group it would be 1.062 (1.009, 1.118), Figure 8.5. The original ORs per 10 cm^3 were 1.017 (.968, 1.069) in the group aged under 50 and 1.058 (1.008, 1.110) for the women 50 or more.

FIGURE 8.5: Comparison of the ORs before and after adjustment

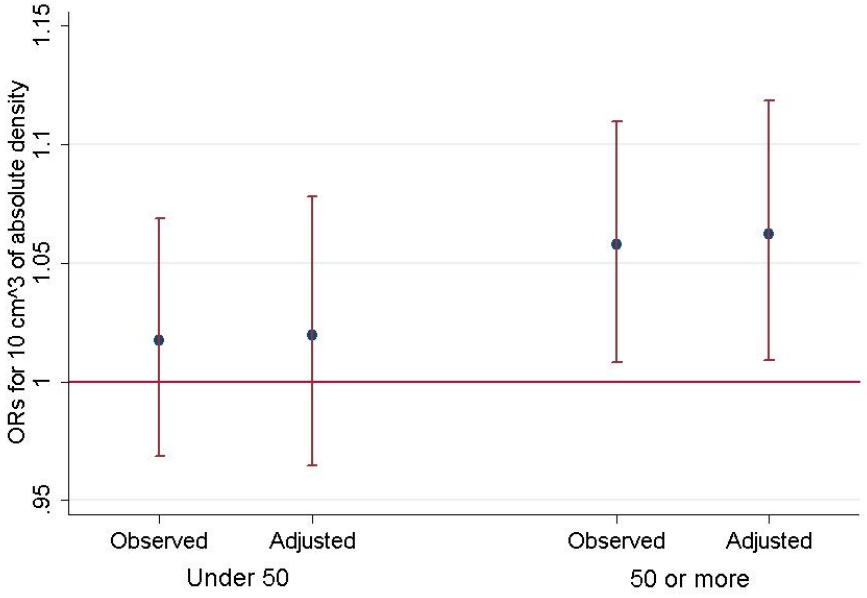


Table 8.8 compares the association of Quantra density estimates with breast cancer risk [Chapter 5] with that of dense area assessed with interactive thresholding by Stone *et al.* [266].

From equation (8.7), we obtained that an ICC in Quantra density estimates of .83 (.30, 1) would be needed to give an adjusted association with breast cancer risk similar to Stone *et al.*'s findings, whereas in our data the ICC was .89 (.87, .92).

$$ICC = \lambda = \frac{\tilde{\beta}}{\beta^*} = \frac{log(1.18)}{log(1.22)} = \frac{.1653}{.1989} = .8315$$

That is, the correction for measurement error, as estimated from repeat Quantra measures, is not sufficient to give results compatible with Stone's area based estimates. In other words we would need a more substantial measurement error in our data for the

TABLE 8.8: ORs estimating breast cancer risk associated with increasing quintiles of mammographic density

Measuring method	Odds Ratio	[95% Conf.	Interval]	P>z
Quantra case-control data [Chapter 5], *unadjusted*				
dense volume	1.18	1.03	1.36	.020
Stone's and colleagues' data (interactive thresholding) [266]				
dense area	1.22	1.11	1.33	<.001

adjusted estimates to replicate Stone and colleagues's findings. However, the difference in coefficients per quintile is modest, and it might be due to variation in the population, rather than a truly poorer risk association for Quantra.

8.3.3 Discussion

Estimates of association between risk and Quantra absolute density did not change significantly after correction per measurement error. ORs per 10 cm^3 of dense tissue went from 1.017 to 1.020 in the under 50 group and from 1.058 to 1.062 in those aged 50 or more. There is an important qualification to interpretation of these results. The relatively high ICC's, 89% and 93% in subjects aged respectively under 50 and 50 or more, suggests that there is strong repeatability and reliability, and, therefore, low measurement error. However, it may not be error around the quantity at which it is aimed, that is to say, although we get very similar measurements when we repeat the method on the same individual, it may not be measuring density, or the aspect of density which is relevant to breast cancer risk. Thus both before and after correction for measurement error, we still have a relatively weak relationship of the measure with breast cancer risk. There are very strong correlations between density for left and right breast, and between density measures for the same breast from two different views, and as we have seen here, between successive measures taken a year or so apart. All this is evidence that Quantra is measuring some aspect of breast appearance with considerable precision, but it might not be capturing the aspect which most strongly relates to breast cancer risk. In this case, help from statistical methods is limited and it is the density estimation algorithm that needs further correction. Studies [48, 50, 248] comparing Quantra volumetric assessments to MRI breast density [47] observed a low correlation between measures of absolute density [50, 248] and also suggested that Quantra significantly overestimates

density in low dense breasts, in comparison to MRI [48]. These findings give us a better understanding of the accuracy of Quantra density estimates, nevertheless should be validated in a larger datasets.

The comparison with Stone's and colleagues' findings also highlighted that measurement error is lower than one would expect in order to observe an association between density and breast cancer risk as strong as recorded in literature [6]. This should further encourage improvement to the algorithm for breast density assessment. That said, the risk of breast cancer assessed by quintile is not radically different between Stone *et al.*'s area based estimate and our Quantra volume estimate.

From this, doubts arise again as to whether a single radiological view of the breast could be sufficient to construct a volumetric measure of density, since it seems to be a very ambitious aim. The use of breast MRI and the development of breast tomosynthesis and automated tomosonography [268–270] could lead to opportunities for volumetric density estimation which uses more complete three-dimensional information in construction of the density measure.

There is another possible explanation for the slightly poorer risk prediction than area-based measures even after correction for measurement error. The correction method assumes random, non-differential measurement error acting in an additive manner. If the measurement error is systematically related to the true value (for example, overestimated at low values, underestimated at high), our correction would be inadequate [271]. However, the relatively low variation within persons and between occasions suggests that it is not so much measurement error that is the problem, rather it is the phenomenon being measured.

Currently, several other novel fully-automated volumetric approaches [36–42, 236] have been or are undergoing validation. When evaluating them it is important to consider the potential measurement error and take action to adjust for it. In particular in regression analysis, regression coefficients of error-prone variables are usually attenuated, and treatment effects are potentially estimated with bias in either direction when we include error-prone covariates [272]. We used the method of moment correction [257], but other methods are available, for instance regression calibration [273] that estimates the unknown true value X by developing and fitting a calibration model for on the observed covariates. It was not possible to apply regression calibration to our data because it requires an external dataset comprising details on the true, or gold standard, measures of dense volume, and not only there was no such dataset available for this analysis, but also to date there is no agreement on what is the most reliable measure of fibroglandular volume.

8.4 Relationship between two- and three-dimensional density measures

The high intraclass correlations and the surprisingly low within-screenee variation in Quantra measures observed in the previous section, suggest that classical random measurement error is not responsible for the relatively modest associations with breast cancer risk.

The range of Quantra measures between screenees also seems small, but of course this will depend on the population under study, in particular the age range, as well as on the properties of the measurement method. If there really is considerably lower variability between women's Quantra densities than between their densities as measured by other methods, this may indicate that an aspect of individual breast composition which reflects individual risk of breast cancer is not captured by the Quantra measure.

In our data the Quantra and BIRADS assessments were not strongly related [Chapter 5]. Given the processing of the digital mammography, visual assessment of these mammograms for density is fraught with difficulties of interpretation. We can, however, obtain some indirect evidence, by comparing the average and variation of Quantra densities in our population with visual density assessment from film mammography by individual year of age. The conditioning on individual year of age renders the populations more comparable.

Percent breast density was visually assessed on 10,046 women aged 50 and over, in the CADET1 study of computer aided detection in breast screening. The subjects attended breast screening in Aberdeen and Manchester in the late 1990's [274]. Breast density was assessed by each of the eight radiologists involved in the study using a linear analogue scale. For further details of the density component of the study see Duffy *et al.* [4].

Since the visual assessment is based on area, and Quantra on volume, we need to transform one of these. As in the previous section, we adopt the approximating assumption that volumes are spherical. In particular if the dense portion of the breast were spherical with radius r_1 and the total breast volume were spherical with radius r_2, the percentage area and volume are

$$A = \frac{r_1^2}{r_2^2}$$

and

$$V = \frac{r_1^3}{r_2^3}.$$

From this

$$V = A^{\frac{3}{2}}.$$

TABLE 8.9: Characteristics of percent density in the two studies [Chapter 6] [4], according to age

Age	Visually assessed percent dense area			Transformed percent dense area		Automated percent dense volume		
	Obs.	Mean (SD)	Range	Mean (SD)	Range	Obs.	Mean (SD)	Range
50	736	39 (22)	0-96	27 (21)	0-94	13	25 (14)	10-62
51	845	40 (23)	0-98	28 (22)	0-97	6	28 (8)	17-42
52	706	38 (22)	0-96	26 (21)	0-94	6	19 (6)	12-29
53	581	37 (21)	0-96	25 (20)	0-94	6	25 (7)	15-32
54	577	35 (21)	0-96	24 (19)	0-94	3	15 (2)	13-16
55	516	33 (22)	0-98	22 (20)	0-97	4	19 (6)	12-24
56	508	33 (21)	0-98	22 (19)	0-97	10	16 (6)	11-31
57	541	32 (21)	2-94	21 (20)	0.3-91	7	16 (8)	10-33
58	595	31 (19)	0-96	19 (17)	0-94	3	17 (4)	14-21
59	681	32 (20)	0-90	20 (18)	0-85	8	22 (16)	9-55
60	681	30 (19)	0-90	19 (17)	0-85	8	23 (9)	11-39
61	674	30 (18)	0-96	18 (16)	0-94	4	20 (7)	12-29
62	682	30 (19)	0-92	19 (17)	0-88	4	20 (7)	13-27
63	660	29 (19)	0-92	18 (17)	0-88	3	17 (4)	14-21
64	689	30 (19)	0-96	19 (17)	0-94	2	22 (5)	18-25
Overall	9672	33 (21)	0-98	22 (19)	0-97	86	21 (10)	9-62

The situation is complicated by the fact that neither the breast nor the dense portion of the breast are spherical. However, since one measure is in two dimensions and the other in three, one might expect to achieve a comparison approximately on the same scale by first expressing the visual percent dense area as a proportion (between 0 and 1 rather than 0 and 100), raising it to the power of $\frac{3}{2}$ and converting back to a percentage.
Table 8.9 shows the mean, SD and absolute range of densities in the two populations by individual year of age, for the visual percent area density, the volumetric percent density and the former transformed as noted above. Because the CADET1 mammograms were taken within the UK National Breast Screening Programme, which at the time had a policy of inviting ages 50-64, very few mammograms for individual years of age greater than 64 were available. We therefore show results for ages 50-64 only.

The numbers of subjects in each age group are small in the Quantra dataset. However, certain observations are clear: first, the average transformed percent areas are similar in absolute size to the average Quantra volumes; second, that all three measures decline in mean value with age; and third, that both overall and for each year of age, the transformed area-based measures are considerably more variable between subjects than

the Quantra volumetric estimates. Overall, the coefficients of variation were 0.64 for visual percent density, 0.86 for transformed visual percent density and 0.48 for Quantra volumetric percent density. In the Quantra case-control dataset [Chapter 5], breast density in absolute terms (cm^3) showed a stronger association with breast cancer risk than percent breast density, therefore it could be interesting to explore also the relationship between absolute dense volume and absolute dense area. Unfortunately these data were not available in the CADET1 study. Another observation is that spherical approximation may not be the most appropriate solution to transform dense volume into dense areas.

While not conclusive, these observations are consistent with the visual percent dense area representing a wider variability among women, and this may contain significant individual risk information, in addition to an inevitable quantity of purely random error. Further support for this comes from a study which calculated area and volumetric densities on the same subjects, using a different volumetric measurement tool, the Standard Mammographic Form [275]. In this study in 1,830 women without breast cancer the mean percent dense area using the Cumulus interactive threshold method [256] was 24 with an SD of 18. The average percent volume using the Standard Mammographic Form was 24 again, but with an SD of 7.

8.5 Comparison of density measures in their association with breast cancer risk

In this thesis we had the opportunity of analysing the association of breast density and breast cancer risk using a variety of measures (visual and Cumulus assessments in Chapter 2, Cumulus percent and absolute estimates in Chapter 4, Quantra and BIRADS in Chapter 5 and Volpara in Chapter 7) in a number of studies. In the present section, we investigate their performances in terms of association with breast cancer risk.

In order to promote a more straight-forward interpretation of the results we used standardised ORs. In other words, every measure of density was divided by its standard deviation so that all the density estimates and ORs were expressed on the same scale, i.e. per unit of standard deviation. BIRADS was excluded from this analyses because of its categorical nature. The standardised ORs presented in Table 8.10 are all age-adjusted (either by analysis or matching). In case of percent density in IBIS-I we also adjusted for BMI, as BMI confounding effect may lead to underestimating the association between percent density and breast cancer risk [12, 25]. Absolute density estimates instead do

not require this adjustment, being less affected by changes in BMI [33, 75] [Chapters 4, 5 and 7]. We included Kallemberg *et al.*'s results for an external comparison of our results with a more extensive dataset (N=53,975) [276].
We also compared the ability of the density measures to discriminate cases from controls, using the areas under the ROC curve (AUCs).

TABLE 8.10: Standardised measure ORs for developing breast cancer after age-adjustment

Density measure	OR	[95% Conf.	Interval]	P>z	AUC
IBIS-I case-control study, N=1065 (case=123) [Chapter 2]					
Visual assessment (21-category)	1.32	1.08	1.61	.006	.58
Visual assessment (21-category)*	1.49	1.20	1.86	<.001	.60
Cumulus percent dense area	1.23	1.01	1.51	.043	.57
Cumulus percent dense area*	1.35	1.08	1.69	.009	.58
FH01 case-control study, N=292 (cases=101) [Chapter 4]					
Cumulus absolute dense area	1.24	.99	1.57	.064	.56
Quantra, N=400 (cases=200) [Chapter 5]					
Automated absolute dense volume	1.26	1.02	1.55	.030	.57
Volpara, N=366 (cases=182) [Chapter 7]					
Automated absolute dense volume	1.26	1.01	1.56	.038	.55
Kallenberg *et al.*'s Volpara study, N=53,975 [276]					
Automated absolute dense volume†	1.25	1.17	1.33	*not available*	-

Notes: ORs are expressed per unit of standard deviation;
(*) adjusted for age and BMI;
(†) adjusted for age and total breast volume;

Table 8.10 reports the results of this comparison. Similarities, both in terms of standardised ORs and AUCs, among the different methods are noticeable, especially among the automated volumetric measures and the computer-assisted Cumulus. All the AUCs have low values indicating an overall poor discriminatory ability of the methods here compared. It should be noted though that in general predictive models for breast cancer struggle to get near 1 [167], and also there has been discussion on whether the AUC is the most appropriate statistic to evaluate the predictive ability of a model [172, 173]. Nevertheless in our analyses, visual classification appeared as the best predictor, both in terms of standardised OR and AUC, especially when adjusting also for BMI.
This finding would suggest that there is still something in the parenchymal patterns that is missed by computerised techniques, but that appears more obvious to a trained

eye. It could be speculated that the risk factor is not just pure dense tissue but it depends also on the texture or other aspects of pattern on the mammogram. Nevertheless there is no doubt that, for a better use of information on density in a clinical setting, an automated measure would be preferable and more practical. For this reason several alternative automated methods have been proposed. Nielsen *et al.* developed a software that extracts textural information from all pixels of segmented breast images, yielding a mammographic texture resemblance marker which was strongly related to breast cancer risk [56]. Their findings showed that this measure of mammographic texture was independent of mammographic density and that it may provide additional information on breast cancer risk. Likewise Schmidt and colleagues designed CIRRUS, a fully-automated predictor of breast cancer risk, that tries to identify the features in a mammographic image that best predict breast cancer risk [57]. Heine *et al.* had a simpler but still effective approach [58]. They worked on an algorithm for estimating the "variation measure", i.e. the standard deviation of calibrated pixel values, that appeared to have an association with breast cancer risk at least as strong as percent density assessed with Cumulus [58]. That is, the more variable the pixels values, the higher the risk.

8.6 Conclusions

In this chapter we have discussed inter- and intra-reader agreement in visual and computer-assisted density assessments. We evaluated the measurement error in a fully-automated volumetric density assessment software tool and attempted to address this issue mathematically with Rosner's method [257–259]. Then, we studied the relationship between two- and three-dimensional density estimates. Finally, we compared different methods for assessing density according to their association with breast cancer risk, on an equivalent scale, and their accuracy in discriminating cases from controls.

In conclusion we observed ICCs in excess of 80% in both visual and Cumulus density estimates when performed by trained readers, which gives some confidence in the reproducibility of results in studies based on film mammograms. Secondly, according to our results, Quantra estimates have a limited measurement error, and they suggest that it might be productive to revise the algorithm, as it may not be measuring the tissue component most relevant to breast cancer risk. Volumetric estimates appeared to have a much lower variability than two-dimensional estimates, and this is expected to affect the discriminatory power of density in risk prediction contexts. Finally, in our comparison,

visual density estimates resulted the method more strongly associated with breast cancer risk, suggesting that there are some aspects of mammographic appearance relevant for breast cancer development, that are missed by computerised methods.

Further, all these assessments are based on mammography which is designed to detect tumours, not to measure density. For instance, during mammographic examinations we compress the breast and consequently the dense tissue is also compressed, possibly making its measurement more difficult. Hence, it may be that a tool specifically designed to measure the amount of fibroglandular tissue in the breast would improve noticeably both the density and the risk-prediction estimates. Also alternatives to mammography examinations, such as MRI, automated ultrasound, electrical impedance or tomosynthesis, should also be considered for deriving density indicators [43–50, 54, 59–61, 248]. Although these examinations are often more time and resource consuming, they could provide more accurate 3D estimates of breast composition. It may be that dual MRI measures of fibroglandular volume and background parenchymal enhancement have complementary relationships to breast cancer risk. Both are positively associated with risk [49]. The effect of background parenchymal enhancement on risk was observed to remain significant after adjustment for fibroglandular tissue volume but not vice versa in a study with 39 breast cancer cases and 116 controls [54]. However, the ORs for fibroglandular tissue volume remained substantially higher than unity, and it may be that larger numbers of cases would show independent effects of both measures. If this were the case, MRI-based risk determination from breast composition might well be a major improvement upon mammographic measures.

As a matter of fact, methods for density estimation from these new 3D breast imaging systems, in particular tomosynthesis, are currently being developed. The research on this topic has just started moving its first steps and currently there is scarce published evidence of their association to risk, which is the ultimate validation of their usefulness in clinical practice.

An important implication of Table 8.10 is that the automated and semi-automated methods all seem to have similar propensity for risk prediction. Thus, if they were to be used as part of individual risk prediction management or surveillance planning, it would be reasonable to use any of the measures. For example, the fully automated methods are almost as strongly associated with risk as the more labour-intensive semi-automated methods. It is desirable, however, to identify the additional risk information picked up by expert visual assessment and incorporate this into the automated methods.

Conclusions

Chapter 9

Conclusions

9.1 Introduction

Mammographic density has been known as a risk factor for breast cancer since 1976, when Wolfe first suggested an association between parenchymal patterns and risk of developing a tumour in the breast [5]. In the past decades mammographic density has been the object of many studies that have enriched our knowledge on the subject, although the causal pathway linking enhanced fibroglandular tissue and risk of breast cancer has not yet been fully explained. As summarised in the literature review in Chapter 1, mammographic density is a strong independent risk factor for breast density in terms of both relative and attributable risk and appears to have potential for improving not only screening and risk management programs, but also selection for chemoprevention trials and other disease-control policies. Despite this, its use is still currently limited mainly because of a lack of a practical standardised method of measuring density.

Hence, this research project focused on providing further insights of mammographic density as a biomarker of breast cancer risk, its potential in risk-prevention and disease-control trials, and its association with tumour sub-types and other risk factors. Additionally, the project investigated the two main automated volumetric methods for assessing density on the market (Quantra and Volpara).

9.2 Part I: Mammographic density as a biomarker for breast cancer risk

In Chapter 2 we investigated whether mammographic density reduction is a valid surrogate marker, including whether it fulfilled Prentice's criteria [185], in the IBIS-I risk-prevention trial, using a nested case-control dataset. The results (PIG=89%) suggested that a reduction in density of at least 10% could be an effective indicator of the success of tamoxifen chemoprevention.

Chapter 3 presents the results of analyses on 601 post-menopausal women enrolled in the ATAC study. On these data we observed that higher levels of mammographic density were not associated with enhanced risk of recurrence. Similarly, and in contrast to the IBIS results, change in the proportion of dense tissue, over 1, 2 or 5 years, did not predict the outcome of the treatment. We also investigated the relationship between density and tumour characteristics, and found that the associations were mostly not significant. However results suggested that subjects with poorly differentiated primary invasive breast cancer experienced smaller density reductions in the years following diagnosis. Furthermore, higher baseline levels of cell proliferation, indicated by levels of Ki-67, were associated with lesser significant changes in breast composition.

Chapter 4 focused on a different population, that of a case-control study of 298 women aged between 40 and 49 enrolled in the FH01 study because of their enhanced familial risk of breast cancer. In this dataset absolute density (cm^2) was more effective than percent density in discriminating cases and controls. In addition, absolute dense area was not associated with most of the risk factors, such as age at menarche or parity, that percent dense area was, making it a simpler risk-predictor and less affected by confounding factors. Moreover, the association between breast cancer risk and absolute density was not altered after adjusting for Tyrer-Cuzick 10-year risk estimates. This result supports the belief that mammographic density could add information to the existing risk-prediction models for breast cancer even in populations already known to be at increased risk and should be useful in the design of more tailored screening programs.

The studies analysed in this part of the thesis differ significantly from one another because of design, purpose and population. Mammographic density at baseline was a strong predictor for risk of developing breast cancer in both IBIS-I and FH01 case-control

studies. Eligible subjects for both studies were at enhanced familial risk, and FH01 focused on women aged 40-49. Our findings suggest that breast density has the potential to provide an additional discriminatory power in the standard risk-prediction models for breast cancer, and thus support the belief that density should be introduced in such models to improve the identification of subjects at enhanced risk, who could benefit from preventive interventions or extra screening examinations (e.g. magnetic resonance imaging or MRI, digital breast tomosynthesis and ultrasound). A possible future tactic might be to use a woman's first mammogram in the programme (at age 47-52 in the UK programme) to provide risk information, contribute to individual surveillance planning, and possibly inform primary prevention intervention.

Our findings from the IBIS-I study indicate that mammographic density reduction is a good predictor of the efficacy of tamoxifen therapy. In practical terms, mammographic examination, including density assessment, should be carried out at baseline, then repeated regularly from a year after beginning the treatment, so that changes in density can be assessed and incorporated into the decision making process for continuing this therapy or switching to a different treatment. For instance, a woman might switch to an aromatase inhibitor in order to achieve a better outcome, if density indicated non-response to tamoxifen.

There was no association between density and risk of recurrence in the ATAC data. This could be evidence of the fact that breast density may not be an important factor once the neoplastic process has started [63, 82]. In addition to this, most of the subjects involved in the study underwent treatments such as chemotherapy or radiotherapy which are known to influence breast composition, not only in terms of appearance but also of susceptibility to recurrence. Moreover these treatments may modify response of the breast radiological appearance to hormonal treatment, since mammographic density reduction was not an indicator of the success or failure of the tamoxifen therapy, unlike in the IBIS-I and other studies [17–19].

These possibilities for using density in surveillance, prevention and disease-control programs are contingent on an accepted and practicable, in terms of resource use, method for assessing density. Most popular techniques currently in use present limitations and are often too resource-intensive for routine population-based programmes. For this reason automated methods of density measurement were evaluated in the second part of this thesis.

9.3 Part II: Automated volumetric assessment of mammographic density

Chapter 5 describes the findings from a study on 400 age-matched cases and controls, evaluating the risk-predictive potential of density assessed with Quantra, a novel automated volumetric method, and comparing it with BIRADS, a well-established method based on visual classification. The main result is that absolute density (cm^3) was the most promising predictor for breast cancer, performing better than percent density. Surprisingly, there was no overall association between Quantra measures and BIRADS. Among cases, unexpectedly, BIRADS category increased with increase in non-dense and total breast volume, and as percent density correspondingly declined. Since in this study the association between BIRADS and breast cancer risk was weaker than in previous literature [6], we speculated that visual classification might be affected by the use of processed digital mammograms, that have already been altered to lesser the effect of density on diagnostic accuracy.

In Chapter 6 the analyses involved 332 healthy women, aged 34-81, who underwent two mammographic examinations, 1 to 5 years apart. On both occasions density was assessed with Quantra, and we investigated the change in breast composition between screens and according to age at baseline. The main findings suggested that the effect of time and age was especially evident on non-dense volume (cm^3), which increased steadily, whereas no substantial change was seen in percent and absolute density. These results within subject may be partly due to the limited time interval between screens, but they support the hypothesis that for risk assessment, absolute density should be preferred to percent density because it appears more stable. The lack of substantial change in absolute density across ages (cross-sectional analysis) is in contrast to previous literature [88] and may indicate an area where the Quantra algorithm could be improved.

Most of the mammograms from the Quantra case-control study were reassessed using Volpara, another novel fully-automated volumetric measure. These data were then analysed focusing on the relationship between Volpara density measures and breast cancer risk, on their association with BIRADS classification and on a comparison with Quantra measures. Results are reported in Chapter 7. In brief, findings from this dataset were very similar to those from the previous Quantra case-control study. Volpara absolute density was a strong breast cancer risk predictor, that should be preferred to percent density, and there was no strong relationship between Volpara and BIRADS measures.

In addition to this, the association between breast cancer risk and Volpara absolute density was similar in strength to that observed for Quantra absolute density, as the analyses with standardised odds ratios highlighted. In general the two measures appeared strongly related.

In both Quantra datasets and in the one including Volpara assessments we also considered the relationship between breast composition and area of residence to verify the hypothesis of an "urban effect". Results showed no significant difference, between women living in London or in other UK regions, in terms of amount of fibroglandular tissue (cm^3), but London residents had consistently lower amounts of fatty tissue (non-dense breast volume). This suggests a lower body mass index in London residents when compared to subjects from another UK region, confirming previous findings [249].

As it was highlighted previously, it is crucial for introducing density assessment for risk-prediction in routine population-based screening that we have a standardised, reliable and precise measuring method, feasible to be performed in such a context, i.e. one that is not onerous in terms of time, labour and costs. For this reason, Chapter 8 addresses some of the measurement and risk prediction issues raised regarding both visual assessments and volumetric measures.
In this chapter we estimated inter- and intra-reader agreement in visual and computer-assisted density assessments, observing over 80% reliability in both visual and Cumulus density estimates, when performed by trained readers, that is reassuring about the reproducibility of data in studies based only on film mammograms.
We, then, evaluated measurement error in Quantra fully-automated volumetric density assessments. Our findings indicated that Quantra estimates have limited measurement error, although it is not clear what they are measuring; thus our results might suggest further development of the algorithm. Then, we studied the relationship between two- and three-dimensional density estimates. Results showed that volumetric estimates appeared to have a much lower variability than bi-dimensional estimates, and this may affect the discriminatory power of density in risk prediction contexts. Finally, in our comparison of different density measuring methods, visual assessment appeared as the strongest predictor of breast cancer risk. This suggests that there are aspects of the mammographic appearance relevant to breast cancer risk that are missed by the computerised methods but not by a trained eye. For this reason, several alternative automated methods [56–58], that have recently been proposed, consider features of mammographic appearance other than density, for instance textural information or variation in pixel

values.

The comparison of risk prediction propensities was based on standardised odds ratios, that computed all the methods "fairly", on the same scale, and AUCs. One notable finding was that the association of the volumetric density measures with risk was only slightly weaker than that of the more resource-intensive interactive thresholding method, Cumulus. Since the latter is considered the gold standard by many researchers, the implication of this is that for only a small loss in predictive potential, a significant gain in resources could be achieved by use of the fully automated volumetric methods. With the growing prevalence of digital mammography, it would appear reasonable, if density is to be measured routinely in the screening programme, to use one of the fully automated methods.

That said, the notably better prediction from visual assessment suggests that there is room for improvement in automated assessment. Also a phenomenon we observed in both volumetric automated methods, Quantra and Volpara, is the limited variability in their density estimates [Chapter 8]. Further analyses on the correlation between non-dense and total breast volume revealed that between these two measures exists an almost deterministic relationship (Pearson's correlation coefficient: .99 for Quantra and 1.00 for Volpara). This result adds further incentive for continuing evolution of the automated algorithms.

9.4 Implications and policy for future research

In the 1980s, the American Cancer Society recommended a baseline mammogram for women at average risk who were aged 35-40 years to provide a comparison image that would be available when regular screening began at age 40 or older. This recommendation was dropped in 1992 after a consensus meeting reviewed the evidence on screening recommendations and agreed that there was little evidence to support a benefit of the baseline screening before age 40 years [277]. Since the harms of such a strategy might exceed the benefits, instead we suggest to exploit the potential for density assessed at the first available mammogram (around age 50 in the UK). Over the past two decades research has led to more detailed and accurate information on predictors of breast cancer risk, and made them available more easily and at a lower cost. Hence now we should probably reconsider the idea of a baseline risk-assessment for breast cancer, based on

known risk factors, such as age at menarche, parity and family history, as well as mammographic density and possibly SNPs, when a women is invited for her first screen examination at around age 50. This could help design optimal tailored screening plans, varying starting age, interval and possibly adopting new methods of examination, such as MRI and tomosynthesis [150], as supplementary or alternative investigations. This could also contribute to decisions in relation to primary prevention. A reasonable approach might be to pilot such an early risk assessment including mammographic density, in an experimental design similar to the introduction of the breast screening age extension in the UK.
The results from the standardised odds ratio analysis suggest that it would be safe to use one of the fully automated methods in the first instance. However, evaluation of this should be built into such a pilot, and in parallel research should continue on further development of fully automated methods.

A risk assessment, based on density, would also facilitate decision making when considering primary prevention treatments, such as tamoxifen. In this case a mammographic density assessment, a year into the treatment, is recommended to compare with baseline and verify whether the subject is responding positively to the therapy or would benefit more from a different prevention regimen, e.g. anastrozole. On the basis of results here and elsewhere, mammographic density reduction does not seem to be useful as a surrogate marker in case of treatment of existing breast cancer with aromatase inhibitors, however it has been suggested that a measure of the amount of fibroglandular tissue obtained from breast MRI may be more sensitive to treatment-induced changes in this case [200]. Although MRI is unlikely to be used in general population surveillance in the near future, there is no reason why its use should not be expanded in follow-up of cases of breast cancer. Whether density might be a marker of primary prevention effectiveness with aromatase inhibitors should be investigated in the IBIS-2 study in the future. Findings from the ATAC trial [278] revealed that another useful biomarker suggesting a positive response to anastrozole therapy might be the appearance of new vasomotor or joint symptoms within the first three months of treatment. Further studies are needed to confirm or dismiss these hypotheses.

In any case, these tailored strategies for breast screening and disease-control can be rendered more attractive by improving the methods of measuring density. Quantra and Volpara absolute dense volume estimates appear promising for predicting risk, without being significantly affected by measurement error; nevertheless both algorithms would benefit from further improvement. Automated approaches considering textural features

could also represent a valid alternative. Further studies with richer and larger datasets are needed.
A fully validated objective means of assessing density and an agreed and clear policy regarding how to use this detail in clinical practice should be the primary goals of research in this field. Without this knowledge informing of their density women undergoing screening could cause more harm than good, thus the introduction of Breast Density Inform Law, which is currently spreading across the US, may be premature. Firstly, a standardised automated density assessment tool is essential in order to ensure that all women access the same information. Secondly, an accurate risk predicting system, incorporating density, is required to be able to discuss effectively with the subject about their risk and evaluate their options, avoiding unnecessary anxiety, especially until there is an established health policy strategy comprising additional examinations or preventive treatments.

Although other technologies such as MRI clearly have great potential in breast composition determination, resources arguably prohibit their use in mass screening, although this may not always be the case. Also, their use in high risk groups means that such populations might have individual risk more accurately predicted by incorporation of MRI-based measures of breast composition in risk assessment. As noted in Chapter 8, it may be that the dual effect of background parenchymal enhancement and fibroglandular tissue volume can substantially improve our current mammographic density as a risk predictor. Confirming or refuting this should be a target for future studies.
One technology which has potential to radically improve automated volumetric density is digital breast tomosynthesis [279, 280]. Unlike MRI, the incorporation of tomosynthesis into mass screening is feasible, and the technology can in principle provide much more three-dimensional information on breast composition than two-view mammography. It would therefore seem timely to develop and evaluate methods of density measurement from three-dimensional tomosynthesis data.

As stated above, the priority in research on breast density and its clinical applications should be to identify the most appropriate standardised method to assess it. For this reason in Chapter 8 we presented a comparison of the methods available in this thesis and their association to risk, bearing in mind that this analyses had the major weakness of not being performed on the same dataset. Thus an ideal future study would require a case-control dataset, where cases are screen-detected, to limit potential masking bias, and density is assessed using several methods, primarily: visual classification (21 categories), Cumulus that would be used as current reference, secondly automated methods

such as Quantra and Volpara, and, possibly, the newly presented methods based on pattern or variation measures. It would be interesting also to have data for at least a portion of these women on fibroglandular volume and background parenchymal enhancement measured with MRI, in order to see which of the afore-mentioned methods is the more accurate, to clarify what each method is actually measuring and to understand how they differ. Nevertheless, for practical reasons, the priority must be to identify the density measure with the strongest association with risk, as there is still an ongoing debate about the measure of breast composition in terms of risk prediction and association with pathology. Such a study should be feasible in the near future as there are currently really large studies, such as PROCAS and TOMMY, collecting multiple density measures using different tools. These have great promise in terms of answering this question.

Bibliography

[1] J. Tyrer, S.W. Duffy, and J. Cuzick. A breast cancer prediction model incorporating familial and personal risk factors. *Statistics in medicine*, 23(7):1111–1130, 2004.

[2] W.E. Barlow, E. White, R. Ballard-Barbash, P.M. Vacek, L. Titus-Ernstoff, P.A. Carney, J.A. Tice, D.S.M. Buist, B.M. Geller, R. Rosenberg, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. *Journal of the National Cancer Institute*, 98(17):1204–14, 2006.

[3] J.A. Tice, S.R. Cummings, E. Ziv, and K. Kerlikowske. Mammographic breast density and the gail model for breast cancer risk prediction in a screening population. *Breast cancer research and treatment*, 94(2):115–122, 2005.

[4] S.W. Duffy, I.D. Nagtegaal, S.M. Astley, M.G. Gillan, M.A. McGee, C.R. Boggis, M. Wilson, U.M. Beetles, M.A. Griffiths, A.K. Jain, et al. Visually assessed breast density, breast cancer risk and the importance of the craniocaudal view. *Breast Cancer Res*, 10(4):R64, 2008.

[5] J.N. Wolfe. Breast patterns as an index of risk for developing breast cancer. *American Journal of Roentgenology*, 126(6):1130–7, 1976.

[6] V.A. McCormack and I. dos Santos Silva. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiology Biomarkers & Prevention*, 15(6):1159–69, 2006.

[7] G. Ursin, L. Hovanessian-Larsen, Y.R. Parisky, M.C. Pike, and A.H. Wu. Greatly increased occurrence of breast cancers in areas of mammographically dense tissue. *Breast Cancer Res*, 7(5):R605–R608, 2005.

[8] N. Boyd, L. Martin, A. Gunasekara, O. Melnichouk, G. Maudsley, C. Peressotti, M. Yaffe, and S. Minkin. Mammographic density and breast cancer risk: evaluation of a novel method of measuring breast tissue volumes. *Cancer Epidemiology Biomarkers & Prevention*, 18(6):1754–62, 2009.

[9] P.P. Provenzano, D.R. Inman, K.W. Eliceiri, J.G. Knittel, L. Yan, C.T. Rueden, J.G. White, and P.J. Keely. Collagen density promotes mammary tumor initiation and progression. *BMC medicine*, 6(1):11, 2008.

[10] K. Ghosh, L.C. Hartmann, C. Reynolds, D.W. Visscher, K.R. Brandt, R.A. Vierkant, C.G. Scott, D.C. Radisky, T.A. Sellers, V.S. Pankratz, et al. Association between mammographic density and age-related lobular involution of the breast. *Journal of Clinical Oncology*, 28(13):2207–2212, 2010.

[11] C.M. Vachon, H. Sasano, K. Ghosh, K.R. Brandt, D.A. Watson, C. Reynolds, W.L. Lingle, P.E. Goss, R. Li, S.E. Aiyar, et al. Aromatase immunoreactivity is increased in mammographically dense regions of the breast. *Breast cancer research and treatment*, 125(1):243–252, 2011.

[12] C.M. Vachon, C.C. Kuni, K. Anderson, V.E. Anderson, and T.A. Sellers. Association of mammographically defined percent breast density with epidemiologic risk factors for breast cancer (united states). *Cancer Causes and Control*, 11(7): 653–662, 2000.

[13] C.M. Vachon, C.H. Van Gils, T.A. Sellers, K. Ghosh, S. Pruthi, K.R. Brandt, and V.S. Pankratz. Mammographic density, breast cancer risk and risk prediction. *Breast Cancer Res*, 9(6):217, 2007.

[14] N.F. Boyd, G.A. Lockwood, J.W. Byng, D.L. Tritchler, and M.J. Yaffe. Mammographic densities and breast cancer risk. *Cancer Epidemiology Biomarkers & Prevention*, 7(12):1133–44, 1998.

[15] N.F. Boyd, J.W. Byng, R.A. Jong, E.K. Fishell, L.E. Little, A.B. Miller, G.A. Lockwood, D.L. Tritchler, and M.J. Yaffe. Quantitative classification of mammographic densities and breast cancer risk: results from the canadian national breast screening study. *Journal of the National Cancer Institute*, 87(9):670–5, 1995.

[16] C. Byrne, C. Schairer, J. Wolfe, N. Parekh, M. Salane, L.A. Brinton, R. Hoover, and R. Haile. Mammographic features and breast cancer risk: effects with time, age, and menopause status. *Journal of the National Cancer Institute*, 87(21): 1622–9, 1995.

[17] J. Cuzick, J. Warwick, E. Pinney, S.W. Duffy, S. Cawthorn, A. Howell, J.F. Forbes, and R.M.L. Warren. Tamoxifen-induced reduction in mammographic density and breast cancer risk reduction: a nested case–control study. *Journal of the National Cancer Institute*, 103(9):744–52, 2011.

[18] J. Kim, W. Han, H. Moon, S.K. Ahn, H. Shin, J. You, S. Han, S. Im, T. Kim, H.R. Koo, et al. Breast density change as a predictive surrogate for response to adjuvant endocrine therapy in hormone receptor positive breast cancer. *Breast Cancer Res*, 14(4):R102, 2012.

[19] J. Li, K. Humphreys, L. Eriksson, G. Edgren, K. Czene, and P. Hall. Mammographic density reduction is a prognostic marker of response to adjuvant tamoxifen therapy in postmenopausal patients with breast cancer. *Journal of Clinical Oncology*, 31(18):2249–2256, 2013.

[20] K.L. Ko, I.S. Shin, J.Y. You, S. Jung, J. Ro, and E.S. Lee. Adjuvant tamoxifen-induced mammographic breast density reduction as a predictor for recurrence in estrogen receptor-positive premenopausal breast cancer patients. *Breast cancer research and treatment*, pages 1–9, 2013.

[21] I.T. Gram, E. Funkhouser, and L. Tabár. The tabar classification of mammographic parenchymal patterns. *European journal of radiology*, 24(2):131–136, 1997.

[22] S.W. Duffy, R.W. Jakes, F.C. Ng, and F. Gao. Interaction of dense breast patterns with other breast cancer risk factors in a case–control study. *British journal of cancer*, 91(2):233–236, 2004.

[23] N.F. Boyd, B. O'sullivan, J.E. Campbell, E. Fishell, I. Simor, G. Cooke, and T. Germanson. Mammographic signs as risk factors for breast cancer. *British journal of cancer*, 45(2):185–93, 1982.

[24] N.F. Boyd, H.M. Jensen, G. Cooke, and H.L. Han. Relationship between mammographic and histological risk factors for breast cancer. *Journal of the National Cancer Institute*, 84(15):1170–9, 1992.

[25] N.F. Boyd, L.J. Martin, L. Sun, H. Guo, A. Chiarelli, G. Hislop, M. Yaffe, and S. Minkin. Body size, mammographic density, and breast cancer risk. *Cancer Epidemiology Biomarkers & Prevention*, 15(11):2086–92, 2006.

[26] N.F. Boyd, H. Guo, L.J. Martin, L. Sun, J. Stone, E. Fishell, R.A. Jong, G. Hislop, A. Chiarelli, S. Minkin, et al. Mammographic density and the risk and detection of breast cancer. *New England Journal of Medicine*, 356(3):227–236, 2007.

[27] J. Gao, R. Warren, H. Warren-Forward, and J.F. Forbes. Reproducibility of visual assessment on mammographic density. *Breast cancer research and treatment*, 108 (1):121–127, 2008.

[28] M.J. Yaffe. Measurement of mammographic density. *Breast Cancer Res*, 10(3): 209, 2008.

[29] J.N. Wolfe, A.F. Saftlas, and M. Salane. Mammographic parenchymal patterns and quantitative evaluation of mammographic densities: a case-control study. *American Journal of Roentgenology*, 148(6):1087–92, 1987.

[30] J.W. Byng, M.J. Yaffe, R.A. Jong, R.S. Shumak, G.A. Lockwood, D.L. Tritchler, and N.F. Boyd. Analysis of mammographic density and breast cancer risk from digitized mammograms. *Radiographics*, 18(6):1587–98, 1998.

[31] J.J. Heine, M.J. Carston, C.G. Scott, K.R. Brandt, F.F. Wu, V.S. Pankratz, T.A. Sellers, and C.M. Vachon. An automated approach for estimation of breast density. *Cancer Epidemiology Biomarkers & Prevention*, 17(11):3090–7, 2008.

[32] N.F. Boyd, L.J. Martin, M. Yaffe, and S. Minkin. Mammographic density. *Breast Cancer Res*, 11(Suppl 3):S4, 2009.

[33] J. Stone, J. Ding, R. Warren, S. Duffy, and J. Hopper. Using mammographic density to predict breast cancer risk: dense area or percentage dense area. *Breast Cancer Research*, 12(6):R97, 2010.

[34] M. Jeffreys, R. Warren, G. Davey Smith, and D. Gunnell. Breast density: agreement of measures from film and digital image. *British journal of radiology*, 76 (908):561–3, 2003.

[35] J.A. Harvey. Quantitative assessment of percent breast density: analog versus digital acquisition. *Technology in cancer research & treatment*, 3(6):611–6, 2004.

[36] Z. Aitken, V.A. McCormack, R.P. Highnam, L. Martin, A. Gunasekara, O. Melnichouk, G. Mawdsley, C. Peressotti, M. Yaffe, N.F. Boyd, et al. Screen-film mammographic density and breast cancer risk: a comparison of the volumetric standard mammogram form and the interactive threshold measurement methods. *Cancer Epidemiology Biomarkers & Prevention*, 19(2):418–28, 2010.

[37] M. Jeffreys, R. Warren, R. Highnam, and G.D. Smith. Breast cancer risk factors and a novel measure of volumetric breast density: cross-sectional study. *British journal of cancer*, 98(1):210–216, 2007.

[38] K. Hartman, R. Highnam, R. Warren, and V. Jackson. Volumetric assessment of breast tissue composition from ffdm images. *Digital Mammography*, pages 33–39, 2008.

[39] S. Malkov, J. Wang, K. Kerlikowske, S.R. Cummings, and J.A. Shepherd. Single x-ray absorptiometry method for the quantitative mammographic measure of fibroglandular tissue volume. *Medical physics*, 36:5525–36, 2009.

[40] O. Pawluczyk, B.J. Augustine, M.J. Yaffe, D. Rico, J. Yang, G.E. Mawdsley, and N.F. Boyd. A volumetric method for estimation of breast density on digitized screen-film mammograms. *Medical physics*, 30:352–64, 2003.

[41] J. Kaufhold, J.A. Thomas, J.W. Eberhard, C.E. Galbo, and D.E.G. Trotter. A calibration approach to glandular tissue composition estimation in digital mammography. *Medical physics*, 29:1867–80, 2002.

[42] J.A. Shepherd, K. Kerlikowske, L. Ma, F. Duewer, B. Fan, J. Wang, S. Malkov, E. Vittinghoff, and S.R. Cummings. Volume of mammographic density and risk of breast cancer. *Cancer Epidemiology Biomarkers & Prevention*, 20(7):1473–82, 2011.

[43] J. Chen. Quantitative imaging of breast density. *OMICS Journal of Radiology*, 2013.

[44] Nancy A Lee, Henry Rusinek, Jeffrey Weinreb, Ramesh Chandra, Hildegard Toth, Cory Singer, and Gillian Newstead. Fatty and fibroglandular tissue volumes in the breasts of women 20-83 years old: comparison of x-ray mammography and computer-assisted mr imaging. *AJR. American journal of roentgenology*, 168(2): 501–506, 1997.

[45] Jun Wei, Heang-Ping Chan, Mark A Helvie, Marilyn A Roubidoux, Berkman Sahiner, Lubomir M Hadjiiski, Chuan Zhou, Sophie Paquerault, Thomas Chenevert, and Mitchell M Goodsitt. Correlation between mammographic density and volumetric fibroglandular tissue estimated on breast mr images. *Medical physics*, 31(4):933–942, 2004.

[46] Deborah J Thompson, Martin O Leach, Gek Kwan-Lim, Simon A Gayther, Susan J Ramus, Iqbal Warsi, Fiona Lennard, Michael Khazen, Emilie Bryant, Sadie Reed, 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 Res*, 11(6):R80, 2009.

[47] Catherine Klifa, Julio Carballido-Gamio, Lisa Wilmes, Anne Laprie, John Shepherd, Jessica Gibbs, Bo Fan, Susan Noworolski, and Nola Hylton. Magnetic resonance imaging for secondary assessment of breast density in a high-risk cohort. *Magnetic resonance imaging*, 28(1):8–15, 2010.

[48] Despina Kontos, Ye Xing, Predrag R Bakic, Emily F Conant, and Andrew DA Maidment. A comparative study of volumetric breast density estimation in digital mammography and magnetic resonance imaging: Results from a high-risk population. In *SPIE Medical Imaging*, pages 762409–762409. International Society for Optics and Photonics, 2010.

[49] M.C. Pike and C.L. Pearce. Mammographic density, mri background parenchymal enhancement and breast cancer risk. *Annals of Oncology*, 24(suppl 8):viii37–viii41, 2013.

[50] Jeff Wang, Ania Azziz, Bo Fan, Serghei Malkov, Catherine Klifa, David Newitt, Silaja Yitta, Nola Hylton, Karla Kerlikowske, and John A Shepherd. Agreement of mammographic measures of volumetric breast density to mri. *PloS one*, 8(12): e81653, 2013.

[51] S. Ciatto, D. Bernardi, M. Calabrese, M. Durando, M.A. Gentilini, G. Mariscotti, F. Monetti, E. Moriconi, B. Pesce, A. Roselli, et al. A first evaluation of breast radiological density assessment by quantra software as compared to visual classification. *The Breast*, 2012.

[52] R. Highnam, S. Brady, M. Yaffe, N. Karssemeijer, and J. Harvey. Robust breast composition measurement-volpara tm. *Digital Mammography*, pages 342–349, 2010.

[53] Saskia van Engeland, Peter R Snoeren, Henkjan Huisman, Carla Boetes, and Nico Karssemeijer. Volumetric breast density estimation from full-field digital mammograms. *Medical Imaging, IEEE Transactions on*, 25(3):273–282, 2006.

[54] V. King, J.D. Brooks, J.L. Bernstein, A.S. Reiner, M.C. Pike, and E.A. Morris. Background parenchymal enhancement at breast mr imaging and breast cancer risk. *Radiology*, 260(1):50–60, 2011.

[55] J. Li, L. Szekely, L. Eriksson, B. Heddson, A. Sundbom, K. Czene, P. Hall, and K. Humphreys. High-throughput mammographic-density measurement: a tool for risk prediction of breast cancer. *Breast Cancer Research*, 14(4):R114, 2012.

[56] M. Nielsen, G. Karemore, M. Loog, J. Raundahl, N. Karssemeijer, J.D.M. Otten, M.A. Karsdal, C.M. Vachon, and C. Christiansen. A novel and automatic mammographic texture resemblance marker is an independent risk factor for breast cancer. *Cancer Epidemiology*, 35(4):381–387, 2011.

[57] D.F. Schmidt, E. Makalic, J. Stone, R.M.L. Warren, and J.L. Hopper. Some machine learning approaches to discovering information in mammograms. In *Proceedings of the 6^{t}h International Workshop on Breast Densitometry and Breast Cancer Risk Assessment, San Francisco, CA, USA*, 2013.

[58] J.J. Heine, C.G. Scott, T.A. Sellers, K.R. Brandt, D.J. Serie, F. Wu, M.J. Morton, B.A. Schueler, F.J. Couch, J.E. Olson, et al. A novel automated mammographic density measure and breast cancer risk. *Journal of the National Cancer Institute*, 104(13):1028–1037, 2012.

[59] Y Zou and Z Guo. A review of electrical impedance techniques for breast cancer detection. *Medical engineering & physics*, 25(2):79–90, 2003.

[60] R.I. Freimanis and M. Yacobozzi. Breast cancer screening. *NC Med J*, 75(2): 117–120, 2014.

[61] Despina Kontos, Predrag R Bakic, Ann-Katherine Carton, Andrea B Troxel, Emily F Conant, and Andrew DA Maidment. Parenchymal texture analysis in digital breast tomosynthesis for breast cancer risk estimation: a preliminary study. *Academic radiology*, 16(3):283–298, 2009.

[62] E. Warner, G. Lockwood, D. Tritchler, and N.F. Boyd. 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*, 16(1):67–72, 1992.

[63] C.M. Vachon, K. Ghosh, and K.R. Brandt. Mammographic density: Potential as a risk factor and surrogate marker in the clinical setting. *Current Breast Cancer Reports*, 5(3):183–193, 2013.

[64] J. Cuzick. Assessing risk for breast cancer. *Breast Cancer Res*, 10(suppl 4):S13, 2008.

[65] R.L. Egan and R.C. Mosteller. Breast cancer mammography patterns. *Cancer*, 40 (5):2087–2090, 1977.

[66] L. Ma, E. Fishell, B. Wright, W. Hanna, S. Allan, and N.F. Boyd. Case-control study of factors associated with failure to detect breast cancer by mammography. *Journal of the National Cancer Institute*, 84(10):781–5, 1992.

[67] M.T. Mandelson, N. Oestreicher, P.L. Porter, D. White, C.A. Finder, S.H. Taplin, and E. White. Breast density as a predictor of mammographic detection: comparison of interval-and screen-detected cancers. *Journal of the National Cancer Institute*, 92(13):1081–7, 2000.

[68] C.H. van Gils, J.D.M. Otten, A.L.M. Verbeek, and J.H.C.L. Hendriks. Mammographic breast density and risk of breast cancer: Masking bias or causality? *European journal of epidemiology*, 14(4):315–320, 1998.

[69] N.F. Boyd. Mammographic density and risk of breast cancer. *2013 ASCO Educational Book*, pages e57–62, 2013.

[70] N.F. Boyd, C. Greenberg, G. Lockwood, L. Little, L. Martin, D. Tritchler, J. Byng, and M. Yaffe. Effects at two years of a low-fat, high-carbohydrate diet on radiologic features of the breast: results from a randomized trial. *Journal of the National Cancer Institute*, 89(7):488–96, 1997.

[71] S.W. Duffy. Epidemiology of female breast cancer. *Breast Cancer (ed. Mitchell, M.J.)*, pages 1–12, 2010.

[72] J. Cuzick, J. Warwick, E. Pinney, R.M.L. Warren, and S.W. Duffy. Tamoxifen and breast density in women at increased risk of breast cancer. *Journal of the National Cancer Institute*, 96(8):621–8, 2004.

[73] J.T. Schousboe, K. Kerlikowske, A. Loh, and S.R. Cummings. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. *Annals of internal medicine*, 155(1):10–20, 2011.

[74] N. Houssami and K. Kerlikowske. The impact of breast density on breast cancer risk and breast screening. *Current Breast Cancer Reports*, 4(2):161–168, 2012.

[75] C.S. Wong, G.H. Lim, F. Gao, R.W. Jakes, J. Offman, K.S. Chia, and S.W. Duffy. Mammographic density and its interaction with other breast cancer risk factors in an asian population. *British journal of cancer*, 104:871–4, 2011.

[76] C.M. Vachon, K.R. Brandt, K. Ghosh, C.G. Scott, S.D. Maloney, M.J. Carston, V.S. Pankratz, and T.A. Sellers. Mammographic breast density as a general marker

of breast cancer risk. *Cancer Epidemiology Biomarkers & Prevention*, 16(1):43–9, 2007.

[77] J. Cuzick, D. Berridge, and J. Whitehead. Mammographic dysplasia as entry criterion for breast cancer prevention trials. *Lancet*, 337(8751):1225, 1991.

[78] J. Chen, D. Pee, R. Ayyagari, B. Graubard, C. Schairer, C. Byrne, J. Benichou, and M.H. Gail. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *Journal of the National Cancer Institute*, 98(17):1215–26, 2006.

[79] H. Darabi, K. Czene, W. Zhao, J. Liu, P. Hall, K. Humphreys, et al. Breast cancer risk prediction and individualised screening based on common genetic variation and breast density measurement. *Breast Cancer Res*, 14(1):R25, 2012.

[80] L. Yaghjyan, G.A. Colditz, L.C. Collins, S.J. Schnitt, B. Rosner, C. Vachon, and R.M. Tamimi. Mammographic breast density and subsequent risk of breast cancer in postmenopausal women according to tumor characteristics. *Journal of the National Cancer Institute*, 103(15):1179–1189, 2011.

[81] S. Antoni, A.J. Sasco, I. dos Santos Silva, and V. McCormack. Is mammographic density differentially associated with breast cancer according to receptor status? a meta-analysis. *Breast cancer research and treatment*, 137(2):337–347, 2013.

[82] E.S. Hwang, D.L. Miglioretti, R. Ballard-Barbash, D.L. Weaver, and K. Kerlikowske. Association between breast density and subsequent breast cancer following treatment for ductal carcinoma in situ. *Cancer Epidemiology Biomarkers & Prevention*, 16(12):2587–2593, 2007.

[83] L.A. Habel, A.M. Capra, N.S. Achacoso, A. Janga, L. Acton, B. Puligandla, and C.P. Quesenberry. Mammographic density and risk of second breast cancer after ductal carcinoma in situ. *Cancer Epidemiology Biomarkers & Prevention*, 19(10): 2488–2495, 2010.

[84] A.H. Olsen, K. Bihrmann, M.B. Jensen, I. Vejborg, and E. Lynge. Breast density and outcome of mammography screening: a cohort study. *British journal of cancer*, 100(7):1205–1208, 2009.

[85] G.L. Gierach, L. Ichikawa, K. Kerlikowske, L.A. Brinton, G.N. Farhat, P.M. Vacek, D.L. Weaver, C. Schairer, S.H. Taplin, and M.E. Sherman. Relationship between mammographic density and breast cancer death in the breast cancer surveillance consortium. *Journal of the National Cancer Institute*, 104(16):1218–1227, 2012.

[86] B.S. Wiseman and Z. Werb. Stromal effects on mammary gland development and breast cancer. *Science*, 296(5570):1046–9, 2002.

[87] M.H. Barcellos-Hoff and D. Medina. New highlights on stroma-epithelial interactions in breast cancer. *Breast Cancer Res*, 7(1):33–36, 2005.

[88] S.W. Hutson, P.N. Cowen, and C.C. Bird. Morphometric studies of age related changes in normal human breast and their significance for evolution of mammary cancer. *Journal of clinical pathology*, 38(3):281–7, 1985.

[89] A.M. Oza and N.F. Boyd. Mammographic parenchymal patterns: a marker of breast cancer risk. *Epidemiologic reviews*, 15(1):196–208, 1993.

[90] N. Boyd, L. Martin, J. Stone, L. Little, S. Minkin, and M. Yaffe. A longitudinal study of the effects of menopause on mammographic features. *Cancer Epidemiology Biomarkers & Prevention*, 11(10):1048–1053, 2002.

[91] V.A. McCormack, N.M. Perry, S.J. Vinnicombe, and I. dos Santos Silva. Changes and tracking of mammographic density in relation to pike's model of breast tissue aging: a uk longitudinal study. *International Journal of Cancer*, 127(2):452–461, 2010.

[92] N.F. Boyd, J.M. Rommens, K. Vogt, V. Lee, J.L. Hopper, M.J. Yaffe, and A.D. Paterson. Mammographic breast density as an intermediate phenotype for breast cancer. *The lancet oncology*, 6(10):798–808, 2005.

[93] A. Stuedal, H. Ma, L. Bernstein, M.C. Pike, and G. Ursin. Does breast size modify the association between mammographic density and breast cancer risk? *Cancer Epidemiology Biomarkers & Prevention*, 17(3):621–7, 2008.

[94] L. Yaghjyan, G.A. Colditz, and K. Wolin. Physical activity and mammographic breast density: a systematic review. *Breast cancer research and treatment*, 135(2): 367–380, 2012.

[95] C.G. Woolcott, L.S. Cook, K.S. Courneya, N.F. Boyd, M.J. Yaffe, T. Terry, R. Brant, A. McTiernan, H.E. Bryant, A.M. Magliocco, et al. Associations of overall and abdominal adiposity with area and volumetric mammographic measures among postmenopausal women. *International Journal of Cancer*, 129(2): 440–8, 2011.

[96] M.C. Pike, M.D. Krailo, B.E. Henderson, J.T. Casagrande, and D.G. Hoel. hormonalrisk factors,breast tissue ageand the age-incidence of breast cancer. *Nature*, 303:767–70, 1983.

[97] M. Lokate, R.K. Stellato, W.B. Veldhuis, P.H.M. Peeters, and C.H. van Gils. Age-related changes in mammographic density and breast cancer risk. *American journal of epidemiology*, 2013.

[98] K.H. Johnson and P.S. Millard. Oral contraceptives and breast cancer. *The Journal of family practice*, 43(4):340–1, 1996.

[99] E. Sala, R. Warren, J. McCann, S. Duffy, R. Luben, and N. Day. High-risk mammographic parenchymal patterns, hormone replacement therapy and other risk factors: a case-control study. *International journal of epidemiology*, 29(4): 629–36, 2000.

[100] P.M. Vacek and B.M. Geller. A prospective study of breast cancer risk using routine mammographic breast density measurements. *Cancer Epidemiology Biomarkers & Prevention*, 13(5):715–22, 2004.

[101] N.F. Boyd, L.J. Martin, Q. Li, L. Sun, A.M. Chiarelli, G. Hislop, M.J. Yaffe, and S. Minkin. Mammographic density as a surrogate marker for the effects of hormone therapy on risk of breast cancer. *Cancer Epidemiology Biomarkers & Prevention*, 15(5):961–6, 2006.

[102] N. Hou, S. Hong, W. Wang, O.I. Olopade, J.J. Dignam, and D. Huo. Hormone replacement therapy and breast cancer: Heterogeneous risks by race, weight, and breast density. *Journal of the National Cancer Institute*, 105(18):1365–1372, 2013.

[103] G.A. Greendale, B.A. Reboussin, S. Slone, C. Wasilauskas, M.C. Pike, and G. Ursin. Postmenopausal hormone therapy and change in mammographic density. *Journal of the National Cancer Institute*, 95(1):30–7, 2003.

[104] C.M. Vachon, T.A. Sellers, R.A. Vierkant, F.F. Wu, and K.R. Brandt. Case-control study of increased mammographic breast density response to hormone replacement therapy. *Cancer Epidemiology Biomarkers & Prevention*, 11(11):1382–8, 2002.

[105] C.M. Rutter, M.T. Mandelson, M.B. Laya, and S. Taplin. Changes in breast density associated with initiation, discontinuation, and continuing use of hormone replacement therapy. *JAMA: the journal of the American Medical Association*, 285(2):171–6, 2001.

[106] N.F. Boyd, J. Stone, L.J. Martin, R. Jong, E. Fishell, M. Yaffe, G. Hammond, and S. Minkin. The association of breast mitogens with mammographic densities. *British journal of cancer*, 87(8):876–882, 2002.

[107] J. Eng-Wong, J. Orzano-Birgani, C.K. Chow, D. Venzon, J. Yao, C.E. Galbo, J.A. Zujewski, and S. Prindiville. Effect of raloxifene on mammographic density and breast magnetic resonance imaging in premenopausal women at increased risk for breast cancer. *Cancer Epidemiology Biomarkers & Prevention*, 17(7):1696–701, 2008.

[108] A.L. Eilertsen, L. Sandvik, B. Steinsvik, and P.M. Sandset. Differential impact of conventional-dose and low-dose postmenopausal hormone therapy, tibolone and raloxifene on c-reactive protein and other inflammatory markers. *Journal of Thrombosis and Haemostasis*, 6(6):928–934, 2008.

[109] M. Freedman, J.S. Martin, J. O'Gorman, S. Eckert, M.E. Lippman, S.C.B. Lo, E.L. Walls, and J. Zeng. Digitized mammography: a clinical trial of postmenopausal women randomly assigned to receive raloxifene, estrogen, or placebo. *Journal of the National Cancer Institute*, 93(1):51–6, 2001.

[110] C.M. Vachon, J.N. Ingle, V.J. Suman, C.G. Scott, H. Gottardt, J.E. Olson, and P.E. Goss. Pilot study of the impact of letrozole vs. placebo on breast density in women completing 5 years of tamoxifen. *The Breast*, 16(2):204–210, 2007.

[111] T. Cigler, D. Tu, M.J. Yaffe, B. Findlay, S. Verma, D. Johnston, H. Richardson, H. Hu, S. Qi, and P.E. Goss. 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 research and treatment*, 120(2):427–435, 2010.

[112] C.M. Vachon, V.J. Suman, K.R. Brandt, M.L. Kosel, A.U. Buzdar, J.E. Olson, F. Wu, L.M. Flickinger, G. Ursin, C.R. Elliott, et al. Mammographic breast density response to aromatase inhibition. *Clinical Cancer Research*, 19(8):2144–2153, 2013.

[113] M.N. Pollak. Endocrine effects of igf-i on normal and transformed breast epithelial cells: potential relevance to strategies for breast cancer treatment and prevention. *Breast cancer research and treatment*, 47(3):209–217, 1998.

[114] V.A. McCormack, M. Dowsett, E. Folkerd, N. Johnson, C. Palles, B. Coupland, J.M. Holly, S.J. Vinnicombe, N.M. Perry, and I. dos Santos Silva. Sex steroids, growth factors and mammographic density: a cross-sectional study of uk post-menopausal caucasian and afro-caribbean women. *Breast Cancer Res*, 11(3):R38, 2009.

[115] R.M. Tamimi, S.E. Hankinson, G.A. Colditz, and C. Byrne. Endogenous sex hormone levels and mammographic density among postmenopausal women. *Cancer Epidemiology Biomarkers & Prevention*, 14(11):2641–7, 2005.

[116] R.M. Tamimi, C. Byrne, G.A. Colditz, and S.E. Hankinson. Endogenous hormone levels, mammographic density, and subsequent risk of breast cancer in postmenopausal women. *Journal of the National Cancer Institute*, 99(15):1178–87, 2007.

[117] E. Ziv, J. Tice, R. Smith-Bindman, J. Shepherd, S. Cummings, and K. Kerlikowske. Mammographic density and estrogen receptor status of breast cancer. *Cancer Epidemiology Biomarkers & Prevention*, 13(12):2090–5, 2004.

[118] H. Ma, J. Luo, M.F. Press, Y. Wang, L. Bernstein, and G. Ursin. Is there a difference in the association between percent mammographic density and subtypes of breast cancer? luminal a and triple-negative breast cancer. *Cancer Epidemiology Biomarkers & Prevention*, 18(2):479–85, 2009.

[119] J.R. Pasqualini, G. Chetrite, C. Blacker, M.C. Feinstein, L. Delalonde, M. Talbi, and C. Maloche. Concentrations of estrone, estradiol, and estrone sulfate and evaluation of sulfatase and aromatase activities in pre-and postmenopausal breast cancer patients. *Journal of Clinical Endocrinology & Metabolism*, 81(4):1460–4, 1996.

[120] W. Yue, R.J. Santen, J.P. Wang, C.J. Hamilton, and L.M. Demers. Aromatase within the breast. *Endocrine-Related Cancer*, 6(2):157–64, 1999.

[121] J.S. Varghese, P.L. Smith, E. Folkerd, J. Brown, J. Leyland, T. Audley, R.M.L. Warren, M. Dowsett, D.F. Easton, and D.J. Thompson. The heritability of mammographic breast density and circulating sex-hormone levels: two independent breast cancer risk factors. *Cancer Epidemiology Biomarkers & Prevention*, 21 (12):2167–2175, 2012.

[122] L.J. Martin and N. Boyd. Potential mechanisms of breast cancer risk associated with mammographic density: hypotheses based on epidemiological evidence. *Breast Cancer Res*, 10(1):1–14, 2008.

[123] N.F. Boyd, G.S. Dite, J. Stone, A. Gunasekara, D.R. English, M.R.E. McCredie, G.G. Giles, D. Tritchler, A. Chiarelli, M.J. Yaffe, et al. Heritability of mammographic density, a risk factor for breast cancer. *New England Journal of Medicine*, 347(12):886–894, 2002.

[124] A.B. Crest, E.J. Aiello, M.L. Anderson, and D.S.M. Buist. Varying levels of family history of breast cancer in relation to mammographic breast density (united states). *Cancer Causes and Control*, 17(6):843–850, 2006.

[125] J.S. Pankow, C.M. Vachon, C.C. Kuni, R.A. King, D.K. Arnett, D.M. Grabrick, S.S. Rich, V.E. Anderson, and T.A. Sellers. Genetic analysis of mammographic breast density in adult women: evidence of a gene effect. *Journal of the National Cancer Institute*, 89(8):549–56, 1997.

[126] M. Kataoka, A. Antoniou, R. Warren, J. Leyland, J. Brown, T. Audley, and D. Easton. Genetic models for the familial aggregation of mammographic breast density. *Cancer Epidemiology Biomarkers & Prevention*, 18(4):1277, 2009.

[127] C.M. Vachon, T.A. Sellers, E.E. Carlson, J.M. Cunningham, C.A. Hilker, R.L. Smalley, D.J. Schaid, L.E. Kelemen, F.J. Couch, and V.S. Pankratz. Strong evidence of a genetic determinant for mammographic density, a major risk factor for breast cancer. *Cancer research*, 67(17):8412–8, 2007.

[128] G. Ursin, E.O. Lillie, E. Lee, M. Cockburn, N.J. Schork, W. Cozen, Y.R. Parisky, A.S. Hamilton, M.A. Astrahan, and T. Mack. The relative importance of genetics and environment on mammographic density. *Cancer Epidemiology Biomarkers & Prevention*, 18(1):102–12, 2009.

[129] D.F. Easton, K.A. Pooley, A.M. Dunning, P.D.P. Pharoah, D. Thompson, D.G. Ballinger, J.P. Struewing, J. Morrison, H. Field, R. Luben, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, 447 (7148):1087–1093, 2007.

[130] F. Odefrey, J. Stone, L.C. Gurrin, G.B. Byrnes, C. Apicella, G.S. Dite, J.N. Cawson, G.G. Giles, S.A. Treloar, D.R. English, et al. Common genetic variants associated with breast cancer and mammographic density measures that predict disease. *Cancer research*, 70(4):1449–58, 2010.

[131] S. Lindström, C.M. Vachon, J. Li, J. Varghese, D. Thompson, R. Warren, J. Brown, J. Leyland, T. Audley, N.J. Wareham, et al. Common variants in znf365 are associated with both mammographic density and breast cancer risk. *Nature genetics*, 43(3):185–187, 2011.

[132] K.N. Stevens, S. Lindstrom, C.G. Scott, D. Thompson, T.A. Sellers, X. Wang, A. Wang, E. Atkinson, D.N. Rider, J.E. Eckel-Passow, et al. Identification of a novel percent mammographic density locus at 12q24. *Human Molecular Genetics*, 2012.

[133] J.S. Varghese, D.J. Thompson, K. Michailidou, S. Lindström, C. Turnbull, J. Brown, J. Leyland, R.M.L. Warren, R.N. Luben, R.J. Loos, et al. Mammographic breast density and breast cancer: evidence of a shared genetic basis. *Cancer research*, 72(6):1478–1484, 2012.

[134] L.E. Kelemen, T.A. Sellers, and C.M. Vachon. Can genes for mammographic density inform cancer etiology? *Nature reviews. Cancer*, 8(10):812–23, 2008.

[135] I. Dumas and C. Diorio. Polymorphisms in genes involved in the estrogen pathway and mammographic density. *BMC cancer*, 10(1):636, 2010.

[136] R.M. Tamimi, D. Cox, P. Kraft, G.A. Colditz, S.E. Hankinson, and D.J. Hunter. Breast cancer susceptibility loci and mammographic density. *Breast Cancer Res*, 10(4):R66, 2008.

[137] M.A. Helvie, M.A. Roubidoux, B.L. Weber, and S.D. Merajver. Mammography of breast carcinoma in women who have mutations of the breast cancer gene brca1: initial experience. *American Journal of Roentgenology*, 168(6):1599–602, 1997.

[138] J. Chang and M. Berer. Mammography in asian patients with brca1 mutations. *Lancet*, 353(9169):2070–2071, 1999.

[139] Z. Huo, M.L. Giger, O.I. Olopade, D.E. Wolverton, B.L. Weber, C.E. Metz, W. Zhong, and S.A. Cummings. Computerized analysis of digitized mammograms of brca1 and brca2 gene mutation carriers1. *Radiology*, 225(2):519–26, 2002.

[140] M. Tilanus-Linthorst, L. Verhoog, I.M. Obdeijn, K. Bartels, M. Menke-Pluymers, A. Eggermont, J. Klijn, H. Meijers-Heijboer, T. van der Kwast, and C. Brekelmans. A brca1/2 mutation, high breast density and prominent pushing margins of a tumor independently contribute to a frequent false-negative mammography. *International journal of cancer*, 102(1):91–95, 2002.

[141] G. Mitchell, A.C. Antoniou, R. Warren, S. Peock, J. Brown, R. Davies, J. Mattison, M. Cook, I. Warsi, D.G. Evans, et al. Mammographic density and breast cancer risk in brca1 and brca2 mutation carriers. *Cancer research*, 66(3):1866–72, 2006.

[142] C. Byrne, C. Schairer, L.A. Brinton, J. Wolfe, N. Parekh, M. Salane, C. Carter, and R. Hoover. Effects of mammographic density and benign breast disease on breast cancer risk (united states). *Cancer Causes and Control*, 12(2):103–110, 2001.

[143] L.C. Hartmann, T.A. Sellers, M.H. Frost, W.L. Lingle, A.C. Degnim, K. Ghosh, R.A. Vierkant, S.D. Maloney, V.S. Pankratz, D.W. Hillman, et al. Benign breast disease and the risk of breast cancer. *New England Journal of Medicine*, 353(3): 229–237, 2005.

[144] C.L. Carter, D.K. Corle, M.S. Micozzi, A. Schatzkin, and P.R. Taylor. A prospective study of the development of breast cancer in 16,692 women with benign breast disease. *American journal of epidemiology*, 128(3):467–77, 1988.

[145] A.C. Degnim, D.W. Visscher, H.K. Berman, M.H. Frost, T.A. Sellers, R.A. Vierkant, S.D. Maloney, V.S. Pankratz, P.C. de Groen, W.L. Lingle, et al. Stratification of breast cancer risk in women with atypia: a mayo cohort study. *Journal of clinical oncology*, 25(19):2671–7, 2007.

[146] N.F. Boyd, H.M. Jensen, G. Cooke, H.L. Han, G.A. Lockwood, and A.B. Miller. Mammographic densities and the prevalence and incidence of histological types of benign breast disease. *European journal of cancer prevention*, 9(1):15–24, 2000.

[147] J.A. Tice, E.S. OMeara, D.L. Weaver, C. Vachon, R. Ballard-Barbash, and K. Kerlikowske. Benign breast disease, mammographic breast density, and the risk of breast cancer. *Journal of the National Cancer Institute*, 2013.

[148] E. Sala, R. Warren, J. McCann, S. Duffy, N. Day, and R. Luben. Mammographic parenchymal patterns and mode of detection: implications for the breast screening programme. *Journal of Medical Screening*, 5(4):207, 1998.

[149] S.Y.H. Chiu, S. Duffy, A.M.F. Yen, L. Tabár, R.A. Smith, and H.H. Chen. Effect of baseline breast density on breast cancer incidence, stage, mortality, and screening parameters: 25-year follow-up of a swedish mammographic screening. *Cancer Epidemiology Biomarkers & Prevention*, 19(5):1219–28, 2010.

[150] J.S. Drukteinis, B.P. Mooney, C.I. Flowers, and R.A. Gatenby. Beyond mammography: New frontiers in breast cancer screening. *The American journal of medicine*, 2013.

[151] V. Corsetti, N. Houssami, M. Ghirardi, A. Ferrari, M. Speziani, S. Bellarosa, G. Remida, C. Gasparotti, E. Galligioni, and S. Ciatto. Evidence of the effect of adjunct ultrasound screening in women with mammography-negative dense breasts: Interval breast cancers at 1 year follow-up. *European Journal of Cancer*, 47(7): 1021–1026, 2011.

[152] E. Warner, K. Hill, P. Causer, D. Plewes, R. Jong, M. Yaffe, W.D. Foulkes, P. Ghadirian, H. Lynch, F. Couch, et al. Prospective study of breast cancer incidence in women with a brca1 or brca2 mutation under surveillance with and without magnetic resonance imaging. *Journal of Clinical Oncology*, 29(13):1664–9, 2011.

[153] A.J. Rijnsburger, I.M. Obdeijn, R. Kaas, M. Tilanus-Linthorst, C. Boetes, C.E. Loo, M.N.J.M. Wasser, E. Bergers, T. Kok, S.H. Muller, et al. Brca1-associated breast cancers present differently from brca2-associated and familial cases: long-term follow-up of the dutch mrisc screening study. *Journal of Clinical Oncology*, 28(36):5265–73, 2010.

[154] M.O. Leach, C.R. Boggis, A.K. Dixon, D.F. Easton, R.A. Eeles, D.G. Evans, F.J. Gilbert, I. Griebsch, R.J. Hoff, P. Kessar, et al. Maribs study group. screening with magnetic resonance imaging and mammography of a uk population at high familial risk of breast cancer: a prospective multicentre cohort study (maribs). *Lancet*, 365(9473):1769–1778, 2005.

[155] R.Z. Bigenwald, E. Warner, A. Gunasekara, K.A. Hill, P.A. Causer, S.J. Messner, A. Eisen, D.B. Plewes, S.A. Narod, L. Zhang, et al. Is mammography adequate for screening women with inherited brca mutations and low breast density? *Cancer Epidemiology Biomarkers & Prevention*, 17(3):706–11, 2008.

[156] E. Wenkel, M. Heckmann, M. Heinrich, S.A. Schwab, M. Uder, R. Schulz-Wendtland, WA Bautz, and R. Janka. Automated breast ultrasound: Lesion detection and bi-rads classification-a pilot study. In *RöFo. Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*, volume 180, pages 804–808. Thieme, 2008.

[157] K.M. Kelly, J. Dean, W.S. Comulada, and S. Lee. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *European radiology*, 20(3):734–742, 2010.

[158] K.M. Kelly and G.A. Richwald. Automated whole-breast ultrasound: advancing the performance of breast cancer screening. In *Seminars in Ultrasound, CT, and MRI*, volume 32, pages 273–280. Elsevier, 2011.

[159] W.A. Berg, Z. Zhang, D. Lehrer, R.A. Jong, E.D. Pisano, R.G. Barr, M. Böhm-Vélez, M.C. Mahoney, W.P. Evans, L.H. Larsen, 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*, 307(13):1394–1404, 2012.

[160] E.D. Pisano, C. Gatsonis, E. Hendrick, M. Yaffe, J.K. Baum, S. Acharyya, E.F. Conant, L.L. Fajardo, L. Bassett, C. D'Orsi, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. *New England Journal of Medicine*, 353(17):1773–1783, 2005.

[161] M.H. Gail, L.A. Brinton, D.P. Byar, D.K. Corle, S.B. Green, C. Schairer, and J.J. Mulvihill. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute*, 81(24):1879–86, 1989.

[162] J.P. Costantino, M.H. Gail, D. Pee, S. Anderson, C.K. Redmond, J. Benichou, and H.S. Wieand. Validation studies for models projecting the risk of invasive and total breast cancer incidence. *Journal of the National Cancer Institute*, 91(18): 1541–8, 1999.

[163] E.B. Claus, N. Risch, and W.D. Thompson. The calculation of breast cancer risk for women with a first degree family history of ovarian cancer. *Breast cancer research and treatment*, 28(2):115–120, 1993.

[164] E.B. Claus, N. Risch, and W.D. Thompson. Autosomal dominant inheritance of early-onset breast cancer. implications for risk prediction. *Cancer*, 73(3):643–651, 1994.

[165] C.J. van Asperen, M.A. Jonker, C.E. Jacobi, J.E.M. van Diemen-Homan, E. Bakker, M.H. Breuning, J.C. van Houwelingen, and G.H. de Bock. Risk estimation for healthy women from breast cancer families. *Cancer Epidemiology Biomarkers & Prevention*, 13(1):87–93, 2004.

[166] D. Ford, D.F. Easton, D.T. Bishop, S.A. Narod, and D.E. Goldgar. Risks of cancer in brca1-mutation carriers. *The Lancet*, 343(8899):692–695, 1994.

[167] E. Amir, DG Evans, A. Shenton, F. Lalloo, A. Moran, C. Boggis, M. Wilson, and A. Howell. Evaluation of breast cancer risk assessment packages in the family history evaluation and screening programme. *Journal of medical genetics*, 40(11): 807–14, 2003.

[168] S.J. Schonfeld, D. Pee, R.T. Greenlee, P. Hartge, J.V. Lacey, Y. Park, A. Schatzkin, K. Visvanathan, and R.M. Pfeiffer. Effect of changing breast cancer

incidence rates on the calibration of the gail model. *Journal of Clinical Oncology*, 28(14):2411–7, 2010.

[169] J.C. Boughey, L.C. Hartmann, S.S. Anderson, A.C. Degnim, R.A. Vierkant, C.A. Reynolds, M.H. Frost, and V.S. Pankratz. Evaluation of the tyrer-cuzick (international breast cancer intervention study) model for breast cancer risk prediction in women with atypical hyperplasia. *Journal of Clinical Oncology*, 28(22):3591–6, 2010.

[170] V.S. Pankratz, L.C. Hartmann, A.C. Degnim, R.A. Vierkant, K. Ghosh, C.M. Vachon, M.H. Frost, S.D. Maloney, C. Reynolds, and J.C. Boughey. Assessment of the accuracy of the gail model in women with atypical hyperplasia. *Journal of Clinical Oncology*, 26(33):5374–9, 2008.

[171] J.A. Tice, S.R. Cummings, R. Smith-Bindman, L. Ichikawa, W.E. Barlow, and K. Kerlikowske. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Annals of internal medicine*, 148(5):337–47, 2008.

[172] M.S. Pepe, H. Janes, G. Longton, W. Leisenring, and P. Newcomb. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American journal of epidemiology*, 159(9):882–90, 2004.

[173] M.J. Pencina, R.B. D'Agostino Sr, R.B. D'Agostino Jr, and R.S. Vasan. Evaluating the added predictive ability of a new marker: from area under the roc curve to reclassification and beyond. *Statistics in medicine*, 27(2):157–172, 2008.

[174] K. Visvanathan, R.T. Chlebowski, P. Hurley, N.F. Col, M. Ropka, D. Collyar, M. Morrow, C. Runowicz, K.I. Pritchard, K. Hagerty, et al. American society of clinical oncology clinical practice guideline update on the use of pharmacologic interventions including tamoxifen, raloxifene, and aromatase inhibition for breast cancer risk reduction. *Journal of Clinical Oncology*, 27(19):3235–58, 2009.

[175] National Istitute for Health and Clinical Excellence (NICE). Familial breast cancer: classification and care of women at risk of familial breast cancer in primary, secondary and tertiary care (partial update of cg14). http://guidance.nice.org.uk/CG41, 2006.

[176] M. Sperrin, L. Bardwell, J.C. Sergeant, S. Astley, and I. Buchan. Correcting for rater bias in scores on a continuous scale, with application to breast density. *Statistics in medicine*, 2013.

[177] NHS. Breast screening programme annual review. http://www.cancerscreening.nhs.uk/breastscreen/publications, 2010.

[178] P. Skippage, L. Wilkinson, S. Allen, N. Roche, M. Dowsett, and R. Hern. Correlation of age and hrt use with breast density as assessed by quantra. *The breast journal*, 19(1):79–86, 2013.

[179] L. Ellison-Loschmann, F. McKenzie, R. Highnam, A. Cave, J. Walker, and M. Jeffreys. Age and ethnic differences in volumetric breast density in new zealand women: A cross-sectional study. *PloS one*, 8(7):e70217, 2013.

[180] J. Cuzick, J. Forbes, R. Edwards, M. Baum, S. Cawthorn, A. Coates, A. Hamed, A. Howell, and T. Powles. First results from the international breast cancer intervention study (ibis-i): a randomised prevention trial. *Lancet*, 360(9336):817–824, 2002.

[181] J. Cuzick, J.F. Forbes, I. Sestak, S. Cawthorn, H. Hamed, K. Holli, and A. Howell. Long-term results of tamoxifen prophylaxis for breast cancer96-month follow-up of the randomized ibis-i trial. *Journal of the National Cancer Institute*, 99(4): 272–282, 2007.

[182] M. Baum, A.U. Budzar, J. Cuzick, J. Forbes, J.H. Houghton, J.G. Klijn, and T. Sahmoud. Anastrozole alone or in combination with tamoxifen versus tamoxifen alone for adjuvant treatment of postmenopausal women with early breast cancer: first results of the atac randomised trial. *Lancet*, 359(9324):2131–2139, 2002.

[183] J. Cuzick, I. Sestak, M. Baum, A. Buzdar, A. Howell, M. Dowsett, and J.F. Forbes. Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 10-year analysis of the atac trial. *The lancet oncology*, 11(12):1135–1141, 2010.

[184] SW Duffy, J Mackay, S Thomas, E Anderson, THH Chen, I Ellis, G Evans, H Fielder, R Fox, G Gui, et al. Evaluation of mammographic surveillance services in women aged 40–49 years with a moderate family history of breast cancer: a single-arm cohort study. *Health Technology Assessment*, 17(11), 2013.

[185] R.L. Prentice. Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in medicine*, 8(4):431–440, 1989.

[186] A. Schatzkin and M. Gail. The promise and peril of surrogate end points in cancer research. *Nature Reviews Cancer*, 2(1):19–27, 2002.

[187] A. Howell, J. Cuzick, M. Baum, A. Buzdar, M. Dowsett, J.F. Forbes, G. Hoctin-Boes, J. Houghton, G.Y. Locker, J.S. Tobias, et al. Results of the atac (arimidex, tamoxifen, alone or in combination) trial after completion of 5 years adjuvant treatment for breast cancer. *Lancet*, 365(9453):60–62, 2005.

[188] "FH01 collaborative teams". Mammographic surveillance in women younger than 50 years who have a family history of breast cancer: tumour characteristics and projected effect on mortality in the prospective, single-arm, fh01 study. *LANCET ONCOLOGY*, 11(12):1127–1134, 2010.

[189] L.S. Freedman, B.I. Graubard, and A. Schatzkin. Statistical validation of intermediate endpoints for chronic diseases. *Statistics in medicine*, 11(2):167–178, 1992.

[190] N.F. Boyd, L.J. Martin, M.J. Yaffe, and S. Minkin. Mammographic density and breast cancer risk: current understanding and future prospects. *Breast Cancer Research*, 13(6):223, 2011.

[191] L.J. Martin, S. Minkin, and N.F. Boyd. Hormone therapy, mammographic density, and breast cancer risk. *Maturitas*, 64(1):20–26, 2009.

[192] A.J.J. Wood and C.K. Osborne. Tamoxifen in the treatment of breast cancer. *New England Journal of Medicine*, 339(22):1609–1618, 1998.

[193] Y. Qu. Evaluation of a surrogate marker: validity and efficiency. *Statistics in medicine*, 2012.

[194] Y. Qu and M. Case. Quantifying the effect of the surrogate marker by information gain. *Biometrics*, 63(3):958–960, 2007.

[195] C.H. Mason and W.D. Perreault Jr. Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research*, pages 268–280, 1991.

[196] A. Alonso, G. Molenberghs, T. Burzykowski, D. Renard, H. Geys, Z. Shkedy, F. Tibaldi, J. C. Abrahantes, and M. Buyse. Prentice's approach and the meta-analytic paradigm: A reflection on the role of statistics in the evaluation of surrogate endpoints. *Biometrics*, 60(3):724–728, 2004.

[197] B. Ettinger, D.M. Black, B.H. Mitlak, R.K. Knickerbocker, T. Nickelsen, H.K. Genant, C. Christiansen, P.D. Delmas, J.R. Zanchetta, J. Stakkestad, et al. Reduction of vertebral fracture risk in postmenopausal women with osteoporosis treated with raloxifene. *JAMA: the journal of the American Medical Association*, 282(7): 637–645, 1999.

[198] L. Pearman, R. Kagan, J. Arsenault, and D. Muram. The effects of raloxifene on mammographic breast density: a review of clinical trials. *Menopause*, 17(3): 654–659, 2010.

[199] A. McTiernan, C.Y. Wang, B. Sorensen, L. Xiao, D.S.M. Buist, E.J.A. Bowles, E. White, M.A. Rossing, J. Potter, and N. Urban. No effect of aspirin on mammographic density in a randomized controlled clinical trial. *Cancer Epidemiology Biomarkers & Prevention*, 18(5):1524–1530, 2009.

[200] T. Cigler, H. Richardson, M.J. Yaffe, C.J. Fabian, D. Johnston, J.N. Ingle, E. Nassif, R.L. Brunner, M.E. Wood, J.L. Pater, 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 research and treatment*, 126(2):453–461, 2011.

[201] V. Assi, J. Warwick, J. Cuzick, and S.W. Duffy. Clinical and epidemiological issues in mammographic density. *Nature reviews Clinical oncology*, 9(1):33–40, 2011.

[202] A.M. Brewster, G.N. Hortobagyi, K.R. Broglio, S. Kau, C.A. Santa-Maria, B. Arun, A.U. Buzdar, D.J. Booser, V. Valero, M. Bondy, et al. Residual risk of breast cancer recurrence 5 years after adjuvant therapy. *Journal of the National Cancer Institute*, 100(16):1179–1183, 2008.

[203] G.J.R. Porter, A.J. Evans, E.J. Cornford, H.C. Burrell, J.J. James, A.H.S. Lee, and J. Chakrabarti. Influence of mammographic parenchymal pattern in screening-detected and interval invasive breast cancers on pathologic features, mammographic features, and patient survival. *American Journal of Roentgenology*, 188 (3):676–683, 2007.

[204] T. Cil, E. Fishell, W. Hanna, P. Sun, E. Rawlinson, S.A. Narod, and D.R. McCready. Mammographic density and the risk of breast cancer recurrence after breast-conserving surgery. *Cancer*, 115(24):5780–5787, 2009.

[205] T.M. Prowell, A.L. Blackford, C. Byrne, N.F. Khouri, M. Dowsett, E. Folkerd, K.S. Tarpinian, P.P. Powers, L.A. Wright, M.G. Donehower, et al. Changes in breast density and circulating estrogens in postmenopausal women receiving adjuvant anastrozole. *Cancer Prevention Research*, 4(12):1993–2001, 2011.

[206] K. Kerlikowske and A.I. Phipps. Breast density influences tumor subtypes and tumor aggressiveness. *Journal of the National Cancer Institute*, 103(15):1143–1145, 2011.

[207] K. Ghosh, K.R. Brandt, T.A. Sellers, C. Reynolds, C.G. Scott, S.D. Maloney, M.J. Carston, V.S. Pankratz, and C.M. Vachon. Association of mammographic density with the pathology of subsequent breast cancer among postmenopausal women. *Cancer Epidemiology Biomarkers & Prevention*, 17(4):872–879, 2008.

[208] J. Ding, R. Warren, A. Girling, D. Thompson, and D. Easton. Mammographic density, estrogen receptor status and other breast cancer tumor characteristics. *The breast journal*, 16(3):279–289, 2010.

[209] S.M. Conroy, I. Pagano, L.N. Kolonel, and G. Maskarinec. Mammographic density and hormone receptor expression in breast cancer: The multiethnic cohort study. *Cancer Epidemiology*, 35(5):448–452, 2011.

[210] A.I. Phipps, D.S.M. Buist, K.E. Malone, W.E. Barlow, P.L. Porter, K. Kerlikowske, E.S. O'Meara, and C.I. Li. Breast density, body mass index, and risk of tumor marker-defined subtypes of breast cancer. *Annals of epidemiology*, 22(5): 340–348, 2012.

[211] K. Heusinger, S.M. Jud, L. Häberle, C.C. Hack, B.R. Adamietz, M. Meier-Meitinger, M.P. Lux, T. Wittenberg, F. Wagner, C.R. Loehberg, et al. Association of mammographic density with hormone receptors in invasive breast cancers: Results from a case-only study. *International Journal of Cancer*, 131(11):2643–2649, 2012.

[212] M. Dowsett, C. Allred, J. Knox, E. Quinn, J. Salter, C. Wale, J. Cuzick, J. Houghton, N. Williams, E. Mallon, et al. Relationship between quantitative estrogen and progesterone receptor expression and human epidermal growth factor receptor 2 (her-2) status with recurrence in the arimidex, tamoxifen, alone or in combination trial. *Journal of clinical oncology*, 26(7):1059–1065, 2008.

[213] J. Cuzick, M. Dowsett, S. Pineda, C. Wale, J. Salter, E. Quinn, L. Zabaglo, E. Mallon, A.R. Green, I.O. Ellis, 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. *Journal of Clinical Oncology*, 29(32):4273–4278, 2011.

[214] J. Cuzick. A wilcoxon-type test for trend. *Statistics in medicine*, 4(4):543–547, 1985.

[215] M. Dowsett, I.E. Smith, S.R. Ebbs, J.M. Dixon, A. Skene, C. Griffith, I. Boeddinghaus, J. Salter, S. Detre, M. Hills, et al. Short-term changes in ki-67 during

neoadjuvant treatment of primary breast cancer with anastrozole or tamoxifen alone or combined correlate with recurrence-free survival. *Clinical Cancer Research*, 11(2):951s–958s, 2005.

[216] S.J. Vinnicombe, A.D. MacVicar, R.L. Guy, J.P. Sloane, T.J. Powles, G. Knee, and J.E. Husband. Primary breast cancer: mammographic changes after neoadjuvant chemotherapy, with pathologic correlation. *Radiology*, 198(2):333–340, 1996.

[217] J.A. Harvey, R.J. Santen, G.R. Petroni, V.E. Bovbjerg, M.E. Smolkin, F.S. Sheriff, and J. Russo. Histologic changes in the breast with menopausal hormone therapy use: correlation with breast density, estrogen receptor, progesterone receptor, and proliferation indices. *Menopause*, 15(1):67–73, 2008.

[218] J.F. de Souza Sales Jr, C. Cabello, M. Alvarenga, R. Zocchio Torresan, and G. Mendes Duarte. Nonproliferative epithelial alteration and expression of estrogen receptor and ki67 in the contralateral breast of women treated with tamoxifen for breast cancer. *The Breast*, 16(2):197–203, 2007.

[219] J.S. Mandelblatt, K.A. Cronin, S. Bailey, D.A. Berry, H.J. de Koning, G. Draisma, H. Huang, S.J. Lee, M. Munsell, S.K. Plevritis, et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Annals of internal medicine*, 151(10):738–747, 2009.

[220] A. Maurice, D.G. Evans, J. Affen, R. Greenhalgh, S.W. Duffy, and A. Howell. Surveillance of women at increased risk of breast cancer using mammography and clinical breast examination: Further evidence of benefit. *International Journal of Cancer*, 131(2):417–425, 2012.

[221] P.D.P. Pharoah, N.E. Day, S. Duffy, D.F. Easton, and B.A.J. Ponder. Family history and the risk of breast cancer: A systematic review and meta-analysis. *International Journal of Cancer*, 71(5):800–809, 1997.

[222] R. Warren. Hormones and mammographic breast density. *Maturitas*, 49(1):67–78, 2004.

[223] C.M. Checka, J.E. Chun, F.R. Schnabel, J. Lee, and H. Toth. The relationship of mammographic density and age: implications for breast cancer screening. *American Journal of Roentgenology*, 198(3):W292–W295, 2012.

[224] R. Gabe, S.W. Duffy, J. Mackay, E. Anderson, S. Duffy, T. Ellis, G. Evans, H. Fielder, J. Gray, G. Gui, et al. The challenge of evaluating annual mammography screening for young women with a family history of breast cancer. *JOURNAL OF MEDICAL SCREENING*, 13(4):177–182, 2006.

[225] N.F. Boyd, G.A. Lockwood, J.W. Byng, L.E. Little, M.J. Yaffe, and D.L. Tritchler. The relationship of anthropometric measures to radiological features of the breast in premenopausal women. *British journal of Cancer*, 78(9):1233, 1998.

[226] G. Haars, P.A.H. van Noord, C.H. van Gils, D.E. Grobbee, and P.H.M. Peeters. Measurements of breast density: no ratio for a ratio. *Cancer Epidemiology Biomarkers & Prevention*, 14(11):2634–2640, 2005.

[227] J. Stone, R.M.L. Warren, E. Pinney, J. Warwick, and J. Cuzick. Determinants of percentage and area measures of mammographic density. *American journal of epidemiology*, 170(12):1571–1578, 2009.

[228] K. Kerlikowske, D. Grady, J. Barclay, E.A. Sickles, and V. Ernster. Effect of age, breast density, and family history on the sensitivity of first screening mammography. *Jama*, 276(1):33–38, 1996.

[229] B.L. De Stavola, I.H. Gravelle, D.Y. Wang, D.S. Allen, R.D. Bulbrook, I.S. Fentiman, J.L. Hayward, and M.C. Chaudary. Relationship of mammographic parenchymal patterns with breast cancer risk factors and risk of breast cancer in a prospective study. *International journal of epidemiology*, 19(2):247–254, 1990.

[230] G. Maskarinec, I. Pagano, G. Lurie, and L.N. Kolonel. A longitudinal investigation of mammographic density: the multiethnic cohort. *Cancer Epidemiology Biomarkers & Prevention*, 15(4):732–739, 2006.

[231] C.G. Woolcott, K. Koga, S.M. Conroy, C. Byrne, C. Nagata, G. Ursin, C.M. Vachon, M.J. Yaffe, I. Pagano, and G. Maskarinec. Mammographic density, parity and age at first birth, and risk of breast cancer: an analysis of four case–control studies. *Breast cancer research and treatment*, 132(3):1163–1171, 2012.

[232] A.Y. El-Bastawissi, E. White, M.T. Mandelson, and S.H. Taplin. Reproductive and hormonal factors associated with mammographic breast density by age (united states). *Cancer Causes & Control*, 11(10):955–963, 2000.

[233] J. Warwick, E. Pinney, R.M.L. Warren, S.W. Duffy, A. Howell, M. Wilson, and J. Cuzick. Breast density and breast cancer risk factors in a high-risk population. *The breast*, 12(1):10–16, 2003.

[234] American College of Radiology (ACR). Illustrated breast imaging reporting and data system (bi-rads), 4th edn.

[235] S. Obenauer, K.P. Hermann, and E. Grabbe. Applications and literature review of the bi-rads classification. *European radiology*, 15(5):1027–1036, 2005.

[236] J. Diffey, A. Hufton, and S. Astley. A new step-wedge for the volumetric measurement of mammographic density. *Digital Mammography*, pages 1–9, 2006.

[237] P.L. Skippage, L.S. Wilkinson, S.D. Allen, N. Roche, M. Dowsett, and R. AHern. Correlation of age and hrt with breast density as assessed by quantra. In *Royal College of Radiology Breast Group, Belfast, UK*, 2009.

[238] P.L. Skippage, L.S. Wilkinson, S.D. Allen, N. Roche, M. Dowsett, and R. AHern. Measuring breast density using quantraon full field digital mammography. In *Royal College of Radiology Breast Group, Belfast, UK*, 2009.

[239] P.L. Skippage, L.S. Wilkinson, S.D. Allen, N. Roche, M. Dowsett, and R. AHern. Correlation of age and hrt with breast density as assessed by quantra. In *RSNA 2009, Chicago, IL, USA*, 2009.

[240] K. Pinker, N. Perry, S. Milner, K. Mokbel, and S. Duffy. Validation of a new automated volumetric breast density measurement system as a marker of breast cancer risk. In *RSNA 2009, Chicago, IL, USA*, 2009.

[241] E. Rafferty, A. Smith, and L. Niklason. Comparison of three methods of estimating breast density: Bi-rads density scores using full field digital mammography, bi-rads density scores using breast tomosynthesis, and volumetric breast density. In *RSNA 2009, Chicago, IL, USA*, 2009.

[242] N. Tuncbilek, A. Sezer, U. Uğur, U. Ulaş, and ʻOzdemir H. Urut, U. Coşar R. Qualitative and quantitative analysis of fibroglandular tissue in the digital environment. In *10th National Congress of Breast Diseases, Izmir, Turkey*, 2009.

[243] N. Perry, S. Milner, K. Mokbel, S. Duffy, and K. Pinker. A new automated volumetric breast density measurement system confirms higher breast density associated with urban women. In *European Congress of Radiology, Vienna, Austria*, 2010.

[244] D. Tzias, S. George, L. Wilkinson, R. Mehta, C. Lobo, A. Hainsworth, and A. Sharma. Correlation of ethnicity with breast density as assessed by quantra. *Breast Cancer Research*, 13(Suppl 1):O5, 2011.

[245] J. Ding, R. Warren, I. Warsi, N. Day, D. Thompson, M. Brady, C. Tromans, R. Highnam, and D. Easton. Evaluating the effectiveness of using standard mammogram form to predict breast cancer risk: case-control study. *Cancer Epidemiology Biomarkers & Prevention*, 17(5):1074–1081, 2008.

[246] N.M. Perry, P.C. Allgood, S.E. Milner, K. Mokbel, and S.W. Duffy. Mammographic breast density by area of residence: possible evidence of higher density in urban areas. *Current Medical Research and Opinion®*, 24(2):365–368, 2007.

[247] J.A. Harvey, C.C. Gard, D.L. Miglioretti, B.C. Yankaskas, K. Kerlikowske, D.S.M. Buist, B.A. Geller, T.L. Onega, et al. Reported mammographic density: Film-screen versus digital acquisition. *Radiology*, 266(3):752–758, 2013.

[248] J. Wang, A. Aziz, D. Newitt, B. N. Joe, N. Hylton, and J. A. Shepherd. Comparison of hologics quantra volumetric assessment to mri breast density. In *Breast Imaging*, pages 619–626. Springer, 2012.

[249] S. Scholes, A. Prescott, and M. Bajekal. Health and lifestyle indicators for strategic health authorities, 1994-2002. *Available at Department of Health website as:*
http://www. dh. gov. uk/PublicationsAndStatistics/Publications/PublicationsStatisti cs/PublicationsStatisticsArticle/fs/en, 2004.

[250] K. Kerlikowske, D. Grady, J. Barclay, V. Ernster, S.D. Frankel, S.H. Ominsky, and E.A. Sickles. Variability and accuracy in mammographic interpretation using the american college of radiology breast imaging reporting and data system. *Journal of the National Cancer Institute*, 90(23):1801–1809, 1998.

[251] C. Balleyguier, S. Ayadi, K. Van Nguyen, D. Vanel, C. Dromain, and R. Sigal. Birads classification in mammography. *European journal of radiology*, 61(2):192–194, 2007.

[252] D. Graham-Rowe. Risk analysis: A dense issue. *Nature*, 485(7400):S60–S61, 2012.

[253] Matakina Technology Limited. Volpara, July 2013. URL http://www.volparadensity.com/.

[254] A.F. Saftlas, R.N. Hoover, L.A. Brinton, M. Szklo, D.R. Olson, M. Salane, and J.N. Wolfe. Mammographic densities and risk of breast cancer. *Cancer*, 67(11): 2833–2838, 1991.

[255] C. Byrne. Studying mammographic density: implications for understanding breast cancer. *Journal of the National Cancer Institute*, 89(8):531–532, 1997.

[256] J.W. Byng, N.F. Boyd, E. Fishell, R.A. Jong, and M.J. Yaffe. The quantitative analysis of mammographic densities. *Physics in medicine and biology*, 39(10):1629, 1999.

[257] B. Rosner, WC Willett, and D. Spiegelman. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in medicine*, 8(9):1051–1069, 1989.

[258] B. Rosner, D. Spiegelman, and W.C. Willett. Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. *American journal of epidemiology*, 136(11):1400–1413, 1992.

[259] D. Spiegelman, A. McDermott, and B. Rosner. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *The American journal of clinical nutrition*, 65(4):1179S–1186S, 1997.

[260] J.M. Bland and D.G. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, 327(8476):307–310, 1986.

[261] L.I. Lin. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45(1):255–268, 1989.

[262] L.I. Lin. A note on the concordance correlation coefficient. *Biometrics*, 56:324–325, 2000.

[263] R. Müller and P. Büttner. A critical discussion of intraclass correlation coefficients. *Statistics in medicine*, 13(23-24):2465–2476, 1994.

[264] J.J. Heine, K. Cao, and J.A. Thomas. Effective radiation attenuation calibration for breast density: compression thickness influences and correction. *Biomedical engineering online*, 9(1):73, 2010.

[265] M.G.J. Kallenberg, C.H. van Gils, M. Lokate, G.J. den Heeten, and N. Karssemeijer. Effect of compression paddle tilt correction on volumetric breast density estimation. *Physics in Medicine and Biology*, 57(16):5155, 2012.

[266] J. Stone, J. Ding, R.M.L. Warren, and S.W. Duffy. Predicting breast cancer risk using mammographic density measurements from both mammogram sides and views. *Breast cancer research and treatment*, 124(2):551–554, 2010.

[267] O.C. Ukoumunne. A comparison of confidence interval methods for the intraclass correlation coefficient in cluster randomized trials. *Statistics in medicine*, 21(24): 3757–3774, 2002.

[268] H.J. Shin, H.H. Kim, J.H. Cha, J.H. Park, K.E. Lee, and J.H. Kim. Automated ultrasound of the breast for diagnosis: interobserver agreement on lesion detection and characterization. *American Journal of Roentgenology*, 197(3):747–754, 2011.

[269] D. Bernardi, S. Ciatto, M. Pellegrini, V. Anesi, S. Burlon, E. Cauli, M. Depaoli, L. Larentis, V. Malesani, L. Targa, et al. Application of breast tomosynthesis in screening: incremental effect on mammography acquisition and reading time. *British Journal of Radiology*, 85(1020):e1174–e1178, 2012.

[270] N. Duric, N. Boyd, P. Littrup, M. Sak, L. Myc, C. Li, E. West, S. Minkin, L. Martin, M. Yaffe, et al. Breast density measurements with ultrasound tomography: A comparison with film and digital mammography. *Medical physics*, 40(1):013501, 2013.

[271] S.A. Bashir and S.W. Duffy. The correction of risk estimates for measurement error. *Annals of Epidemiology*, 7(2):154, 1997.

[272] Y. Guo, R.J. Little, and D.S. McConnell. On using summary statistics from an external calibration sample to correct for covariate measurement error. *Epidemiology*, 23(1):165, 2012.

[273] L.A. Stefanski and R.J. Carroll. Covariate measurement error in logistic regression. *The Annals of Statistics*, 13(4):1335–1351, 1985.

[274] F.J. Gilbert, S.M. Astley, M.A. McGee, M.G.C. Gillan, C.R.M. Boggis, P.M. Griffiths, and S.W. Duffy. Single reading with computer-aided detection and double reading of screening mammograms in the united kingdom national breast screening program1. *Radiology*, 241(1):47–53, 2006.

[275] J.J. Ding. Case-control study on the effectiveness of using standard mammogram form to predict breast cancer risk. PhD Thesis. *University of Cambridge*, page 164, 2007.

[276] M.G.J. Kallenberg, J. Holland, J.O.P. Wanders, G.H. van Gils, and N. Karssemeijer. Association between automated, volumetric breast density measures and breast cancer in a large screening population. In *Proceedings of the 6$^{\sharp}$h International Workshop on Breast Densitometry and Breast Cancer Risk Assessment, San Francisco, CA, USA*, 2013.

[277] B.C. Yankaskas, S. Haneuse, J.M. Kapp, K. Kerlikowske, B. Geller, D.S.M. Buist, et al. Performance of first mammography examination in women younger than 40 years. *Journal of the National Cancer Institute*, 102(10):692–701, 2010.

[278] J. Cuzick, I. Sestak, D. Cella, and L. Fallowfield. Treatment-emergent endocrine symptoms and the risk of breast cancer recurrence: a retrospective analysis of the atac trial. *The lancet oncology*, 9(12):1143–1148, 2008.

[279] P. Skaane, A.I. Bandos, R. Gullien, E.B. Eben, U. Ekseth, U. Haakenaasen, M. Izadi, I.N. Jebsen, G. Jahr, M. Krager, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology*, 267(1):47–56, 2013.

[280] S. Ciatto, N. Houssami, D. Bernardi, F. Caumo, M. Pellegrini, S. Brunelli, P. Tuttobene, P. Bricolo, C. Fantò, M. Valentini, et al. Integration of 3d digital mammography with tomosynthesis for population breast-cancer screening (storm): a prospective comparison study. *The lancet oncology*, 14(7):583–589, 2013.