Beyond breast density - Novel uses of automated mammographic analysis in breast cancer screening

Susan M Hudson

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy of the University of London

September 2023

Department of Non-communicable Disease Epidemiology

Faculty of Epidemiology and Population Health

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

No funding received

DECLARATION OF OWN WORK

I, Susan Hudson, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Susan Hudson

September 2023

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

DECLARATION OF OWN WORK

All students are required to complete the following declaration when submitting their thesis.

Please note: Assessment misconduct includes any activity that compromises the integrity of your research or assessment of is will be considered under the Assessment Irregularity Policy. This includes plagiarism, cheating and failure to follow correct progression and examination procedures.

Please see the following documents for further guidance:

- Research Degrees Handbook
- Assessment Irregularities Policy

Supervisors should be consulted if there are any doubts about what is permissible.

1. STUDENT DETAILS

Student ID Number	12406683	**Title**	Ms
First Name(s)	Susan M		
Surname/Family Name	Hudson		
Programme of Study	PhD		
LSHTM Email (if this is no longer active, please provide an alternative)	Susan.Hudson@lshtm.ac.uk; sue.hudson@pasconsulting.co.uk		

2. TITLE OF THESIS

Title of Thesis	Beyond breast density - Novel uses of automated mammographic analysis in breast cancer screening.

3. DECLARATION

I have read and understood the LSHTM's definition of plagiarism and cheating. I declare that this thesis is my own work, and that I have acknowledged all results and quotations from the published or unpublished work of other people.

I have read and understood the LSHTM's definition and policy on the use of third parties (either paid or unpaid) who have contributed to the preparation of this thesis by providing copy editing and, or, proof reading services. I declare that no changes to the intellectual content or substance of this thesis were made as a result of this advice, and, that I have fully acknowledged all such contributions.

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 or infringe any third party's copyright or other intellectual property right.

Student Signature	[redacted]
Date	

ABSTRACT

Introduction: The global burden of breast cancer is increasing, however secondary prevention via population-based mammographic screening has proven effective in reducing mortality in high income countries. Concomitant with screening are various drawbacks including over diagnosis and false positives, hence new methods for improving screening performance are vital if the balance of "benefit versus harm" is to be improved. The emergence of automated tools for large-scale objective image analysis offers new possibilities. The overall aim of this thesis was to look at three ways that such tools might be used for improving breast screening performance.

Objectives: (1) To assess whether automatically estimated volumetric breast measurements could be used as a proxy for BMI (a confounder in the association between breast density and breast cancer risk) in screening settings where BMI data are not available. (2) To use a novel, automatically estimated measurement of left breast versus right breast fluctuating asymmetry (FA), in breast volume and mammographic density, to determine whether FA is associated with cancer detection at screening or the occurrence of cancer in the interval between screens (i.e. interval cancers). (3) To assess whether variations in, objectively measured, mammographic compression force, pressure and paddle tilt are associated with screening performance.

Methods: For objective (1) data from a previously-conducted UK case-control study (414 cases/685 controls) and a Norwegian cohort study (657 cases/61,059 non-cases) were pooled using fixed-effect models (Study I). For objectives (2) and (3) four studies (II-V) were designed, requiring data collation and image analysis of over 90,000 screens (from which 904 cancer cases were detected) at a UK population screening programme. Cross-sectional designs were used for assessing the association between relevant exposures and breast cancer detection at screening, and nested case-control designs for assessing the association of exposures with the occurrence of an interval cancer.

Results: 1) Study I confirmed that, in screening age women, non-dense breast volume is strongly and positively correlated with BMI ($r=0.74$, $p<0.0001$) and it showed empirically for the first time, that using breast volume estimates in place of BMI leads to minimal difference in the association between % mammographic density and breast cancer risk (pooled RR 1.51(95%CI 1.41-1.61) in both cases). 2) Studies II and III showed that mammographic density FA is common and that women with highest FA were more likely to be diagnosed with cancer at screening (OR=1.26 (95% CI 1.07-2.27) for top versus bottom third of the distribution; P for linear trend=0.012). Similarly, women in the top third of FA were at higher risk of interval cancer (OR=1.68 (95% CI 0.97-2.92)). 3) Studies IV and V found that breast compression thickness decreased with increased compression force and pressure, but that increasing pressure above ~15kPa resulted in minimal further reduction. Compression pressure was negatively associated with detection of cancer at screening (OR=0.74 (95% CI 0.60- 0.92) for the top versus bottom third of the pressure distribution).

Conclusions: These findings showed that automated mammogram analysis tools have the potential to be used in novel ways in breast cancer screening and, more widely, in breast cancer risk assessment even in high-volume screening settings where it is not feasible to routinely collect BMI data. FA, a novel automated measurement, may help to identify women at higher breast cancer risk and those more likely to have an interval cancer. Finally, my thesis challenges the view that using 'as much force as tolerated' during mammography is the best strategy and suggests that there are more subtle associations between breast compression technique and screening outcomes.

ACKNOWLEDGEMENTS

My heartfelt gratitude goes to all who have supported me during this odyssey, which started in a little office in Tooting where Dr Louise Wilkinson and I spent many hours enthusing over ideas and projects and realising that as ever there are far more questions than answers. Louise you are a star as everyone knows who has been lucky enough to work with you. My thanks also to Dr Sam Heller who offered important professional insights and Drs Ralph Highnam and Chris Tromans at Matariki for their support and guidance in setting up the Volpara package at SWLBSS and for answering my annoying questions.

A huge 'thank you' is reserved for my supervisor Professor Isabel dos Santos Silva for taking me on, despite my age, and who has taught me so much. I have felt privileged to have you as my supervisor and I know that you have found time for me when you really have none. Likewise, to Professor Bianca de Stavola who agreed to join my supervisory team and guide me through the more complex statistical issues with patience and humour. My respect for you both is unbounded. My appreciation also goes to everyone who has helped on the way including the lovely team at SWLBSS, which became my second home for a while.

Thank you to my all my friends but especially Sam who was there from the beginning, supported my endeavours, was interested enough to hear the technical details and made me go for it. Finally, my eternal love and thanks to my family: Wolter, Tom, Helen, and Laura for unerringly and unnervingly taking it as read that I would get there in the end. Thank you to my lovely, intelligent sisters and my Dad, for their encouragement and to my Mum who sadly missed the end of the journey but who would have definitely made a better job of it than me had she born in a different generation.

I hope that the papers that we have published for this thesis[1] provide a couple more 'bricks in the knowledge wall' for breast cancer screening and that the skills and knowledge that I have acquired keep me useful for a few more years. Meanwhile I will keep on asking those annoying questions because you are never too old to learn.

[1] For the title of this thesis, I acknowledge the paper by Gastounioti et al which adopted a similar title (1).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

TABLE OF ABBREVIATIONS

Acronym and abbreviation	Definition
ANOVA	Analysis of variance
AUC	Area under the Receiver Operator Characteristic curve
BC	Breast Cancer
BD	Breast Density (as a general concept)
%BD	Percent breast density = Breast dense volume (or area) / Total breast volume (or area)
BF%	Body fat %. Relative proportion of the body mass that is made up of fat tissue.
BMI	Body mass index (weight kg/height m^2)
iBMI	Inverted BMI
BV	Breast Volume cm^3
CC	Cranio-caudal mammogram view
CELBSS	Central and East London breast screening service
CI	Confidence Interval
CGHFBC	Collaborative Group on Hormonal Factors in Breast Cancer
CT	Computerised tomography
DBT	Digital breast tomosynthesis
DCIS	Ductal carcinoma in-situ
DICOM	Digital Imaging and Communications in Medicine (an agreed protocol for information exchange in medical imaging)
DV	Dense volume i.e. absolute volume of dense tissue
ER+ ER-	Estrogen receptor status
FA	Fluctuating Asymmetry
FGV	Fibro-glandular Tissue Volume cm^3. Stromal and epithelial tissue
FN	False negative
HIC	High Income Countries
IARC	International Agency for Research on Cancer (WHO)
IC	Interval cancer
ICC	Intra-class correlation coefficient
IMD	Index of Multiple Deprivation
IQR	Interquartile range
LMIC	Low and Middle Income Countries
LOWESS	Locally Weighted Scatterplot Smoothing
MET	Metabolic equivalent of task
MGD	Mean glandular dose of radiation in mGy
MHT	Menopausal Hormone Treatment
mGy	Milligray. A unit of absorbed radiation equal to 0.001 gray. 1 gray is the dose of 1 joule of energy absorbed per kilogram of matter
MLO	Mediolateral oblique mammogram view

MRI	Magnetic resonance imaging
%MD	Percentage mammographic breast density
NBSS	National Breast Screening System (Computer System)
NDV	Non-dense volume cm^3
NHS	National Health Service
NHSBSP	England and Wales National Health Service Breast Screening Programme
OR	Odds ratio
OPERA	Odds per adjusted standard deviation
PACS	Picture Archiving and Communication system
PAF	Population Attributable Fraction
PPV	Positive predictive value
PR + PR-	Progesterone receptor status
RR	Relative risk
SD	Standard deviation
SGH	St George's University Hospitals NHS Foundation Trust
SWLBSS	South West London Breast Screening Service
Volpara	Volpara® Density™ Trademark of Matakina Technology
VBD	Volumetric Breast Density (VBD) – From Volpara
VDG	Volumetric Density Grade (VDG) – From Volpara
WHO	World Health Organisation

STRUCTURE OF THIS THESIS

Overview

Chapter 1 gives an overview of the background and scope of this thesis, looking at the burden of breast cancer (BC) and the main factors that determine who is most at risk. It covers the strategies for BC prevention focusing on mammographic screening with its benefits and harms.

Chapter 2 provides a literature review covering specific factors relevant to my study aims, that make mammographic screening less effective and potentially more harmful.

The aims of this thesis are laid out in Chapter 3 'Research question' together with a summary of the rationale behind these aims.

Chapter 4 covers methods, including the data collection that was used for this thesis. It also provides the justification for the automated tool selection.

Chapters 5, 6 and 7 contain the five publications that make up the main body of the work.

Chapter 8 provides an overarching discussion of the body of work and summarises the overall conclusions.

Research publications

Paper I ***-*** *Adjusting for BMI in analyses of volumetric mammographic density and breast cancer risk*

Sue Hudson, Kirsti Vik Hjerkind, Sarah Vinnicombe, Steve Allen, Cassia Trewin, Giske Ursin, Isabel Dos-Santos-Silva, Bianca L De Stavola

Breast Cancer Research December 2018

Paper II ***-*** *Ethnic and age differences in right-left breast asymmetry in a large population-based screening population*

Sue M Hudson, Louise S Wilkinson, Rachel Denholm, Bianca L De Stavola, Isabel Dos-Santos-Silva

British Journal of Radiology October 2019

Paper III ***-*** *Left-right breast asymmetry and risk of screen-detected and interval cancers in a large population-based screening population*

Sue M Hudson, Louise S Wilkinson, Bianca L De Stavola, Isabel Dos-Santos-Silva

British Journal of Radiology August 2020

Paper IV - *To what extent are objectively measured mammographic imaging techniques associated with compression outcomes*

Sue M Hudson, Louise S Wilkinson, Bianca L De Stavola, Isabel Dos-Santos-Silva

British Journal of Radiology March 2023

Paper V – *Are mammography image acquisition factors, compression pressure and paddle tilt, associated with breast cancer detection in a large population-based screening programme?*

Sue M Hudson, Louise S Wilkinson, Bianca L De Stavola, Isabel Dos-Santos-Silva

British Journal of Radiology June 2023

1 CHAPTER 1 INTRODUCTION

In this chapter I start by looking at the general burden of breast cancer (BC) and the main factors that determine who is most at risk. Next, I give a broad overview of strategies for BC prevention and then move on to focus on mammographic breast cancer screening as a key strategy in high income countries. This will entail a review of the balance of benefits and harms that arise from population-based mammographic screening programmes based on findings from the most relevant systematic reviews on this topic.

1.1 Breast cancer incidence and mortality

The global burden of breast cancer (BC) is high, with BC the most frequently diagnosed cancer and the most common cause of death from cancer in women worldwide, accounting for an estimated 2.3 million new cases and 690,000 deaths in 2020 (2). BC incidence has increased over the last decades (Figure 1.1) and Sung et al in the review of Global Cancer Statistics for 2020 concluded that the upward tendency is set to continue slowly in N. America, Europe and Oceania and more rapidly in less developed countries and in high-income Asian countries (2).

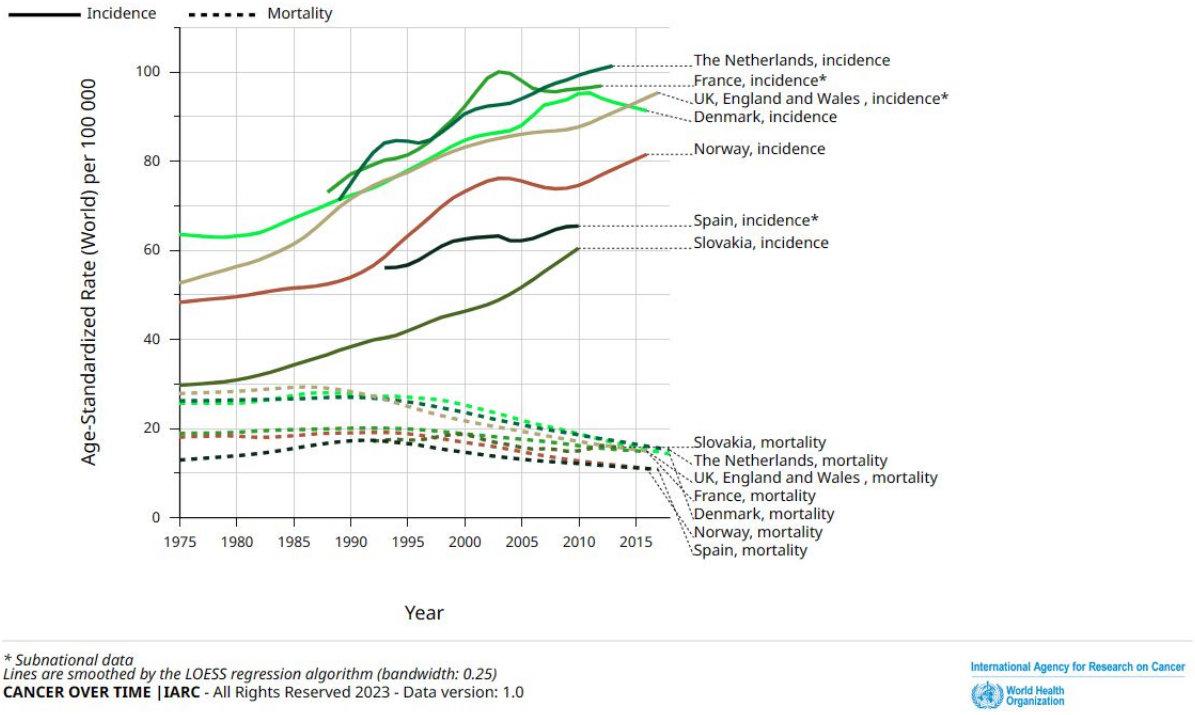


Figure 1.1 Trends in incidence of female BC and BC mortality in selected European countries : age-standardised rate (World standard population) per 100,000 woman-years(3).

More timely detection and more effective treatments have improved survival rates in High Income Countries (HIC) and consequently mortality rates have decreased markedly since the early 1990s.

However, in most less developed countries mortality from breast cancer is still rising, reflecting increases in the incidence of the disease as well as lack of improvements in survival which in most low-resource settings has remained poor. The increase in incidence rates can partly be explained by increased prevalence of risk factors such as late age at first pregnancy, lower parity, increases in the use of exogenous sex hormones (i.e. oral contraceptives and menopausal hormone therapy) and increases in body mass index (BMI) and, in developed countries, also by an increase in the number of cancers detected through organised screening, a number of which would have remained undiagnosed in the absence of screening. The age-adjusted incidence rates in the UK and other developed countries are expected to remain relatively stable or increase very slowly over the coming years (4), but with an aging population in the UK the absolute numbers of cancers diagnosed in the screening age range (50-70) is set to increase up to the mid 2040's (5).

1.1 Breast cancer risk factors

Sex and age are the two most important risk factors with over 99% of BC being diagnosed in women and 81% of cancers in the UK in 2011 being diagnosed in women aged 50 and over (6). BC risk doubles approximately every 10 years up to the age of 50 and increases at a slightly reduced rate after this (7). The aetiology of BC is generally well-understood, and it has become clear that BC is a multi-factorial disease with different risk factors acting, and their effects accumulating and interacting at different stages throughout a woman's lifetime, as illustrated in Figure 1.2. It is not the purpose of this thesis to review each risk factor in depth but in this section, I summarise the key BC risk factors in addition to age and sex, using data from a number of sources including predominantly large cohort studies, systematic reviews with meta-analyses of published data, and pooled analyses of individual-level data from studies.

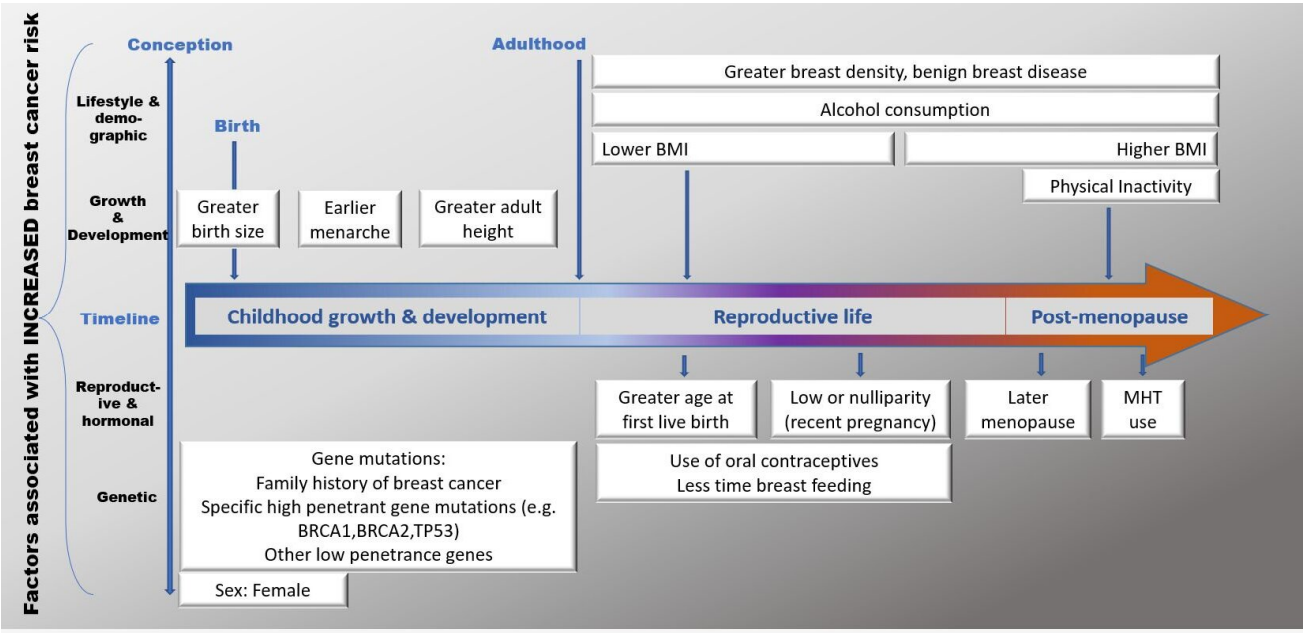


Figure 1.2 Main breast cancer risk factors acting cumulatively throughout a woman's lifetime

Early-life risk factors

Many factors affect a woman's risk of BC at different stages of her life, some factors even acting in utero e.g. there is some evidence that BC risk is elevated in women who were exposed in utero to diethylstilboestrol (DES) given to their mothers to prevent pregnancy complications (8). The Collaborative Group on Pre-Natal Risk Factors and Subsequent Risk of Breast Cancer (CGPFBC) reanalysed the pooled individual data and concluded that both birth weight and birth length were positively associated with BC risk (9). Further pooled data studies have shown that childhood development is associated with BC risk, with increased BC risk associated with earlier age of menarche (10) and greater attained adult height, which is largely determined by growth up to puberty (11) .

Table 1.1 Summary of early life factors

Risk Factor	**Study type**	**Estimated magnitude of relative BC risk**
Birth size (weight and length)	Pooled analysis of individual level data from 32 studies (22,058 cases) by the CGPFBC. Studies based on birth records rather than recalled measurements.	RR per one SD [=0.5 kg] increment in birth weight: 1.06; 95% CI 1.02–1.09) (9). RR per one SD increment in birth length: 1.06 [95% CI 1.03–1.10] and 1.09 [95% CI 1.03–1.15].
Age at menarche	CGHFBC pooled analysis of individual-level data from 117 epidemiological studies, including 118,964 women with invasive breast cancer and 306,091 controls (none of whom had used MHT)	RR increased by 1·050 (95% CI 1·044–1·057; p<0·0001) for every year younger at menarche (10).
Adult height	Study using pooled data from seven American and European prospective cohort studies (337,819 women and 4,385 incident invasive BC cases). Controlled for diet and reproductive factors.	Height is an independent risk factor for postmenopausal BC; pooled RR for women over 1.75m compared to <1.60m RR 1.22 (95% CI 0.94 – 1.76) P-trend <0.001. In premenopausal women evidence is less clear (11).

Abbreviations: BC, breast cancer; CI, confidence interval; CGPFBC Collaborative Group on Pre-Natal Risk Factors and Subsequent Risk of Breast Cancer; CGHFBC Collaborative Group on Hormonal Factors in Breast Cancer; MHT, menopausal hormone treatment; RR, relative risk; SD Standard deviation.

Reproductive and hormonal BC risk factors

A large number of BC risk factors are related to hormonal and reproductive history and the evidence has been weighed by a wide range of experts working in the Collaborative Group on Hormonal

Factors in Breast Cancer (CGHFBC). This group, set up in 1992, conduct pooled-analyses of individual-level data from most of the studies conducted worldwide on risk factors for breast cancer in women. Whilst initially focused on reproductive and hormonal factors, they also consider BMI, physical activity levels, and tobacco and alcohol use. A list of the main hormonal and reproductive risk factors together with the risk estimated by the CGHFBC and other large-scale reviews/cohort studies as specified is given below to provide an indication of the relative magnitude of each risk:

Table 1.2 Summary of hormonal and reproductive BC risk factors

Risk Factor	**Study type**	**Estimated magnitude of relative BC risk**
Parity	CGHFBC pooled analysis of individual-level data on breastfeeding patterns from 47 epidemiological studies in 30 countries. 50,302 women with invasive breast cancer and 96,973 controls.	RR decreased by 7% (5.0-9.0; p<0.0001) for each birth (12).
Breast feeding	As above	RR of breast cancer decreased by 4.3% (95% CI 2.9-5.8; p<0.0001) for every 12 months of breastfeeding (12). This effect was independent of other reproductive and demographic factors.
Age at first live birth	Early studies including a large 1996 Swedish cohort study of 12,782 women suggested that long-term BC risk decreased with earlier age at first birth. This is supported by analysis of the Nurses' Health Study in the USA (91,523 women followed up to age 70) (13). More recent meta-analysis suggests this association is limited to hormone receptor-positive cancers (14).	RR decreased by 13% (95% CI 8-19; P<0.001) for each 5-year decrement in age at first birth (15)
Age at menopause	As above	RR increased by (1·029, 1·025–1·032; p<0·0001), for every year older at menopause. This effect was independent of age at menarche (10).
Oral contraceptive use	CGHFBC pooled analyses of 54 (primarily US and European) studies including 53,297 women with BC and 100,239 without BC.	Combined oral contraceptives are associated with small increased RR of BC during use 1.24 (1.15-1.33). 1-4 years after stopping RR = 1.16 (1.08-1.23) and 5-9 years after stopping RR = 1.07 (1.02-1.13), p<0.01 in all cases). No significant excess risk of having

		breast cancer diagnosed 10 or more years after stopping use (16).
MHT use	A large prospective cohort study called known as the "One Million Women study" was conducted in 2002 in the UK breast screening age group (50-69) and includes 1,038,114 white women, 5,877 South Asian women and 4914 Black women) (17).	Oestrogen only use for 10 years increases risk by 18%. Combined oestrogen–progestogen use for >5 years increases risk by 63% (17).

Abbreviations: BC, breast cancer; BMI, body mass index; CI, confidence interval; CGHFBC Collaborative Group on Hormonal Factors in Breast Cancer; HR, hazard ratio; MHT, menopausal hormone treatment; OR, odds ratio; RR, relative risk

For ethical reasons most of the evidence comes from observational studies rather than from RCTs. Thus, biases including unmeasured and residual confounding cannot be ruled out as potential explanations for the observed associations. However, it is noteworthy that for most reproductive and lifestyle factors the magnitude of the BC risk estimates, have been remarkably consistent across a large number of studies of different designs, conducted in different populations and over distinct period of times. Such studies are likely to have been affected by different sources of biases and confounding structures. The CGHFBC concluded that many of the reproductive risk factors are independent of ethnicity, year of birth and lifestyle factors although some effects such as the associations between menarche, menopause and postmenopausal BC risk were attenuated by BMI (10).

Life-style risk factors

A list of the main lifestyle risk factors together with the risk estimated by the CGHFBC and other large-scale reviews/cohort studies as specified is given below to provide an indication of the relative magnitude of each risk:

Table 1.3 Summary of lifestyle BC risk factors

Risk Factor	**Study type**	**Estimated magnitude of relative BC risk**
Alcohol consumption	Analyses of 58,515 cases and 95,067 controls from 53 studies. Using data gathered by CGHFBC. The findings were corroborated by re-analysis of the One Million Women study (see Table 1.2).	CGHFBC: RR of BC increases by 7.1% (95% CI 5.5 – 8.7%; P<0.00001) for each additional unit (10g) of alcohol consumed on a daily basis (18). Million women study: BC risk increased by 12% (95% CI 9 – 14%; Ptrend < 0.001) for each additional 10g per day (19).
Physical inactivity	Systematic review up to 2017. Found 126 observational cohort studies with	Significant inverse associations for physical activity and post-

	125,900 BC cases. A meta-analysis included 9 studies on vigorous activity, 21 studies that considered recreational activity, 9 studies on occupational activity and 6 on total activity. Studies used different physical activity metrics, so the meta-analyses simply compared high versus low levels. Only 6 studies allowed dose response analysis.	menopausal breast cancer risk were observed. In meta-analysis of the 6 dose response studies, the summary RR was 0.98 (95% CI 0.97–0.99) per 10 metabolic equivalent of task (MET)-hour/week of recreational physical activity (20).
Body mass index	(i) Large prospective cohort of over 5 million UK adults (~9% of the UK population) identified through primary care records. In all, 34,707 BC cases occurred during follow-up from 1987 to 2012. Data available on potential confounders (e.g. age, smoking, alcohol use and diabetes) and on menopausal status as an effect modifier (21). Cox regression models were fitted to examine associations between BMI and risk of site-specific cancers including BC. (ii) A later global systematic review with meta-analysis of data from 31 cohort studies corresponding to 3 million women from America, Europe and Asia.	(i) UK analysis estimated HR per 5/kg/m^2 increase in BMI was 1.05 (95% CI 1.03-1.08) in postmenopausal women and 0.89 (95% CI 0.86-0.94) in premenopausal women. There were no differences between the group who had a history of smoking and never smokers. ~5.1% of post-menopausal cases in the UK could be attributed to obesity (21). ii) The global meta-analysis reported a pooled RR = 1.03 (95% CI: 1.02–1.05) per 1 kg/m^2 increment (22).

Abbreviations: BC, breast cancer; BMI, body mass index; CI, confidence interval; CGHFBC Collaborative Group on Hormonal Factors in Breast Cancer; HR, hazard ratio; MET, metabolic equivalent of task; OR, odds ratio; RR, relative risk

Life-style risk factors are often more difficult to assess reliably and most, such as alcohol consumption, only play a minor part in increasing BC risk. The CGHFBC concluded that ~4% of BC incidence in developed countries was attributable to alcohol consumption and that this figure was much lower (0.6%) in developing countries. Smoking was found to have little or no independent effect on the risk of developing BC (18). Alongside these studies there have been many cohort and case-control studies that have looked for an association between specific foods and nutrients and BC risk, but no consistent or conclusive findings have been reported (23). The difficulties of determining the independent effects of different lifestyle risk factors are illustrated by the case of alcohol and tobacco. A high proportion of the tobacco users also consumed alcohol and once the effects of alcohol were controlled for there was no significant increase in risk associated with tobacco use but it is difficult to adjust properly because the potential confounding variable is subject to measurement error and residual confounding may arise. The most conclusive evidence came from a study of those never exposed to alcohol (never users) where residual confounding could be eliminated. Residual confounding is also an issue for other risk factors such as inactivity and BMI. The meta-analyses of high versus low activity present evidence that higher physical activity is

inversely associated with both risk of pre- and risk of postmenopausal breast cancers across all measures of activity in women of normal BMI but even the largest systematic reviews with meta-analyses acknowledge that difficulties lie in the way that physical activity is measured, the paucity of data that is viable for dose-response analysis and the bias of the study groups towards high income countries (20).

Breast density and BC risk

As long ago as 1976 Wolfe showed that a higher amount of white 'mammographic dense' region on the mammogram was associated with increased risk of cancer diagnosis and 'masking' of cancers (24). This white region represents the epithelial and connective (stromal) tissue in the breast, but breast tumours also appear as white areas on the mammogram. In contrast, as fat is a radiolucent (non-dense) tissue it appears as dark areas on a mammogram (see Figure 2.2). Subsequently many studies on breast tissue composition, as assessed by mammography, have confirmed that breast density (BD) is an important phenotypic biomarker for increased BC risk. In 2006 McCormack et al carried out a systematic review of 42 studies and found that, after adjustment for other risk factors, BD is more strongly associated with BC than most other risk factors. Women in the highest breast density category have 4.6 times the risk of those in the lowest density category (25). Huo et al. (2014) carried out an update to this review which confirmed the findings and Petterson et al. confirmed earlier findings that BD is an independent BC risk factor (26). BD offers more potential for prevention than many other risk factors for several reasons. First, BD is one of the few known risk factors for breast cancer that is potentially modifiable. Second, elevated BD is relatively common. Third, studies have shown that BD tracks through a woman's adult life (27) making it possible to identify high-risk women in young adulthood where preventive measures are likely to be more effective (28). Fourth, it is now logistically feasible to routinely perform valid automated BD assessments to all participants in large screening programmes. Thus, in Chapter 2 we present a wider discussion of BD as a risk factor for BC, the role it plays in tumour masking in mammography, and the potential for BD assessment in improving breast screening performance.

Genetic BC risk factors

In addition to hormonal and reproductive risk factors there is an ever-evolving body of knowledge about the role of genetic factors on the development of BC. The CGHFBC examined data from 52 studies conducted world-wide including 58,209 cases and 101,986 controls, stratifying by age, menopausal status, parity and number of sisters. This pooled analysis found that family history of breast disease in first degree relatives (mothers, sisters, daughters) was one of the strongest risk factors for BC. Compared to women who had no affected first degree relative those with one, two,

and three or more affected first-degree relatives experienced, respectively, an approximate two-fold (risk ratio: 1.80; 99% CI 1.69-1.91), three-fold (2.93; 2.36-3.64) and four-fold (3.90 (2.03-7.49) increase in the risk of developing BC. In comparison to lifestyle and reproductive factors, this represents a relatively strong risk factor, which was found to be independent of reproductive history. To put this into context however, 89% of the cases in the study did *not* have an affected first-degree relative and because the overall incidence of BC is relatively low, most women who have first degree relatives with BC will never be diagnosed with BC themselves.

In the UK, average lifetime risk of BC is estimated to be 13%. A moderate lifetime risk is regarded as 17%-30%, and high lifetime risk is considered by the National Institute for Health Care Excellence (NICE) to be 30% or higher (29, 30). Specific gene mutations (*BRCA1, BRCA2, TP53, PTEN)* are associated with a very high lifetime risk (>30%-90%) (31). Other mutations (CHEK2, ATM, NBS1, RAD50, BRIP1, PALB2) are associated with moderate (17% - 30%) life-time risk. But only around half the cases of familial BC can be explained by mutations in these known genes (32). An increasing number (>200 to date) of common low-penetrance genetic variants, e.g. single-nucleotide polymorphisms (SNPs), have also been identified, each of which is associated with a small increase in BC risk. However, women who carry a large number of these risk alleles will have in a high genetic risk of BC as their effects will be multiplicative. Therefore, there is a possibility that genetic estimates will become ever more refined and the proportion of cases attributable to genetic causes may grow as more SNPs are identified.

Risk factors and breast cancer sub-types

Breast cancer is a highly heterogeneous malignancy with distinct subtypes. These subtypes are commonly grouped into four categories based on the immunohistochemical expression of hormone receptors: oestrogen receptor positive (ER+), progesterone receptor positive (PR+), human epidermal growth factor receptor positive (HER2+), and triple-negative (TNBC), which is characterized by the lack of expression of any of these receptors There is increasing evidence that the associations between BC risk and reproductive risk factors may vary by receptor subtype. A recent case control study of 4,748 cases (in largely white population) from the Norwegian breast screening programme (2006 – 2014), found that reproductive factors are associated with all BC sub-types to some degree but more strongly with luminal-like (i.e., ER+ and PR+) cancers (33). This supports earlier meta-analyses by the CGHFBC, which found that effect of menopause on BC risk was stronger for ER+ disease than for ER- disease ($p<0{\cdot}01$ for both comparisons) (10). Breast density, itself a risk factor for breast cancer, has also been shown to be associated with hormonal risk factors. However, there is no evidence that the association between BD and BC risk varies by tumour

receptor sub-type (34). There is a need for a better understanding of the associations between reproductive factors, BD and BC sub-types because treatment opinions differ depending upon the cancer sub-types and some sub-types are associated with more aggressive disease. A better understanding on the impact of lifestyle factors on BC sub-types along the life course is also needed (20).

Risk factors and ethnic factors

Many studies are biased towards women in HIC and there is also unfortunately a bias towards studies in largely White populations, but consideration of ethnic risk factors is important in the design of screening programmes that are to be appropriate for the whole population. A retrospective study of over 600,000 cases in the USA between 1979 and 2008 found that BC incidence rates in the White population were lower than for Black women to age ~40, thereafter rates in white women are higher (35). There is further evidence from a large study of age-specific incidence rates, using population registry data from the USA that BC sub-type may also be associated with ethnicity. It was found that in the Black population a lower proportion of BC are of the ER+,PR+ phenotypes (36) relative to White women of the same age. A smaller (n=2,915) study in the UK similarly found that Black women developed cancer at a younger age and had a higher risk of TNBC BC phenotypes (37, 38). Gathani et al (2014) however reanalysed data (1,038,114 White women, 5,877 South Asian women and 4,914 Black women) from the prospective "One Million Women" study in the UK and concluded that that age specific differences in BC incidence were largely a result of the differences in risk factors (38). On the other hand, Januszewski et al. point out that recruitment into the One Million Women study was restricted to women of screening-age and ethnic minorities were relatively underrepresented. They concluded that the evidence is not yet definitive and that the situation is multi-facetted emphasising the importance of differences in tumour biology in terms of implications for survival rates (39).

Average age of BC detection may be lower in China than in the USA. Song et al used data from the National Central Cancer Registry in China and compared this to the US Surveillance, Epidemiology, and End Results (SEER) Program database for the same time period. They reported that after adjustment for population age structure, the median age of diagnosis was approximately 5 years earlier in China (40). Interestingly, however, for women of Chinese ethnicity who lived in the USA (selected from the SEER database), the average age at diagnosis was similar to the US White population (41). Similar studies include Ziegler et al. who found that in Asian-American women who were recent immigrants, there was a lower incidence of BC compared to 2nd generation immigrants, albeit in a relatively small case-control study (42). Overall studies suggest that the association of

ethnicity with BC risk is not straightforward and may be modified by acculturation whereby over two or more generations the risk gradients disappear. A wider mix of ethnicity and locations for study populations is required to ensure that findings are more generalisable.

BC risks summary

In conclusion we have a good knowledge of the aetiology of BC and have identified a wide range of associated risk factors as discussed above. This knowledge has led to the development of sophisticated risk assessment models, such as the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA)(43), the Gail model (44) and the Tyrer-Cuzick (IBIS) breast cancer risk prediction model (45). These tools were originally designed for use in genetic clinics, to which women with a family history of breast cancer are referred, but there is increasing interest in their use in general population risk stratification. After age (and sex), familial history is the strongest risk factor for BC however, only around 3-5% of all breast cancers can be attributable to specific known high-lifetime and medium-lifetime risk gene mutations (46) because their prevalence in the general population is low. Unfortunately, even the most sophisticated risk prediction models currently have limited discriminatory power in the population as a whole (AUC is currently around 69% at best in younger women and ~65% in older women (47)) and even were it possible to identify those most at risk, many of the risk factors will have accumulated over the woman's lifetime and not be amenable to change.

1.2 Strategies for breast cancer prevention

Mortality can be reduced by primary (i.e., by preventing disease occurrence), secondary (i.e., through early diagnosis and treatment of asymptomatic (screening) and symptomatic (downstaging) cancers) or tertiary prevention (i.e., through better treatment and management of the disease). There are a number of strategies in each category that are potentially relevant for use in the UK.

Primary prevention strategies

Primary prevention could potentially be improved by modification of lifestyle choices. A recent review suggests that over 50% of BCs in the USA could be prevented through lifestyle choices and use of chemoprevention for higher risk groups (48), whilst some have estimated that almost 40% of deaths in the UK could be addressed by prevention (49). Unfortunately, most risk factors account, each one individually, for a small PAF, for example Bhaskaran et al estimated that in the UK, even if BMI were reduced across the whole population such that no women were overweight or obese, it would only reduce the number of new BC cases per year by around 5% (21) and eliminating all alcohol consumption across the population in the UK, whilst helpful, would only reduce the number of new cases per year by around 4% (18). This means that a more holistic approach is required. Many

of the BC risk factors are also risk factors for other non-communicable diseases including other cancers and cardio-vascular disease therefore there is a clear justification for addressing these risks through generic public health strategies which promote changes in lifestyle behaviours such as diet, exercise and alcohol consumption. The incentives to change for an individual are much stronger in this global context.

Other preventative strategies would rely on changing reproductive and childrearing behaviours and most are not desirable/amenable to change. For example, whilst much larger average family size and longer periods of breast feeding would reduce BC risk, they would have potentially detrimental effects on child health and family economics which in themselves are detrimental to women and children.

For women at highest risk because they are carriers of high-risk mutations, personalised chemoprevention, and for a small number, prophylactic surgery may be viable strategies e.g., among BRCA1 and BRCA2 mutation carriers the cumulative risk to age 80 years is estimated to be ~90% and ~41% respectively (50, 51). In the UK NICE recommends tamoxifen, anastrozole or raloxifene to be offered to women at over 30% lifetime risk depending upon menopausal status and comorbidities (30) and recommends that elective prophylactic mastectomy be discussed with women with BRCA1, BRCA2 or TP53 mutations as an appropriate strategy for BC prevention. This approach however is only applicable to very few women since the percentage of BC attributable to high-penetrance genes is relatively small and furthermore the potential coverage of this strategy is restricted since it requires personalised risk assessment to identify those most at risk. In future, polygenic risk scores have the potential to further refine risk stratification – in conjunction with existing risk assessment methods based on clinical, family history and BD information – to ensure that prevention measures and clinical recommendations are tailored as much as possible to a woman's risk, but their utility is still being assessed.

In the UK at present, genetic testing, accompanied by counselling, is only offered to those most likely to be at increased risk based on familial history as specified by guideline 164 published by NICE (30). Based on these guidelines, a NHSBSP High Risk screening programme was rolled out from 2011 for those in the very risk categories based on previous radiotherapy, or familial/genetic risk (the NHSBSP classify a genetic or familial lifetime risk of >40% as very high risk in contrast to NICE where >30% lifetime risk is considered high risk). Only those who are referred from primary care for counselling are assessed however and the full extent of the penetrance of these genes is not known. Consequently, at present the possibilities for breast cancer prevention strategies based on genetic testing remain limited.

Secondary prevention strategies

The most appropriate secondary prevention strategy is largely dependent upon the healthcare options available in different countries. In low and middle income (LMIC) countries the delay between palpable tumour detection and diagnosis can be, on average, 6 or more months in comparison to around 30 days in HIC (52). Tumours tend to be of larger size and higher grade when they are eventually diagnosed in LMIC, and studies have shown that time from onset of symptoms to start of treatment greater than three months is associated with poorer outcomes (53). Although there are some differences in cancer phenotype between women in different ethnic groups, the vast majority of the cancers diagnosed in LMIC are of the ER+, PR+ phenotype, amenable to treatments and not any more aggressive than those detected in HIC. An effective strategy for LMIC would be first to address downstaging of symptomatic women by reducing the delay between first symptoms and diagnosis rather than by any screening for asymptomatic disease as, in any case, their health systems are already stretched by the increasing burden of symptomatic women and have no resources to cope with the extra burden that would be created by the introduction of screening (i.e., screening the large pool of eligible women in the population, diagnostic assessment of suspicious cases, and management of confirmed asymptomatic cases). In HIC, initial reductions in mortality occurred through downstaging of symptomatic disease due to improvements in breast cancer awareness by the women and the healthcare professionals, coupled with improvements in treatment. Further reductions in mortality can be achieved by population-based screening using mammography to detect asymptomatic disease. This strategy assumes that early treatment is effective in reducing mortality. To understand how effective screening is, it is important to estimate the relative reduction in mortality attributable to early detection through screening as opposed to reduction in mortality over time that is related to improvements in treatment and population health.

Screening is always associated with harms as well as benefits, as discussed later in this chapter. To maximize the benefits, whilst minimizing the harms, screening is targeted at those in the population who are most likely to get the disease and who can be shown to benefit from early diagnosis through more effective treatment. This can be achieved through stratified screening for asymptomatic disease using risk assessment tools or specific risk factors. To an extent, breast screening in the UK is already stratified because it only invites women in a higher risk age band (typically 50-70) that can benefit most from early detection. A small number of women, with an identified increased genetic risk, are also invited in the UK via the Family History screening protocol that involves more frequent screening from a younger age using supplementary MRI imaging. More subtle means of population stratification are possible, and trials based on more frequent screening for women with higher breast density are in progress (54).

Tertiary prevention strategies

The final risk reduction strategy comprises a range of tertiary treatments. Treatment of BC has improved greatly over the past 50 years, thanks to advances in surgery, radiotherapy and chemotherapy and the fact that these disciplines are drawn together via a multi-disciplinary team approach. Biomarkers for targeted adjuvant therapies, such as aromatase inhibitors for hormone receptor-positive breast cancers have been identified and these have been highly successful in reducing mortality. Breast cancer survival brings with it a range of challenges which include both treatment related consequences (e.g. cardiovascular problems and fertility impairment after adjuvant treatments), and body-image and mental health issues. In HIC it is widely recognised that the long term impacts of surviving cancer can be significant and in the UK and the National Cancer Survivorship Initiative was mandated to work with charities to improve delivery of a more holistic approach to cancer care to improve outcomes after treatment (55).

In summary great improvements in early diagnosis of symptomatic disease (downstaging), detection of asymptomatic disease (screening) and effective treatments (including hormonal and adjuvant therapies) have been made over the last 30 years which have played a large role in the reduction of mortality from BC seen in developed countries and illustrated in Fig 1.1 (56, 57). However, primary prevention has faced serious challenges since the disease is multi-factorial, with risk factors acting, and their effects accumulating, throughout a woman's lifetime. Furthermore, most of the non-genetic, and potentially modifiable, risk factors are either lifestyle behaviours (alcohol consumption, exercise, post-menopausal obesity) or family planning choices (e.g. age at first pregnancy, number of children). Indeed, there are strong health and socioeconomic reasons that argue against some of these changes (e.g. fewer pregnancies is associated with a decline in infant mortality). Moreover, most known risk factors are associated with only a small increase in a woman's lifetime risk of developing BC it is difficult to make a persuasive case for such changes simply on the basis of BC prevention. However, primary prevention may play a role in BC control as part of a broader population-based strategy aimed at controlling the increasing burden from non-communicable diseases by tackling shared risk factors, including risk factors for breast cancer (e.g. excess BMI, alcohol intake, physical inactivity).

Primary prevention for high-risk groups, e.g. with a strong family history including those with an identified gene mutation such as BRCA1 or BRCA2, is easier to implement and includes personalised chemoprevention, prophylactic mastectomy and/or more intensive screening. However, the number of such women is very low with 95% of BCs being sporadic, i.e. non-familial (51) (58).

Secondary prevention strategies are viable through screening for asymptomatic disease in developed countries, where incidence is high and access to diagnostic facilities and effective treatment is universal. BC cancer screening, currently mainly through mammography, offers a tenable approach to reducing mortality from BC through early detection of asymptomatic (non-palpable) tumours followed by early and effective treatment.

1.3 Population-based mammographic breast cancer screening

Early detection is key to reducing mortality from BC, therefore many high income countries have introduced breast screening programmes whereby asymptomatic women are invited for screening mammograms at regular (normally 2-3 yearly) intervals with the aim of detecting BCs before they become symptomatic (e.g. palpable) (59).

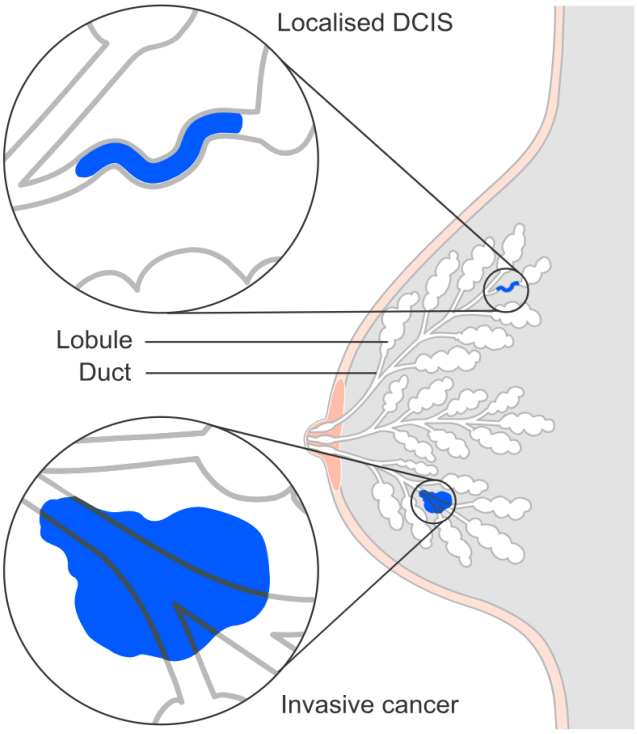


Figure 1.3 DCIS and Invasive Breast Cancer (60)

Most BCs start in the breast ducts and whilst they remain within the duct are classified as ductal carcinoma in situ (DCIS). Around 20% of the cancers detected at mammographic screening in the UK are DCIS (61), and these can be treated before they break out of the duct and invade the surrounding tissue to become invasive. Once they are invasive there is a chance that the cells can spread to lymph nodes and other regions of the body. However, only a proportion of the DCIS cases would ever have progressed to become invasive whilst others would have remained in situ for the rest of a woman's life. Unfortunately, screening tests cannot differentiate between these cases. Most of the remaining 80% of screen-detected cancers are identified before they have become

symptomatic and the rationale for screening is that they can be treated in a more effective and less invasive manner than would be the case if they were detected later.

In England and Wales, all women aged 50 to 70 years are invited once every 3 years to undergo standard 2-view mammography of each breast as part of the National Health Breast Screening Programme (NHSBSP) (62). All images are double read and any woman with a suspicious mammogram is recalled for further assessment (i.e. additional imaging, clinical examination and, if needed, needle biopsy). Cancer cases and equivocal outcomes are then referred to a breast surgeon.

Over 2 million women are screened each year in England and Wales, and over 19,000 cancers were detected in 2018-2019 (63). It is estimated that the five-year survival figures for all women diagnosed with BC is ~85% (64) whilst survival rates for screen-detected BC are higher ~97% (65). In addition, the quality of life for women with their BC detected through screening is higher than for those with BC diagnosed symptomatically (66). In 2018 in England ~33% of female BC cases were diagnosed through the breast screening programme and over half (55%) of the early stage (Stage I) cancers were detected through the screening pathway (67).

Unfortunately, however, mammography is not a perfect (100% valid) screening tool and around 25% of all BC diagnosed in women screened in the UK are diagnosed symptomatically in the period between screens, i.e. following a negative screen (68). Interval cancers tend to have a poorer prognosis than screen-detected cancers (69) sometimes because they are more aggressive but also because they may have become larger, to a palpable size, during the interval since screening.

Interval cancer can be classified by cause:

- Reader misjudgement (‘false negatives’) – ‘suspicious’ or ‘uncertain’ on a review of the screening images.
- Masked on the screening image (‘occult’)
- Cancers that have developed rapidly since screening and so could never have been detected by the screening mammogram (‘true intervals’).

Interval cancers indicate how well a screening programme is performing (70) and reviewing interval cancers can help us understand why some cancers are missed in screening (71) (72). A recent audit in the NHSBSP found that around ~25% of the interval cancers reported were true false negatives i.e. reader misjudgement (73).

1.4 Breast cancer screening benefits and harms

There is an ongoing debate about the actual benefits of mammographic screening in terms of lives saved versus the extent of harm caused. The chief considerations are summarised in Table 1.4 below:

Table 1.4 Overview of potential benefits and harms

Potential Benefits:	**Potential harms:**
Mortality Reduction through earlier diagnosis and treatment.	*Mammography exposes women to ionising radiation.*
Morbidity reduction through earlier diagnosis and less invasive treatments of smaller cancers.	*Over-diagnosis as some of those diagnosed would not have had BC in their lifetime.*
Logistic and economic benefits of reduced disease burden.	*Treatment-related morbidity, i.e. the harms of unnecessary cancer treatments.*
	Psychological harms of False Positives.
	False Negatives, i.e. cancer missed at screening
	Economic costs of tests on women who would never have had cancer diagnosed.

In order to address this question, an independent committee, led by Professor Michael Marmot, was jointly commissioned by Cancer Research UK and the UK National Cancer Directorate to review all the evidence in the context of the UK. This panel produced a wide-ranging report in 2012 and 2013 (74, 75) which included meta-analyses of 11 RCTs (mainly >25 years old) and ~20 more contemporary observational studies. The controversy over benefits and harms is not only a UK issue and a similar review with a global context was carried out by the World Health Organisation (WHO) and reported in June 2015 (46, 76). This review accessed 20 cohort and 20 case-control studies all conducted in the developed World.

The appropriate measure of screening benefit is reduction in mortality rather than cancer-detection per se. Both Marmot and the WHO found that breast screening reduced mortality by ~20% in women aged 50-69 who were invited to screening (the “intention-to-treat” group) and the WHO report concluded that screening had a higher reduction in mortality risk (~40%) in those who actually attended screening. In common with a number of other observational studies a recent (2021) case-control study of >8000 cases, in London, indicated that attending mammography screening led to a

mortality reduction of 39% in those who had attended at least one screening and 26% in all those who had been invited for screening (i.e. the intention to treat group) (77).

The second potential benefit of screening is morbidity reduction which is achieved by earlier diagnosis and less invasive treatments. There is less research on this, but the benefits of earlier diagnosis can theoretically be measured in terms of quality-adjusted life years (QALYs). Morton et al. in a meta-analysis of UK data found a small difference in estimated QALYs between women diagnosed at screening (12.93 years) and these diagnosed symptomatically (13.08 years) but the number of qualifying studies was small (78). There are also possible psychological benefits of less invasive treatments not considered by this analysis.

The benefits of breast screening are relatively clear, but the main debate is over the extent of harm that is caused to healthy women. Although there is some potential harm from mammography itself, the amount of radiation received is very small (79, 80). Warren et al. estimated that in the UK the ratio of saved lives to cancers induced is between 156 : 1 and 312 : 1 depending upon breast size (with larger breasts undergoing more radiation) (80), therefore the benefits of screening far outweigh the risks in this area. The chief problem lies with over-diagnosis (81) which can be defined as breast cancers detected at screening which would never have been found or never caused harm if women had not been screened (76). Differentiating between true cancer lesions and 'over-diagnosed' lesions is currently impossible and this is particularly a problem in cases of DCIS, many of which will never develop into a full-blown invasive cancer during a woman's lifetime. This is an inevitable effect of screening because once detected it is impossible to tell from the cancer histopathology exactly which cancers would have become symptomatic in a women's lifetime. ~20% of the cancers detected at screening are DCIS and some of these cases will never progress outside the confines of the duct. In other cases, the sojourn time (between the point that the tumour becomes detectable and the point at which it is palpable) can be long. If screening detects too many of these slow growing cancers, then it may be doing more harm than good. This is a controversial area not least because it is difficult to estimate the extent to which over-diagnosis actually occurs.

The independent reviewers compared cumulative incidence of breast cancer in women screened and not-screened in the limited number of RCTs available that allowed for a sufficient lead time post the end of screening. The difficulties are that these studies are not just old but also most RCTs resulted in screening being subsequently offered to all women skewing results in the un-screened arm. Others have also measured cumulative incidence in a screened population and compare to that expected in the absence of screening, but it is important that sufficient lead-time should be considered. The problem is that this relies on projections based on pre-screening era trends (82).

Alternatively, because incidence peaks as screening is introduced, it is possible to measure a compensatory drop in the following period but again this is hard to interpret. Consequently, there remains great controversy about the extent of over-diagnosis with estimates ranging from as low as 5% (83), through 11% excess incidence (74) to more than 30% (84), (85). The consensus from the Marmot review was that the extent of over-diagnosis is around 11% whilst the WHO estimated that it was around 6.5% but both acknowledged that this estimate was relevant for high income countries only. In absolute terms, for women screened every 3 years between the ages of 50 and 70, approximately 3 women are over-diagnosed for every 1 life saved in the UK screening programme (74). A consequence of over diagnosis is that women receive unnecessary and harmful treatments including surgery, chemotherapy, radiotherapy and adjuvant hormone therapy. All of these are associated with small risks of increased mortality but perhaps more significantly increased morbidity and psychological harms.

A further potential harm comes from false negative results at screening. Cancer is successfully detected in almost 9 per 1000 women screened in the UK (86), but around 3 per 1000 women screened still present as "interval cancers" i.e. symptomatically in the period between routine screens (68). Interval cancers tend to have a poorer prognosis than screen-detected cancers (69) and there is a possibility that the false reassurance offered to women with a false negative screening result may actually delay diagnosis. Although I could find no conclusive findings, the literature that exists suggests that there is no evidence that false reassurance plays an important role in delaying BC diagnosis (87), (88) but more research is probably warranted in this area.

There is more clear-cut evidence for the psychological harms of false positive results at screening i.e. women recalled for extra assessment tests but found to be cancer free. The assessment process involves extra tests for women who have equivocal initial mammogram findings. In 2017-2018 in the NHSBSP, 3.8% of those screened were recalled for extra tests (84,559 women). Of these only 19,558 had cancer meaning that 65,000 women (>75% of the women assessed) underwent unnecessary tests and stress. Around half of the women assessed underwent biopsy procedures as well as additional imaging (89). A systematic review carried out by Bond et al. (2013) found that the psychological effects of this may endure for 3 years (90). A consequence is that women who were assessed may be slightly more likely to miss their next screen, despite the fact that they have an elevated risk of a cancer being detected in the next screening round (90).

Screening programmes can be assessed in terms of the direct harms and benefits to the women screened but there is also a need to justify the use of limited resources. It is not within the scope of this thesis to look at cost effectiveness of screening but a 2017 systematic review on cost

effectiveness of breast screening by Morton et al. found that, after accounting for unnecessary treatment of over-diagnosed women, BC screening in the UK was still moderately cost-effective but pointed out that this might change over time as common drug treatments lose their patents and become cheaper (78). They concluded that the relative costs of treating a screen detected cancer are ~14% lower than the costs of treating a symptomatic cancer in the UK NHS system. Marmot et al. also considered whether recent improvements to treatments makes screening irrelevant but concluded that the benefits of screening and better treatments are likely to be independent (74).

Summary

It was not within the scope of this thesis to repeat a full review of the harms and benefits of breast screening in HIC, but a review of key literature shows that the consensus at present is that breast screening in the UK is justified for women aged 50-70 years. In terms of reductions in BC mortality and, thus, absolute number of lives saved. Marmot et al. (74) estimated that, for a UK screening programme, a total of 180 women need to be screened every three years for 20 years, between the ages of 50 and 70, for one life to be saved. At the same time however 3 women would be diagnosed and treated who did not need to be diagnosed because their cancer would not have become symptomatic in their lifetime. Based on volumes that were screened in 2012-2013 they estimated that around 1,300 lives were saved per year in the UK and extrapolating to current screening volumes this number is likely to be between, 1,400 and 1,500 lives a year. It is however also clear that it is not a given fact that mammographic screening is always beneficial to all women in the screening age range and it is clear that in LMIC there are likely to be better strategies for reducing BC mortality. This will become more evident as new treatments evolve and the tools for tailoring screening according to a woman's risk becomes more feasible. It is therefore imperative that we constantly look at new ways of "tilting the balance" between the benefits and harms of breast screening and, in Chapter 2, I carry out a literature review covering specific factors that make screening less effective and potentially more harmful to some women, with a view to finding new ways of addressing some of these failings.

2 CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

This chapter provides the findings of a review of certain factors that may affect the delicate balance between the benefits and harms of BC screening.

This chapter starts by providing an overview of the process of mammography and the different measurements that can be derived from mammographic images. I go on to outline the methods that have evolved for estimating BD using mammographic images.

Next, I summarise current research on how breast density measurements may be used to improve the effectiveness of mammographic screening, i.e. by allowing more targeted, stratified screening. As body adiposity is crucial to the interpretation of breast density, I include findings of my systematic review of the association between adiposity (as reflected by body mass index (BMI)) and BC risk and how this variable acts as a confounder in the association between BD and BC.

I then summarise the findings from published systematic reviews on how breast density affects the effectiveness of mammographic breast screening by masking tumours and increasing the number of FNs. I include a systematic review of a little explored feature of breast density i.e. left-breast versus right-breast asymmetry, as a possible factor in tumour masking and also as a potential inherent BC risk factor.

The final section is a summary of findings from my systematic review on aspects of breast compression in the mammographic screening process, because this is a key factor for successful imaging as poor imaging may limit screening sensitivity thereby increasing the number of FNs. For well-researched subjects such as the association between BD and BC, the review relied upon published systematic reviews only but for three specific novel topic areas I carried out a separate systematic literature review.

2.2 Literature search methodology

For each of the specific literature searches I conducted, studies were identified by searching electronic databases, relevant technical manuals, professional guidelines, and radiological conference proceedings. In addition, references and citations within identified papers of interest were reviewed for additional resources. The following main sources were included:

Table 2.1 Key data sources for literature searches

Source	Description	Notes
Embase (Classic+Embase)	Medicine and biomedicine electronic database.	Biased towards Europe, journals
Medline (Pubmed)	Medicine and biomedicine electronic database.	Biased towards N. America, journals
Biology Browser	Electronic database.	Included for fluctuating asymmetry topics
Global Health	Public Health electronic database	Breast screening topics
Healthcare Management Information Consortium (HMIC)	Public Health official publications and grey literature	Screening topics in general
Volpara	Volpara Publications Science Hub	Includes manuals and more obscure conference proceedings and posters with references to Volpara and related technology
Conferences	European Congress of Radiology (ECR) Radiological Society of North America (RSNA)	Radiological conference proceedings. May contain poster information that is not widely published.

The scope was limited to English language resources and, depending upon topic, date limits were set to exclude literature that was unlikely to be relevant e.g. because the relevant technology did not exist prior to that date. Searches were initially carried out in August 2015 to January 2016 and updated in line with recent publications between 2019 and 2022.

Search terms were designed for each topic area, based on specific research questions and these are detailed in appendices A, B and C as they relate to the three systematic search topics. For each search there were several steps involving sub-searches, and these are also described in the appendices along with the search results. Abstracts were reviewed and results were filtered and deduplicated. The refined selections were downloaded to Endnote for full text retrieval and management and detailed review. Most of the retrieved studies were not comparable and therefore no meta-analyses could be performed, but the key literature was summarised into structured review tables for each topic, including effect measures (OR, RR, etc) where available, for ease of comparison.

2.3 Mammography as a screening tool

Mammographic screening is an imaging technique that uses x-rays to image the breast tissue using purpose-built mammography machines. The objective is to differentiate small, non-palpable lesions from healthy breast tissue. X-ray attenuation is different for fatty, fibroglandular and cancerous tissue. The contrast between different areas on the image is used by the film reader to identify different types of tissue on the mammogram but unfortunately both cancers and dense breast tissue (see 1.2 and 2.4) are radiopaque and hard to differentiate. Although the radiation risks associated with screening mammography are relatively low, the regular exposure of well-women to potentially harmful x-ray exposures also means that dose should be kept to the lowest level required to obtain a diagnostic quality image.

Mammography involves compressing the breast between a detector plate and a transparent paddle such that the breast is immobilised, and the thickness of tissue minimised without causing unnecessary pain. The force applied to achieve this compression can be monitored by the practitioner[2]. A tilting or hinging paddle may optionally be used to adjust the angle of the top paddle away from horizontal to reflect the natural shape of the compressed breast (see Figure 2.1 for an example of a mammogram being taken in the cranio-caudal (CC) view). For the medio lateral oblique view (MLO) the arm of the gantry is rotated such that is approximately parallel with the pectoral muscle. Both CC and MLO images of each breast are carried out at each screening in the NHSBSP, because whilst the MLO view includes the majority of the breast tissue up to and including the edge of the chest wall, it is less good at imaging the upper inner portion of the breast which the CC view covers more successfully. Thus, a comprehensive 4-view imaging set is required to adequately cover all of the breast tissue (Figure 2.2). In the NHSBSP automated exposure control (AEC) is used to terminate the exposure when sufficient radiation has been delivered based on a low-exposure pre-scan which estimates breast thickness and density and hence exposure level required.

[2] The terms practitioner and mammographer are used interchangeably in this thesis. Whilst most breast screening mammographic practitioners are also qualified mammographers some are specially trained assistant practitioners.

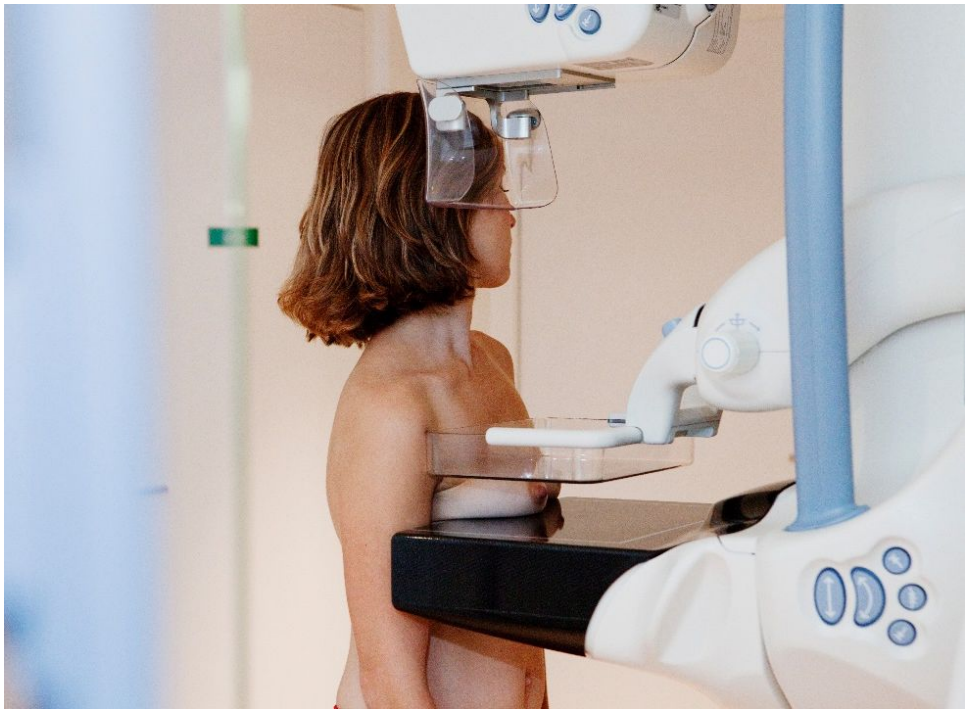


Figure 2.1 Typical mammographic imaging of a cranio-caudal view (Adobe licence)

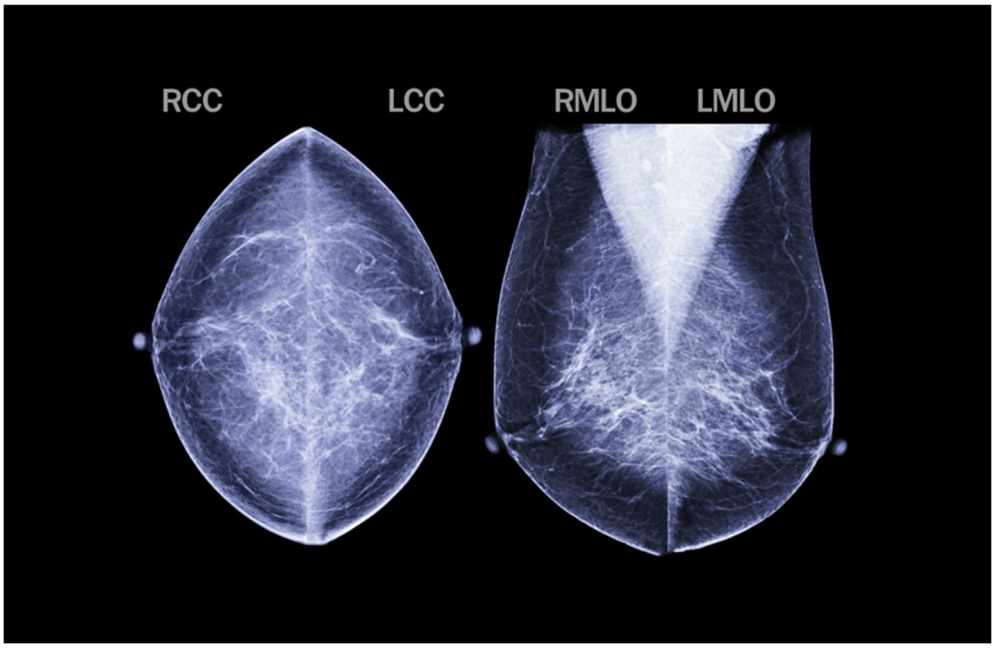


Figure 2.2 Right and Left breast images from a cranio-caudal (CC) view and medio-lateral oblique (MLO) view (Adobe licence)

2.4 Estimating breast density from x-ray images

Breast Density (BD) represents the amount of epithelial and connective (stromal) tissue (fibro-glandular volume (FGV)) in the breast and, as this tissue is radiopaque, it appears as white cotton-like areas on a mammogram and is referred to as mammographic density; in contrast, as fat is a radiolucent (non-dense) tissue it appears as dark areas on a mammogram (Figure 2.3).

Different density patterns were first classified by Wolfe in 1976 (24), who noted an association between 'prominent' ductal patterns on the mammogram and cancer risk. An alternative qualitative pattern-based system was also developed by Tabár in 1997, which identified 4 different categories of parenchymal pattens (91) .

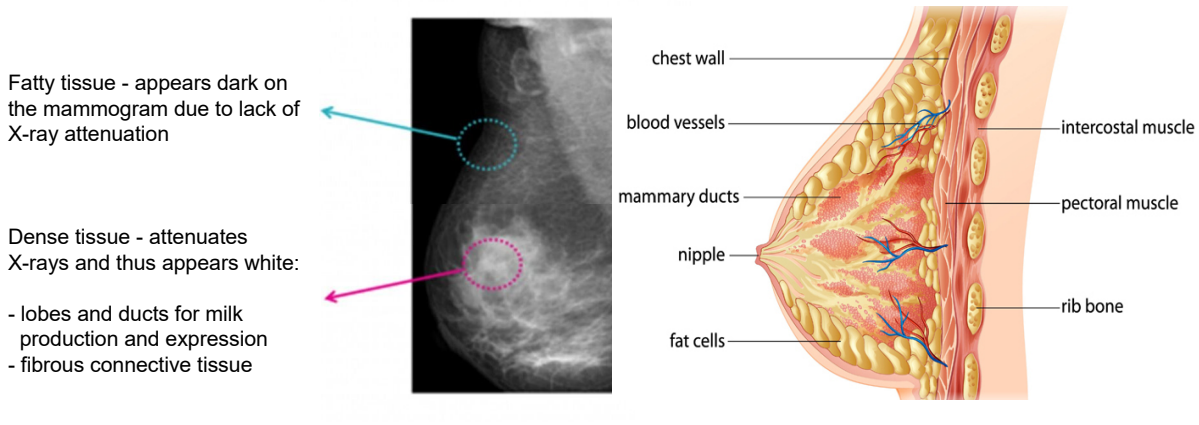


Figure 2.3 Mammographic Density (Courtesy of Volpara Solutions)

The BI-RADS classification, developed by the American College of Radiology, aimed to standardise the way that density is reported using the relative amount of dense tissue in the breast (%BD). The 4th edition of the BI-RADS density classification was released in 2003 and incorporated a quantitative scale depending on the %BD: category 1 (<25% fibroglandular tissue), category 2 (25-50%), category 3 (50-75%) and category 4 (>75%) (92) (93). The 5th Edition released in 2013 reverted to a qualitative scale in order to focus on the presence of dense tissue that may obscure lesions i.e., the risk of masking rather than inherent susceptibility to BC. Unfortunately, all these classifications are subjective and prone to inter-reader variability (94, 95).

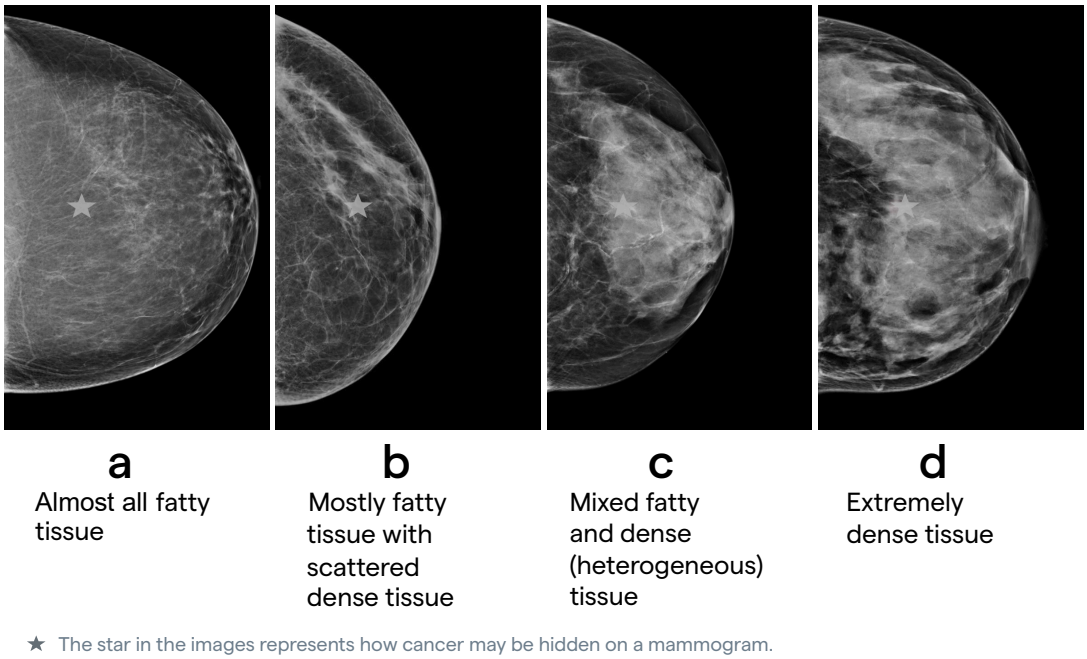


Figure 2.4 Typical patterns of mammographic density in Western women (Courtesy of Volpara Solutions)

Until relatively recently the 'gold-standard' for measuring density was Cumulus, a semi-automated method using full-field digital mammography (FFDM) (or digitised analogue images), developed at the University of Toronto (96), but unfortunately this is was a labour intensive method because it relies in readers defining the breast edge and areas of high density. Studies show this is a highly reproducible method (96, 97) but unfortunately it is not practical for application across the high volumes seen in a screening context.

FFDM has replaced analogue film across the whole of the NHSBSP since 2011 and automated BD calculation software can potentially be used to analyse these digital images, allowing BD and other breast volumetric measurements to be calculated for *every* woman screened.

The fully automated tools fall into 2 main groupings depending upon whether an area or volume-based approach is adopted (See Figure 4.3 and section 4.3 for an overview of the technology as it relates to this thesis). Area-based tools suffer from measurement error because they cannot estimate the thickness of breast tissue, this means that a highly compressed breast can appear less dense than the same breast that has undergone less compression. Volumetric estimates do not suffer from this problem, but they do require the raw image for processing which may not always be available. Different terminology has been used to discuss these different measurements which can lead to some confusion. However, for the purposes of this thesis the abbreviations and meanings below are assumed where possible.

Table 2.2 Breast density terminology - Conventions and abbreviations

Abbreviation	Definition
Breast density **(**BD)	Breast Density (as a general concept) based on the amount of fibro-glandular tissue in the breast, or on the identification of specific parenchymal pattens, as seen on a mammographic or MRI image.
Percentage breast density (%BD)	A measure of relative breast density estimated from a mammographic or MRI image. Percent breast density = Breast dense volume (or area) / Total breast volume (or area). In some papers this is alternatively referred to as %MD (% mammographic density).
Dense area (DA)	The total *area* of the mammographic image that appears as white cotton-like areas. Estimates the total area of fibro-glandular tissue on the image.
Non dense area (NDA)	Estimated total area of mammographic image that is dark i.e. fatty tissue cm^2.
Dense volume (DV)	Mammographic dense *volume* as estimated from mammographic images. Estimates the absolute (total) volume of fibro-glandular tissue in the breast.
Non-dense volume (NDV)	Estimate of the volume of fatty tissue in the breast cm^3.
Breast volume (BV)	Estimated overall total breast volume

Footnote: Note that some of the papers included in my thesis may deviate slightly from this convention depending upon the conventions of the publication and reviewers.

2.5 Risk prediction and stratification using breast density

Women with higher BD have not only an elevated BC risk but also a greater chance that their cancer will be missed when reading a mammogram because, like dense tissue they appear as radio-opaque areas on a mammographic image. One possible approach to maximizing the benefits of mammographic screening whilst reducing its harms would be to use BD alone, or in conjunction with information on other risk factors, to tailor screening according to a woman's BC risk. This would mean inviting women with high BD (higher risk) to more frequent screening whilst those with low BD (lower risk) could be invited less frequently.

The link between BD, estimated using either 2-dimensional (area-based) or 3-dimensional (volumetric) approaches, and BC risk is a well-researched area, with findings consistently showing a strong association. McCormack et al (25) carried out a systematic review and meta-analysis, which included 42 eligible studies, and estimated that women in the densest category (i.e. with radio-dense tissue occupying >75% of the total breast area on a mammogram) have almost a 5-fold increase in the risk of BC (RR= 4.64 (3.64-5.91)) relative to those with the least dense category (<5% of radio-dense tissue); they found no evidence for effect modification by other risk factors. Boyd et al. (98) concluded that after adjustment for other risk factors, BD is more strongly associated with BC

than most other risk factors. A meta-analysis by Cummings et al. in 2009 found a RR = 4.20, (95% CI = 3.61 to 4.89), for >75% BD versus <5% BD (99). In 2014, a systematic review by Huo et al., updated the findings from earlier analyses by considering a further 18 USA, Australian and European studies on the association between BD and BC risk, and found that the RR of high versus low density were similar to the earlier findings (RR of the highest density category versus the lowest ranging between 2.45 to 5.34). Huo et al. also identified key future research areas including how to deploy automated density tools in the clinical environment (100). All these studies also found that age and BMI were important confounders in the association between BD and BC.

BD is a strong risk factor for both invasive and in-situ BC but the biological pathways through which this operates are not completely understood (101, 102). At the simplest level a large amount of epithelial tissue in dense breasts, which is associated with a higher cell proliferation rate, results in a higher probability of somatic mutations leading to the development of cancer. BD also includes connective tissue (stroma) which increases the abundance of collagen and other proteins, which encourage the growth and migration of epithelial cells (103-105).

Even though the magnitude of the risk specifically associated with BD is lower than that associated with age or high-penetrance genes, it may account for a large number of cases because it is relatively common, with around 10% of women of screening age having very dense breasts (25, 106). McCormack et al. estimated that the population attributable fraction (PAF) of BD>50% was 23.2% for post-menopausal women and even higher (42.8%) for pre-menopausal women (25). Thus, BD is an important biomarker of subsequent risk of BC with the potential for being useful for risk stratification in a screening context. Longitudinal studies of changes in %BD in 645 UK women by McCormack et al. (2010)(27) show that rate of change (decrease) in %BD over a series of screening examinations (2 to 5 screens) was consistent across all women no matter what their initial density category. This was confirmed by Krishnan et al (2017) who found that over the course of 4-5 mammograms (mean time between mammograms 2.2 years) in 970 Australian women, the area-based mammographic measures that predict BC were highly correlated (28). A study by Yaghjyan et al. (2013) (107) found that associations between BD and BC risk persist for up to 10 years after the initial mammogram (107). These studies suggest that a one-off breast density estimate might be a useful predictive tool for stratification or that if used in early adulthood may be used as part of a preventative strategy.

However, even if shown to be effective in large RCTs, stratification methods based on BD would be difficult to implement across a population-based programme where millions of women are invited

each year and to be practical this would require a cost-effective, validated, and automated tool for estimating the breast density of each woman screened.

2.6 Breast density and mammographic screening performance, the effect of masking

As early as 1985 it was hypothesised by Whitehead et al (108) that mammographic parenchymal patterns could be associated with masking of breast tumours but Boyd et al. (2007) (97) were the first to undertake (three) large nested case control studies with an extended length of follow up (8 years), that examined the association of mammographic density with screen-detected, interval and symptomatic cancers. Boyd et al. found that women in the highest category of BD had greater odds of cancer in the year following a 'normal' mammogram than women in the lowest density category (OR of 17.8 (95% CI 4.8 – 65.9)) (97) thus showing that mammographic sensitivity decreases as BD increases. The most obvious explanation for this is that radiopaque dense tissues may overlay cancers, which also appear radiopaque thus 'masking' cancers from the point of view of the image reader. Pisano in 2008 went on to estimate that, amongst pre- and peri-menopausal women, mammography sensitivity was less than 60% in women with dense breasts, compared to 86% in fatty breasts but the differences were less significant in older women (109) and Wanders et al. in 2017 were able to repeat this in a large cohort study using automated breast density measurements on over 100,000 women (667 screen detected and 234 interval cancers) and reported similar findings based on 2 year screening intervals, finding a 61% sensitivity in the highest density BIRADS category in comparison to 86% in the least dense category (110). Thus, there is compelling evidence that mammography is less effective for women with dense breasts and that suggests screening using mammography is compromised for such women and likely to be particularly ineffective in younger pre-menopausal women who have higher breast density.

Carney et al. (2003) found, in a prospective cohort study of over 300,000 North American women, that increased BD (assessed by radiologist on BIRADS scale) decreases the specificity of mammography findings. Specificity increased from 89.1% in women with extremely dense breasts to 96.9% in women with almost entirely fatty breasts (111), meaning that women with dense breasts are subject to relatively more false positive screens and undergo more invasive testing. Krishnan et al. (2016) differentiated between 'inherent' (screen-detected) cancer risk which was best explained by %BD (or dense area) adjusted for age and BMI (112), whereas 'masking risk' (interval cancers) was best explained by a relative measure of density i.e. %BD adjusted for age, and the association was stronger.

Despite the strong evidence for decreased sensitivity and decreased specificity in women with high breast density, as Huo et al pointed out, there is currently a shortage of evidence on how

mammographic density assessment should be used in the context of screening programmes (100). Huo's systematic review in 2014 concluded that although increased BD is positively associated with BC risk there is no evidence that it is associated with higher mortality, and it is not clear how BD should be used in practice since BD assessment methods were not specifically designed with risk stratification in mind. In a later systematic review into the use of supplemental screening for women with dense breasts, Melnikow et al. (2016) concluded that "Studies identifying more accurate and reproducible methods of identifying women with dense breasts are needed" (113). Importantly they found no studies that looked at long term outcomes and mortality. Early findings from the Dutch 'DENSE' RCT which offers supplemental MRI imaging for women with very dense breasts, suggest that this may reduce the risk of an interval cancer being detected between screens by up to 50% (114) but the number of cases reported on so far was small and the additional costs of MRI scanning are high and may not be practical for a population-based screening programme. Recently (2019) the BRAID trial has commenced in the UK, which is a randomised, multi-centre study assessing the impact of supplementary imaging for women with dense breasts (using one of abbreviated MRI, whole breast ultrasound or contrast enhanced spectral mammography as supplementary methods) in the UK NHSBSP (54). At the time of this thesis however there is no clear alternative to using mammography for large population-based breast screening programmes despite its acknowledged weaknesses.

2.7 Adiposity, BMI, breast density and breast cancer risk

The association between BMI, BD and BC risk

As previously discussed (section 1.2), BC risk increases with BMI in post-menopausal women. Although the exact causal pathways are subject to some debate (26) it is thought that the conversion of androgens produced by the supra-renal glands into oestrogens by enzymes in the fat tissue is responsible, because it occurs on a larger scale in women with high BMI. Systematic reviews such as those by McCormack et al. (2006) (25) and the Huo et al. (2014) critical review (100) have established that BMI, together with age, is an important confounder in the relationship between BD and BC risk, because it is inversely associated with %BD and it is also an independent risk factor for BC (Figure 2.5).

Boyd pointed out that studies of the association between %BD and BC should therefore adjust for this 'negative confounding' of BMI to avoid underestimating the magnitude of the %BD effect (115).

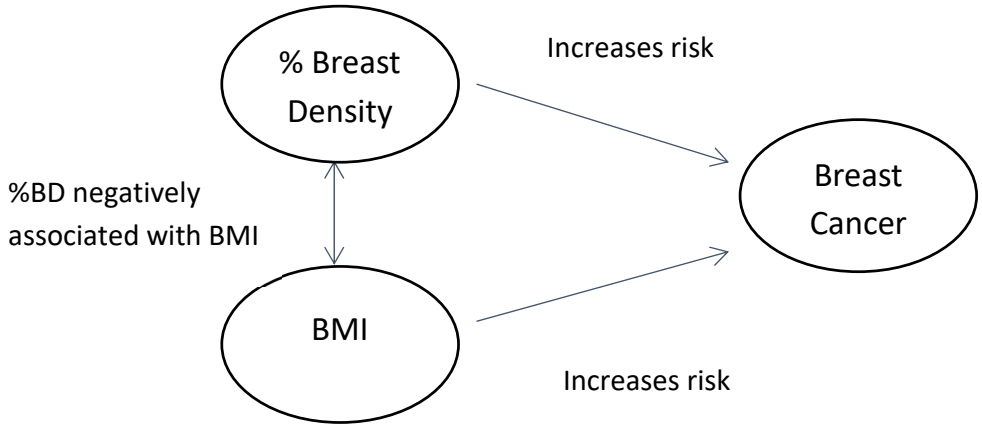


Figure 2.5 Associations between BD, BMI and BC in postmenopausal women

As discussed in Section 1.1, %BD is increasingly used as an intermediate phenotype in epidemiological studies, and it also offers the potential to be used in breast cancer prevention strategies. Any population-based risk stratification incorporating BD estimates should, therefore, also adjust for BMI, but unfortunately BMI is not routinely recorded in population-based screening programmes in the UK. Therefore, a simple, reliable, readily available proxy for BMI is required. This is a little explored topic therefore a structured literature search was conducted to find resources that have examined practical methods for evaluating BMI in large scale settings. Of particular interest is the possible association between BMI and breast measurements estimated from an x-ray image (fatter larger breasts being associated with higher adiposity). It is possible that automated breast measurements made on an x-ray image could act as a proxy for BMI where BMI itself cannot be measured. Appendix A provides details of the method, search terms and structured summary table for a systematic review on this topic: the findings are discussed and summarised below.

Body fat and BMI

Accurate techniques exist for estimating the percentage of a person's body mass that is composed of fatty tissue (BF%), including the use of a series of dual energy X-ray absorptiometry (DXA) body scans for estimating the relative % of fat in the body based on differential attenuation of x-rays. Another accurate method is densitometry (hydrostatic underwater weighing) which can be used to calculate whole body density and hence the proportion of, lower density, fat. Whilst such methods are useful as a reference for calibration (116), they are impractical and inappropriate in screening settings. Alternative indirect methods for estimation of %BF include skinfold measurements (where the practitioner must be skilled to achieve reliable results) and bioelectrical impedance using hand-held devices, but the simplest, and most widely used, method is to measure both height and weight to calculate BMI (BMI (kg/m^2)).

Various studies have investigated the validity of BMI as a proxy for adiposity. Deurenberg at al. (116) found that BMI performs well, after adjustment for age and sex in comparison to the gold standard of DXA in European populations but a further study found that participants of Asian ethnicity had a higher BF% for the same BMI than Caucasian participants (117) and Rush et al. (2007) (118) confirmed this finding. The relationship may also vary with age, for example older persons may show a higher BF% compared with younger persons with comparable BMIs (119). Others have suggested that the relationship is non-linear in those with low BMI and may only work well within a higher range of BMI values although most studies are restricted to children and it is not clear whether this is true in adults (120). It has been suggested that inverted BMI (iBMI) (cm^2/kg) may be a more useful predictor of adiposity than BMI (BMI (kg/m2) because it is normally distributed (121) and the relationship between BF% and BMI may be non-linear, but that of iBMI and BF% is linear (122). However, iBMI is not widely used in risk assessment models and is less intuitive.

Skin folds and impedance methods for assessing adiposity are likely to be impractical or unacceptable in a routine breast screening setting, and in the UK NHSBSP, measurement of height and weight to get BMI has been rejected as impractical given the fact that the whole mammographic screening process is expected to last around 10 minutes. The alternative is to collect information on BMI as part of the self-completed screening questionnaire but this does rely on self-reporting which has been found to introduce measurement errors, and if these are differential, even biases (123). Furthermore, women who self-report their height and weight are likely to be a biased sample.

Despite its known drawbacks BMI is the most widely used estimate of adiposity in BC epidemiological studies and risk assessment models incorporate BMI as a risk factor (43-45). Easily available, automated estimates of BMI for all breast screening participants would clearly confer advantages for risk assessment and stratification and for epidemiological studies.

Area-based breast composition measurements and BMI

A consistent, strong positive correlation has been found between non-dense area of the breast (NDA) as estimated using 2-D area-based analysis of mammographic images and BMI, with correlation coefficients of +0.59 reported by Boyd et al (115) and Lokate et al. (124) and +0.62 by Pettersson et al (125). For postmenopausal women %BD is negatively correlated with BMI as shown in Figure 2.5, with reported correlations for %BD using area-based estimates ranging between -0.38 (115) and -0.61 (126). The correlation between *absolute* area-based breast density (DA) and BMI is also weakly negative ranging from -0.21 (124) to -0.32 (127) (128).

Area-based assessments of breast volume and breast-tissue composition have their drawbacks; in particular, they are 2-dimensional and based on a dichotomous classification of fat/no-fat. Total amount of breast tissue, which is highly variable depending on, for example, the compression used at imaging (129) and the relative size of a woman's breasts, cannot be assessed by area-based tools as the thickness of the breast cannot be estimated. Hence, it has been acknowledged that volumetric measurements of breast volume and breast-tissue composition may be preferred (130, 131).

Automated volumetric breast measurements and BMI

Recent development of automated digital image assessment software means that volumetric breast density assessment is now more widely available. Volumetric assessment tools exist which are both, robust (132) and reliable (133). Volumetric %BD is negatively correlated with BMI with reported correlations around -0.50 (128), similar to area-based methods. However in contrast to 2-D methods, Shepherd et al (130) and Schetter et al (128) both found a moderate weak positive correlation between volumetric density (DV) and BMI (0.44 and 0.41 respectively).

The relationship between BMI and volumetric breast measurements is less well understood than area-based associations, and consequently the importance of BMI as a potential confounder in a volumetric-based density study is less certain. It is plausible that breast non-dense volume (NDV) (i.e. fat in the breast) or its correlate, total breast volume (BV) could be used as a proxy for BMI. Until recently the process of estimating BV and NDV has been labour intensive but by using automated 3-D assessment tools they can now be calculated from digital mammographic images.

Interestingly both NDV and BV have been explicitly used as a proxy for BMI in analyses of mammographic density and BC risk in studies where BMI data are not available (134, 135) but the validity of this approach had never been empirically tested at the time of this thesis. It would be a useful contribution to our understanding if the validity of this approach could be confirmed.

2.8 Anthropometrical asymmetry, breast development and breast cancer

Introduction

In order to improve the balance between benefits and harms in breast screening, we require new methods for ever refining the stratification of screening. One little explored potential breast cancer risk factor is asymmetry in left side versus right side breast volume (BV) and mammographic density volume (DV) and in the following section, I discuss the reasons why this may be of interest. A structured literature search was conducted to find relevant resources and Appendix B provides

details of the method, search terms and structured summary table for a systematic review on this topic: the findings are discussed and summarised below.

Asymmetry, development and health

There is extensive biological knowledge that points towards 'fluctuating asymmetry' (FA) being a useful measure of phenotypic and genetic quality. Parsons et al (1990) reported that increased FA, i.e. increased anthropometrical asymmetry in paired features, is a common response to increased stress during development (136), hypothesising that higher FA was a reflection of poorer developmental stability at the molecular, chromosomal and epigenetic level. Thornhill and Moller (1997) identified that FA was associated with a number of chromosomal abnormalities including Fragile-X syndrome and neural tube defects (137), and Milne (2003) found that higher FA (in six traits including ear breadth, ear length and wrist breadth) was associated with a significant increase in women reporting that they had two or more identified health conditions, although the study population was young (aged 26) and restricted to women of White ethic groups (138).

Studies of dermatoglyphics have shown that increased asymmetry in hand patterns is also associated with increased risk of several diseases including breast cancer (139). Manning and Leister (2001) hypothesised that high 2nd digit to 4th digit ratio (2D:4D) was associated with higher in-utero exposure to oestrogen and found evidence that high second to fourth digit ratios are associated with early onset of BC (140) which was confirmed by a later study (141). Bunevicius carried out a systematic review on this topic in 2018 and found 19 relevant studies. Meta analysis confirmed that women with high 2nd digit to 4th digit ratio (2D:4D), thought to be associated with lower exposure or sensitivity to prenatal testosterone and/or higher levels in utero oestrogen levels, had increased risk of early-onset BC(142).

Moller et al. (1995) showed that, as in other paired features, breast FA is related to both fecundity and general health. However, like many early studies on breast FA their study group was restricted to women who were candidates for plastic surgery and was therefore not a representative sample of the population in general. A later study, conducted in a UK breast screening population, by Manning et al. however confirmed that increased breast FA was correlated not only with age, height and parenchymal type but also with reproductive factors such as parity, age at first birth and age at menopause; they reasoned that an individual's ability to tolerate exposure to oestrogens, particularly during periods of growth, may be reflected in a higher degree of homeostasis and thus bilateral symmetrical development of paired organs such as the breasts (143).

Breast asymmetry and BC risk

Exposure to endogenous and exogenous sex hormones is recognized to be important in FA, breast development and also in the pathogenesis of breast cancer (12, 16, 17, 144, 145), with the effect of many reproductive factors on breast cancer risk, e.g. early age at menarche and late age at menopause, being mediated by circulating levels of these hormones (146). There is also some evidence from a meta-analysis of 32 studies by dos-Santos-Silva et al. that even pre-natal exposure to high levels of sex hormones may increase the risk of breast cancer e.g. studies have reported positive associations between breast cancer risk and birth size, pre-eclampsia and multiple births, all possible markers of raised, in-utero, exposure to oestrogens (9). It is therefore plausible that breast FA, being similarly associated with reproductive and hormonal factors, could be a biomarker of BC.

As discussed in section 1.2, BC is a heterogeneous disease with different tumour subtypes classified on the basis of gene expression or hormone status. Identification of a tumour subtype is clinically relevant because subtypes are associated with differential treatment options and prognoses. Current research indicates that reproductive factors are to some extent associated with all subtypes but with the strongest associations seen for luminal-like subtypes (33, 147). It is therefore plausible, given the association of FA with reproductive factors, that FA may also be more strongly associated with particular *sub-types* of BC.

Historically, a large number of studies have found that there is a slight but significant predominance of left-sided breast cancer (148-150) and also that the left breast is on average larger than the right breast (151, 152). A possible explanation for this phenomenon is that a larger breast reflects an increase amount of epithelial tissue and stroma (i.e. radio-dense tissue on a mammographic image) may create an environment with higher aromatase activity (153, 154) which would favour tumour development but the exact mechanisms for this remain a subject of debate. My literature search found very few studies that have examined the association between BV asymmetry, breast DV asymmetry and breast cancer risk and to our knowledge none have looked at the association between asymmetry and subtypes of breast cancer. Scutt et al. in 1997 carried out a small study in the UK breast screening population, using visually assessed mammographic breast size (BV) asymmetry estimates (~250 cases; ~250 age-matched controls) and found that absolute BV asymmetry was positively associated with a cancer detected at that screen (155). Scutt et al. (2006) also found an association between BV asymmetry and medium-term risk of breast cancer diagnosis (mean time to diagnosis 6.44 years) after adjustment for known risk factors (BMI, age at menarche) and absolute breast size (156). Interestingly they found no correlation between tumour size and BV asymmetry, which suggests that the association is not simply explained by the tumour being

responsible for the increased breast volume in the affected breast. Eltonsy et al. (2007) used a computerised algorithm to estimate BV asymmetry from screening mammographic images (280 screen-detected cancer cases; 82 controls). They found that mean absolute BV asymmetry, adjusting for BV, was significantly higher in cancer patients (157). Only limited research has looked at the association between volumetric measurements of breast asymmetry and FNs at screening. Kayar et al. (2015) used physical breast measurements (251 cases; 466 controls) from a Turkish outpatient (non-screening) clinic, to identify a 'pathological breast asymmetry ratio' (158). They found that left breast:right breast BV ratio of >±20% was associated with an increased risk of breast cancer being diagnosed within one year of the examination (158).

Findings to date are therefore limited but consistent with BV asymmetry being associated with the presence of a breast cancer (155, 157-159) as well as with a higher risk of having a breast cancer diagnosed in the short- and medium-term (156). There is also limited evidence that asymmetry in mammographic density, might be associated with higher short-term (160-162), and medium-term (163) risk of being diagnosed with breast cancer.

We require new methods for refining the stratification of breast screening and it is possible that breast asymmetry (left side versus right side) in breast volume (BV) and mammographic density volume (DV) may be a factor that could contribute to more subtle risk assessments in screening populations and potentially act as a biomarker for different subtypes of BC. There is, as yet however, very little research in this area.

2.9 Mammography technique and screening outcomes

Not all mammographic images are equally good and any factor which leads to a sub-optimal image is potentially detrimental to cancer detection and makes mammographic screening less effective. Of particular interest in this thesis is the extent to which objectively measured mammography compression techniques (such as the force used) are associated with both measurable outcomes of mammography (such as compression thickness), and also with screening performance outcomes, in particular missed cancers i.e., FNs at screening which result in interval cancers being diagnosed. The key factors that are historically thought to influence the ability to detect cancers are summarised in Table 2.3.

Table 2.3 Image acquisition parameters that may influence image quality

Factor	**Description**	**Availability of empirical data**
Positioning	All breast tissue must be imaged. Mammographers are taught that elevating the inframammary fold will bring the breast and any lesion closer to the imaging receptor enhancing image quality.	Qualitative information only at the time of this thesis. New tools for automated positioning assessment currently being introduced by manufacturers.
Breast **Thickness**	Thinner breast tissue reduces radiation dose and improves the perceived image contrast, enhancing lesion visibility.	Average distance between detector plate and paddle is held in Digital Imaging and Communications in Medicine (DICOM) tags on mammography unit.
Compression **Force**	The breast is compressed to reduce thickness and required dose and to reduce movement and potential blur. A correctly compressed breast is less likely to suffer from tissue superimposition, which can obscure cancers.	DICOM tag.
Mean Glandular **Dose**	Mammography systems use automated exposure control, which calculates the required KV, exposure time and the appropriate anode/filter based on breast thickness and a short low dose pre-exposure. Typical dose is 4.5 absorbed dose in milligray (mGy) for a 2-view screening exam of both breasts (164). Dose should be kept to the minimum to avoid the harmful effects of x-ray exposure.	DICOM tag.
Pressure Applied (kPa)	Pressure applied. Typically expressed in kilo Pascals (kPa) determined from the recorded compression force (N) per unit of contact area between breast and detector plate (m^2). Represents a measure of force applied relative to the size of the breast.	Calculated using DICOM tag and area measurement (automated or labour-intensive manual)
Paddle Tilt	The angle of the plate from the horizontal in degrees. Depends upon whether a fixed or tilting/flexible paddle is used. Influences how much thicker the breast is at the chest wall than at the nipple as a result of the slant. A thicker breast at the chest wall will have a lower contrast.	DICOM tag.

A systematic review was conducted to search for literature on mammographic technique, screening outcomes and BC detection. The search terms, approach and summarised findings are documented in Appendix C and discussed below.

The review found that until relatively recently, (2015 onwards), little quantitative research had been published on 'how' a mammographic image is taken. Hence empirical knowledge about the relationship between mammography technique, the 'quality' of the image and screening

performance is limited. Most of these studies took place in parallel with the studies conducted in this thesis and were not available at the time of my study design. Earlier studies tended to be descriptive and only with advent of automated image analysis tools, has it been practical to process enough images to give studies sufficient power to detect the differences in outcomes associated with relatively small variations in objectively measured imaging technique parameters.

Standards and Consistency in mammography

Mammography guidelines and standards have been developed so that best practice can be shared and adhered to. The existing guidance for NHSBSP mammographers is comprehensive in terms of how to guide women through the mammography process and has a high-quality information on positioning and system calibration. Salvagnini et al. found that 'adequate' compression is important alongside positioning as a key factor in achieving a good image (165) because it helps to reduce movement (blur), separate overlying tissues and also reduce thickness, thereby improving tumour conspicuity (165, 166). During compression the breast volume is not reduced but the breast tissue is stretched and spread into a thinner layer. Yaffe et al. (2011) found that as compression reduces the absorbed radiation dose during the screening procedure it decreases the risk of x-ray scatter which can result in a 'noisy' image which is more difficult to read (167).

Despite its acknowledged importance, quantitative guidance on compression is extremely limited. No specific guidance was found on the use of flexible versus rigid compression paddles. It is well accepted that an image does not need to be perfect it simply needs to be 'good enough' to allow cancers to be detected, however the definition of 'good enough' is subjective. At the time that the data in this thesis was collected, good practice in the NHSBSP included the mammographers reviewing a sample of their images on the PGMI (Perfect, Good, Moderate, Inadequate) scale on an ad hoc basis, but this system is subjective and studies by Boyce et al. (2012) concluded that even in the UK and Norway, countries where the system had been used for a number of years, inter-rater agreement on the PGMI scale was poor (168).

At the time of the data collection for this thesis, the NHSBSP standards also stated that the force measured by the x-ray machine should not exceed 20daN (169), although in practice this is an almost meaningless guideline since 20daN is rarely, if ever, reached in screening mammography in the NHSBSP. European guidelines available at that time were also largely subjective, "The radiographer must ensure that the breast is properly compressed, but no more than is necessary to achieve good image quality" (170) and subsequent updates refer only to a 'standard' compression force of 10daN although it is not clear how widely this was expected to be applied (171).

Different European screening programmes have started to become more specific in their guidance e.g. the Norwegian programme now has a guideline that the force range should be between 11 and 18daN but this is clearly a wide range of acceptable values (172). NHSBSP compression guidelines however remain largely subjective. Updated guidelines were issued by the NHSBSP in December 2017 that set out a more structured approach to image quality audit requiring that a minimum of 20 images per mammographer are audited every 2 months at service level, using an image quality assessment tool. However the guidelines largely concentrate on checking for correct positioning and contain only two (subjective) checks for image sharpness and movement including the directive to ensure 'adequate compression to hold breast firmly / no movement' (173). The reluctance to offer specific empirical guidance is perhaps understandable given the lack of an evidence base to back up more objective guidance.

By collecting technical parameters on the way that images were acquired from NHSBSP screening centres (between 2004 and 2010), Mercer et al (174) demonstrated poor consistency in force and dose between mammographers but found no relationship between experience of the mammographer and the force used. This study was repeated across 3 different NHSBSP screening centres and analysis of variance (ANOVA) of mean compression force values of practitioners demonstrated a significant difference ($p < 0.0001$) between screening centres, which suggests that there may be local conventions in the way that mammographers approach mammography (129). The same authors also found significant between-practitioner differences when examining the compression force applied by different mammographers between sequential screens on the same woman (175). Such findings are not limited to the UK with high levels of compression force variability found in the Netherlands and USA (176), Norway (177, 178), Australia (179), Sweden (180) and Ghana (181) suggesting that mammographers may have a preferred individual approach to the task. Ng (2017) carried out an international review of compression forces across 17 countries and found that women in the Netherlands received the greatest average force (13 daN) and Switzerland the lowest (6.6 daN) (182). This shows that despite attempts to standardise mammography, there is widescale evidence of inconsistency in empirical compression measurements across different mammographers, screening services and internationally. There is clearly scope for greater empirical study into the impact that this may have on cancer detection and screening performance not just in the UK but in a wider geographical context.

The importance of breast compression as a key factor in 'good' image acquisition

Compression force can be monitored by the mammographer as it is applied to the breast during mammography and the objective is to reduce the thickness of the breast such that the image quality

is improved, and the dose minimised. Helvie et al (183) in 1994 found a positive association between thickness and dose and a negative association between thickness and image quality measures (sharpness and contrast), in analogue film mammography, by comparing CC to MLO views. These findings have since been generally confirmed using full field digital mammography (FFDM) but some have suggested that the association is more subtle. Saunders et al in 2008 (166) carried out simulations using breast phantoms and found a linear association between scatter of radiation and breast compression thickness but interestingly noted that the impact on tumour conspicuity was not linear; an increase in thickness from 6 to 6.75 cm could be acceptable, with little impact on image quality or dose. They concluded that breast compression plays a less important role in lesion conspicuity for digital mammography than for analogue mammography. This experiment was extended by Salvagnini (2016) (165) using 520 images with 4 different thickness levels. The images were matched on density and read by 4 different experienced radiologists. Salvagnini found that tumour detectability (based on area under the curve (AUC) comparisons) fell as breast thickness increased. These experiments highlight some of the difficulties in designing standards for breast compression. Whilst there is an association between thickness and dose and thickness and conspicuity, the association may not be linear and there was even some suggestion from Saunders (166) that at highest compression levels the breast was distorted resulting in a higher proportion of the breast in high dose areas. Another problem is that such experiments are necessarily carried out on breast phantoms because of the ethical issues associated with subjecting study participants (especially those who have already been diagnosed with a cancer) to unnecessary radiation if repeat compressions are carried out at different levels of force and thickness. Hence there is a need for more long-term observational studies on larger data sets from real-life settings.

Force and pressure as potential guidelines in mammography

As noted earlier most objective standards in mammographic compression have chosen to specify an indicative level, or range, of force that should be applied to the breast during mammography. Although force is measured by the mammography unit and can be viewed and adjusted by the practitioner (Figure 2.6), Murphy et al. (2015) in a qualitative study into mammographer's views and behaviours, found that practitioners prefer to use their own intuitive judgement to make finer adjustments to force rather than rely on machine readouts (184). This hints at some of the barriers any introduction of objective standards may face.

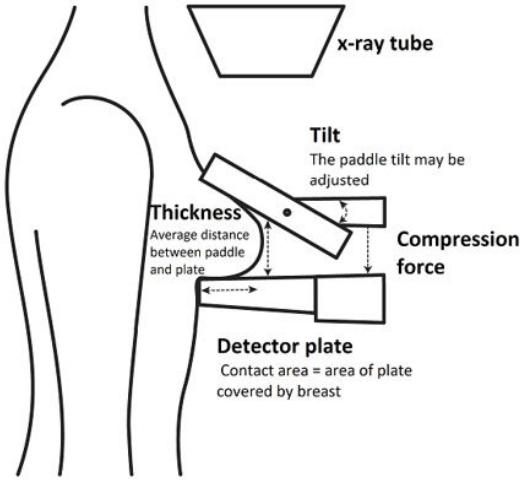


Figure 2.6 Compression of the breast during CC image acquisition schematic

Poulos et al (2003) in an observational study using repeat compressions on a sample of 114 women in Australia found that there was no linear relationship between applied compression force and compressed breast thickness (185). These findings, which were the first to suggest that force may not be the best dimension for calibrating compression, were subsequently corroborated in a number of larger studies including that by Branderhorst et al (2015)(176) who compared over 37,000 images from a breast screening programme in the Netherlands and 7,000 from USA and observed that thickness was correlated with breast volume and that force distribution was characterised by large variation. Waade (2017) in a study over >17,000 women in Norway confirmed that there was a negligible correlation between compression force and breast thickness (r=0.186) (172). These studies all suggest that breast volume (BV) rather than force is the key determinant of compressed thickness in breast screening settings.

The degree of force required to achieve a thin layer of breast at the plate is therefore correlated to the size of the breast and in practice there appears to be a strong observed correlation between breast size and thickness as illustrated by the heatmaps below Figure 2.7 (taken from initial work on this thesis, see also Chapters 4 and 7) but very weak correlations between thickness and force.

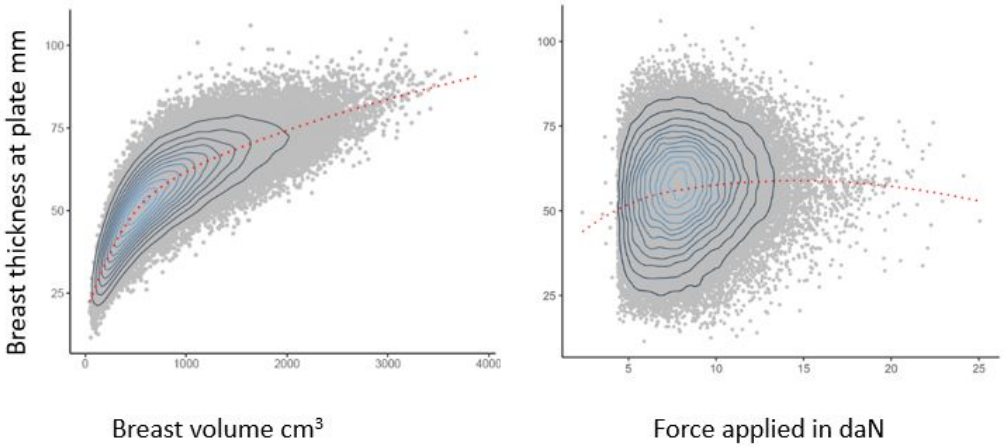


Figure 2.7 Heatmaps showing correlation between breast size and thickness during compression

It is therefore logical that force standardised protocols may result in a relative under-compression of larger breasts and greater resulting breast thickness. This led to suggestions from de Groot (2015) that 'pressure', (force divided by the area of contact between breast and detector plate) may be a better measure of compression, because it takes account of the size of the breast as well as the force applied (186, 187).

Moshina (2018) also showed that individual breast characteristics, such as %BD, are also correlated with compression force, pressure and thickness (188) which adds weight to the argument that force alone is a very crude parameter upon which to base a standard for breast compression. It is important to highlight that, until very recently, there was no possibility of assessing pressure in real-time during the mammographic process and similarly the only estimate of breast volume available during the imaging process, was that based on the mammographer's own observations. The introduction of new tools for measuring and monitoring mammography in real-time, means that, in future, it may be possible and practical to utilise pressure measurements in a screening environment.

Force and pressure and cancer detection

In order to validate any proposed new breast compression standards, there is a need for a better understanding of the relationship between compression, as measured by force and/or pressure and cancer detection, particularly in the context of screening performance. Since work on this thesis started, other studies have begun to explore the association between compression force, pressure and screening performance. The first research in this area on a large study population, was undertaken in 2016 in the Dutch breast screening programme using automated measurements from over 100,000 screens. Holland et al. challenged the view that greater compression pressure is more

effective, finding improved positive predictive value (PPV) and cancer detection when 'moderate' pressure was applied (189). The same authors also later found that screening sensitivity (based on interval cancers) was significantly lower in the highest pressure compression quintiles (190) . A similar Norwegian study by Moshina et al. (2017) also found that compression pressure was positively associated with interval cancer (191). However, a recent UK study by Hill et al. (2022), which used a different design, appears to contradict these findings. They found that pressure measured at initial screen was a significant predictor of interval versus screen detected cancers, with higher pressure being associated with a lower risk of interval cancers (192). The studies were not directly comparable but suggest that the exact nature of the relationship between pressure and cancer detection is still not well understood.

It was not the purpose of this thesis to look at the role that positioning plays in successful mammography but any approach that ignores positioning, risks compromising the performance of screening, as illustrated by Taplin et al 2002 (193) who found that interval cancer cases failed the positioning criteria significantly more than screen detected cancer cases and emphasised the importance of positioning for detection. Generally, there is a shortage of empirical evidence regarding positioning although manufacturers are developing real-time tools to objectively assess positioning during screening, and these may be deployed in future.

Paddle design and paddle tilt

The optional, flexible paddle, was introduced by equipment manufacturers to make the process of mammography more comfortable, allowing the practitioner to 'tilt' the paddle using a hinging mechanism, to accommodate the shape and size of the breast. The effectiveness of these flexible paddles for pain reduction was initially queried by Broeders (2015) following an observational study in the Netherlands on 288 women whereby compressions were undertaken using first flexible, then rigid paddles with no difference in pain reported (194). This finding was also confirmed more recently (2019) by Moshina et al. in a larger study (n=4675) in Norway (195). Interestingly Broeders also noted that when 3 radiologists assessed the quality of the resulting images in the Dutch study, they found that rigid paddles showed more breast tissue because the flexible paddles tended to push more tissue towards the chest wall, the rigid paddle images also had better contrast. This was a relatively small study, and the image quality assessments were subjective but it suggested that the assumption that flexible paddles can be used without drawbacks should be challenged (194). There may be other technical issues associated with paddles for example in a calibration exercise mammography machine readout for breast thickness was less accurate when a flexible paddle was used than when rigid paddles were used (196) and Ma et al (2021) recently suggested that flexible

paddles require a longer settling time before the image exposure to avoid blur which may be detrimental to image quality (197).

Pain during breast compression

In parallel with the role that pressure plays in reducing breast thickness, some studies have looked at the possibility of pressure-based protocols for reducing pain in mammography. de Groot (2013) experimented with a pressure standard of 10kPa in a small study on 196 women attending the Dutch breast screening programme. They suggested that a pressure standard may reduce the number of women reporting severe pain without compromising the image quality as assessed by a team of experienced radiologists (198). There is some evidence from a systematic review by Whelehan (2013) (199) and a subsequent UK study by Meyer et al. (2014) (200) to suggest that women who experience more pain at mammography are less likely to re-attend with potential consequences for breast screening effectiveness, but beyond this we know very little about whether the lack of compression standardization or the selection of flexible or rigid paddle type has any consequences for screening programme effectiveness.

Conclusion

Whilst acknowledging that there have been some important recent innovations since work on this thesis began, it is still hard to disagree with Hogg's assertion in 2013 that "given that mammography is well established there is surprisingly little published empirical research into techniques for performing it" (201). There is a clear need for more research in this area, focusing upon the association between mammographer's imaging choices and the cancer detection performance of screening programmes.

2.10 New technology and new opportunity

Recent technological developments offer new opportunities to improve and tailor our screening capability. These developments provide large-scale automated empirical data on mammographic density, breast composition and also on the image acquisition parameters that may be associated with image quality and cancer conspicuity. Chapter 4 Methods section 4.3, provides a summary of comparative findings on different methods of density assessment and the justification for the use of the Volpara (Volpara Health Technologies, Wellington, New Zealand), tool used for the main studies in this thesis.

Most personalized BC risk models now include BD as a factor (47, 202) and in the majority of US states, it has become a legal requirement for clinicians to inform women of their BD (203), in large part due to the "Are You Dense" campaign. In future BD may be used to stratify women with a

higher inherent risk and masking risks offering these women a more frequent or supplemental screen. Thus, a better understanding of what these tools can offer in practical settings is required.

The scope of this thesis is to carry out studies, using this technology in a large population-based screening setting where even small improvements to the sensitivity and specificity of mammographic screening can make a positive difference to the balance of benefits and harms associated with population-based breast screening programmes.

3 CHAPTER 3 RATIONALE AND RESEARCH QUESTION

3.1 Rationale

The current academic consensus is that breast cancer screening is an important strategy for tackling the burden of breast cancer in developed countries, enabling earlier diagnosis of non-symptomatic tumours followed by early treatment. Whereas treatments have improved greatly since organised breast screening began in the late 1980's, little has changed in the way that screening mammography is undertaken notwithstanding the change from single to double reading and the migration from analogue to digital imaging. Measurable improvements in performance have been limited. Despite this, mammography is currently still the best tool that we have for screening in large population programmes, but the balance of benefits and harms of mammographic screening is constantly, and rightly under scrutiny. It is imperative that we continually look for areas of possible improvement, even a modest reduction in the number of false negatives for example would have a non-negligible impact on the performance of mammographic screening given the large numbers of women invited in population-based programmes.

It is the overall aim of this thesis to look at the potential role of automated breast composition measurement tools in improving the performance of breast screening programmes. These tools are now sufficiently evolved so that real-time estimates of breast composition can be provided in screening settings. This opens the possibility for both interventions during the mammographic acquisition process and also for larger scale population risk stratification based on breast composition measurements. Large scale research into the use of breast density assessment is already being undertaken but it is the aim of this thesis to look at some more novel ways that these measurements could be used as tools in the quest to 'tilt the balance' between the benefits and harms of the breast screening programme in developed countries.

3.2 Research questions

The performance of mammographic screening can potentially be improved by better targeted screening programmes. This may involve increasing the frequency of screening for those at greater than population risk or those at risk of more aggressive cancers. Alternatively, or additionally, it may involve supplemental screening with other modalities such as MRI, for those who have greater risk of tumour masking as a result of increased breast density. The stratification of population-based screening programmes would rely on assessment of women's risk of developing BC based on BD in conjunction with information on other genetic and non-genetic risk factors (e.g., age, family history, reproductive history, BMI). However, the ability to perform BD assessments within screening

programmes is compromised in the UK by lack of information on BMI for women screened. Therefore, my first research question is:

1. "Can automatically calculated breast volume or non-dense volume estimates be used as a proxy for BMI in settings where BMI measurements are not available?"

Screening performance can be improved by reducing the number of cancers missed by screening, i.e. false negatives, such that: a) there is an improvement in BC mortality because more lives are saved through earlier diagnosis b) there is an improvement in morbidity because treatments are less invasive though earlier intervention. To be effective this must be achieved without increasing the number of women undergoing unnecessary extra testing because of a recall for assessment at screening.

My second research question focusses on a novel breast comparison made possible by new automated breast composition tools. I ask whether left breast versus right breast asymmetry analysis can be used to identify women at higher risk of BC or to identify those at risk of having their cancer missed during mammography:

2. "Is left-right breast asymmetry in breast volume or breast density associated with diagnosis of screen-detected cancers or false negatives at screening and is higher asymmetry a cue that the cancer may be of a more aggressive subtype?"

Thirdly, this thesis looks the process of mammographic image acquisition. By improving our understanding of the associations between objectively measurable image acquisition parameters (e.g., compression pressure, force and paddle tilt) and screening outcomes, mammography guidelines might be improved benefiting screening performance.

3. "Are objectively measurable breast compression techniques associated with risk of false negative outcomes at screening?"

These studies have been made possible by the availability of tools for large-scale, automated, real-time image analysis. These tools make it feasible, for the first time, to automatically analyse all images taken at breast screening and thus open-up the possibility of new screening interventions. Tools were initially introduced for the purpose of automatically estimating breast density but as a 'by-product' they also provide objective high-volume data on breast imaging compression parameters and also breast composition data from which asymmetry measures can be calculated.

My studies were among the earliest that utilise these by-products in novel ways beyond their original purpose, which was to estimate breast density.

Specific objectives within these three research questions, and how they were addressed, are outlined at the start of the relevant chapters, i.e. Chapters 5, 6 and 7.

4 CHAPTER 4 METHODS

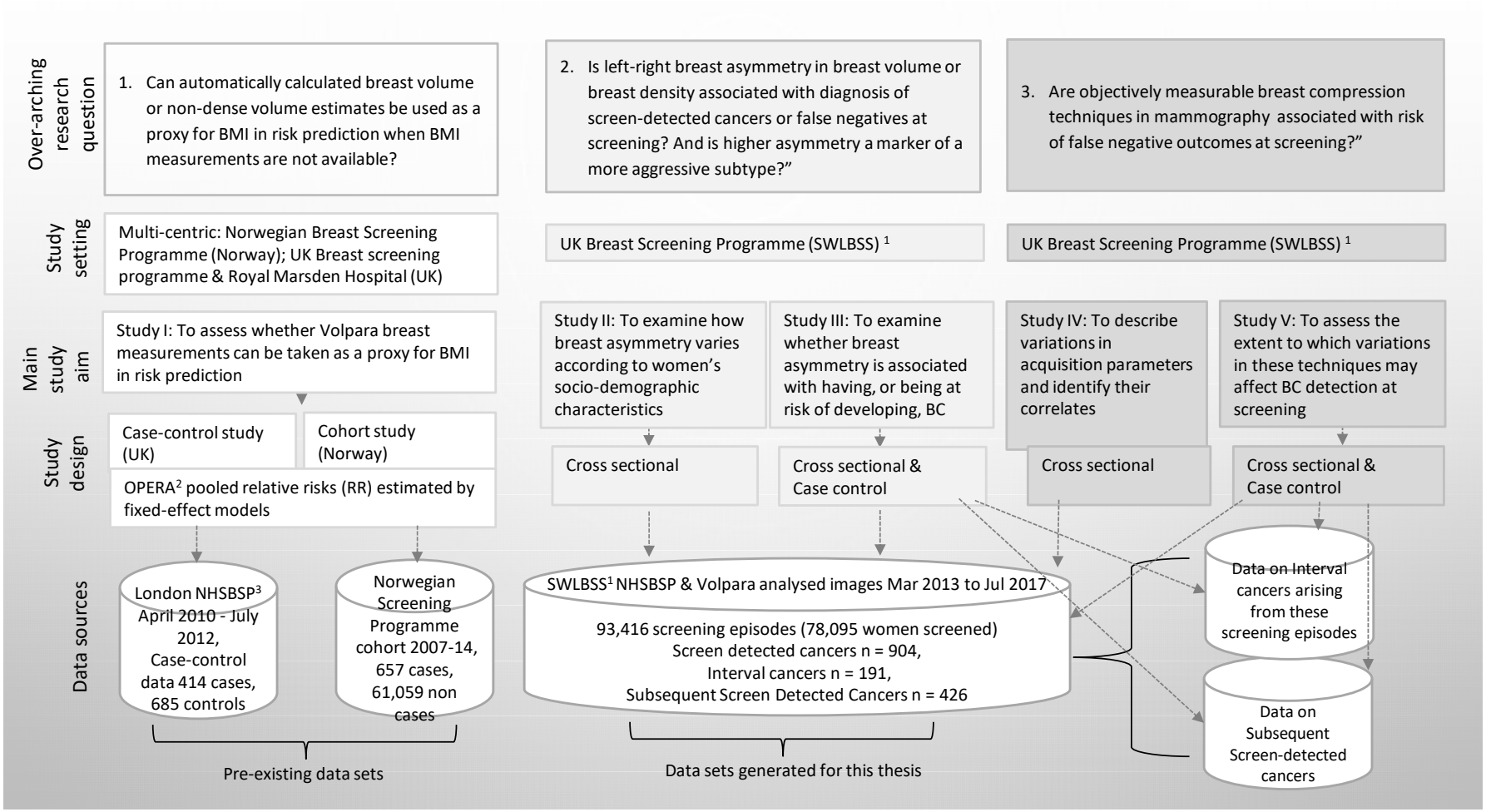
4.1 Overview of aims and studies and the data sources required

In order to address the aims of this thesis, large-scale data collection, sampling and cleaning was required for different study objectives and a specialist tool was required for analysing mammographic images. Figure 4.1 provides an overview of the studies conducted and how the research questions (aims), studies, methods and datasets are related for the purposes of this thesis.

4.2 Data sources and collection methods and timelines

For aim 1, which addresses the question of whether automatically calculated breast volume estimates can be used as a proxy for BMI, I examined previously collected data from two studies. The first was a UK case control study and the second a Norwegian cohort study. Further details are provided in section 4.8 and Chapter 5.

For aim 2 and 3 (studies II to V in this thesis) a large-scale data collection exercise was designed to provide sufficient evidence to be gathered across an ethnically and socially diverse screening population to inform the studies. The data capture was designed to include all the key exposure, confounder and outcome measures that were required for the planned research (section 4.6) and involved setting up the data capture from scratch at the start of the thesis. It was important to ensure that these data could be collected without disturbance to the day-to-day screening mammography process. Data from approximately 100,000 screens were required and, at the time these studies were designed, there had been no UK-based data collection and analysis using automated mammographic analytical software on this scale. The data collection overview and timescales are summarised in Figure 4.2.



Footnotes:

[1] SWLBSS South West London Breast Screening programme run by the St Georges Hospital NHS Foundation Trust.
[2] Odds expressed per adjusted standard deviation (OPERA) to allow comparison of same analysis on two different study populations.
[3] NHSBSP National Health Service Breast Screening programme for England and Wales. Mandates standardised data collection for all screening episodes.

Figure 4.1 Overview of aims, studies, methods and data sources

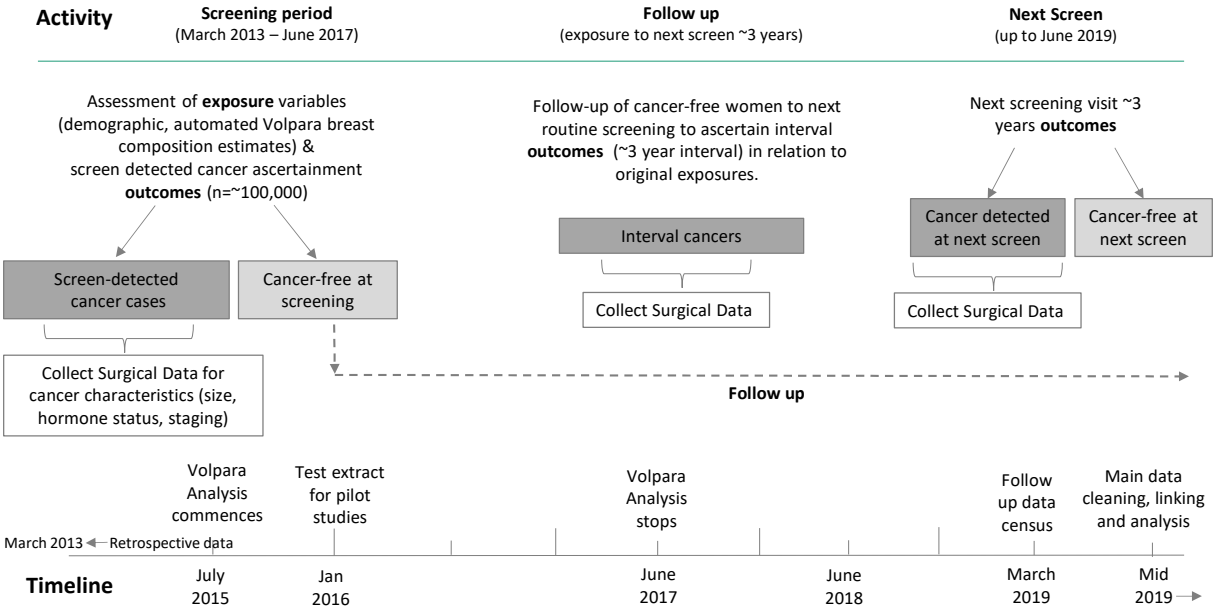


Figure 4.2 Timeline for data collection of exposures and outcomes (including cancer diagnosis) for studies II-V

4.3 Selection of an automated analysis tool for measurement of exposures of interest

Requirements

Mammographic screening in the UK utilises full-field digital mammography (FFDM) and the resulting images may be processed using automated algorithms, which provide comprehensive image acquisition data as well as volumetric estimates of breast composition. However, currently, the screening programme does not use mammographic density as a diagnostic aid, and participants are not informed on whether they have dense breasts.

Breast screening is a high-volume, time-limited procedure, with each examination allocated approximately 10-12 minutes. For large-volume studies such as those in this thesis, a tool was required that could be introduced without any disturbance to the standard mammographic screening process. The key outputs required were objective mammographic estimates of breast volume (BV), non-dense volume (NDV) and dense volume (DV) together with mammographic density percentage (%BD) and a conversion to the BI-RADS 4th and 5th editions scale (see section 4.6). To calculate asymmetry measures, these values were required for each individual mammographic view and each breast (i.e., at least 4 sets of values for each mammographic examination) with these individual values being also used to calculate an average score for the whole screening round examination. Similarly, to assess the associations between objective image acquisition parameters and outcomes, values for force, contact area at detector plate, thickness, dose and paddle tilt were required for each breast side and mammographic view. Some of these acquisition values are already

available in the Digital Imaging and Communications in Medicine image header (DICOM an agreed protocol for information exchange in medical imaging) but it facilitates retrieval of the data for analysis if they are also available in one consolidated output file. The selected tool was required to perform well in terms of its validity (i.e., accuracy in estimating breast tissue volumes against a 'gold standard' and ability to predict subsequent BC risk), reliability and reproducibility and preferably be FDA approved.

Technology overview

A review of the technology available at the conception of this thesis was conducted (in 2014-15), to identify tools that provided quantitative results automatically in real time using a standalone package that had also been validated and FDA approved at that time. Figure 4.3 summarises the major breast density assessment tools available at the time the studies in this thesis were designed and undertaken. At that time there were four main choices of automated volumetric assessment tools available: SXA (204) uses a single x-ray absorptiometry method but requires the inclusion of a breast-tissue phantom in the imaging process as a reference for calibration. Cumulus V uses x-ray attenuation parameters and thickness to assess dense volume (205). For each mammography system however prior calibration is required, which makes them more difficult to use in a high throughput setting. Quantra is produced by Hologic (Hologic Inc, MA, USA) (206) and Volpara is produced by Volpara Health Technologies, New Zealand (207) and both products are commercially available. These methods work in similar ways using pixel intensity in the raw images and known x-ray attenuation properties. They output both %BD, absolute DV and categories mapped to the BI-RADS classifications offering reporting flexibility. Volpara also reports a wide range of image acquisition parameters such as force, paddle tilt and dose. Quantra and Volpara do not require phantoms or individual calibration, which makes them more practical in a large-scale screening environment.

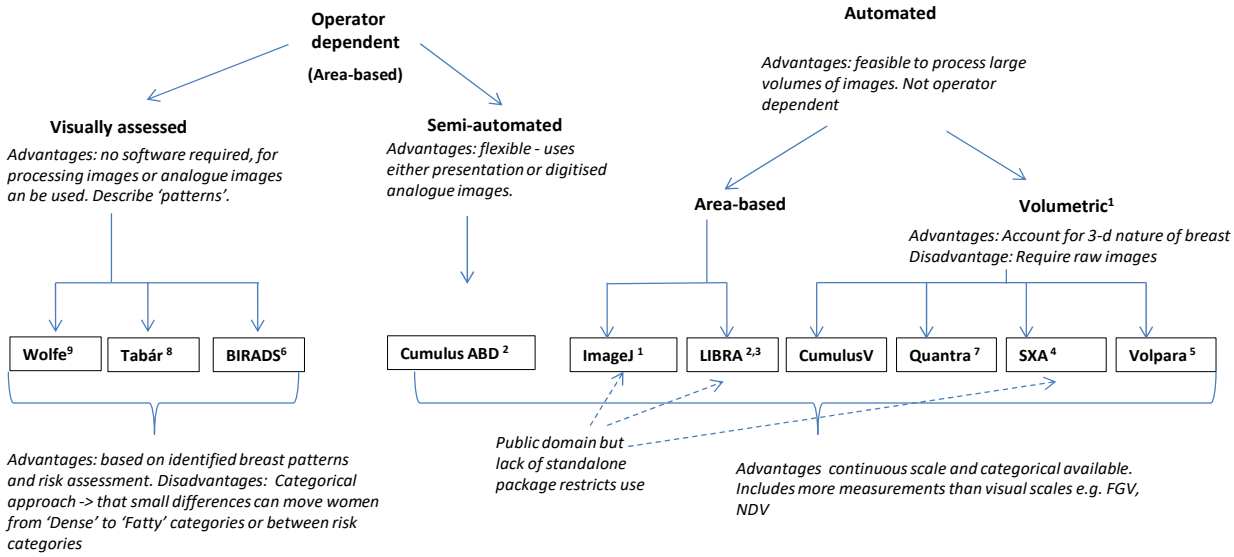


Figure 4.3 Types of image analysis

Footnotes:
[1] ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA. Public domain Java software [2] Active threshold method Cumulus software, University of Toronto, Toronto, Ontario, Canada). [3] Laboratory for Breast Radiodensity Assessment (LIBRA) – Publicly available. [4] SXA single energy x-ray absorptiometry SXA analysis software package (UCSF, San Francisco, CA) [5] Volpara® Density™ Trademark of Matakina Technology. [6] American College of Radiology (American College of Radiology 2013) [7] Quantra™ Hologic Inc. 35 Crosby Drive Bedford, MA – can report area or volumetric density. [8] Tabár groups based on anatomic-mammographic correlations. Scale I-III High risk to IV and V Low risk. [9] Wolfe determined by mammographic parenchymal pattern - Wolfe classification.

It was concluded that the algorithm Volpara® DensityTM version 1.5.11, (Matakina Technology Limited, Wellington, New Zealand) (208) (Volpara), which processes raw digital mammographic images using a fully automated algorithm, was logistically well suited to our objectives and was furthermore available for use in the screening location where the studies were to be conducted.

Comparative literature review Volpara

A more detailed literature search, using the search term "Volpara" was conducted in 2015 for all available literature on more scientific aspects of this particular algorithm, at the time of study design. 58 references were initially found, and Appendix D Table D.1 provides a summary table of relevant findings. This search was repeated in 2019 but research findings post-2015 were not available when initially selecting the research tools.

Early studies by Brand (2014)(209) and Alonzo-Proulx (2015)(133) found that Volpara gave objective and reliable volumetric BV, plate contact area and DV estimates, this was later supported by Holland et al (2016)(210) who found that Volpara showed higher inter-exam agreement on the Breast Imaging and Reporting Data System (BI-RADS) density scale. The Volpara grade (algorithm version 1.4 and 1.5.1) showed good agreement with radiologists' performance using the BI-RADS 4th Edition classification in several different settings (211-213). Wang et al in 2013 (214) compared SXA,

Quantra and Volpara (1.4.3) with MRI assessments of density, which are considered the 'ground truth' for assessing accuracy of these estimates. They found that Volpara showed the strongest correlation to MRI. Gubern-Merida et al (2014) (215) similarly found a strong correlation between Volpara (1.4.3) estimates and MRI measurements.

Later case-control studies by Astley et al (2017,2018) (216) that looked at the association between different mammographic density measures and breast cancer detection after adjusting for all available reproductive, BMI and age risk factors, found that visually aided assessment of BD (double reading with Cumulus), demonstrated the strongest relationship with BC detected at screening and at subsequent screen or as an interval cancer, but that Volpara %BD provided the strongest relationship amongst the automated measures included in the study (Volpara, Quantra, Cumulus, Densitas).

In conclusion, visually aided techniques using Cumulus have been shown to be good predictors of BC risk and some studies found that they were better at discriminating cases from controls than Volpara (216, 217), whereas other studies found that Volpara performed better in terms of discrimination (132). The semi-automated tools may have performed differently in different settings because they depend upon operator training and are hence subjective, they were also found to be time consuming and not practical in a real-time screening setting. Volpara is objective has been found to be practical in clinic settings (0% failure rate to process images in one study where other tools systematically rejected images (132)), in addition it was one of the few automated tools available at the time of this study that had FDA approval on all the platforms used by the NHSBSP.

The literature review found that there were also some potential drawbacks with Volpara. Seo et al. reported that despite generally good agreement with visual BI-RADS there was decreased reliability in breasts with scattered density (218). There are also suggestions that the when using flexible paddles, the thickness (196) and dose (219) calculations may be compromised. The automated algorithm is not also designed to make estimates in situations where there were not exactly four images taken (208), which restricts its usefulness for analysing very large breasts when extra (mosaic) images are required. However, on balance it was found to be a good, objective and comprehensive tool and available and practical for the study setting of my research.

Summary

In summary, our decision to implement Volpara (version 1.5.1) was based on independent validation of the algorithm that had been peer reviewed at the time our work began. It was considered that Volpara outperformed subjective measurements in terms of reliability and reproducibility and critically allowed us to assess large volumes of women in real-time (133, 210, 214, 220, 221) providing that raw images are available. It performed as well or better than similar tools in terms of reliability and reproducibility of volumetric estimates and showed a good correlation with the BI-RADS (4th Edition) standard. Therefore, we concluded that this algorithm provided a realistic tool for screening settings and that, with certain known caveats Volpara (version 1.5.1) could be used to estimate all the main parameters that were required for our studies including the BI-RADS 4th edition categories. Furthermore, it provided comprehensive data on individual images (left and right breast) and different mammographic views (CC and MLO) which allowed for the straightforward calculation of novel data measures which were required for this thesis.

How does Volpara work?

The software estimates the relative volume of fibroglandular tissue in comparison to the whole breast. It first identifies the breast area (and removes the pectoral muscle in the MLO view) then it calculates contact area (cm2) between breast and detector plate. It uses an internal reference value by identifying a pixel value that represents 100% fatty tissue and uses this as a relative measure for all other pixels. The algorithm uses the known x-ray attenuation coefficients for fat and dense tissue at a given particular thickness, for given x-ray tube and voltage combinations. It calculates thickness of dense material at any point over the detector plate according to the absorbed radiation for that pixel using these tables. Total dense volume is estimated by integration of thickness over the image. The breast volume is estimated by thickness multiplied by contact area.

These calculations are used to provide estimates (in cm3) of the volume of the breast (BV) and the volume of the radio-dense tissue (DV). Estimates are provided separately for each of the four (left/right CC and MLO) images and as an average across all four images. The non-dense volume (NDV) is estimated by the algorithm (as BV - DV) for each image and as an overall average and the % dense volume estimated as the ratio of DV to BV. The algorithm provides score of 1 to 4 which is an estimate of the BI-RADS (Breast Imaging Reporting and Data System) 4th Edition classification for mammographic density. The 4th Edition categorises the breast density depending on quartiles of, visually assessed, fibroglandular tissue: 1, almost entirely fatty; 2, scattered areas of fibroglandular density; 3, heterogeneously dense; and 3: extremely dense (222). Volpara also provides an estimated value for the 5th edition of BI-RADS, published in 2013, which redefined the density

categories to exclude percentages of dense area and instead to describe the distribution on the basis of possibility of having an obscured lesion: a, the breasts are almost entirely fatty; b, scattered areas of fibroglandular density; c, the breasts are heterogeneously dense, which may obscure small masses; d, The breasts are extremely dense, which lowers the sensitivity of mammography (92).

Volpara also reports quantitative information on 'how' the image was taken including force, dose, thickness, and paddle tilt, for all individual sides and views. Paddle tilt is reported as a computed slant angle (degrees from horizontal) and computed slant in mm i.e., the difference in thickness between the breast at the chest wall and at the nipple as a result of the slant.

4.4 Assessing the validity and reliability of asymmetry estimates

The technology review found that Volpara results have a high correlation with volumetric measurements from more sophisticated imaging methods such as CT and MRI and are highly repeatable. However, the literature search found no research into the validity and reliability of breast *asymmetry* measures derived from Volpara estimates. Since objective mammographic (Left versus Right side) asymmetry estimates are an important exposure of interest for this thesis, a preliminary investigation was undertaken to gain a better understanding of the validity and reliability of asymmetry estimates. The objective was to determine whether any variation observed in the measurements of asymmetry is due to measurement bias or a random effect. The study and findings are detailed in Appendix E. The study concluded that asymmetry studies should exclude cases where asymmetry measurements are likely to be biased by known technical and clinical issues including:

- Technical recalls – imaging deemed not adequate by film reader,
- Examinations where just one side was imaged or the CC images were missing,
- Examinations where the Volpara algorithm rejected either the CC image or the MLO image (on the basis of its own internal checks (208))
- Examinations taken at non-routine events (e.g., assessment clinics),
- Examinations on women who had previously been diagnosed with BC,
- Examinations where multiple >4 (mosaic images) were required.

There was no evidence for systematic bias in the asymmetry estimates calculated using the outputs from the Volpara algorithm, but individual estimates of asymmetry were not as reliable as measures of breast density and breast volume reported in similar studies.

4.5 Study population for studies based in South West London

The study participants for the main studies (Chapter 6 and 7) undertaken as part of this thesis were women who underwent routine screening mammography as part of the England and Wales National Health Service Breast Screening Programme (NHSBSP) at the South-West London Breast Screening Service (SWLBSS) based in the St George's University Hospitals National Health Service (NHS) Foundation Trust. They were resident in the catchment of the SWLBSS which covers the following London boroughs: Wandsworth, Merton, Sutton, Croydon, Richmond and Kingston (see purple circle in Figure 4.4) . Participants were screened at least once during the period 01/03/2013 to 20/06/2017.

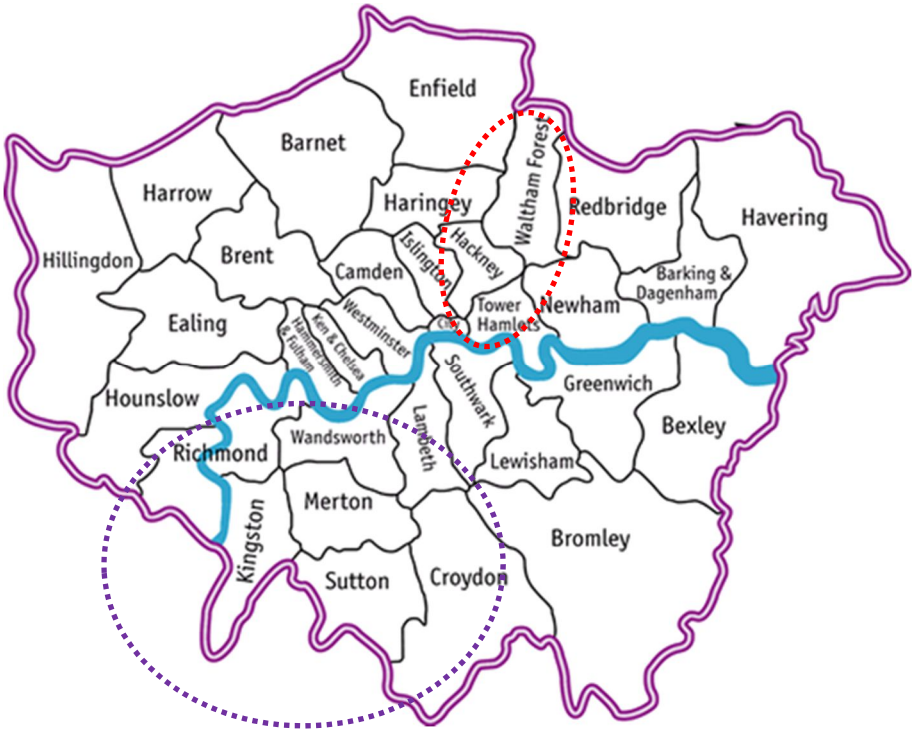


Figure 4.4 Boroughs of London indicating approximate catchments of South-West London Breast Screening Service and of Central and East London BSS

" Map of Greater London and the London Boroughs " by Ross Burgess used under CC BY 4.0 / Highlighter lines added to original.

The NHSBSP population-based mammographic screening programme invites all women aged 50-70 years once every 3 years and has a coverage of ~75% (223). The study population also included a small number of younger women (aged 29-45) who had been identified as having a higher risk of breast cancer and therefore were invited for screening on an annual basis (29), plus any women over 73 years who had optionally contacted the service for a self-referred screening appointment. All women were asymptomatic at the time of screening. Data are routinely collected as part of standard screening protocol; of particular interest are data associated with known risk factors for

breast cancer (age, ethnicity, previous breast cancer, previous biopsy for benign lesions) and clinical information on outcome of this screening, any related surgical interventions and pathology including tumour size, grade, and hormone receptor status. Data for other known breast cancer risk factors (e.g., parity, duration of breast feeding, age at menarche, body mass index (BMI), family-history of breast cancer) are not collected in a systematic way across the NHSBSP screening programme and thus were unavailable.

4.6 Exposures and outcomes of interest

Data on exposures and outcomes were collected for each screening event and additional data were recorded to allow the exclusion of ineligible women from the studies. Data were derived from two main sources, firstly the screening administrative system i.e., the National Breast Screening System (NBSS) which records data on the women invited for screening and their screening outcomes, (Tables 4.1, 4.4).

Table 4.1 NBSS screening data record of exposures

NBSS data field	**Description**	**Exposure (E) Confounder (C) or exclusion**
Screening Date	Date of screening procedure. Key linked to a screening image in Volpara.	
Age At First offered appointment	Woman's age in full years at date of first offered appointment. (Used so that age at screen is recorded without breaching PID rules[a]).	E/C
Date BC Diagnosed	Exists if woman has *previous* breast cancer diagnosed.	Exclusion
Episode Character	Type of screening event: First Call (not screened before) Higher Risk Non-routine Recall (unusual, for early recalls) Routine Recall (woman previously screened)) Self-Referral (normally older women)	E
Ethnic Origin [b]	A code from code 2001 ONS code list or prefer not to say. Collected at screening appointment.	E/C
Partial Mammography	Y if mammography was partial i.e. examination not completed.	Exclusion parameter
Prevalent Incident Status	Whether screen was first (P) or subsequent (I) screen	E/C
Special Appointment	Y or N or blank – used to indicate that a woman had unusual screen (possibly disabled or with implants)	Exclusion parameter
IMD 2010 [c]	IMD deprivation score of LSOA where woman lives based on 2010 deprivation levels. Low score = low deprivation	E

Footnotes:
[a]Patient Identifiable Data (PID) was removed from the data collection.
[b]Ethnicity was categorised according to the Census classification and summarised as, “Asian” (Indian, Pakistani or Bangladeshi or other), “Black-African”, “Black-British or Caribbean or other”, “Chinese”, “Mixed” (White and Black, White and Asian or any other mixed), “White” (British or Irish or other) and “Other” (224).
[c] Index of Multiple Deprivation was derived in a two-stage process via woman’s postcode which was linked to a Lower Super Output Area (LSOA) a demographic region with a population of ~1,500 that has a coherent socio-economic profile. The LSOA code was linked to an Index of Multiple Deprivation (225).

Secondly data was extracted from the SWLBSS Picture Archiving and Communication system (PACS) that records the technical details of each imaging event and the digital image files. As a joint project between the author, the SWLBSS Picture Archiving and Communication system (PACS) team, the SWLBSS IT team and the Volpara technical support team, a new data collection process was set up in 2015. Raw images from NHSBSP standard 2-view FFDM mammography were stored on the SWLBSS PACS. The images were taken on NHSBSP approved mammogram machines, which were almost exclusively Hologic Lorad Selenias (Hologic Company, Bedford, MA, USA) equipped with an x-ray tube with W(tungsten) anode material combined with Rh (rhodium) and Ag (silver) filter materials. A small number of images (~350) were taken on a Siemens machine. The make of the mammogram machine and its location were recorded in the database. All Hologic machines were equipped with Hologic flexible compression paddles. All raw mammographic images from 5 main screening locations were analysed using Volpara (version 1.5.11, macro version 5.0) (208), which provided automated estimates of all key breast volumetric estimates as shown in Table 4.2 itemised by view (CC or MLO) and side (L and R) plus averaged values for each side, view and overall examination level. Information about the imaging process were also recorded in the Volpara database including compression exposures (force applied, pressure and paddle tilt) and outcomes, (compression thickness and dose), see Table 4.3.

Table 4.2 Volpara breast volumetric estimates as exposures

Measurement	**Description**	**Volpara name**
Breast Volume (BV)	Estimated total breast volume cm^3.	BV
Dense Volume (DV)	Estimate of volume of non-fatty tissue in the breast cm^3.	Fibro-glandular Tissue Volume (FGV)
Volumetric Percentage Breast Density (BD%)	Ratio of fibro-glandular tissue volume to the whole breast volume as %.	Volumetric Breast Density (VDB)
Non-dense Volume (NDV)	Estimate of fatty tissue in the breast cm^3.	NDV

Table 4.3 Volpara Imaging parameters as exposures/outcomes

Measurement	Description	Volpara name
Breast Thickness (mm)	Derived from DICOM tags on mammography unit. The thickness of the breast impacts on the perceived image contrast.	Hmm
Compression Force (N)	The median compression force applied to the breast during exposure in Newtons (N). Derived from DICOM tags on mammography unit.	FN
Mean Glandular Dose	Mean glandular dose (MGD) i.e. manufacturers estimate of absorbed dose to the breast in milligray (mGy) (226).	MGD
Pressure Applied (kPa)	The median pressure applied in kilo Pascals (kPa) determined from the recorded compression force (N) per unit of contact area (m2) calculated by the VolparaDensity algorithm.	PkPa
Tilt (degrees)	The angle of the compression paddle from the horizontal in degrees.	SlantDeg
Mammographer ID	The ID of the practitioner taking the mammogram as recorded but anonymised before data was analysed	

Raw images were continuously processed using the Volpara algorithm between 01/06/2015 and 20/06/2017 and the archive of raw images from 01/03/2013 to 01/06/2015 were also processed by the algorithm during evenings and weekends. An intermediate Volpara datafile extract was made in January 2016 for women screened in the period 01/03/2013 to 09/07/2015 in order to carry out initial descriptive and reliability analyses (Paper II and Appendix E) in advance of the case-control studies.

Image analysis using Volpara ceased on 20/06/2017 and the final data extract included 94,408 screening examinations. All data cleansing, deduplication and linkage to national data sets was carried out by the author.

Outcome ascertainment

For each completed set of analysed images (representing one screening visit or 'examination'), the mammogram record was linked to the relevant NBSS record to ascertain the outcome of the screening (recalled for assessment, cancer, normal, interval cancer) and the laterality, subtype and size of any cancer detected. For all screen detected and interval cancers, surgical data is entered into the NBSS system, although there is a time lag between diagnosis and final surgical record entry and the data for interval cancers is less complete and less comprehensive than for screen-detected cancers.

All cancers detected at screening are registered immediately in the NBSS system, but collection of non-screening cancers relies on a national register, the English National Cancer Online Registration

Environment (ENCORE), from which a download to screening services is periodically made available. This is a high-quality source but the interval between reporting to the ENCORE system and registration in NBSS is dependent on local resourcing outside the control of my work. Therefore, it was not possible to design cohort studies based upon this data collection. It is also recognised that potentially there may be a small number of women who develop BC after leaving the UK and hence who are not recorded in ENCORE (lost to follow up) and these women may be predominantly of non-White ethnicity.

The following outcomes were ascertained and linked to the relevant baseline screening episode where baseline exposures were measured.

Table 4.4 Key Screening Outcomes – from the National Breast Screening System (NBSS)

Outcome	**Description**
Cancer detected at screen	Y or N
Diagnosis date	Date the screen detected cancer was diagnosed
Assessed at screen	If woman had any assessment tests at the initial screen Y or N
Needle biopsy?	Y if fine needle aspiration or wide bore needle or vacuum aided excision were carried out at assessment
ER Status	Oestrogen hormone receptor status of any cancer detected
PR Status	Progesterone hormone receptor status
HER2 Status	Human Epidermal Growth Factor Receptor 2 hormone receptor status
Interval cancer date of diagnosis	Date the interval cancer was diagnosed
Invasive tumour size	In mm
Non-invasive tumour size	In mm
Interval cancer type	One of the following codes (from code list ICT): 0 = Unclassifiable 1 = Normal/benign 2 = Uncertain 3 = Suspicious
Interval cancer size mm	Size of the lesion identified in the previous screening mammogram (if the interval cancer type is 3 – Suspicious)
Tumour laterality	Left Right or Bilateral

A pragmatic cut-off date of March 2019 was identified and at this date all available surgical data were extracted for screen-detected cancer cases using standard NBSS reports. The time lag following diagnosis gave enough time for most of the cases to have completed their initial treatment and surgical and pathology records to have been updated. The data was cleaned and linked to the original baseline screening record collected at the time exposure measures were taken and all anonymised data was stored in a Master data file in Microsoft Access® on the SWLBSS server.

At this cut-off date the NBSS database was scanned for any subsequent interval cancers that had been reported for women in the baseline dataset i.e. for women who had exposures recorded and went on to develop symptomatic BC in the interval before their next screen. These were identified and extracted using NBSS standard reports. Also, at this date the database was scanned for any subsequent screens that had been carried out on women who were in the baseline data set. 'Next round screen detected cancers' i.e., subsequent cancers were identified and extracted using special data extracts created using a reporting tool available to NBSS users. These data were stored as Microsoft Excel® files on the SWLBSS server.

Out of a total of 93,416 screening examinations, a total of 904 screen-detected cancer cases were diagnosed at the original, baseline screen. 191 interval cancers, i.e., those that are diagnosed in the interval between screens had been recorded at the census date of March 2019 and 426 cancer cases had been diagnosed in subsequent screens for women (who previously had a baseline screen) by that census date. A further interrogation of the NBSS data was made in June 2019 to collate latest surgical data entered for all cancer cases in the study population.

4.7 Eligibility and exclusion criteria for data collected specifically for this thesis (studies II-V)

For studies II to V (described in Chapters 6 and 7) a total of 101,757 examinations (Volpara 'studies') were analysed using the Volpara algorithm. (An examination or Volpara study includes the whole set of images pertaining to one imaging event). Not all examinations however were eligible for every objective in this thesis. Examinations were excluded if the event was identified as a non-screening event e.g., an assessment image or a training or symptomatic use of the machine (n=7,349). Examinations were excluded if the episode type was missing because this made it impossible to differentiate between assessment and screening records (n=992). This left 93,416 contemporaneous screening examinations which took place during the study period (Figure 4.5).

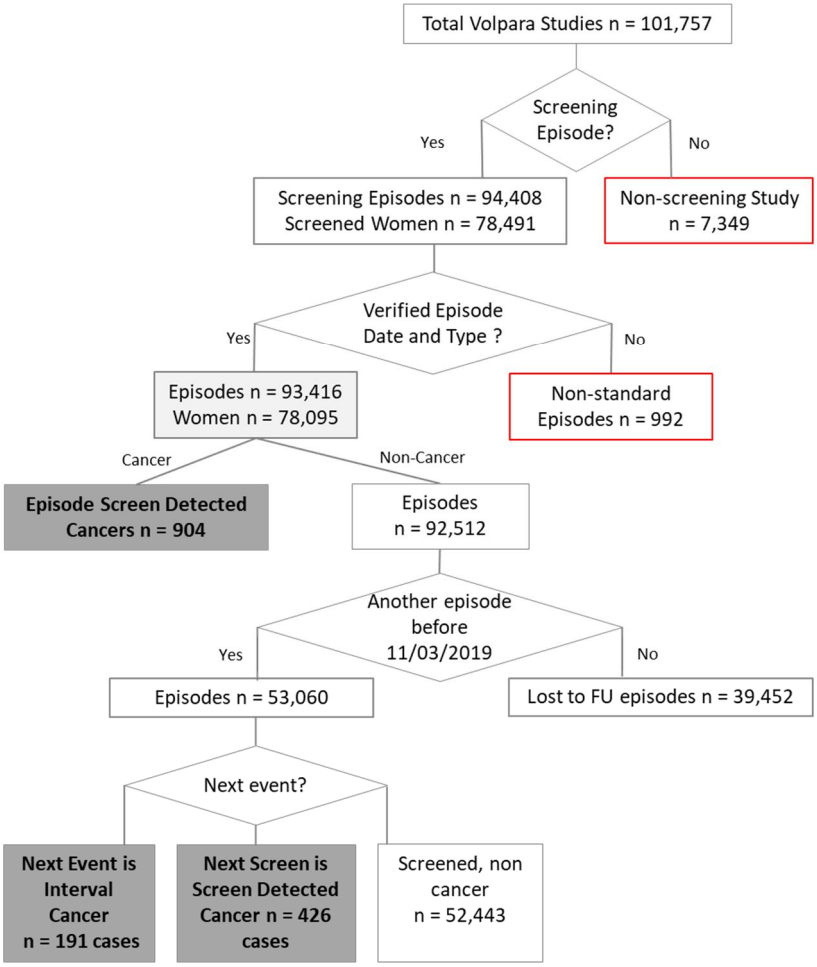


Figure 4.5 All examinations processed by Volpara algorithm – images acquired between March 2013 and July 2017. Exclusions in red.

4.8 Study populations for Norwegian Cohort study and UK Case control study

For the study on proxy measurements for BMI (Chapter 5), I arranged access to previously collected data from 2 different European countries. The study participants were drawn from a Norwegian cohort study and from a UK-based case control study. The Norwegian cohort were women invited for breast screening by BreastScreen Norway, the 2-yearly Norwegian Breast Screening programme, administered by the Cancer Registry of Norway and living in one of 4 Norwegian counties (Hordaland, Rogaland, Akershus, and Trøndelag). Between August 2007 and August 2014. 61,716 women, who attended for screening, completed a questionnaire that recorded exposure to standard BC risk factors (at the first screen) and current exposures at subsequent screens. All women in the study had standard two-view (mediolateral oblique and craniocaudal) FFDM of each breast. Exposures recorded included age, country of birth (Norway/non-Norway), menopausal status, parity, age at menarche, MHT use (yes/no), breastfeeding (yes/no), education level and BMI. In addition, all

women attending for their first screen had their breast volumetric measurements including absolute (DV) and %BD estimated using Volpara software carried out by the Norwegian team (see Table 4.3).

The UK cases were women who were newly diagnosed with BC in the Royal Marsden Hospital London, between April 2010 and July 2012. Controls were women who attended routine 3-yearly screening at the Central and East London Breast Screening Service (CELBSS), (see Figure 4.4 marked in red), during the same period and were found to be breast cancer free. CELBSS is part of the NHSBSP (see section 4.5) and exposures were ascertained as described in section 4.6 above and itemised in Table 4.2. Additional exposures to BC risk factors were collected by questionnaire at the time of screening for controls and after diagnostic confirmation for cases. Exposures recorded included age, ethnicity (summarised as White/non-White), menopausal status, parity, age at menarche, MHT use (yes/no), breastfeeding (yes/no), education level and BMI. All cases and controls had had their breast volumetric measurements including absolute (DV) and %BD estimated using Volpara software, by the UK study team (see Table 4.3).

4.9 Study designs overview

The detailed objectives of each study area are described in relevant chapters (5,6 and 7) and the relevant study designs are described, which take into account the nature of the data available and the timing of the different exposures and outcome data collection.

4.10 Ethical approval

For Study I The UK study was approved by all relevant ethics committees (Research Ethics Committees from the Royal Marsden Hospital, the Barts and the London NHS Trust, and the London School of Hygiene and Tropical Medicine). Participants provided written informed consent. The Norwegian study was approved by the Regional Committee for Medical and Health Research ethics in the South-East Health Region of Norway.

Studies II to V were carried out on fully anonymous, routinely collected data only, held in accordance with the NHS Cancer Screening Programmes Confidentiality and Disclosure Policy 2011 (227). The NHSBSP has section 251 support under the NHS Act 2006. The research protocol was approved by all relevant ethics committees (Research Ethics Committees from St George's University Hospitals NHS Foundation Trust, and the London School of Hygiene and Tropical Medicine). During the period of this work, the author held an honorary contract at SGH.

5 CHAPTER 5 Automated Breast Measurements as a proxy for BMI

5.1 Introduction

Mammographic breast density is a strong BC risk factor and offers the potential (in combination with other risk factors) to target screening strategies. It has traditionally been assessed based on the relative amount of dense *area* (white pixels) and fat (translucent pixels) on the mammogram. %BD calculated in this way is negatively correlated with BMI (127) which is itself a BC risk factor (positively associated with risk in post-menopausal women but negatively associated with risk in pre-menopausal women) (21). Therefore, any study that fails to control for BMI is likely to produce a biased estimate of the magnitude of the breast density – breast cancer risk association (115). Likewise, any risk assessment tool that attempts to stratify large screening populations based on routinely collected data will be compromised if BMI data is not available. In the NHSBSP no data on height, weight or BMI is routinely collected and no alternative measures of adiposity are recorded.

This study was designed to ask if, should BMI data not be available, a suitable adjustment can be made by using another characteristic of the mammographic image? The literature search (Chapter 2) found strong evidence that area-based estimates of breast size are positively correlated with BMI and have the potential to be used as a surrogate for BMI. There was a consensus that volumetric measurements may be a better proxy but there had been no formal empirical study that tested the validity of this approach. This study was therefore designed to assess whether automated measurements, such as non-dense volume (NDV) or breast volume (BV), could act as a proxy for BMI in a BMI-adjusted model that predicts BC, if BMI measurements were missing and to assess how well this proxy-adjusted density measure predicts BC.

5.2 Aims and Objectives

Aims:

To examine whether BMI is an important confounder of the relationship between *volumetric* breast density (absolute or percentage), as estimated by Volpara, and breast cancer risk.

To examine whether volumetric BV and NDV estimates can be used as a proxy for BMI when data on the latter are not available.

Primary Objectives:

1) To compare the strength of the associations of absolute breast dense volumes (DV) and %BD with the odds of BC with and without adjustment for BMI.

2) To compare the strength of the associations of volumetric DV and %BD with the odds of BC adjusting for BMI or one of its potential proxies, i.e. BV or NDV.

Secondary Objectives:

1) To describe the distributions of BMI and its potential proxies including BV and NDV using controls from a UK study
2) To describe the correlations between BMI and its potential proxies (NDV, BV) and breast density %BD, DV.
3) To compare the strength of association between BMI and potential proxies

The primary objectives are addressed in Research Paper I (section 5.3). This paper is supplemented by unpublished descriptive analyses that address secondary objectives not explicitly covered by the published paper (section 5.4).

5.3 Paper I Adjusting for BMI in analyses of volumetric mammographic density and breast cancer risk (228)

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	12406683	**Title**	Ms
First Name(s)	Susan M		
Surname/Family Name	Hudson		
Thesis Title	Beyond breast density – Novel uses of automated mammographic analysis in breast cancer screening		
Primary Supervisor	Professor Isabel dos Santos Silva		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Breast Cancer Research		
When was the work published?	2018		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	n/a		
Have you retained the copyright for the work?*	**No**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I was responsible for the conceptualisation of the work, development of the methodology, curating and analysing the UK data set and providing guidance for the analysis of the Norweigian dataset and for writing the publication.

SECTION E

Student Signature	[redacted]
Date	

Supervisor Signature	
Date	

Hudson et al. Breast Cancer Research (2018) 20:156
https://doi.org/10.1186/s13058-018-1078-8

Breast Cancer Research

RESEARCH ARTICLE Open Access

Adjusting for BMI in analyses of volumetric mammographic density and breast cancer risk

Sue Hudson[1*], Kirsti Vik Hjerkind[2], Sarah Vinnicombe[3], Steve Allen[4], Cassia Trewin[2], Giske Ursin[2], Isabel dos-Santos-Silva[1] and Bianca L. De Stavola[5]

Abstract

Background: Fully automated assessment of mammographic density (MD), a biomarker of breast cancer risk, is being increasingly performed in screening settings. However, data on body mass index (BMI), a confounder of the MD–risk association, are not routinely collected at screening. We investigated whether the amount of fat in the breast, as captured by the amount of mammographic non-dense tissue seen on the mammographic image, can be used as a proxy for BMI when data on the latter are unavailable.

Methods: Data from a UK case control study (numbers of cases/controls: 414/685) and a Norwegian cohort study (numbers of cases/non-cases: 657/61059), both with volumetric MD measurements (dense volume (DV), non-dense volume (NDV) and percent density (%MD)) from screening-age women, were analysed. BMI (self-reported) and NDV were taken as measures of adiposity. Correlations between BMI and NDV, %MD and DV were examined after log-transformation and adjustment for age, menopausal status and parity.
Logistic regression models were fitted to the UK study, and Cox regression models to the Norwegian study, to assess associations between MD and breast cancer risk, expressed as odds/hazard ratios per adjusted standard deviation (OPERA). Adjustments were first made for standard risk factors except BMI (minimally adjusted models) and then also for BMI or NDV. OPERA pooled relative risks (RRs) were estimated by fixed-effect models, and between-study heterogeneity was assessed by the I^2 statistics.

Results: BMI was positively correlated with NDV (adjusted $r = 0.74$ in the UK study and $r = 0.72$ in the Norwegian study) and with DV ($r = 0.33$ and $r = 0.25$, respectively). Both %MD and DV were positively associated with breast cancer risk in minimally adjusted models (pooled OPERA RR (95% confidence interval): 1.34 (1.25, 1.43) and 1.46 (1.36, 1.56), respectively; $I^2 = 0\%$, $P > 0.48$ for both). Further adjustment for BMI or NDV strengthened the %MD–risk association (1.51 (1.41, 1.61); $I^2 = 0\%$, $P = 0.33$ and 1.51 (1.41, 1.61); $I^2 = 0\%$, $P = 0.32$, respectively). Adjusting for BMI or NDV marginally affected the magnitude of the DV–risk association (1.44 (1.34, 1.54); $I^2 = 0\%$, $P = 0.87$ and 1.49 (1.40, 1.60); $I^2 = 0\%$, $P = 0.36$, respectively).

Conclusions: When volumetric MD–breast cancer risk associations are investigated, NDV can be used as a measure of adiposity when BMI data are unavailable.

Keywords: BMI, Breast cancer, Breast density, Mammographic density, OPERA

* Correspondence: susan.hudson@lshtm.ac.uk
[1]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK
Full list of author information is available at the end of the article

Introduction

Mammographic density captures the amount of (radio-)dense tissue in the breast. Mammographic density, for a woman's age and body mass index (BMI), is a well-established breast cancer risk factor [1, 2]. This biomarker of risk is being increasingly used as an intermediate phenotype in epidemiological studies. It also offers the potential for breast cancer prevention strategies, including screening, to be tailored according to a woman's individual risk, in combination with other non-genetic and genetic risk factors.

Mammographic density has traditionally been assessed as the absolute or relative amount (as percentage of the total breast size) occupied by the dense tissue which appears on a mammographic image as white "cotton-like" patches. Percent mammographic density (%MD) is negatively correlated with BMI (reported correlation coefficients ranging between −0.41 and −0.61 [3, 4]), which itself is a breast cancer risk factor (positively associated with risk in post-menopausal women but negatively associated with risk in pre-menopausal women) [5]. Therefore, it is essential to adjust for BMI (as well as age) and any study of mammographic percent density that fails to do so will lead to confounded estimates of the association between percent density and risk [6].

The recent introduction of full-field digital mammography (FFDM), paralleled by the development of fully automated digital image assessment software, has meant that mammographic density assessment is now routinely performed in many screening settings, thus providing a unique opportunity for the conduct of large-scale studies on this biomarker of risk. However, a common barrier to such investigations is the lack of information on a woman's BMI as data on this variable are rarely collected at screening.

Most of the fully automated density assessment methods developed for FFDM attempt to estimate from the two-dimensional images, the volume of radio-dense tissue (DV) as well as the volume of non-dense (fat) tissue (NDV) and the total volume of the breast (BV) which, in Western populations, is highly correlated with NDV. Hence, NDV, or its BV correlate, has been used as a proxy for BMI in analyses of mammographic density and breast cancer risk in studies where BMI data are not available [7, 8]. However, the validity of such an approach has never been tested empirically. The aim of this study is to assess whether NDV can be used as a proxy for BMI when assessing associations between volumetric estimates of mammographic density, as derived from two-dimensional images, and breast cancer risk.

Methods

Study participants

The present analysis was conducted within two studies: a case control study from the UK and a cohort study from Norway.

UK study

The study methodology of the UK case control study is described in detail elsewhere [9]. In short, cases (n = 414) were women with newly diagnosed breast cancer at the Royal Marsden Hospital (RMH), London, between April 2010 and July 2012. Controls (n = 685) were women screened and found to be breast cancer-free at the Central and East London Breast Screening Service (CELBSS) in the same time period. The CELBSS invites women between 50 and 70 years of age for mammographic screening once every 3 years as part of the English National Health Service Breast Screening Programme. Women over 70 years can optionally contact the service for a self-referred appointment every 3 years.

Data on breast cancer risk factors, including age, ethnicity, parity, menopausal status, use of oral contraceptives, use of hormone therapy and self-reported height and weight, were collected by a self-administered questionnaire at the time of screening for controls and within 15.5 months of the diagnostic mammography for cases. BMI was calculated as weight in kg/(height in m)^2. Ethnicity was categorised in accordance with the census classification as "White", "Black" (African or Caribbean), "Asian" (Indian, Pakistani or Bangladeshi) and "Other" [10].

Participants underwent full-field digital mammography, with two views (cranio-caudal (CC) and mediolateral-oblique (MLO)), of both breasts. The images were taken on Senographe DS machines (GE Healthcare, Slough, UK). The anonymised raw images were analysed by using Volpara version 1.0 (Matakina Technology Limited, Wellington, New Zealand) [11]. This algorithm provided fully automated estimates of the volumes (all in cm^3) of the total breast (BV), non-dense (fat) tissue (NDV) and dense (fibro-glandular) tissue (DV) separately for each one of the four breast/view images, and percent mammographic density (%MD) was estimated as DV/BV×100.

Norwegian study

The Cancer Registry of Norway is responsible for the administration of BreastScreen Norway (the Norwegian Breast Cancer Screening Program). All women within a targeted age-range of 50–69 years resident in the country are invited to undergo mammography screening every 2 years. From August 2006 to 2014, women who underwent mammographic screening in the nationwide programme were asked to complete a questionnaire on a number of standard breast cancer risk factors and a second questionnaire on current exposure to risk factors. Included in the present study were women who participated in BreastScreen Norway in four counties, had information on volumetric mammographic density from their first mammographic screening between 2007 and 2014, and had completed both questionnaires. However, for the second questionnaire on current exposures, if the

questionnaire or certain values were missing, information from the questionnaire completed at a previous screening round was used (approximately 16.5%). The cohort consisted of 61,716 women, including 657 women who were diagnosed with a first occurrence of breast cancer during a median follow-up from date of screening of 3.84 (interquartile range 2.08, 4.83) years. Women with a previous diagnosis of breast cancer (n = 970), a ductal carcinoma *in situ* (DCIS) diagnosis up to 6 months after the screening date (n = 224) and a bilateral breast cancer (n = 11) were excluded.

In a similar manner to the UK study, all women had standard two-view full-field digital mammography of each breast with Senographe DS or Senographe Essential machines (GE Healthcare) or MDM L50 or MDM L30 machines (Phillips). The raw images were read by Volpara version 1.5.0 (Volpara Health Technologies Limited, Wellington, New Zealand) to obtain, similarly to the UK study, volumetric estimates of BV, NDV, DV and %MD.

Ethical approval

The UK study was approved by all relevant ethics committees (Research Ethics Committees from the Royal Marsden Hospital, the Barts and the London NHS Trust, and the London School of Hygiene and Tropical Medicine). The Norwegian study was approved by the Regional Committee for Medical and Health Research ethics in the South-East Health Region of Norway. In the UK, participants provided written informed consent. In the Norwegian study, in accordance with the Cancer Registry Regulations, returning the questionnaire was considered consent, and information about screening examinations can be used for quality assurance and research if the women have not actively opted out. About 2% of the women attending the programme have opted out.

Statistical methods

Descriptive analysis of UK controls and the full Norwegian cohort included examination of the distributions of BMI and volumetric mammographic measurements. For these analyses, measurements were averaged over the four images (that is, left and right CC and MLO images). Natural-log transformations were applied to average %MD, DV, NDV and BMI to normalise their distributions. Scatter plots and Pearson's correlation coefficients were used to examine BMI associations with %MD, DV, NDV and BV. BMI and each mammographic measure were regressed on age at mammogram, parity and menopausal status using linear regression (including controls only in the UK study and the full cohort in the Norwegian study). Pearson's correlation coefficients between the residuals derived from these models were then calculated (and denoted r) to allow examination of correlations that are not influenced by these variables.

For the UK case control analysis, the average density measures from the CC and MLO images from the unaffected breasts for cases and for a randomly selected breast for controls were used. In order to compare the association of %MD and DV with the odds of breast cancer, after adjusting for different sets of confounders, three different logistic regression models were fitted where these two exposures were first standardised as recommended previously [12].[A] The resulting estimates are referred to as OPERA ORs ("odds ratios per adjusted standard deviation") and are effects per residual standard deviation of the exposure once its association with the confounders is accounted for. Estimation requires first fitting a linear regression model of the exposure on the confounders and then using the standardised residuals derived from this model as the exposure of interest in logistic regression models that include the same confounders. Fifty-one cases and 38 controls (8.1% of the study participants) were excluded from all logistic regression analyses because they were missing at least one of the variables used in the modelling.

The first (minimally adjusted) model controlled for age (continuous), menopausal status (pre-, peri/post-) and parity (yes/no). (Further adjustment for ethnicity, use of exogenous hormones and the other variables listed in Table 1 was also considered but it is not shown as it yielded similar results.) Second, a model was fitted that additionally adjusted for self-reported BMI. Finally, an alternative model was fitted that additionally adjusted for log-transformed NDV in place of BMI. Adjustment for BV instead of NDV was not considered because, albeit this variable is highly correlated with NDV (r = 0.99; P = 0.001), its interpretation is made more difficult by the fact that it reflects both DV and NDV.

In the Norwegian cohort study, average density measures were based on log-transformed average values of the CC and MLO readings from the unaffected breast for cases and from a randomly selected breast for non-cases. Cox regression proportional hazards models were fitted to the cohort data, using age as the time-scale, to evaluate the associations of (log-transformed and standardised as described above for the UK study) %MD and DV with breast cancer risk, expressed in terms of hazard ratios and referred to as OPERA HRs.

Three different models were fitted as in the UK study; the first was minimally adjusted for screening year (categorised using 2-year intervals), menopausal status (pre-, peri-, post-) and parity (yes/no) (further adjustment for country of birth as a proxy for ethnicity did not affect the findings). A second model was additionally adjusted for BMI, and a third model was additionally adjusted for NDV in place of BMI. In all, 10,288 participants, including 99 cases, were excluded from all three models because they missed data for at least one of the variables listed.

[A] Further explanation for the selection of the OPERA method of analysis is outlined in section 5.5 of this thesis.

Table 1 Baseline characteristics of the participants by status in the UK and Norwegian studies[a]

UK case control study			Norwegian cohort study	
	Controls (*n* = 685)	Cases (*n* = 414)	Non-cases (*n* = 61,059)	Cases (*n* = 657)
Age at mammography				
Mean (SD)	59.5 (6.6)	67.5 (12.7)	56.9 (5.74)	57.7 (5.43)
Number	679	412	61,059	657
BMI[b]				
Mean (SD)	26.1 (5.6)	26.4 (4.9)	25.6 (4.2)	25.8 (4.1)
Number	656	368	54,345	589
Ethnicity (UK)/Country of birth (Norway), n (%)				
White/Norway	520 (76.5)	370 (89.4)	56,234 (93.8)	612 (94.2)
Non-white/Outside Norway	160 (23.5)	39 (9.6)	3693 (6.2)	38 (5.8)
Missing	5	5	1132	7
Family history of BC, n (%)				
No	N/A	N/A	45,168 (77.1)	447 (70.0)
Yes	N/A	N/A	13,390 (22.9)	192 (30.0)
Missing			2501	18
Menopausal status[c], n (%)				
Pre- + peri-menopausal	91 (13.3)	55 (13.3)	14,776 (25.2)	141 (22.1)
Post-menopausal	591 (86.7)	358 (86.7)	43,856 (74.8)	496 (77.9)
Missing	3	1	2427	20
Parity, n (%)				
Nulliparous	209 (30.9)	65 (15.9)	4946 (8.5)	57 (9.0)
Parous	467 (69.1)	343 (84.1)	53,563 (91.5)	577 (91.0)
Missing	9	6	2550	23
Age at menarche in years, n (%)				
<13	271 (53.9)	159 (54.1)	16,764 (40.9)	186 (41.9)
14+	232 (46.1)	135 (45.9)	24,202 (59.1)	258 (58.1)
Missing	14	33	4107	43
Hormone therapy use, n (%)				
No	459 (68.8)	246 (63.2)	34,150 (66.2)	305 (55.6)
Yes	208 (31.2)	143 (36.8)	17,418 (33.8)	244 (44.4)
Missing	18	25	9491	108
Educational level, n (%)				
None/primary school	35 (5.2)	17 (6.2)		
Lower secondary			13,772 (23.3)	164 (25.9)
Secondary or higher	641 (94.8)	225 (93.8)	45,457 (76.7)	470 (74.1)
Missing	9	142	1830	23
Breastfeeding among parous women, n (%)				
Yes	358 (76.7)	224 (74.7)	46,107 (99.9)	497 (100)
Missing	3	43	9929	103

Abbreviations: *BC* breast cancer, *BMI* body mass index, *N/A* data not available, *SD* standard deviation

[a]Percentages calculated without missing values

[b]BMI estimated from self-reported height and weight as weight/height^2 (in kg/m^2)

[c]Post-menopausal women defined as those who self-reported natural (cessation of menses for at least 12 months) or surgical menopause, were older than 55 years, or had ever used hormone therapy. Owing to small numbers, pre-menopausal (younger than 55 years and still having regular periods) and peri-menopausal (younger than 55 years and having irregular periods) women were combined into a single category

Three further models were also fitted to the Norwegian data using the full reproductive and lifestyle risk factor questionnaire data collected in this study (that is, screening year category, menopausal status, parity, age at menopause, age at menarche, age at first birth, duration of breastfeeding, use of hormone therapy, family history of breast cancer, education, smoking, alcohol use and physical activity level). In the first model, BMI was omitted; in the second model, BMI was included; in the third model, NDV was used instead of BMI. In total, 25,833 (41.9% of the original cohort) women with missing data on any of the variables examined were excluded to ensure that these additional models were fitted to the same subset of women. Departure from the proportional hazards assumption underlying each of these fitted models was evaluated by using tests based on Schoenfeld residuals. The Akaike information criterion (AIC) corresponding to each multivariable model from the two countries is also reported.

Similar analytical steps were followed to study the associations between BMI and breast cancer risk, and then NDV and breast cancer risk, in both studies, in each case adjusting for age, menopausal status and parity.

Fixed-effects models were used to obtain pooled summary OPERA relative risk (RR) estimates from the two studies. Between-study heterogeneity was assessed by the Q statistic and the I^2 statistic [13].

In all the analyses, we considered statistical significance (two-sided) at a P value of less than 0.05. All analyses were conducted in Stata (IC 14 for the statistical analysis of the UK data and the meta-analysis and IC 15 for the analysis of the Norwegian data) [14].

Results

Study participants

The baseline characteristics of the participants in the two studies are shown in Table 1. In the UK, study cases were, on average, older than controls and more likely to be White. Likewise, cases were slightly older at mammography than non-cases in the Norwegian study. The mean BMI was similar for UK cases and controls and for Norwegian cases and non-cases.

Correlations between BMI and volumetric mammographic measures

The distributions of self-reported BMI and of NDV, the volumetric measurement that reflects the fatty tissue in the breast, were right-skewed in the UK control group and in the full Norwegian cohort (Fig. 1)[B]. Table 2 shows the correlations between each volumetric measure and BMI after adjusting for age, parity and menopausal status. Notably, the two studies yielded very similar results. NDV was highly positively correlated with BMI in both the UK (r = 0.74) and in the Norwegian (r = 0.72) study. In contrast, the correlation between DV and BMI was weakly positive in both the UK study (r = 0.33) and the Norwegian study (r = 0.25). Consequently, %MD was negatively correlated with BMI in both the UK (r = −0.66) and Norwegian (r = −0.57) studies. The correlation between %MD and DV was only moderate after adjustment for age and BMI in the UK (r = 0.33) and Norwegian (r = 0.55) studies (data not shown). Further analyses showed that the findings were robust after stratification by mammographic view, age at mammography and, for the UK study, restricting the analysis to White women (data not shown).

Associations between adiposity measures and breast cancer risk

There were weak positive associations between BMI and breast cancer risk (adjusted for age, menopausal status and parity) in both the UK (OPERA OR 1.10, 95% confidence interval (CI) 0.95, 1.26) and the Norwegian (OPERA HR 1.09, 95% CI 1.01, 1.19) studies. The magnitude of the BMI–risk association was not modified by menopausal status or age in either study (P > 0.30 and P > 0.10, respectively, in models that included interactions with either menopausal status or age), most likely because of the relatively small number of younger (pre-menopausal) women in either study. There was no evidence of an association between NDV and breast cancer risk adjusting for the same covariates as for BMI (UK study OPERA OR 0.96, 95% CI 0.83, 1.11); Norwegian study OPERA HR 1.01, 95% CI 0.93, 1.09).

Associations between relative and absolute volumetric density and breast cancer risk

Figure 2 shows study-specific and pooled summary OPERA estimates of, respectively, %MD and DV with breast cancer risk. In both the UK and Norwegian studies, the minimally adjusted models, which exclude any adjustment for adiposity, show a positive association between %MD and breast cancer risk with no evidence of between-study heterogeneity (I^2 = 0%; P = 0.49). Further adjustment for BMI showed a strengthening of the positive association between %MD and breast cancer risk: the magnitude of the pooled OPERA RR increased from 1.34 (95% CI 1.25, 1.43) to 1.51 (95% CI 1.41, 1.61) upon adjustment for BMI and there was no evidence of between-study heterogeneity for the latter (I^2 = 0%; P = 0.33). Replacing BMI with NDV, as a proxy for level of adiposity, yielded the same strength of association between %MD and breast cancer risk with the pooled OPERA RR increasing to 1.51 (95% CI 1.41, 1.61) and there was no evidence of between-study heterogeneity (I^2 = 0%; P = 0.33).

The association between DV and breast cancer risk was slightly stronger than that found for %MD in the minimally adjusted model (pooled OPERA RR for DV = 1.46,

B Additional analyses including histograms showing the distribution of additional volumetric breast estimates and scatter plots of raw correlations between measurements and BMI are shown in section 5.4 of this thesis

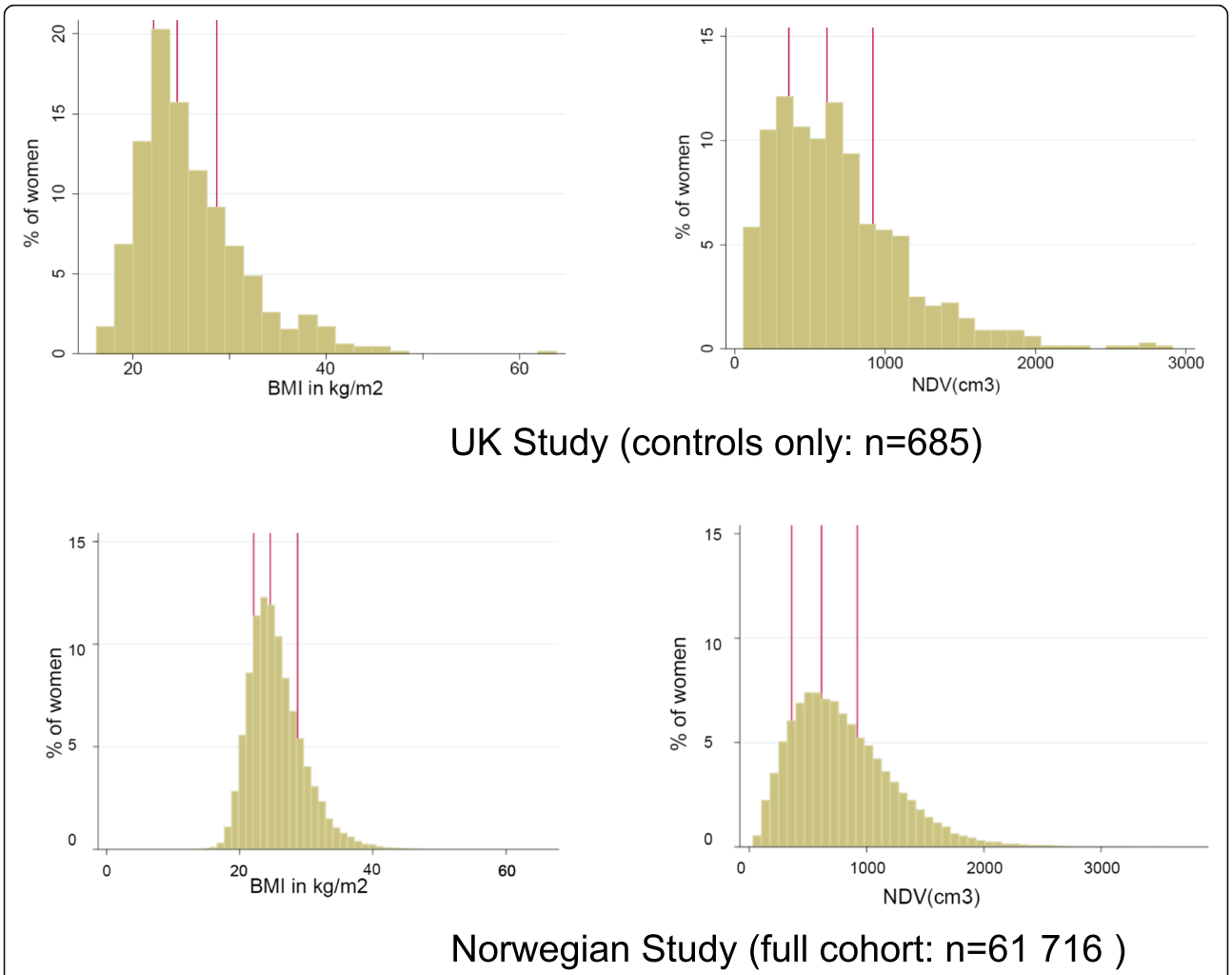


Fig. 1 Distribution of self-reported BMI and measurements of volume of mammographic non-dense tissue in the UK and Norwegian studies. Abbreviations: *BMI* body mass index, *NDV* volume of mammographic non-dense tissue averaged over the cranio-caudal and mediolateral-oblique views from the left and right breasts. Vertical lines represent the median and interquartile range values

Table 2 Pearson's correlation coefficients of mammographic measures[a] with BMI[a] and with log NDV

	Adjusted log BMI		Adjusted log NDV	
	UK case control study controls only[b]	Norwegian cohort study full cohort[c]	UK case control study controls only[b]	Norwegian cohort study full cohort[c]
Adjusted log NDV	0.74	0.72	–	–
Adjusted log %MD	−0.66	−0.57	−0.80	−0.72
Adjusted log DV	0.33	0.25	0.57	0.43

Abbreviations: *BMI* body mass index, *DV* volume of mammographically dense tissue, *NDV* volume of mammographically non-dense tissue, *%MD* percent mammographic density

DV, NDV and %MD averaged over the cranio-caudal and mediolateral-oblique views from the left and right breasts

[a]All mammographic features as well as BMI were regressed on age at mammogram, parity and menopausal status and the residuals from these regressions were used to calculate the correlation coefficients and referred to as "adjusted" measures

[b]n = 646 (women with missing BMI, age, parity, menopausal status or mammographic measurements were excluded and one woman with a BMI greater than 60 was also excluded)

[c]n = 51,427 (women with missing BMI, age, parity, menopausal status or mammographic measurements were excluded or BMI greater than 60 were excluded)

P <0.0001 in all cases

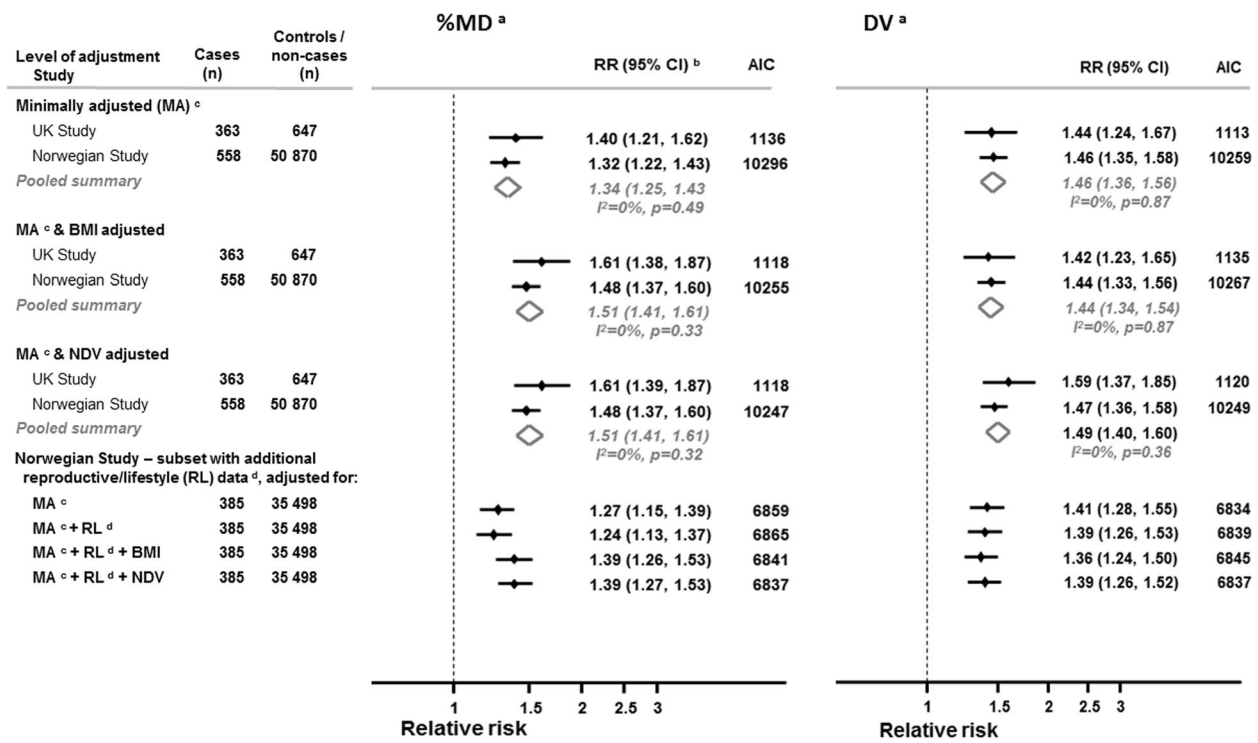


Fig. 2 Mammographic density associations with breast cancer risk with and without adjustment for adiposity in the UK and Norwegian studies. Abbreviations: *BMI* body mass index, *CI* confidence interval, *DV* volume of mammographic dense tissue, *NDV* volume of mammographic non-dense (fat) tissue, *%MD* percent mammographic density. [a]DV, NDV and %MD values are the average from the cranio-caudal and mediolateral-oblique views from the unaffected breast for cases and for a randomly selected breast side for controls, log-transformed. [b]In the UK study, OPERA odds ratios (ORs) were estimated by a logistic regression. In the Norwegian cohort study, OPERA hazard ratios (HRs) were estimated by a Cox regression model in which attained age was taken as the time scale (see Methods section). [c]Minimally adjusted model: analysis adjusted for age, menopausal status and parity in the UK study; analysis adjusted for screening year, menopausal status and parity (see Methods section). [d]Model additionally adjusted for age at menopause, age at menarche, age at first birth, duration of breastfeeding, use of hormone therapy, family history of breast cancer, education, smoking, alcohol use and physical activity level

95% CI 1.36, 1.56) and there was no evidence of between-study heterogeneity (I^2 = 0%; P = 0.87). Adjusting for BMI had little impact on the magnitude of the DV–breast cancer risk association (pooled OPERA RR = 1.44, 95% CI 1.34, 1.54). When BMI was replaced by NDV, the magnitude of the pooled OPERA RR increased only slightly to 1.49 (95% CI 1.40, 1.60) and there was no evidence of between-study heterogeneity (I^2 = 0%; P = 0.36).

The similarity of the estimated RRs for %MD and DV when adjusted for either BMI or NDV indicates that these measures of adiposity lead to equivalent control of confounding. Since the models adjusted for age, menopausal status, parity and NDV have the smallest AIC in both the UK and Norwegian studies, controlling for the NDV, when BMI is self-reported as in these datasets, appears to be (marginally) preferable.

Further analyses of the OPERA estimates in the Norwegian data show that the magnitude of the associations of %MD and DV with breast cancer risk was little changed by adjustment for additional reproductive and lifestyle factors in the subset of women with information on these variables. The addition of BMI to this expanded model strengthened the association between %MD and breast cancer risk; the OPERA HR increased from 1.24 (95% CI 1.13, 1.37) to 1.39 (95% CI 1.26, 1.53). Likewise, replacing BMI with NDV to control for adiposity led to similar estimates (OPERA HR = 1.39, 95% CI 1.27, 1.53) though with a marginally better fitting model using AIC. In contrast, adjustment for BMI or NDV had little effect on the magnitude of the DV–breast cancer risk association (Fig. 2).

Discussion

Main findings

We found that, for screening-aged women, the association between volumetric %MD and breast cancer risk is partly confounded by levels of adiposity and that the two measures of adiposity available in our studies—BMI or NDV—lead to similar adjusted estimates of association. In contrast, when assessing the magnitude of the association between volumetric absolute mammographic density (that is, DV) and breast cancer risk, adjustment for

BMI or NDV had little or no impact on the magnitude of the association. All of these estimates of association are expressed in terms of units per relative standard deviation of the exposure (that is, using the OPERA approach). This allows us to compare estimates while adjusting for different sets of potential confounders, in units that account for the strength of association between the confounders and the exposure of interest [12]. Both the Norwegian cohort and the UK case control analyses found that it is important to adjust for adiposity when the main explanatory variable is volumetric %MD, as estimated from a two-dimensional image; otherwise, the relationship between %MD and breast cancer risk may appear to be weaker. Our results based on AIC suggest that objectively measured NDV may offer a slightly better proxy for adiposity than BMI when comparing the model's specifications in terms of goodness of fit. This may be a consequence of the self-reported nature of the available BMI data, and measurement error may lead to attenuation of the adjustment. Nevertheless, it is unclear the extent to which BMI and NDV capture the same, or different, underlying biological entities.

After adjustment for age, parity and menopausal status, BMI was found to be strongly positively correlated with NDV but strongly negatively correlated with %MD. In contrast, BMI was weakly positively correlated with DV. The observed strong positive BMI–NDV correlation is consistent with findings from area-based mammographic studies [4, 6, 15]. The observed weak positive BMI–DV correlation is also in line with findings from other volumetric density studies [16, 17] but in contrast to those from area-based studies which consistently report a negative correlation [4, 18, 19]. The correlation between DV and %MD after adjustment for age and BMI was not as strong (in either the UK or Norwegian study) as that reported in area-based studies [20].

The present study found a positive, albeit weak, association between BMI and breast cancer risk, reflecting the predominantly peri-/post-menopausal status of the participants, but no association between NDV and breast cancer risk. There is little evidence for an NDV–breast cancer association from volumetric studies to date, but a meta-analysis of data from 13 area-based studies has reported an overall inverse association between mammographic non-dense area and risk [21] albeit with considerable between-study heterogeneity.

Strengths and limitations

Strengths of this investigation include the availability of data from two independent studies of women of screening age. Both studies used the same objective volumetric density assessment method, making the two datasets comparable. In addition, the Norwegian study was population-based and had a very large sample size and detailed data on a wide range of potential confounding variables collected prior to breast cancer diagnosis and therefore was unlikely to have been affected by recall bias. Furthermore, although other studies have assumed that it is reasonable to used NDV as a surrogate for BMI [7, 8], we believe that we are the first to have formally tested this empirically.

A limitation of this investigation is that it relied on self-reported BMI. Previous research suggests that women tend to understate their weight and overstate their height, particularly those who are overweight or obese [22, 23], although a recent study found that women attending BreastScreen Norway reported weight and height within 1 kg/cm of directly measured values [24]. In most population-based screening programmes, however, it is logistically impossible to perform anthropometric measurements when women attend screening. Nevertheless, it would be informative if similar analyses were replicated within a study sample with measured BMI.

We used the OPERA approach to allow comparison across different exposures (that is, effects per residual standard deviation of the exposure once its association with the confounders is accounted for). It is argued that this provides a fairer comparison of the different risk gradients across the different models [12]. However, a common criticism of two-step approaches, such as OPERA, is that the standard errors of the estimated coefficients are underestimated and thus lead to a spurious increase in the precision of the estimated effect sizes [25].

The study was restricted to women of screening age and not generalizable to younger women. There is also evidence that the relationship between percent body fat and BMI is dependent upon ethnicity [26, 27], with Asians having a higher percentage of body fat for any given BMI compared with Caucasians [28]. The relatively small number of non-White women in both studies precluded examination by ethnicity.

Finally, both studies were based on a particular volumetric mammographic density assessment approach. It would be worthwhile to examine the extent to which the present findings can be replicated when alternative methods of assessment of volumetric mammographic density are used.

Conclusions

The availability of fully automated methods to measure mammographic density enables the integration of such measurements within screening programme settings, thus facilitating the conduct of large-scale studies, including research on whether screening should be tailored to a woman's individual risk. A perceived barrier to the conduct of such studies is the lack of information on a woman's BMI. This study shows that the association between DV and breast cancer risk is not confounded by BMI or NDV and hence no adjustment for

these variables is required. In contrast, the association between volumetric %MD and risk is confounded by level of adiposity and adjustment for either BMI or NDV yields similar results. Adjustment for NDV may offer some advantages over BMI as the NDV measurements are objective, being generated by a fully automated algorithm, and thus do not suffer from measurement errors associated with self-reported BMI. Furthermore, in most breast screening settings, it is not feasible to collect BMI data; therefore, NDV values are potentially very valuable because they will be automatically available for every woman screened. Nevertheless, these findings need to be replicated in other populations, particularly among those with a different age and ethnic mix.

Abbreviations
%MD: Percentage mammographic density; AIC: Akaike information criterion; BMI: Body mass index; BV: Breast volume as ascertained from a mammogram; CC: Cranio-caudal view; CELBSS : Central and East London Breast Screening Service; CI: Confidence interval; DV: Dense volume (that is, absolute volume of (radio-)dense tissue seen on a mammogram); FFDM: Full-field digital mammography; HR: Hazard ratio; MLO: Mediolateral oblique mammogram view; NDV: Non-dense volume (that is, volume of non-dense (fat) tissue seen on a mammogram); OPERA: Odds per adjusted standard deviation; OR: Odds ratio; RR: Relative risk

Acknowledgments
Not applicable.

Funding
The original data collection for the UK study was supported by the DaCosta Foundation. The funding bodies played no role in the design, collection, analysis, or interpretation of the data; in the writing of the manuscript; or in the decision to submit the manuscript for publication.

Availability of data and materials
The raw data were generated at the institutions named in the article. The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Authors' contributions
SH and IdSS designed the study. SA and SV recruited the UK participants. SH, KVH and CT performed the statistical analysis with guidance from BDS. SH wrote the first draft of the manuscript. All authors contributed to the interpretation of the results and critically reviewed the draft of the manuscript, read and approved the final version of the manuscript, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Authors' information
SV was previously director of the Central and East London Breast Screening Service (UK).

Ethics approval and consent to participate
See the "Ethical approval" section of this article.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. [2]Cancer Registry of Norway, Institute of Population-based Cancer Research, Oslo, Norway. [3]Division of Imaging and Technology, Ninewells Hospital Medical School, University of Dundee, Dundee DD2 1SY, UK. [4]Royal Marsden NHS Foundation Trust, London SW3 6JJ, UK. [5]Faculty of Population Health Sciences, Institute of Child Health, University College London, London WC1N 1EH, UK.

Received: 23 June 2018 Accepted: 8 November 2018
Published online: 29 December 2018

References
1. McCormack V, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomark Prev. 2006;15:1159–69.
2. Huo CW, Chew GL, Britt KL, Ingman WV, Henderson MA, Hopper JL, Thompson EW. Mammographic density-a review on the current understanding of its association with breast cancer. Breast Cancer Res Treat. 2014;144(3):479–502.
3. Soguel L, Durocher F, Tchernof A, Diorio C. Adiposity, breast density, and breast cancer risk: epidemiological and biological considerations. Eur J Cancer Prev. 2017;26(6):511–20.
4. Baglietto L, Krishnan K, Stone J, Apicella C, Southey MC, English DR, Hopper JL, Giles GG. Associations of mammographic dense and nondense areas and body mass index with risk of breast cancer. Am J Epidemiol. 2014; 179(4):475–83.
5. Bhaskaran K, Douglas I, Forbes H, dos Santos Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults. *Lancet*. 2014;384(9945):755–65.
6. Boyd NF, Martin LJ, Sun L, Guo H, Chiarelli A, Hislop G, Yaffe M, Minkin S. Body size, mammographic density, and breast cancer risk. Cancer Epidemiology Biomarkers and Prevention. 2006;15(11):2086–92.
7. Ellison-Loschmann L, McKenzie F, Highnam R, Cave A, Walker J, Jeffreys M. Age and ethnic differences in volumetric breast density in New Zealand women: a cross-sectional study. PLoS One. 2013;8(7):e70217.
8. van der Waal D, Emaus MJ, Bakker MF, den Heeten GJ, Karssemeijer N, Pijnappel RM, Veldhuis WB, Verbeek AL, van Gils CH, Broeders MJ. Geographic variation in volumetric breast density between screening regions in the Netherlands. Eur Radiol. 2015;25(11):3328–37.
9. Eng A, Gallant Z, Shepherd J, McCormack V, Li J, Dowsett M, Vinnicombe S, Allen S. dos-Santos-Silva I: digital mammographic density and breast cancer risk: a case-control study of six alternative density assessment methods. Breast cancer research : BCR. 2014;16(5):439.
10. 2011 census guidance and methodology [http://www.ons.gov.uk/ons/guide-method/census/2011/index.html].
11. Matakina Technology Ltd.: VolparaDensity™ user manual version 1.5.11. 2014.
12. Hopper JL. Odds per adjusted standard deviation: comparing strengths of associations for risk factors measured on different scales and across diseases and populations. Am J Epidemiol. 2015;182(10):863–7.
13. Higgins JP, Thompson SG: Quantifying heterogeneity in a meta-analysis. (0277–6715 (Print)).
14. StataCorp. Stata statistical software: release 14. College Station, TX: StataCorp LP. In. In: 14 edn; 2015.
15. Lokate M, Peeters PH, Peelen LM, Haars G, Veldhuis WB, van Gils CH: Mammographic density and breast cancer risk: the role of the fat surrounding the fibroglandular tissue. Breast cancer research : BCR 2011, 13(5):R103.
16. Shepherd J, Kerlikowske K, Ma L, Duewer F, Fan B, Wang J, Malkov S, Vittinghoff E, Cummings S. Volume of mammographic density and risk of breast cancer. Cancer Epidemiol Biomark Prev. 2011;20:1473–82.
17. Schetter SE, Hartman TJ, Liao J, Richie JP, Prokopczyk B, DuBrock C, Signori C, Hamilton C, Demers LM, El-Bayoumy K, et al. Differential impact of body mass index on absolute and percent breast density: implications regarding their use as breast cancer risk biomarkers. Breast Cancer Research & Treatment. 2014;146(2):355–63.
18. Irwin ML, Aiello EJ, McTiernan A, Bernstein L, Gilliland FD, Baumgartner RN, Baumgartner KB, Ballard-Barbash R. Physical activity, body mass index, and

mammographic density in postmenopausal breast Cancer survivors. J Clin Oncol. 2007;25(9):1061–6.
19. Sun X, Gierach GL, Sandhu R, Williams T, Midkiff BR, Lissowska J, Wesolowska E, Boyd NF, Johnson NB, Figueroa JD, et al. Relationship of mammographic density and gene expression: analysis of Normal breast tissue surrounding breast Cancer. Clin Cancer Res. 2013;19(18):4972–82.
20. Nguyen TL, Aung YK, Evans CF, Dite GS, Stone J, MacInnis RJ, Dowty JG, Bickerstaffe A, Aujard K, Rommens JM, et al. Mammographic density defined by higher than conventional brightness thresholds better predicts breast cancer risk. Int J Epidemiol. 2017;46(2):652–61.
21. Pettersson A, Graff RE, Ursin G, Santos Silva ID, McCormack V, Baglietto L, Vachon C, Bakker MF, Giles GG, Chia KS, et al. Mammographic density phenotypes and risk of breast cancer: a meta-analysis. J Natl Cancer Inst. 2014;106(5).
22. Stommel M, Schoenborn CA. Accuracy and usefulness of BMI measures based on self-reported weight and height: findings from the NHANES & NHIS 2001-2006. BMC Public Health. 2009;9(1):1–10.
23. Connor Gorber S, Tremblay M, Moher D, Gorber B. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. Obesity reviews : an official journal of the International Association for the Study of Obesity. 2007;8(4):307–26.
24. Tsuruda KM, Sagstad S, Sebuødegård S, Hofvind S. Validity and reliability of self-reported health indicators among women attending organized mammographic screening. Scandinavian Journal of Public Health. 0(0): 1403494817749393.
25. Arnold KF, Ellison G, Gadd SC, Textor J, Tennant P, Heppenstall A, Gilthorpe MS. Adjustment for time-invariant and time-varying confounders in 'unexplained residuals' models for longitudinal data within a causal framework and associated challenges. Stat Methods Med Res. 2018: 962280218756158.
26. Rush EC, Goedecke JH, Jennings C, Micklesfield L, Dugas L, Lambert EV, Plank LD. BMI, fat and muscle differences in urban women of five ethnicities from two countries. International journal of obesity (2005). 2007;31(8):1232–9.
27. Gurrici S, Hartriyanti Y, Hautvast JG, Deurenberg P. Differences in the relationship between body fat and body mass index between two different Indonesian ethnic groups: the effect of body build. Eur J Clin Nutr. 1999; 53(6):468–72.
28. Deurenberg P, Deurenberg-Yap M, Guricci S. Asians are different from Caucasians and from each other in their body mass index/body fat per cent relationship. Obesity reviews : an official journal of the International Association for the Study of Obesity. 2002;3(3):141–6.

5.4 Supplementary descriptive analyses

Additional preliminary analyses were carried out, but not included in the published paper I above. These were conducted among cancer-free women to: a) describe the distributions of BMI and volumetric mammographic measurements in cancer-free women; and b) describe the associations between BMI and its potential proxies (NDV, BV) and breast density %BD, DV.

The analyses were carried out among the control women who were recruited into the UK study, i.e. among women screened and found to be breast cancer-free at the Central and East London Breast Screening Service (CELBSS) between April and 2010 and July 2012. Data on the BC risk factors and breast volumetric measurements that were collected are described in Chapter 4 (and in Paper I).

Descriptive analysis included examination of the distributions of all key measurements. To obtain average %BD, DV, NDV and BV values, Volpara measurements were averaged over 4 images i.e. left and right sides from both CC and MLO images. Natural-log transformations were applied to average %BD, DV, NDV and BV measurements to normalise the distributions. Scatter plots were used on the raw (untransformed) values and linear regression models were used to examine BMI correlations with transformed %BD, DV, NDV and BV. Analyses were repeated stratifying by type of view (CC or MLO); Volpara measurements averaged for the appropriate view. Regression coefficients represent the relative change in volumetric measure (log-transformed %BD, DV, NDV and BV) associated with a 1 standard deviation (SD) unit change in BMI. BV and NDV were regressed on BMI controlling for age at mammogram, parity and menopausal status and ethnicity. Goodness of fit was estimated using the adjusted R^2 value. The analyses were repeated stratifying by view (CC and MLO). Further sensitivity analyses were carried out under the following scenarios: a) exclusion of all women outside of the standard screening age range (50-70), b) restricting the analysis to those with "White" ethnicity.

Distributions of breast density estimates DV and %BD were right skewed as were distributions of BV and NDV; BMI distribution was somewhat right skewed Figure 5.1

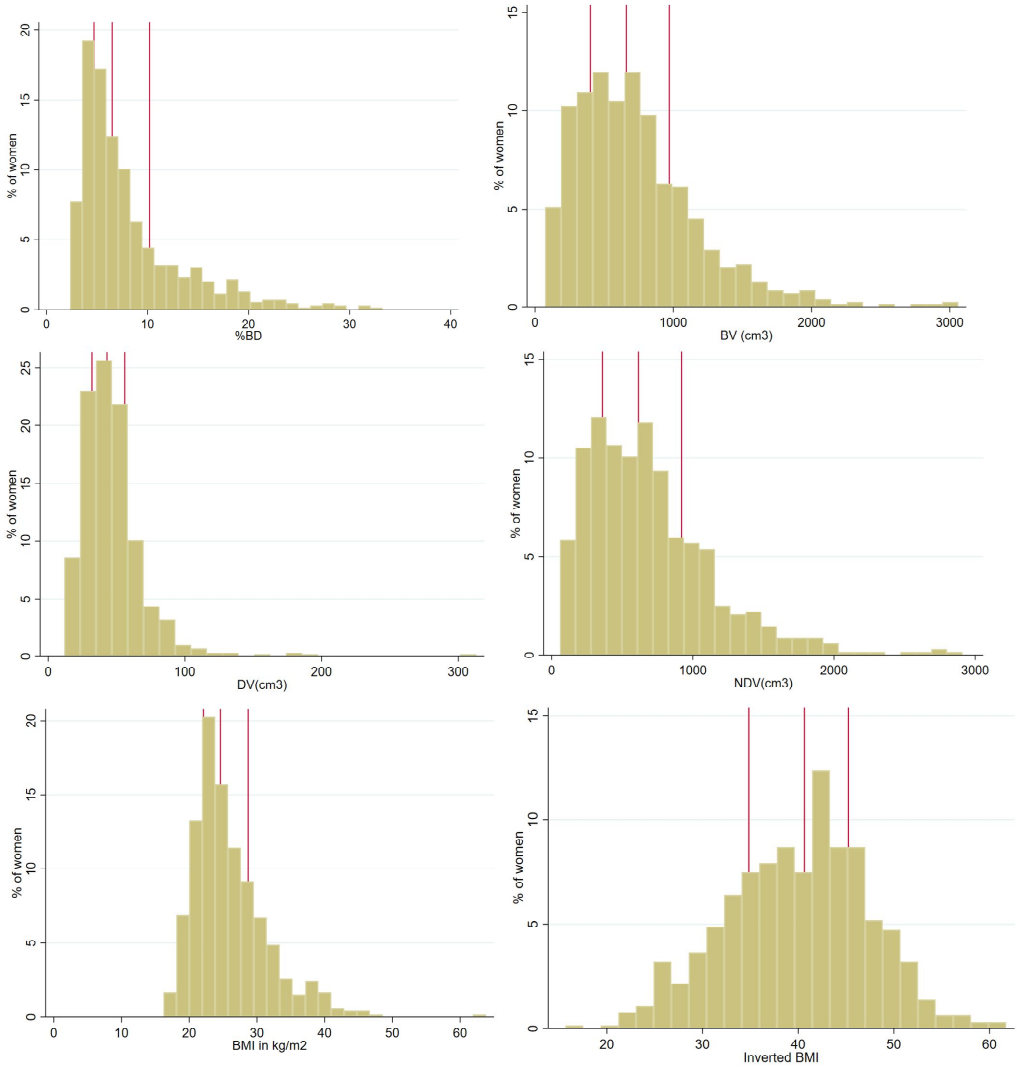


Figure 5.1 Distribution of BMI and breast composition estimates averaged over MLO and CC views from left and right breasts. UK study data

Footnotes: Inverted BMI (iBMI) was calculated as (height in m)^2/weight in kg, giving an approximately normal distribution for BMI. Vertical bars represent median, 25th and 75th quartiles.

The scatter plots Figure 5.2 show that both BV and NDV are strongly positively correlated with BMI. In terms of breast density, %BD was strongly negatively correlated with BMI, conversely a weak positive correlation was observed between absolute density i.e. DV and BMI. Direction and strength of correlations are summarised in Figure 5.3

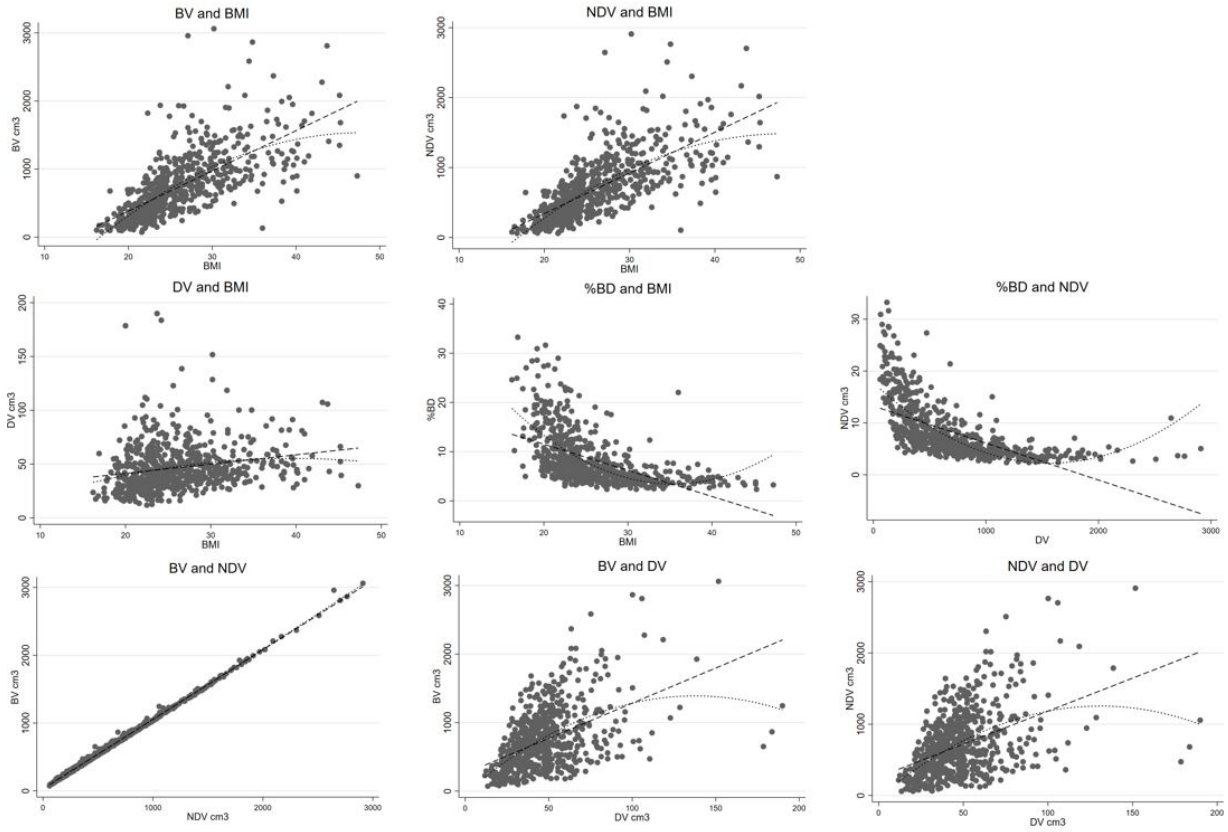


Figure 5.2 Scatter plots showing raw correlations between breast composition estimates and BMI

Notes: Estimates averaged over MLO and CC views from left and right breasts. BMI self-reported. Outlier omitted where BMI>60 (1 woman). UK study data only. Lines show linear and quadratic lines of fit.

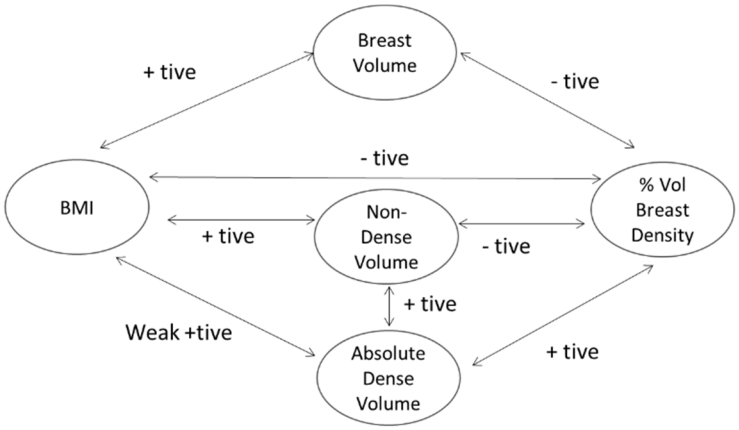


Figure 5.3 Overview of directions and strengths of correlations between breast volumetric estimates and BMI

The linear regression analyses (Table 5.1) show the relative change in BV or NDV associated with 1 SD change in BMI, after adjustment for age and ethnicity. Both BV and NDV were significantly associated with BMI ($p<0.001$) and there was little difference between the strength of association or the R^2 values. Very similar results were found in the analyses stratified by view. For MLO-only views

there was a slightly larger relative change and the R^2 value was slightly larger, although the differences were not statistically significant. The strongest overall association was between BMI and the log transformed NDV from the MLO view (Relative change: 1.59 (95% CI 1.53 to 1.65), p<0.001; adjusted R^2 0.487). These findings were robust to sensitivity analyses based on age and ethnicity (Table 5.1). The adjusted regression analyses confirm that relative volumetric breast density (%BD) is positively correlated with BMI whilst absolute volumetric breast density (DV) is negatively correlated with BMI.

Supplementary findings

These supplementary analyses clarified the associations between volumetric breast measurements and BMI in screening age women. The findings were robust to a range of sensitivity analyses; the magnitude of these associations was little affected by stratification by view or by restricting the age range or by looking at the White ethnic group only. The positive correlation between NVD and BMI is consistent with area-based methods of Baglietto (127), who reported a p=0.62 Spearman rank correlation between non-dense area and BMI. Our finding of a weak positive association between DV and BMI is consistent with the volumetric studies of Shepherd et al (130) and Schetter et al (128) but conflicts with Irwin et al (229) and Sun et al (230) who observed the opposite relationship in area-based studies. In addition we noted a positive association between DV and NDV which is again consistent with Schetter et al (128) but contradicts area-based studies which identify a negative correlation (127). These differences may reflect the difference in methods by which density is calculated between volumetric and area-based systems as explained in detail by Eng et al (132). Area-based automated methods use a cut-off intensity to dichotomize breast pixels as being either dense or non-dense with no gradation whereas volumetric methods quantify a continuous scale for each pixel. The positive correlation between volumetric density and breast volume is explained by a trend towards increased dense tissue as well as increased fat in larger breasts. Our results support Shepherd's view that MLO views may provide a better indication of % body fat (231), since the Volpara measurements derived solely from MLO images had the strongest association with BMI (although the differences between views were not significant at the 5% level). The correlations found in the raw data are of similar strength and the same direction as those found using adjusted data in Paper I.

Overall, the average NDV, from MLO and CC views, was recommended as the preferred proxy for BMI because although the MLO view is a slightly better fit, the differences are small. Using averaged values makes the proxy more robust in situations where individual images may be missing. If NDV is not available, then BV is also a good candidate for a BMI proxy with very similar fit and distribution.

Table 5.1 Volpara breast measurements associations with BMI with adjustment for demographic factors [1]

Average over both CC MLO Views (obs = 651)									
	Relative change	(95% CI)	P value	Adjusted R2		Relative change	(95% CI)	P value	Adjusted R2
Breast volume (BV)	1.53	(1.48 to 1.59)	<0.0001	0.471	**Breast Density (DV)**	1.12	(1.09 to 1.15)	<0.0001	0.111
Non dense volume (NDV)	1.58	(1.52 to 1.64)	<0.0001	0.474	**Breast Density %**	0.73	(0.71 to 0.75)	<0.0001	0.409
MLO View Only (obs = 650)									
	Relative change	(95% CI)	P value	Adjusted R2		Relative change	(95% CI)	P value	Adjusted R2
Breast volume (BV)	1.55	(1.49 to 1.60)	<0.0001	0.483	**Breast Density (DV)**	1.13	(1.10 to 1.17)	<0.0001	0.133
Non dense volume (NDV)	1.59	(1.53 to 1.65)	<0.0001	0.487	**Breast Density %**	0.73	(0.71 to 0.76)	<0.0001	0.407
CC View only (obs = 651)									
	Relative change	(95% CI)	P value	Adjusted R2		Relative change	(95% CI)	P value	Adjusted R2
Breast volume (BV)	1.52	(1.46 to 1.57)	<0.0001	0.448	**Breast Density (DV)**	1.10	(1.07 to 1.14)	<0.0001	0.095
Non dense volume (NDV)	1.57	(1.51 to 1.63)	<0.0001	0.453	**Breast Density %**	0.73	(0.70 to 0.75)	<0.0001	0.390
WOMEN AGED 50-70 ONLY (obs = 613)									
	Relative change	(95% CI)	P value	Adjusted R2		Relative change	(95% CI)	P value	Adjusted R2
Breast volume (BV)	1.53	(1.47 to 1.58)	<0.0001	0.467	**Breast Density (DV)**	1.12	(1.08 to 1.14)	<0.0001	0.102
Non dense volume (NDV)	1.57	(1.51 to 1.64)	<0.0001	0.471	**Breast Density %**	0.73	(0.71 to 0.76)	<0.0001	0.409
WOMEN ETHNICITY = WHITE ONLY (obs = 520)									
	Relative change	(95% CI)	P value	Adjusted R2		Relative change	(95% CI)	P value	Adjusted R2
Breast volume (BV)	1.54	(1.48 to 1.61)	<0.0001	0.470	**Breast Density (DV)**	1.13	(1.09 to 1.17)	<0.0001	0.114
Non dense volume (NDV)	1.59	(1.52 to 1.66)	<0.0001	0.475	**Breast Density %**	0.73	(0.70 to 0.76)	<0.0001	0.405

Footnotes: [1]Volpara volumes from controls averaged over left and right sides and both CC and MLO views (except where explicitly stated). Log transformed. All coefficients are adjusted by age group at mammogram and Ethnicity. White, Black (African or Caribbean) , Asian (Indian, Pakistani, Bangladeshi), Other. Table shows: Relative change in volumetric density measure (BV, NDV, DV or %density) associated with one standard deviation increase in BMI

5.5 Reasons for using the OPERA method for Paper I

Following initial submission for review it was pointed out by reviewers that comparison of risk gradients from regression using cross-sectional standard deviation, although very common in this type of study, may not be the optimum method in situations where models are derived using a different set of confounders and different study participants. Reviewers proposed that analysis should be carried out using Hopper's OPERA (odds ratio per adjusted standard deviation) concept (232, 233), in which the risk gradient is measured on a scale that takes into account other factors adjusted for by design or analysis. Thus, allowing different measures to be compared more appropriately. Hopper points out that this is particularly relevant for studies where different risk factors including mammographic density measurements are being used to assess breast cancer risk. Typically, relative risks are expressed as a reflection of change to the risk factor whilst holding all other risk factors constant (change in risk per one unit standard deviation in the unadjusted risk factor). In the case of OPERA a two stage process is followed. Firstly, the exposures of interest are regressed on potential confounders including age, menopausal status, parity etc to derive residuals that are the effect of the exposure once other factors have been accounted for. These residuals are standardised and used as explanatory variables in regression models to give OR per adjusted standard deviation that can be compared across different models. This was deemed an appropriate method for the analyses in Paper I, because we wished to compare results from 2 different study populations using different main explanatory variables (%BD and DV) with different sets of confounders. In practice the direction and magnitude of the findings were not altered by using the OPERA approach, but this method provides results that can more readily be compared with future studies in this area.

5.6 Summary Review

This study and supplementary analyses were an important first step in showing that objective automated measurements based on mammogram analysis can be used more widely in breast screening and their potential application is not restricted solely to the reporting of breast density. This study shows that if BMI data is not available when assessing risks associated with BD, then a suitable adjustment can be made by using another characteristic of the mammographic image. Automatically estimated NDV was found to be most appropriate proxy for BMI, but BV is also very appropriate. Interestingly, evidence from the OPERA comparisons suggests that the association between BD and BC risk is stronger when automated proxies for adiposity (i.e. mammographic NDV or BV) are used rather than self-reported BMI. The reason for this may be that BMI is subject to reporting errors whereas the mammographic measurements are objective. It was acknowledged in Paper I that there were shortcomings to our studies, mainly because they were limited to screening

age populations. It is known that the association between BMI and BC risk is affected by menopausal status and our findings are therefore restricted to postmenopausal women although there is no reason to assume that automated estimation of NDV and BV would not be appropriate proxies for BMI in younger women. These studies were also restricted to the use of proxies for BMI in the assessment of near term (intrinsic) breast cancer risk but not in models that predict the masking effect of BD, where the role of BMI is less well understood.

6 CHAPTER 6 BREAST FLUCTUATING ASYMMETRY AND BC RISK IN BREAST SCREENING PROGRAMMES

6.1 Introduction

The literature review in Chapter 2 (section 2.8) found a longstanding body of research that supports the theory that fluctuating asymmetry (FA) in paired features is associated with developmental stability, fecundity, phenotypic quality and general health including cancer risk. There was limited research to suggest that left:right asymmetry in breast size may be a risk factor for near and medium term BC and therefore may be a cue, not only to the presence of BC at screening, but also a potential risk factor (in combination with other risk factors) for targeted screening strategies.

The potential for using this novel breast asymmetry measure to improve breast screening performance, depends partly upon the prevalence of breast FA in the screening population at large. Since most early studies were either small, or conducted in a plastic surgery setting, it was therefore pertinent to first conduct a study of the prevalence of breast volumetric asymmetry within a large, ethnically, and socio-economically, diverse breast screening population. Until recently there have been no practical methods for assessing breast asymmetry in large volumes but the introduction of automated mammographic image analysis tools has recently made this feasible. Study II was designed to address these shortcomings in the literature.

The literature search found limited research to suggest that left:right asymmetry in breast *density* may also be associated with higher short and medium term risk of BC but no research was found into whether such asymmetries were associated with particular sub-types of BC, which is plausible given the associations between reproductive and hormonal risk factors and both FA and BC of varying sub-types. Study III was therefore designed to address the research question: “Is left-right breast asymmetry in breast volume or breast density associated with diagnosis of screen-detected cancers or false negatives at screening and is higher asymmetry a cue that the cancer may be of a more aggressive subtype?”. Studies II and II together aimed to further our knowledge about the potential for automated breast FA measurements as a tool in the improvement of breast screening performance.

6.2 Aims and Objectives

Aims:

(i) to examine ethnic, age and socio-economic differences in automated breast measurements (%BD, BV, DV, and left-right asymmetries in these measurements) across an ethnically and socio-economically diverse screening population using routinely collected data;

(ii) to assess the strength of associations between these automated mammographic measurements and the odds of screen-detected and interval breast cancers.

Primary Objectives:

1) Describe Volpara BV, %BD, DV in a diverse UK screening population, by age, ethnicity and Index of Multiple Deprivation (IMD) of area of residence (225).

2) Describe left:right breast asymmetry (i.e. absolute differences) in BV, %BD and DV by age, ethnicity, IMD and BV.

3) Assess the strength of association between age, ethnicity, BV and the outcome measures of asymmetry.

4) Investigate whether asymmetry in %BD/BV/DV is associated with increased odds of cancer detection at the current screen.

5) Investigate whether asymmetry in %BD/BV/DV is associated with increased odds of having an interval cancer.

6) Investigate whether asymmetry in %BD/BV/DV is associated with increased odds of having a cancer detected at the next routine screen.

Secondary Objectives:

7) Investigate the validity of asymmetry measurements using expert opinion from consultant radiologist.

8) Test the reliability of BD and asymmetry measures.

9) Investigate whether asymmetry in %BD/BV/DV/ is associated with increased odds of detecting an ER+ or PR+ cancer or a TNBC at screen at current screen.

Objectives 1 – 3 were addressed in Research Paper II. This paper is supplemented by unpublished supplementary descriptive analyses, not covered by the published paper due to word count constraints, in section 6.4.

Objectives 4 – 6, 9 were addressed in Research Paper III.

Objectives 7 and 8 were addressed by unpublished validation studies, see section 4.4. for summary and Appendix E for details of analysis.

6.3 Paper II Ethnic and age differences in right-left breast asymmetry in a large population-based screening population (234).

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	12406683	**Title**	Ms
First Name(s)	Susan M		
Surname/Family Name	Hudson		
Thesis Title	Beyond breast density – Novel uses of automated mammographic analysis in breast cancer screening		
Primary Supervisor	Professor Isabel dos Santos Silva		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	The British Journal of Radiology		
When was the work published?	2019		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	n/a		
Have you retained the copyright for the work?*	**No**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I was responsible for the conceptualisation of the work, development of the methodology, setting up of the mammogram analysis tools in conjunction with Dr Wilkinson and data collection, cleansing and linking of data sources, data analysis and writing of the publication.

SECTION E

Student Signature	[redacted]
Date	

Supervisor Signature	
Date	

BJR

https://doi.org/10.1259/bjr.20190328

Received: 03 April 2019 | Revised: 26 September 2019 | Accepted: 15 October 2019

Cite this article as:
Hudson SM, Wilkinson LS, Denholm R, De Stavola BL, dos-Santos-Silva I. Ethnic and age differences in right-left breast asymmetry in a large population-based screening population. *Br J Radiol* 2019; **92**: 20190328.

FULL PAPER

Ethnic and age differences in right-left breast asymmetry in a large population-based screening population

[1]SUE M HUDSON, [2]LOUISE S WILKINSON, [3]RACHEL DENHOLM, [4]BIANCA L DE STAVOLA and [1]ISABEL DOS-SANTOS-SILVA

[1]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK
[2]Oxford Breast Imaging Centre, University of Oxford Hospitals NHS Foundation Trust, Oxford, UK
[3]Centre for Academic Primary Care, Bristol Medical School, University of Bristol, Bristol, UK
[4]Population, Policy and Practice Programme, Great Ormond Street Institute of Child Health, University College London, UK

Address correspondence to: Ms Sue M Hudson
E-mail: *susan.hudson@lshtm.ac.uk;sue.hudson@pasconsulting.co.uk*

Objective: Exposure to sex hormones is important in the pathogenesis of breast cancer and inability to tolerate such exposure may be reflected in increased asymmetrical growth of the breasts. This study aims to characterize, for the first time, asymmetry in breast volume (BV) and radiodense volume (DV) in a large ethnically diverse population.
Methods: Automated measurements from digital raw mammographic images of 54,591 cancer-free participants (aged 47–73) in a UK breast screening programme were used to calculate absolute (cm^3) and relative asymmetry in BV and DV. Logistic regression models were fitted to assess asymmetry associations with age and ethnicity.
Results: BV and DV absolute asymmetry were positively correlated with the corresponding volumetric dimension (BV or DV). BV absolute asymmetry increased, whilst DV absolute asymmetry decreased, with increasing age (P-for-linear-trend <0.001 for both). Relative to Whites, Blacks had statistically significantly higher, and Chinese lower, BV and DV absolute asymmetries. However, after adjustment for the corresponding underlying volumetric dimension the age and ethnic differences were greatly attenuated. Median relative (fluctuating) BV and DV asymmetry were 2.34 and 3.28% respectively.
Conclusion: After adjusting for the relevant volumetric dimension (BV or DV), age and ethnic differences in absolute breast asymmetry were largely resolved.
Advances in knowledge: Previous small studies have reported breast asymmetry—breast cancer associations. Automated measurements of asymmetry allow the conduct of large-scale studies to further investigate these associations.

INTRODUCTION

Exposure to endogenous and exogenous sex hormones are recognized to be important in breast development and in the pathogenesis of breast cancer,[1–5] with the effect of many reproductive factors on breast cancer risk, *e.g.* early age at menarche and late age at menopause, being mediated by circulating levels of these hormones.[6] There is also some evidence that prenatal exposure to high levels of sex hormones may increase the risk of breast cancer. Breast cancer risk is elevated in females who were exposed *in utero* to diethylstilboestrol given to their mothers to prevent pregnancy complications[7] and some studies have reported positive associations between breast cancer risk and birth size, pre-eclampsia and multiple births, all possible markers of raised, *in utero*, exposure to oestrogens.[8] It is also thought that an individual's ability to tolerate exposure to oestrogens, particularly during periods of growth, may be reflected in a higher degree of homeostasis and thus bilateral symmetrical development of paired organs such as the breasts.[9] Increased "fluctuating asymmetry," *i.e.* increased anthropometrical asymmetry in paired features, is a common response to increased stress during development[10] and is related to both fecundity and general health.[11–14] For example, studies of dermatoglyphics have shown that increased asymmetry in hand patterns is associated with increased risk of several diseases including breast cancer.[15] Also, females with high second digit to fourth digit ratio (2D:4D) (thought to be associated with lower exposure or sensitivity to prenatal testosterone and/or higher levels *in utero* oestrogen levels) had increased risk of breast cancer[16] and they presented with breast cancer at a younger age.[17,18] An association between left-handedness and increased risk

of breast cancer has also been reported.[19,20] Manning et al showed that increased breast FA was correlated not only with age, height and parenchymal type but also with reproductive factors such as parity, age at first birth and age at menopause.[9]

Only a few small-sized studies, mainly among Caucasians, have so far examined the association between breast size asymmetry and breast cancer risk. Their findings are consistent with asymmetry being associated with the presence of a breast cancer[21–24] as well as with a higher risk of having a breast cancer diagnosed in the short- and medium-term (mean interval between mammography and diagnosis 6.44 years).[25] Mammographic density captures the amount of radiodense tissue in the breast, and there is also some evidence that asymmetry in density might be associated with higher short-term likelihood of being diagnosed with breast cancer.[26–28] It has also been suggested that a slightly larger left breast, with a higher volume of radiodense tissue, may account for the slightly higher frequency of cancers in the left than the right breast although the mechanisms for this are poorly understood.[29–31] Overall, the findings from these studies suggest that asymmetry in breast size and density may reflect underlying biological mechanisms linked to the pathogenesis of breast cancer or may be early consequences of the presence of a tumour. Hence, asymmetry measurements have the potential to be used as risk predictors or diagnostic markers. To our knowledge there is, as yet, no large-scale study of the prevalence of breast volume asymmetry and breast density asymmetry from large population-based studies.

The recent introduction of full-field digital mammography has led to the development of automated algorithms which allow volumetric assessments of both breast size and mammographic density from two-dimensional digital mammographic images. Such automated methods make it feasible to conduct large-scale studies based on objective measurements of bilateral asymmetry in breast size and mammographic density. This study aims to quantify bilateral asymmetry in breast size and mammographic density volume in a very large, and ethnically diverse sample of over 54,000 females who participated in a population-based breast screening programme in England. The findings will provide the first population-based data on the distribution of breast asymmetry, and potential age and ethnic variations.

METHODS

Study participants

The study participants were females resident in one of five London boroughs—Wandsworth, Merton, Croydon, Sutton, Richmond and Kingston—who underwent routine 3-yearly screening mammography as part of the England and Wales National Health Service Breast Screening Programme (NHSBSP) at the South West London Breast Screening Service based in the St George's University Hospitals National Health Service (NHS) Foundation Trust. The NHSBSP is an organized population-based mammographic screening programme, with a call–recall system, which targets females aged 50–70 years and has a coverage of ~75%.[32] Also included were a small number of younger females (aged 29–45) who had been identified as having a higher risk of breast cancer and therefore were invited for screening on an annual basis,[33] plus any females over 73 years who had optionally contacted the service for a self-referred screening appointment. All females were asymptomatic at the time of screening. Participants were screened during the period 01 March 2013 to 18 August 2016. Data on ethnicity were collected as part of the standard screening protocol via a self-completed screening questionnaire. Ethnicity was categorized according to the Census classification and summarized as, "Asian" (Indian, Pakistani or Bangladeshi or other), "Black-African," "Black-British or Caribbean or other," "Chinese," "Mixed" (White and Black, White and Asian or any other mixed), "White" (British or Irish or other) and "Other."[34] Data for other known breast cancer risk factors (*e.g.* parity, duration of breast feeding, age at menarche, body mass index (BMI), family-history of breast cancer) are not collected in a systematic way across the NHSBSP screening programme and thus were unavailable.

Each participant underwent the NHSBSP standard 2-view [craniocaudal (CC) and mediolateral-oblique views (MLO)] mammography of each breast,[35] with the set of four digital raw images being stored on the South West London Breast Screening Service Picture Archiving and Communication system. The images were double read with arbitration by consensus. When participants had multiple screening episodes during the study period, only images from the earliest screen episode were included in the analysis. Raw digital mammographic images were processed via the automated algorithm Volpara® Density™ v. 1.5.11 (Volpara), (Matakina Technology Limited, Wellington, New Zealand)[36]; this algorithm provided fully automated estimates (in cm^3) of the volume of the breast (BV) and the volume of the radiodense tissue (DV) separately for each of the four [left (L) and right (R) breasts/CC and MLO views] images. The screening programme does not use mammographic density as a diagnostic aid, and participants are not informed on whether they have dense breasts.

In all, 66,176 females were screened during the study period. Females were excluded from this analysis if cancer was detected by the current screen (N = 530); if they had a previous history of breast cancer (N = 438); if their screen images were classified as "technical recall," *i.e.* were considered by the reader not to be of high enough quality for diagnosis (N = 26); if they had breast implants; if their standard set of four images (*i.e.* L/R CC and MLO images) was incomplete (N = 9823); and if at least one of the two CC images was rejected by Volpara based on its internal consistency checks (N = 7338). Exclusions were not mutually exclusive, leaving a total of 54,591 females who were eligible for inclusion in the analysis.

Ethical approval

This retrospective study was carried out on fully anonymous, routinely collected data only, held in accordance with the NHS Cancer Screening Programmes Confidentiality and Disclosure Policy 2011. The NHSBSP has section 251 support under the NHS Act 2006. The study was approved by all relevant ethics committees (Research Ethics Committees from St George's University Hospitals NHS Foundation Trust, and the London School of Hygiene and Tropical Medicine).

Statistical methods

For each participant the BV and DV was calculated as the average of the readings obtained from the same side CC and MLO images (*i.e.* CC and MLO views were used to obtain an overall average). Both absolute and relative measures of left-right asymmetry were calculated: *absolute* asymmetry (in cm^3), *i.e.* the unsigned difference between left BV (or DV) and right BV (or DV), and *relative* asymmetry as (|L-R|)/ (*L* + R)/2 expressed as a percentage. Absolute and relative asymmetry were estimated from the CC images only because this view is likely to capture the whole of the breast whilst being less affected than the MLO view by the inclusion of variable amounts of retroglandular fat tissue near the chest wall.[36] (For comparison the equivalent asymmetry measures were also calculated using the MLO views only).

The distributions of absolute and relative asymmetry values were plotted. Natural-log transformations were applied to normalize the distributions of absolute and relative BV and DV asymmetry and quintiles were used to categorize BV and DV into five equally sized categories.

To examine whether age-related variations in breast volume and breast asymmetry differ across the various ethnic groups, medians, 25th and 75th centiles of the distributions of untransformed BV, DV and absolute asymmetry measures were also calculated and plotted separately by 5 year age categories and ethnicity. These were also calculated for each single year of age and plotted after smoothing using a Lowess function (values based on fewer than 20 observations were omitted from the plots). Scatter plots and Spearman correlation coefficients were used to examine the correlations between asymmetry measures and the corresponding volumetric dimension. In order to assess whether allometry is a feature of this relationship (as identified by Manning et al[9]) we regressed log of asymmetry on log of the corresponding volumetric measure.

Linear regression models were used to examine the strength of the associations between each exposure variable—age and ethnicity—and the outcome variables, BV or DV absolute asymmetry, controlling for their respective average volume (BV or DV). Because of the log-transformation, regression coefficients represent the relative change (RC) in absolute asymmetry per one unit change in the exposure category. In all the analyses, we considered statistical significance (two-sided) at p-value < 0.05. All analyses were conducted in Stata (IC 14).[37]

RESULTS

Study participants

The characteristics of the 54,591 participants are shown in Table 1. The majority (~87%) of participants were within the ages of 50–70 years, the age-group targeted by the NHSBSP. Among the 85% of the participants who reported their ethnicity,~76% were White but there were also high numbers of females of Black and Asian ethnicity.

Breast volume, dense volume and absolute asymmetry by age and ethnicity

The median (25th, 75th centiles) BV and DV values for the whole study sample were 757 (496, 1112) cm^3 and 48.9 (36.8, 66.5) cm^3, respectively (Table 1). There was, however, evidence of bilateral asymmetry in BV and DV, with a median (25th, 75th centiles) absolute difference in BV and DV between the two breasts of 60.6 (26.6, 117.8) cm^3 and 5.71 (2.49, 11.27) cm^3, respectively, with the wide interquartile range (IQR) indicating considerable between-woman variation in bilateral asymmetry (Table 1). This difference was seen in every age and ethnic group, albeit with some variations with the smallest median absolute differences seen among Chinese females.

The distributions of BV and DV absolute asymmetry estimates were right skewed and, hence, a log-normal transformation was used to normalize them (Figure 1). The transformed BV and DV asymmetry distributions approximated a normal distribution although both were leptokurtic (kurtosis coefficient: 5.60 and 4.76, respectively) and slightly skewed (skewness coefficient: −1.12 and −0.96, respectively).

Further analyses by age-group show that, on average, BV increased slightly with increasing age up to ages 55–59, declining thereafter (Figure 2). Ethnic variations in BV were much more marked than those observed with age (Figure 3), with BV being, on average, highest among Black Caribbean (median: 956 cm^3) and Black African (960 cm^3) females and lowest among Chinese females (394 cm^3) but with wide between-woman variability being present within each ethnic group. Absolute BV asymmetry showed similar age and ethnicity patterns to those observed for BV (Figures 2 and 3).

In contrast to BV, DV decreased, on average, with increasing age-group from <45 to 70+ years but, similarly to BV, DV was highest among Black Caribbean (median: 58.3 cm^3) and Black African females (56.0 cm^3) and lowest among Chinese females (41.0 cm^3). Absolute DV asymmetry followed a similar pattern to DV, *i.e.* lower values across successive age-groups, and higher among Black African and Black Caribbean females (Figures 2 and 3).

The observed absolute asymmetry in BV and DV reflected that fact that, on average, females had a larger left breast with a larger amount of radiodense tissue. The only exception was that DV was higher in the right breast among Chinese females. [A]

Figure 4, which depicts median single-year-of-age volumetric and asymmetry values by ethnicity, shows that age-related changes in BV varied across the different ethnic groups. Among Asian, Black African and White females, BV increased progressively up to age ~60 years but declined thereafter whilst among Black Caribbean females, BV continued to increase up to age 70 years. In contrast, DV decreased with age in all ethnic groups. There was, however, a marked levelling out after age ~55. BV and DV absolute asymmetry follow the same general pattern as their corresponding underlying volumetric dimension.

Relative asymmetry by age and ethnicity

The magnitude of relative BV asymmetry was similar across all age groups (median overall relative BV asymmetry for all study participants: 2.43% [25th, 75th centiles: (1.15%, 4.19%); Table 1] except that it was slightly higher in the youngest age band

A More detailed presentation of Left to Right breast tissue volumes ratio data by ethnic groups is shown in supplementary material section 6.4 of this thesis.

Table 1. Characteristics of the study participants

			Median (25th and 75th centiles)					
	No.	**Percent**	**BV (cm^3) a**	**DV (cm^3) a**	**BV absolute CC asymmetry b (cm^3)**	**DV absolute CC asymmetry b (cm^3)**	**BV relative CC asymmetry (%) c**	**DV relative CC asymmetry (%) c**
Age at screening (yrs)								
<45	234	0.4	563 (353, 950)	63.8 (47.4, 94.7)	56.7 (24.7, 105.6)	7.79 (3.86, 16.18)	2.87 (1.48, 4.56)	3.26 (1.87, 5.74)
45–49	3297	6.0	727 (450, 1135)	62.0 (45.4, 85.3)	57.5 (24.7, 112.8)	7.46 (3.39, 14.46)	2.42 (1.11, 4.06)	3.40 (1.62, 5.83)
50–54	15,405	28.2	762 (485, 1138)	54.3 (40.6, 75.5)	59.2 (25.8, 115.4)	6.45 (2.79, 12.69)	2.36 (1.12, 4.09)	3.33 (1.54, 5.87)
55–59	12,408	22.7	770 (498, 1148)	48.5 (36.6, 64.6)	61.8 (26.8, 120.4)	5.59 (2.45, 10.86)	2.43 (1.15, 4.21)	3.26 (1.53, 5.82)
60–64	10,440	19.1	767 (515, 1109)	46.1 (35.1, 61.0)	60.1 (26.9, 117.1)	5.21 (2.26, 10.24)	2.41 (1.14, 4.13)	3.18 (1.47, 5.65)
65–69	9483	17.4	751 (506, 1063)	44.0 (33.9, 57.7)	62.5 (27.5, 120.9)	5.04 (2.19, 10.04)	2.50 (1.17, 4.33)	3.23 (1.47, 5.78)
70+	3297	6.0	723 (499, 1014)	42.9 (33.5, 56.1)	63.6 (28.3, 118.9)	5.25 (2.27, 10.10)	2.66 (1.24, 4.52)	3.28 (1.52, 5.79)
Missing	27	0.1						
Ethnic group								
White—British, Irish, Other	35,443	64.9	747 (485, 1098)	47.9 (36.1, 64.9)	59.3 (25.8, 115.5)	5.60 (2.44, 11.13)	2.42 (1.13, 4.18)	3.30 (1.53, 5.84)
Asian d	4829	8.9	718 (508, 1005)	44.8 (34.8, 59.6)	59.4 (27.2, 111.8)	5.02 (2.10, 9.91)	2.43 (1.19, 4.24)	3.17 (1.40, 5.52)
Black—British, Caribbean	2705	5.0	956 (610, 1381)	58.3 (44.6, 77.8)	71.6 (31.1, 136.4)	6.59 (3.02, 12.17)	2.26 (1.04, 4.02)	3.17 (1.50, 5.49)
Black—African	1999	3.7	960 (672, 1347)	56.0 (42.1, 74.0)	81.1 (35.7, 155.5)	6.39 (2.95, 12.65)	2.50 (1.20, 4.14)	3.23 (1.50, 5.64)
Mixed e	1029	1.9	800 (535, 1176)	53.0 (39.4, 71.5)	64.5 (28.4, 124.5)	6.12 (2.71, 11.66)	2.36 (1.13, 4.21)	3.28 (1.50, 5.54)
Chinese	654	1.2	394 (258, 552)	41.0 (29.6, 60.7)	35.1 (16.2, 67.7)	5.03 (2.28, 9.67)	2.71 (1.38, 4.68)	3.38 (1.60, 6.52)
Missing or not reported	7932	14.5	751 (499, 1121)	51.2 (38.5, 70.8)	61.6 (27.5, 119.5)	6.20 (2.71, 12.05)	2.48 (1.19, 4.21)	3.35 (1.56, 5.91)
All females	54,591		757 (496, 1112)	48.9 (36.8, 66.5)	60.6 (26.6, 117.8)	5.71 (2.49, 11.27)	2.43 (1.15, 4.19)	3.28 (1.52, 5.79)

BV, breast volume; CC, craniocaudal; DV, dense volume; MLO, mediolateral oblique.

[a]Calculated from the average BV (or DV) value from the four images: left CC image, right CC image, left MLO image, right MLO image.

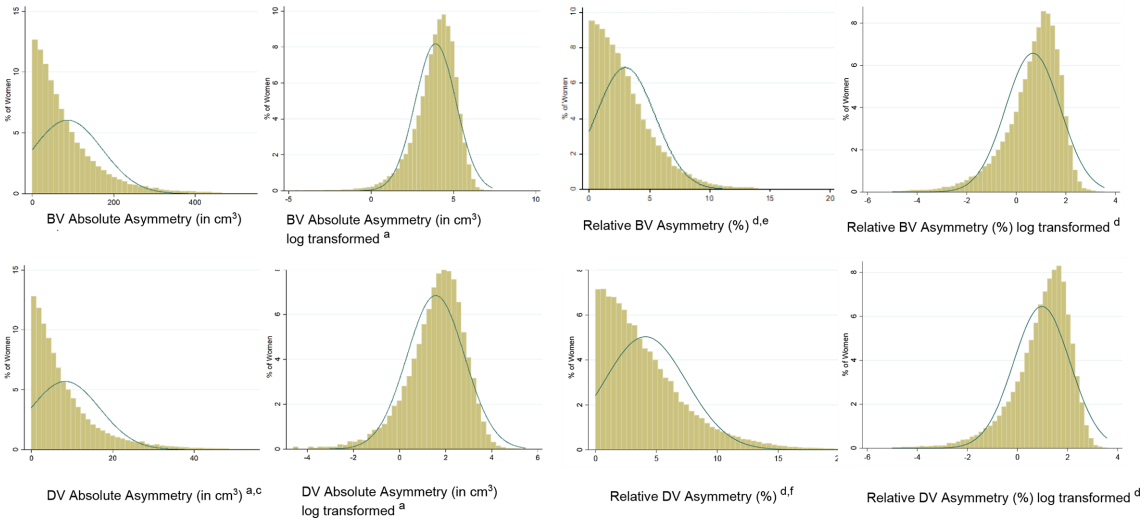
[b]Calculated as the absolute difference between the BV (or DV) value from the left CC image and the BV (or DV) value from the right CC image.

[c]Relative asymmetry estimated as (|L-R|)/ (*L* + R)/2*100, where L and R are volumes from the left and right breasts estimates from the CC views.

[d]Asian includes: British Indian, Pakistani, Bangladeshi, Other Asian excluding Chinese

[e]Mixed includes: White and Black, White and Asian or any other mixed

Figure 1. Distribution of breast tissue absolute and relative asymmetry measurements. (a) Absolute asymmetry derived from absolute difference left and right CC views. (b) Outliers where absolute BV asymmetry >610 cm^3 (10 *mean value) have been omitted to aid clarity (n = 109). (c) Outliers where absolute DV asymmetry >57cm3 (10 * mean value) have been omitted to aid clarity (n = 252). (d) Relative symmetry % derived from (|L-R|)/ (L + R)/2*100, where L and R represent the volumes of the left and right breasts as estimated from the CC views. (e) Outliers where relative BV asymmetry >20% have been omitted to aid clarity (n = 51). (f) Outliers where relative DV asymmetry >20% have been omitted to aid clarity (n = 139). BV,breast volume; CC, craniocaudal; DV, dense volume.



[median 2.87% (1.48%, 4.56%)]. The magnitude of relative BV asymmetry was also similar irrespective of the ethnicity of the participants although slightly higher in the Chinese ethnic group [2.71% (1.38%, 4.68%)].

The magnitude of relative DV asymmetry was similar across all age groups and ethnicities [median overall relative DV asymmetry for all study participants: 3.28% (1.52%, 5.79%)]. Overall age and ethnic variations in relative BV and DV asymmetry were much less marked than those observed for absolute BV asymmetry and absolute DV asymmetry (Figures 2 and 3).

Correlations between absolute asymmetry and volumetric measures

BV and DV absolute asymmetry were moderately positively associated with their corresponding underlying volumetric measure (Spearman correlation coefficient (r): 0.45 and 0.43, respectively; $p < 0.0001$ for both). Regressing log BV asymmetry on log BV revealed negative allometry [coefficient: 0.84; 95% confidence interval 0.83, 0.85)] whilst regressing log DV on log DV revealed slight positive allometry (1.09; 1.07, 1.12). There were no statistically significant differences in the magnitude of these allometry coefficients across the different ethnic groups (data not shown).

Associations between absolute asymmetry and age and ethnicity

The fitted linear regression models showed that BV absolute asymmetry increased with increasing age (in 5 year categories, P for trend (Pt) <0.001; Table 2), and that this trend persisted after adjustment for BV (Pt <0.001). In contrast, DV absolute asymmetry decreased with increasing age (Pt <0.001), but this trend was attenuated upon adjustment for DV (Pt = 0.14; Table 2). Further adjustment for ethnicity affected little the magnitude of the BV or DV absolute asymmetry associations with age (Table 2).

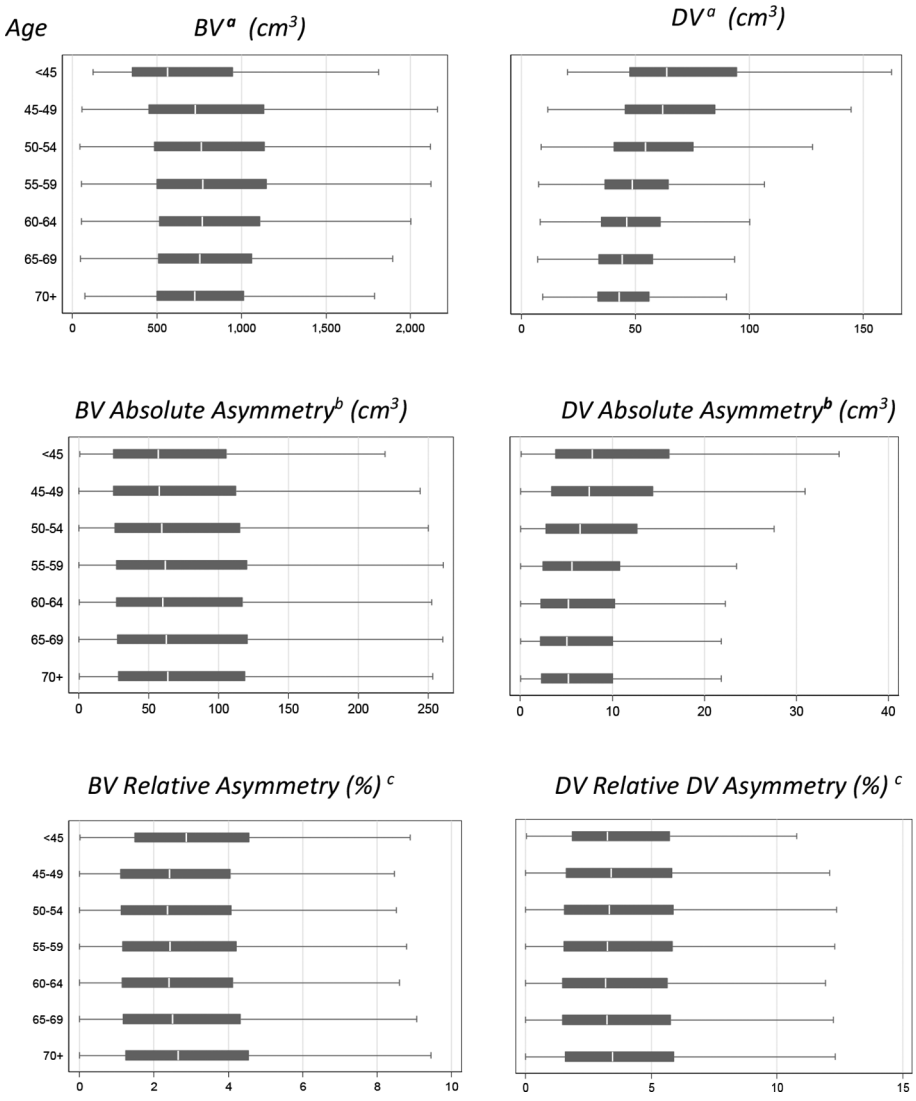
When considering ethnicity on its own, relative to White females (reference group) those of Black Caribbean, Black African and Mixed ethnicity had statistically significantly higher, whilst those of Chinese ethnicity had statistically significant lower, BV absolute asymmetry (Table 2). However, upon adjustment for BV the magnitude of these ethnic differentials was markedly reduced, remaining statistically significant only in Black African females (RC 1.13; 95% CI 1.07, 1.19), while there was borderline evidence of higher BV absolute asymmetry for Asian females (1.04; 1.00, 1.07; Table 2). Similarly, and still relative to White females, DV absolute asymmetry was found to be significantly higher among Black Caribbean and Black African females and significantly lower among Asian and Chinese females in unadjusted analyses. However, these differences remained significant after, adjustment for DV, only for Asian females (0.94; 0.91, 0.98; Table 2). There was no evidence of interaction between age and ethnicity in their effects on BV or DV absolute symmetry (p = 0.69 and p = 0.53, respectively).

DISCUSSION

Main findings

This study of >54,000 females clarifies the associations between absolute breast asymmetry and breast volume, with the findings being broadly consistent with those from a smaller study (n = 500 younger females) by Manning et al which showed that simple linear regression of BV absolute asymmetry (log transformed) on BV gives a significant positive association (our study $r^2 = 0.15$, $p < 0.001$; Manning $r^2 = 0.13$, $p < 0.001$).[9] We also found that absolute DV asymmetry is positively associated with DV. Thus,

Figure 2. Breast tissue volumes and asymmetry measurements by age, medians and IQR. (a) BV and DV are average values estimated from the four mammographic images: left CC image, right CC image, left MLO image, right MLO image. (b) Absolute asymmetry estimated from absolute difference between volume estimates derived from the left and right CC views. (c) Relative asymmetry estimated as (|L-R|)/ (*L* + R)/2*100, where L and R are volume estimates derived from the left and right CC views. Whiskers are calculated as lower adjacent value (*i.e.* smallest observed value ≥ lower quartile +1.5 IQR) and upper adjacent value (*i.e.,* largest observed value ≤ upper quartile +1.5 IQR). BV,breast volume; CC, craniocaudal; DV, dense volume; IQR, interquartile range; MLO,mediolateral oblique.

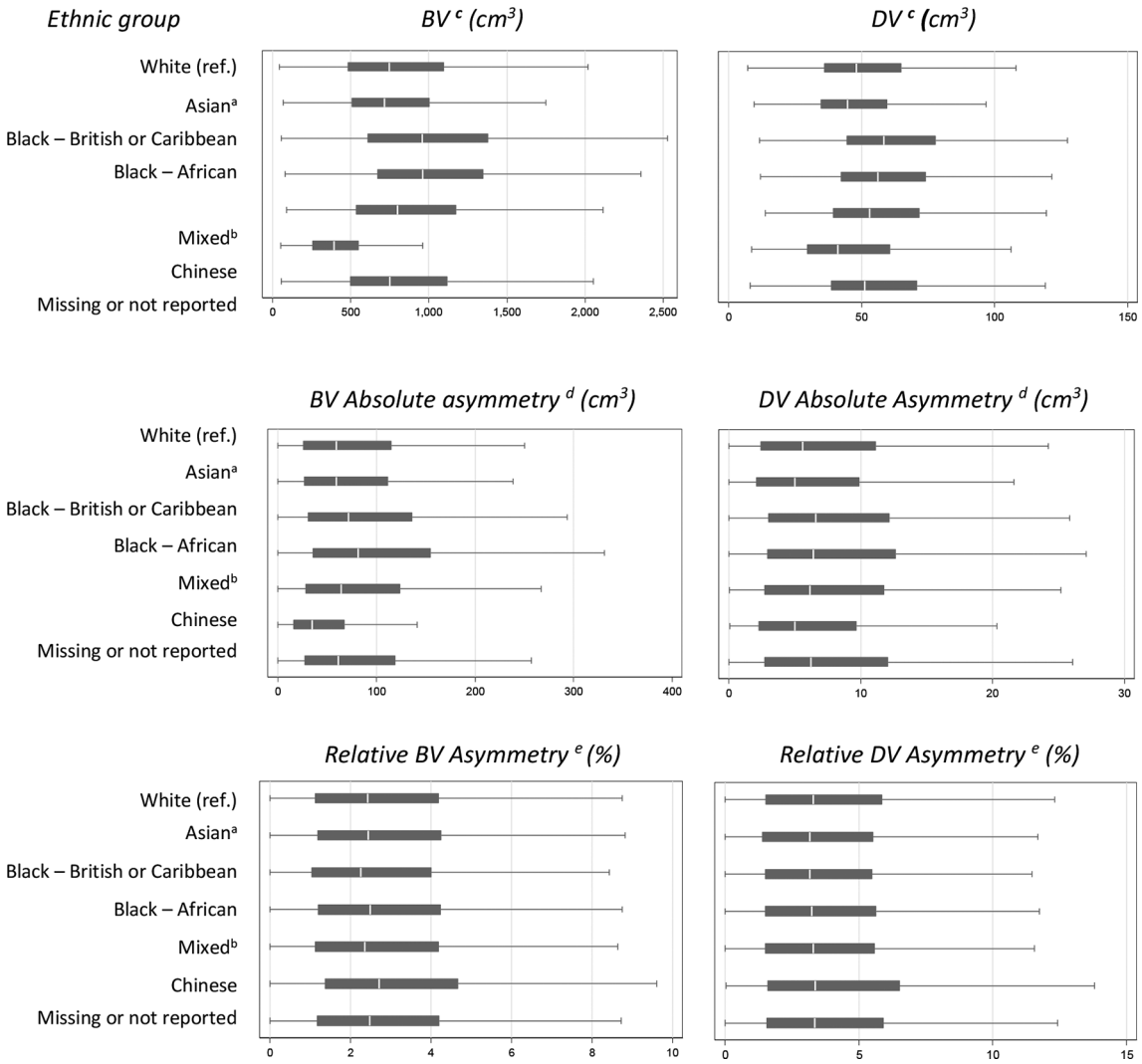


the larger BV (or DV) the higher the magnitude of BV (or DV) absolute asymmetry. This explained, at least in part, the higher levels of BV and DV asymmetry observed in females of Black ancestry as they also had, on average, higher BV and DV. After adjusting for the relevant breast volumetric measure (*i.e.* BV for BV asymmetry, DV for DV asymmetry), the ethnic differences in absolute breast asymmetry observed in the unadjusted analysis were attenuated, indicating that they were largely driven by ethnic differences in breast and dense tissue volumes.

Similar to the findings of Manning et al,[9] our findings showed that the BV absolute asymmetry/BV relationship was negatively allometric across all main ethnic groups, indicating that females with large breasts had a smaller fluctuating asymmetry than expected for their volume. There was, however, evidence that the DV absolute asymmetry/DV relationship was positively allometric.

Like Manning et al we found, using simple linear regression, that BV asymmetry is only weakly positively associated with age (our study $r^2 = 0.004$, $p < 0.001$, Manning $r^2 = 0.019$, $p = 0.02$).[9] The differences in the strength of the association might be explained by the fact that the females in our study were considerably older than those in the study by Manning et al[9] (mean ages 58.57 and 39.85 respectively). We found that DV absolute asymmetry is weakly but negatively associated with age, with these associations being attenuated upon adjustment for DV, indicating that these associations are largely driven by decreasing DV with age.

Figure 3.[B] Breast tissue volumes and asymmetry measurements by ethnicity, medians and IQR. (a) Asian = British Indian, Pakistani, Bangladeshi or other Asian excluding Chinese. (b) Mixed = Mixed White and Black, White and Asian and any other Mixed. (c) BV and DV are average values estimated from the four mammographic images: left CC image, right CC image, left MLO image, right MLO image. (d) Absolute asymmetry estimated from absolute difference between volume estimates derived from the left and right CC views. (e) Relative asymmetry derived from (|L-R|)/ (*L* + R)/2*100, where L and R are volumes from Left and Right CC views. Whiskers are calculated as lower adjacent value (*i.e.* smallest observed value ≥ lower quartile +1.5 IQR) and upper adjacent value (*i.e.* largest observed value ≤ upper quartile +1.5 IQR)



Two earlier studies, one in the USA ($n = 980$)[38] and the other in Switzerland ($n = 87$),[39] focused on the left:right ratio (L:R) in BV. Although such L:R ratio cannot be regarded as a measure of relative asymmetry, it is nevertheless worth noting that their findings are consistent with our finding that, on average, the left BV exceed the right BV by ~4% across the whole breast screening population irrespective of ethnicity and age. There was, however, marked between-woman variability in breast asymmetry among cancer-free, screened females.

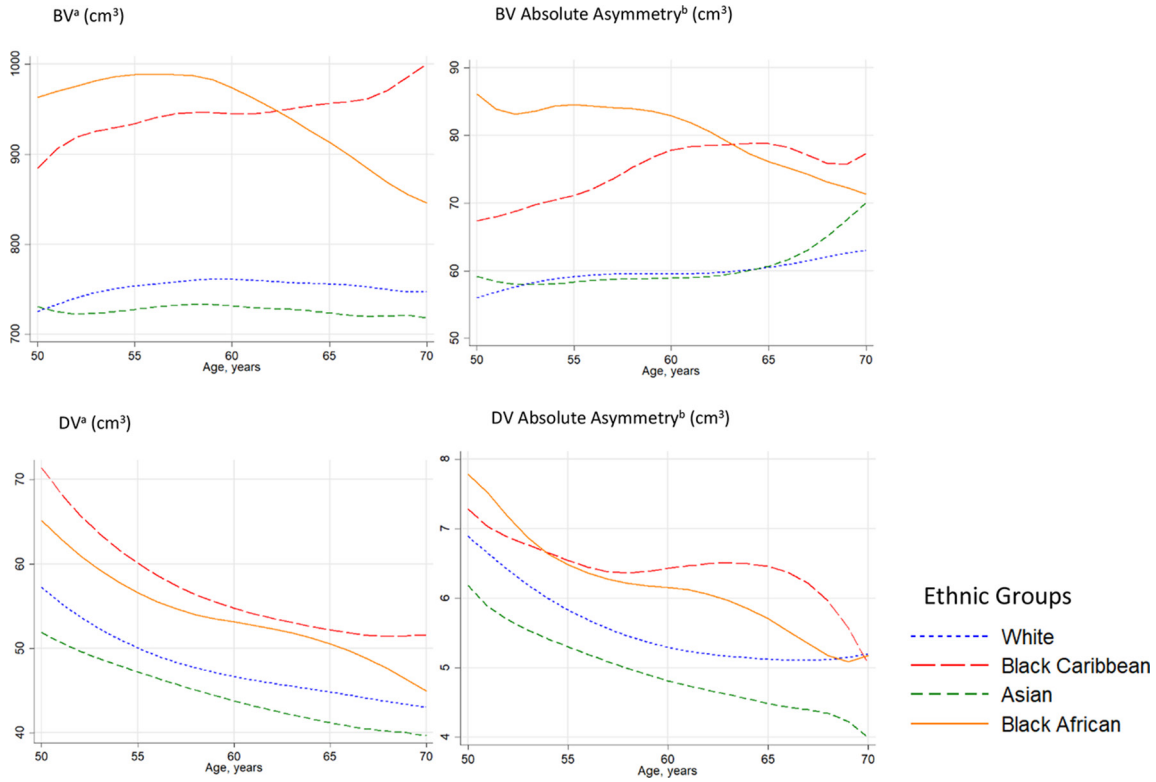
Literature on the prevalence of DV asymmetry is limited. Consistent with our findings Lee et al, in a study of 860 South Korean females, found that the L:R ratio in DV was less than one indicating a greater DV in the right breast,[40] thus challenging the view that the laterality of DV ratio is similar across all ethnic groups. Chen et al[41] on a small sample of 24 Taiwanese females also found that DV, as measured by MRI, was higher in the right than in the left breast.[A]

Strengths and limitations

Strengths of this study include its population-based design, the very large sample size relative to previous studies, and the wide ethnic mix. As the images for both breasts were collected at the same point in time, and under similar technical conditions, within-woman L:R breast comparisons are unlikely to have been biased by anthropometric, reproductive and lifestyle characteristics (*e.g.* BMI, menopausal status) or by differences in image acquisition (*e.g.* differences in mammographic equipment) as these would have affected both breasts similarly. This does not exclude, however, the possibility that the findings may have been affected by within-woman differences

A More details of Left to Right breast volume tissue ratios by ethnic group are shown in supplementary material Section 6.4 of this thesis. B Breast tissue asymmetry measurements by deprivation index are included in supplementary 6.4 .

Figure 4. Breast composition and breast composition asymmetry by age and ethnicity. Median volume, absolute asymmetry in cm^3 in each year band smoothed using Stata Lowess function. [a]BV and DV are average values from the four images: left CC image, right CC image, left MLO image, right MLO image. [b]Asymmetry derived from absolute difference left and right CC views. Year group excluded if fewer than 20 observations in that age group Asian = British Indian, Pakistani, Bangladeshi or other Asian excluding Chinese Chinese ethnicity omitted due to sparsity of data in older females. BV,breast volume; CC, craniocaudal; DV, dense volume; MLO, mediolateral oblique.



in the way the left and right breasts were examined (*e.g.* differences in a female's positioning during mammography). The study relied on an automated method to estimate the volumes of the left and right breasts and the amounts of their radiodense tissues, and thus such objective measurements were not influenced by subject or observer biases. Although the volumetric estimates were derived from two-dimensional images and, hence, may have been affected by errors, these would have affected both breasts similarly.

The study included mostly females of screening age and reflected a mix of ethnic groups living in England. The proportion (15%) of females for whom ethnicity data were missing was relatively low and typical for NHSBSP screening services where collection of self-reported ethnicity data is undertaken.[42] Females with a previous history of breast cancer, or who were diagnosed with cancer at the time of screening, as well as those with breasts implants, were excluded from the study; however, females with other conditions that might have affected their breast size (*e.g.* surgery for non-malignant conditions) could not be excluded as information on these conditions is not routinely collected by the NHSBSP.

A limitation of this study was the lack of data on potential confounders or mediators (*e.g.* BMI, reproductive history) of the age/ethnicity associations with BV and DV asymmetry. Menstrual cyclic variations in breast width asymmetry (measured from CC mammograms) were reported by Manning et al,[43] based on mammograms from 280 premenopausal females, with lowest breast asymmetry occurring around the middle of the cycle (which Scutt & Manning later attributed to ovulation[44]). Although the present study was unable to consider cyclical changes in asymmetry as information on the day of menstrual cycle when the mammogram was taken is not routinely collected by the NSHBSP, the large majority of females screened by the NHSBSP are of post-menopausal age. Nevertheless, future studies of pre-menopausal females should examine cyclic variations in asymmetry and, in particular, whether such variations should be taken into account when assessing asymmetry—breast cancer risk associations.

The study was conducted using one specific algorithm for estimating volumetric breast size and volumetric density. There is no published data specifically on the reliability of asymmetry measures derived from the Volpara volumetric measurements, but the latter have been found to be reliable and repeatable.[45–47] Nevertheless, it would be worthwhile to assess breast asymmetry using other automated methods. Our estimates of BV and DV asymmetry were derived from the CC views of the left and right breasts; however, MLO views produced similar breast asymmetry estimates *[e.g.,* median (IQR) for BV and DV absolute asymmetry for all participants was 60.6 (26.6, 117.8) cm^3 and 5.71 (2.5, 11.3) cm^3, respectively, if derived from the CC views and 65.1 (28.7, 127.0) cm^3 and

Table 2. Linear regression analysis of associations between BC risk factors and asymmetry measures

	BV absolute asymmetry[a] (*N* = 54,591)			DV absolute asymmetry[a] (*N* = 54,591)		
	Unadjusted	Adjusted for BV[b]	Mutually adjusted[c]	Unadjusted	Adjusted for DV[b]	Mutually adjusted[c]
Variable	**RC (95% CI)**	**RC (95% CI)**	**RC (95% CI)**	**RC (95% CI)**	**RC (95% CI)**	**RC (95% CI)**
BV / DV (per quintile)	1.41 (1.40, 1.42)		1.41 (1.40, 1.42)	1.40 (1.39, 141)		1.39 (1.38, 1.40)
	p < 0.001		*<0.001*	*<0.001*		*<0.001*
	$r^2 = 0.15$			$r^2 = 0.14$		
Age[d] (years)						
<45	0.90 (0.76, 1.05)	1.07 (0.92, 1.24)	1.08 (0.93, 1.25)	1.32 (1.12, 1.55)	1.13 (0.97, 1.31)	1.13 (0.91, 1.31)
45–49	0.98 (0.94, 1.03)	1.01 (0.97,1.06)	1.01 (0.93, 1.06)	1.16 (1.10, 1.21)	1.04 (0.99,1.08)	1.03 (0.99,1.08)
50–54 (ref)	1.00	1.00	1.00	1.00	1.00	1.00
55–59	1.04 (1.01, 1.07)	1.03 (1.00,1.05)	1.03 (1.00,1.06)	0.87 (0.84, 0.89)	0.98 (0.95,1.00)	0.98 (0.95,1.01)
60–64	1.04 (1.01, 1.08)	1.03 (1.00,1.06)	1.04 (1.00,1.07)	0.80 (0.78, 0.83)	0.95 (0.93,0.98)	0.96 (0.93,0.99)
65–69	1.04 (1.01, 1.08)	1.06 (1.03,1.09)	1.06 (1.03,1.10)	0.78 (0.76, 0.81)	0.98 (0.05,1.00)	0.98 (0.95,1.01)
70+	1.08 (1.03, 1.13)	1.12 (1.08,1.17)	1.14 (1.09,1.19)	0.82 (0.79, 0.86)	1.05 (1.01,1.10)	1.05 (1.01,1.10)
P for homogeneity	*0.01*	*0.01*	*<0.001*	*<0.001*	*<0.001*	*<0.001*
P trend	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*0.14*	0.20
	$r^2 = 0.004$			$r^2 = 0.009$		
Ethnicity						
White (ref.)	1	1	1	1	1	1
Asian[e]	1.01 (0.97, 1.05)	1.04 (1.00, 1.07)	1.04 (1.01, 1.08)	0.89 (0.85, 0.92)	0.94 (0.91, 0.98)	0.95 (0.91, 0.98)
Black—British/Caribbean	1.20 (1.14, 1.26)	1.00 (0.96,1.05)	1.01 (0.97, 1.06)	1.16 (1.11, 1.22)	0.95 (0.91,1.00)	0.95 (0.91, 1.00)
Black—African	1.39 (1.31, 1.47)	1.13 (1.07,1.19)	1.14 (1.08,1.20)	1.14 (1.08,1.21)	0.98 (0.93,1.03)	0.98 (0.93, 1.03)
Mixed[f]	1.01 (1.02, 1.19)	1.04 (0.97,1.12)	1.04 (0.97,1.12)	1.05 (0.97, 1.14)	0.96 (0.89,1.03)	0.96 (0.89, 1.03)
Chinese	0.60 (0.55, 0.66)	0.94 (0.86,1.03)	0.95 (0.87, 1.04)	0.90 (0.81, 0.99)	1.01 (0.93,1.11)	1.02 (0.93, 1.11)
P for homogeneity	*<0.001*	*<0.01*	*<0.01*	*<0.001*	*0.54*	*0.55*

BC, breast cancer; BV, breast volume; CC, craniocaudal; DV, dense volume.
[a]Absolute breast asymmetry measures are absolute CC asymmetry volumes log transformed.
[b]Adjusted associations: BV asymmetry adjusted by BV category, FG asymmetry by FGV category. Volumetric categories are quintiles of respective volumes.
[c]Mutually adjusted also adjusted for either Age or Ethnicity as appropriate.
[d]Age in 5 year age bands
[e]Asian = British Indian, Pakistani, Bangladeshi
[f]Mixed = mixed White and Black, White and Asian or any other mixed

7.2 (3.2, 14.1) cm^3, respectively, if derived from the MLO views]. Similar associations of these measures with age and ethnicity were also found (data not shown).

Implications

So far, only a few small, studies have examined the relation of breast asymmetry measures with breast cancer. Scutt et al used area-based mammographic breast size (BV) asymmetry measurements from ~250 breast cancer cases and ~250 matched controls, while adjusting for known risk factors and absolute breast size, to show that absolute BV asymmetry at baseline screen was associated, with cancer diagnosis at the baseline screen[21] and also medium-term risk.[22] In a preliminary study, Eltonsy et al examined data from 280 breast cancer cases and 82 controls and found that the mean absolute BV asymmetry, adjusting for BV, was significantly higher in cancer patients.[19] Kayar et al used non-mammographic breast measurements (from Grossman-Rounder Discs) on 251 breast cancer cases and 466 controls from a Turkish outpatient clinic, to propose a 'pathological breast asymmetry ratio', suggesting that a L:R BV ratio of >±20% was associated with an increased risk of breast cancer being diagnosed within one year of the examination.[20]

Zheng et al investigated the relationship between mammographic density percentage (%MD) asymmetry and breast cancer using a bespoke algorithm on mammograms from 230 females with interval cancers (cancers diagnosed between screens) and 230 controls and suggested that as %MD increases there was an increased risk of cancer at both current screen and in the medium term (1–3 years). These models adjusted for subjective breast density category (BIRADS), but not for absolute breast density.[23,24]

The limited available literature suggests that BV and DV asymmetry may have potential value as markers of either the presence of a cancer (diagnostic marker) or the risk of developing cancer in the future (risk predictor). Proper examination of the potential value of these breast asymmetry measures as diagnostic or predictor markers will require the conduct of large-scale and longitudinal studies with objective measurements of breast asymmetry. Objective breast tissue asymmetry estimates can now be obtained using existing fully-automated mammographic volumetric analysis tools and thus can be provided, without additional investigations, for all females attending screening. The availability of such data will facilitate further research into the association between asymmetry and breast cancer, both at the current screen and subsequently, and may potentially provide a practical additional tool for stratifying the screening population in terms of likelihood of having, or risk of developing, breast cancer.

AUTHORS' CONTRIBUTIONS

SMH, IdSS and RD, designed the study; LSW organized the collection of participants' data and provided clinical guidance on the design; SMH performed the statistical analysis with guidance from BDS; SMH wrote the first draft of the manuscript. All authors (SMH, LSW, RD, BDS, IdSS) contributed to the interpretation of the results and critically reviewed the draft of the manuscript; they all read and approved the final version of the manuscript, and they all agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

REFERENCES

1. Breast cancer and hormone-replacement therapy in the Million women study. *The Lancet* 2003; **362**: 419–27. doi: https://doi.org/10.1016/S0140-6736(03)14065-2
2. .Collaborative Group on Hormonal Factors in Breast Cancer Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. *Lancet* 1996; **347**: 1713–27. doi: https://doi.org/10.1016/S0140-6736(96)90806-5
3. Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50 302 women with breast cancer and 96 973 women without the disease. *The Lancet* 2002; **360**: 187–95. doi: https://doi.org/10.1016/S0140-6736(02)09454-0
4. Key T, Appleby P, Barnes I, Reeves G, .Endogenous Hormones and Breast Cancer Collaborative Group Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J Natl Cancer Inst* 2002; **94**: 606–16. doi: https://doi.org/10.1093/jnci/94.8.606
5. Key TJ, Appleby PN, Reeves GK, Travis RC, Alberg AJ, Barricarte A, et al. Sex hormones and risk of breast cancer in premenopausal women: a collaborative reanalysis of individual participant data from seven prospective studies. *Lancet Oncol* 2013; **14**: 1009–19. doi: https://doi.org/10.1016/S1470-2045(13)70301-2
6. Key TJ, Appleby PN, Reeves GK, Roddam AW, Helzlsouer KJ, Alberg AJ, et al. Circulating sex hormones and breast cancer risk factors in postmenopausal women: reanalysis of 13 studies. *Br J Cancer* 2011; **105**: 709–22.
7. Hoover RN, Hyer M, Pfeiffer RM, Adam E, Bond B, Cheville AL, et al. Adverse health outcomes in women exposed in utero to diethylstilbestrol. *N Engl J Med* 2011; **365**: 1304–14. doi: https://doi.org/10.1056/NEJMoa1013961
8. IdS S, De Stavola B, McCormack V. Collaborative group on pre-natal risk F, subsequent risk of breast C: birth size and breast cancer risk: Re-analysis of individual participant data from 32 studies. *PLoS medicine* 2008; **5**: e193.
9. Manning JT, Scutt D, Whitehouse GH, Leinster SJ. Breast asymmetry and phenotypic quality in women. *Evolution and Human Behavior* 1997; **18**: 223–36. doi: https://doi.org/10.1016/S0162-3095(97)00002-0
10. Parsons PA. Fluctuating asymmetry: an epigenetic measure of stress. *Biol Rev Camb Philos Soc* 1990; **65**: 131–45. doi: https://doi.org/10.1111/j.1469-185X.1990.tb01186.x
11. Milne BJ, Belsky J, Poulton R, Thomson WM, Caspi A, Kieser J. Fluctuating asymmetry and physical health among young adults. *Evolution and Human Behavior* 2003; **24**: 53–63. doi: https://doi.org/10.1016/S1090-5138(02)00120-4
12. Jasienska G, Lipson SF, Ellison PT, Thune I, Ziomkiewicz A. Symmetrical women have higher potential fertility. *Evolution and Human*

Behavior 2006; **27**: 390–400. doi: https://doi.org/10.1016/j.evolhumbehav.2006.01.001
13. Thornhill R, Møller AP. Developmental stability, disease and medicine. *Biol Rev Camb Philos Soc* 1997; **72**: 497–548. doi: https://doi.org/10.1017/S0006323197005082
14. Møller AP, Soler M. Thornhill R: breast asymmetry, sexual selection, and human reproductive success. *Evolution and Human Behavior* 1995; **16**: 207–19.
15. Natekar P, DeSouza F. Fluctuating asymmetry in dermatoglyphics of carcinoma of breast. *Indian J Hum Genet* 2006; **12**: 76–81. doi: https://doi.org/10.4103/0971-6866.27790
16. Muller DC, Baglietto L, Manning JT, McLean C, Hopper JL, English DR, et al. Second to fourth digit ratio (2D:4D), breast cancer risk factors, and breast cancer risk: a prospective cohort study. *Br J Cancer* 2012; **107**: 1631–6. doi: https://doi.org/10.1038/bjc.2012.418
17. Manning JT, Leinster SJ. Re: the ratio of 2nd to 4th digit length and age at presentation of breast cancer: a link with prenatal oestrogen? *The Breast* 2001; **10**: 355–7. doi: https://doi.org/10.1054/brst.2001.0284
18. Bunevicius A. The Association of Digit Ratio (2D:4D) with Cancer: A Systematic Review and Meta-Analysis. *Dis Markers* 2018; **2018**: 7698193. doi: https://doi.org/10.1155/2018/7698193
19. Ramadhani MK, Elias SG, van Noord PAH, Grobbee DE, Peeters PHM, Uiterwaal CSPM. Innate left handedness and risk of breast cancer: case-cohort study: table 1. *BMJ* 2005; **331**: 882–3. doi: https://doi.org/10.1136/bmj.38572.440359.AE
20. Fritschi L, Divitini M, Talbot-Smith A, Knuiman M. Left-Handedness and risk of breast cancer. *Br J Cancer* 2007; **97**: 686–7. doi: https://doi.org/10.1038/sj.bjc.6603920
21. Scutt D, Manning JT, Whitehouse GH, Leinster SJ, Massey CP. The relationship between breast asymmetry, breast size and the occurrence of breast cancer. *Br J Radiol* 1997; **70**(OCT): 1017–21. doi: https://doi.org/10.1259/bjr.70.838.9404205
22. Eltonsy HN, Elmaghraby A, Tourassi G. Bilateral breast volume asymmetry in screening mammograms as a potential marker of breast cancer. *Preliminary Experience* 2007;.
23. Kayar R. Cilengiroglu OV: breast volume asymmetry value, ratio, and cancer risk. *Breast Cancer: Basic and Clinical Research* 2015; **2015**: 87–92.
24. Williams AC, Hitt A, Voisin S, Tourassi G. Automated assessment of bilateral breast volume asymmetry as a breast cancer biomarker during mammographic screening. In: *SPIE Medical Imaging: 2013*: International Society for Optics and Photonics; 2013. pp. 86701A–86701.
25. Scutt D, Lancaster GA, Manning JT. Breast asymmetry and predisposition to breast cancer. *Breast Cancer Res* 2006; **8**: R14. doi: https://doi.org/10.1186/bcr1388
26. Zheng B, Sumkin JH, Zuley ML, Wang X, Klym AH, Gur D. Bilateral mammographic density asymmetry and breast cancer risk: a preliminary assessment. *Eur J Radiol* 2012; **81**: 3222–8. doi: https://doi.org/10.1016/j.ejrad.2012.04.018
27. Zheng B, Tan M, Ramalingam P, Gur D. Association between computed tissue density asymmetry in bilateral mammograms and near-term breast cancer risk. *Breast J* 2014; **20**: 249–57. doi: https://doi.org/10.1111/tbj.12255
28. Tan M, Zheng B, Ramalingam P, Gur D. Prediction of near-term breast cancer risk based on bilateral mammographic feature asymmetry. *Acad Radiol* 2013; **20**: 1542–50. doi: https://doi.org/10.1016/j.acra.2013.08.020
29. Senie RT, Saftlas AF, Brinton LA, Hoover RN. Is breast size a predictor of breast cancer risk or the laterality of the tumor? *Cancer Causes Control* 1993; **4**: 203–8.
30. Perkins CI, Hotes J, Kohler BA, Howe HL. Association between breast cancer laterality and tumor location, United States, 1994-1998. *Cancer Causes Control* 2004; **15**: 637–45. doi: https://doi.org/10.1023/B:CACO.0000036171.44162.5f
31. Cheng S-A, Liang L-Z, Liang Q-L, Huang Z-Y, Peng X-X, Hong X-C, et al. Breast cancer laterality and molecular subtype likely share a common risk factor. *Cancer Manag Res* 2018; **10**: 6549–54. doi: https://doi.org/10.2147/CMAR.S182254
32. Health and Social Care Centre. *Breast Screening programme England 2016-2017*. UK: NHS Digital; 2018.
33. National Collaborating Centre for Cancer (UK) Classification and care of people at risk of familial breast cancer and management of breast cancer and related risks in people with a family history of breast cancer. NICE clinical guidelines, no. 164. *In. Cardiff UK* 2013;.
34. Census Guidance and Methodology. 2011. Available from: https://www.ons.gov.uk/census/2011census/2011censusdata/2011censususerguide/variablesandclassifications.
35. The Royal College of Radiologists. *Guidance on screening and symptomatic breast imaging*. London: BFCR; 2013.
36. VolparaDensity™ user manual version 1.5.11. 2014;.
37. StataCorp: **Stata Statistical Software**. *Release 14. College Station*. TX: StataCorp LP; 2015.
38. Senie RT, Rosen PP, Lesser ML, Snyder RE, Schottenfeld D, Duthie K. Epidemiology of breast carcinoma II: factors related to the predominance of left-sided disease. *Cancer* 1980; **46**: 1705–13. doi: https://doi.org/10.1002/1097-0142(19801001)46:7<1705::AID-CNCR2820460734>3.0.CO;2-Q
39. Losken A, Fishman I, Denson DD, Moyer HR, Carlson GW. An objective evaluation of breast symmetry and shape differences using 3-dimensional images. *Ann Plast Surg* 2005; **55**: 571–5. doi: https://doi.org/10.1097/01.sap.0000185459.49434.5f
40. Lee HN, Sohn Y-M, Han KH. Comparison of mammographic density estimation by Volpara software with radiologists' visual assessment: analysis of clinical-radiologic factors affecting discrepancy between them. *Acta Radiol* 2015; **56**: 1061–8. doi: https://doi.org/10.1177/0284185114554674
41. Chen J-H, Chan S, Yeh D-C, Fwu PT, Lin M, Su M-Y. Response of bilateral breasts to the endogenous hormonal fluctuation in a menstrual cycle evaluated using 3D MRI. *Magn Reson Imaging* 2013; **31**: 538–44. doi: https://doi.org/10.1016/j.mri.2012.10.022
42. Jack RH, Møller H, Robson T, Davies EA. Breast cancer screening uptake among women from different ethnic groups in London: a population-based cohort study. *BMJ Open* 2014; **4**: e005586. doi: https://doi.org/10.1136/bmjopen-2014-005586
43. Manning J, Scutt D, Whitehouse GH, Leinster S. Asymmetry and the menstrual cycle. 1996; **17**vol. .
44. Scutt D, Manning JT. Symmetry and ovulation in women. *Hum Reprod* 1996; **11**: 2477–80. doi: https://doi.org/10.1093/oxfordjournals.humrep.a019142
45. Brand JS, Czene K, Shepherd JA, Leifland K, Heddson B, Sundbom A, et al. Automated measurement of volumetric mammographic density: a tool for widespread breast cancer risk assessment. *Cancer Epidemiol Biomarkers Prev* 2014; **23**: 1764–72. doi: https://doi.org/10.1158/1055-9965.EPI-13-1219
46. Alonzo-Proulx O, Mawdsley GE, Patrie JT, Yaffe MJ, Harvey JA. Reliability of automated breast density measurements. *Radiology* 2015; **275**: 366–76. doi: https://doi.org/10.1148/radiol.15141686
47. Holland K, van Zelst J, den Heeten GJ, Imhof-Tas M, Mann RM, van Gils CH, et al. Consistency of breast density categories in serial screening mammograms: a comparison between automated and human assessment. *Breast* 2016; **29**: 49–54. doi: https://doi.org/10.1016/j.breast.2016.06.020

6.4 Supplementary descriptive analyses

Descriptive analysis based on level of socioeconomic deprivation was not included in paper II above, due to word limits and the supplementary tables are included below to satisfy objectives 1 and 2.

The study population and methods for deriving all measures of interest are as described in Paper II. In addition the socioeconomic characteristics of the area of residence of each woman in the study was derived by linking a woman's, routinely collected, postcode to a set of Indices of Deprivation (IMD) 2015, a deprivation index at the lower layer super output area (LSOA), created by the British Department for Communities and Local Government (225). The IMD quintile was derived, ranging from 1 "most deprived" to 5 "least deprived".

To examine whether breast volumetric (BV and DV) and asymmetry (absolute and relative BV and DV asymmetry) characteristics varied across different socioeconomic groups (as characterised by IMD quintile), box plots were showing medians and IQR for each of the IMD quintiles.

To examine whether there was a tendency for the left breast to be larger than the right breast in the different breast volumetric measures (BV, DV) a ratio was derived from Left breast CC view / Right breast CC view for all the different measurements of interest. A forest plot was used to show how this varied across different ethnic groups.

Only 5.2% of particpants were in the most deprived quintile of the national distribution with and equal distribution over the other categories, Table 6.1.

Table 6.1 Supplementary data for characteristics of study participants

	All Women (n=54,591)	
	Frequency	**Percent**
Socioeconomic (IMD) quintile		
1 – Most deprived	2,795	5.2%
2	11,689	21.4%
3	13,311	24.4%
4	13,052	23.9%
5 – Least Deprived	11,965	21.9%
Missing	1,779	3.3%

Median BV was greatest in the most deprived quintile and declined with decreasing levels of deprivation but for all other breast volumetric estimates including asymmetry there was no clear pattern across the deprivation gradient Fig 6.1.

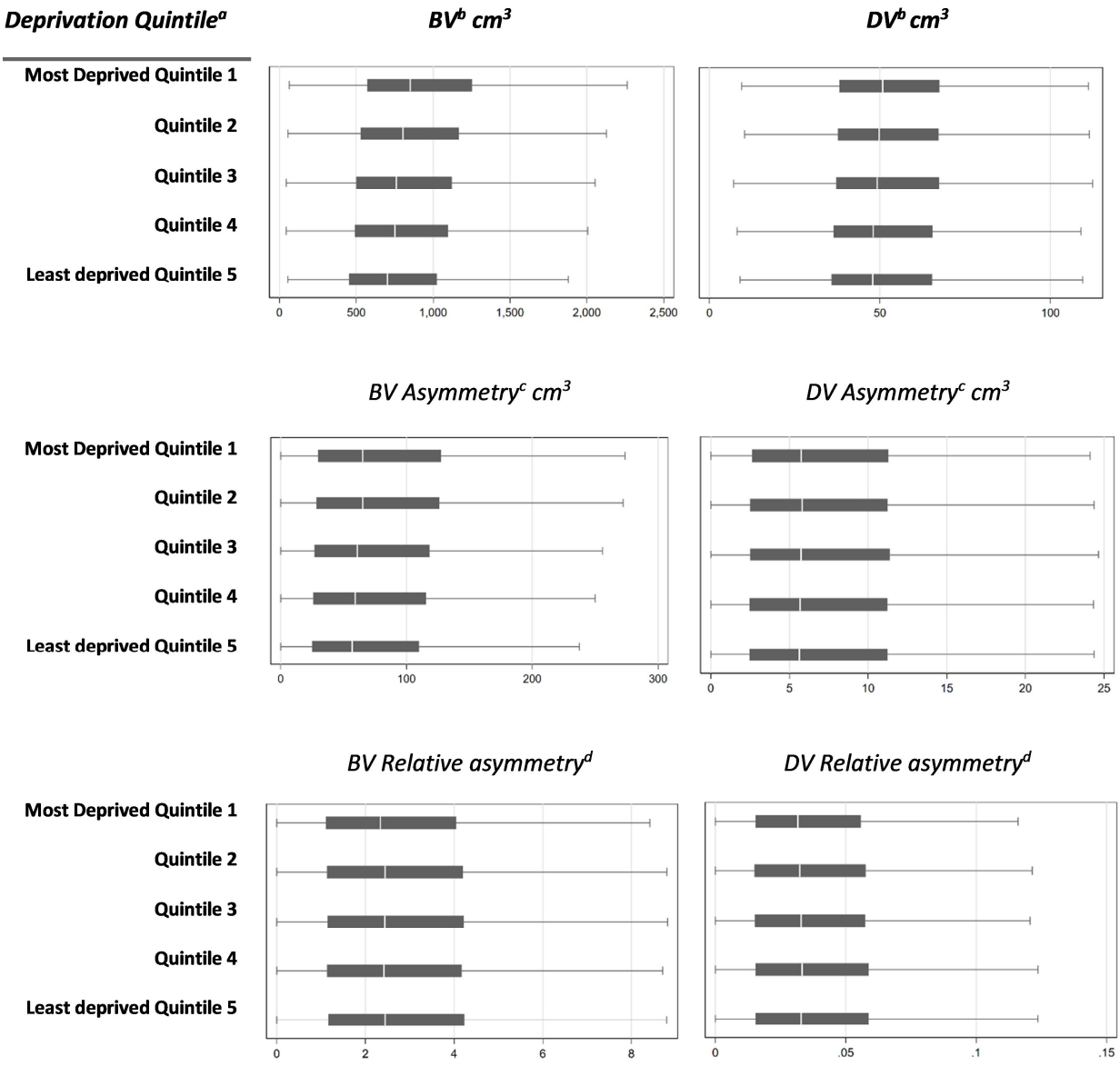


Figure 6.1 Breast tissue volumes and asymmetry measurements by deprivation, medians and IQR

Footnotes:

[a] Quintile of Indices of Deprivation (IMD) 2015, a deprivation index at the lower layer super output area (LSOA), created by the British Department for Communities and Local Government (225).

[b] BV and DV are average values from the 4 images: left CC image, right CC image, left MLO image, right MLO image.

[c] Asymmetry was derived from absolute difference Left and Right CC views.

[d] Relative asymmetry estimated as (|R-L|/(R+L)/2*100), where L and R are volume estimates derived from the left and right side images. Whiskers are calculated as lower adjacent values (i.e., smallest observed value >= lower quartile +1.5 IQR and largest observed value <= upper quartile+1.5 IQR).

The forest plot Fig 6.2 shows that the average (geometric mean) left breast versus right breast volume ratio was 1.04 (95% CI: 1.04, 1.04) and the average DV ratio was 1.03 (95% CI: 1.02, 1.03).

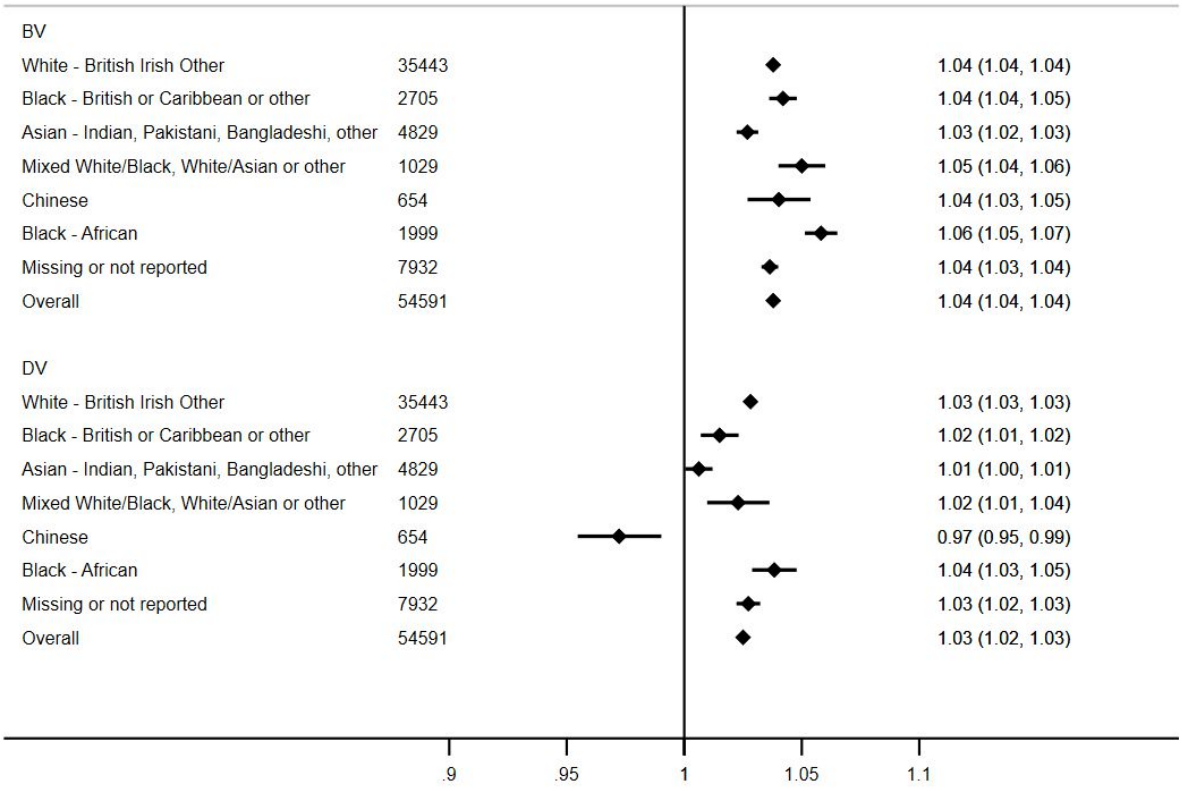


Figure 6.2 Left to Right breast tissue volumes ratios[a] by ethnic group geometric means and 95% CI

Footnotes:

[a] Ratio derived from Left breast CC view / Right breast CC view

[b] Ethnicity groupings as described in Paper II.

The forest plot (Figure 6.2) shows that the observed volumetric ratio patterns are consistent across all ethnic groups with the exception that for women of Chinese ethnicity the DV was on average higher in the right breast with a LR ratio of 0.97 (95% CI: 0.95, 0.99).

Supplementary findings

The supplementary descriptive analyses confirm that whilst BV is somewhat higher in women in the most deprived socioeconomic groups the patterns of BV and DV asymmetry do not differ across different IMD strata. Relative asymmetry is consistent across all IMD levels suggesting that breast asymmetry is independent of socioeconomic conditions.

The supplementary analyses on L:R breast ratio are consistent with previous research and referred to more generally in Paper II. More recently Li et al carried out a large comparative population-based study to compare BD between Chinese and Australian women and, whilst not specifically looking at asymmetry, their results show similar findings to ours based on area-based measurements of breast density(235), with women of Chinese ethnicity having larger average %BD and also a tendency of the right breast to be denser than the left.

6.5 Paper III Left-right breast asymmetry and risk of screen-detected and interval cancers in a large population-based screening population (236)

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	12406683	**Title**	Ms
First Name(s)	Susan M		
Surname/Family Name	Hudson		
Thesis Title	Beyond breast density – Novel uses of automated mammographic analysis in breast cancer screening		
Primary Supervisor	Professor Isabel dos Santos Silva		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	The British Journal of Radiology		
When was the work published?	2020		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	n/a		
Have you retained the copyright for the work?*	**No**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I was responsible for the conceptualisation of the work, development of the methodology, setting up of the mammogram analysis tools in conjunction with Dr Wilkinson and data collection, cleansing and linking of data sources, data analysis and writing of the publication.

SECTION E

Student Signature	[redacted]
Date	

Supervisor Signature	
Date	

BJR

https://doi.org/10.1259/bjr.20200154

Received: 17 February 2020 | Revised: 03 June 2020 | Accepted: 08 June 2020

Cite this article as:
Hudson SM, Wilkinson LS, De Stavola BL, dos-Santos-Silva I. Left–right breast asymmetry and risk of screen-detected and interval cancers in a large population-based screening population. *Br J Radiol* 2020; **93**: 20200154.

FULL PAPER

Left–right breast asymmetry and risk of screen-detected and interval cancers in a large population-based screening population

[1]SUE M HUDSON, BSc MSc, [2]LOUISE S WILKINSON, [3]BIANCA L DE STAVOLA and [1]ISABEL DOS-SANTOS-SILVA

[1]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK
[2]Oxford Breast Imaging Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
[3]Faculty of Pop Health Sciences, Institute of Child Health, University College London, London, UK

Address correspondence to: Ms Sue M Hudson
E-mail: *susan.hudson@lshtm.ac.uk;sue.hudson@pasconsulting.co.uk*

Objectives: To assess the associations between automated volumetric estimates of mammographic asymmetry and breast cancers detected at the same ("contemporaneous") screen, at subsequent screens, or in between (interval cancers).
Methods: Automated measurements from mammographic images (*N* = 79,731) were used to estimate absolute asymmetry in breast volume (BV) and dense volume (DV) in a large ethnically diverse population of attendees of a UK breast screening programme. Logistic regression models were fitted to assess asymmetry associations with the odds of a breast cancer detected at contemporaneous screen (767 cases), adjusted for relevant confounders.
Nested case–control investigations were designed to examine associations between asymmetry and the odds of: (a) interval cancer (numbers of cases/age-matched controls: 153/646) and (b) subsequent screen-detected cancer (345/1438), via conditional logistic regression.
Results: DV, but not BV, asymmetry was positively associated with the odds of contemporaneous breast cancer (P-for-linear-trend (Pt) = 0.018). This association was stronger for first (prevalent) screens (Pt = 0.012). Both DV and BV asymmetry were positively associated with the odds of an interval cancer diagnosis (Pt = 0.060 and 0.030, respectively). Neither BV nor DV asymmetry were associated with the odds of having a subsequent screen-detected cancer.
Conclusions: Increased DV asymmetry was associated with the risk of a breast cancer diagnosis at a contemporaneous screen or as an interval cancer. BV asymmetry was positively associated with the risk of an interval cancer diagnosis.
Advances in knowledge: The findings suggest that DV and BV asymmetry may provide additional signals for detecting contemporaneous cancers and assessing the likelihood of interval cancers in population-based screening programmes.

INTRODUCTION

Breast cancer is the most common female cancer.[1] Mammographic screening programmes, such as the England and Wales Breast Screening Programme (NHSBSP), have been found to reduce mortality through earlier detection.[2] However, although increased amount of radio-dense tissue on a mammogram is associated with an increased risk of developing breast cancer,[3–5] there is strong evidence that it also reduces the effectiveness of breast screening[6–10] by decreasing mammographic sensitivity, as radio-dense tissue may hide cancers. Boyd et al found that females in the highest category of mammographic density (density in >= 75% of mammogram) had greater odds of being diagnosed with cancer in the year following a "normal" mammogram than females in the lowest density category (density in <10% of mammogram) (OR of 17.8 (95% CI 4.8–65.9)).[6] Other research has focused on the texture or type of parenchymal pattern in breast tissue as a risk factor for breast cancer; a review of over 40 research papers concluded that automated analysis of quantitative features in mammographic images may be useful in breast cancer risk assessment and potential stratification for screening but that further research was necessary.[11]

One little explored potential feature is mammographic asymmetry in the total size of the breast and in the size of the radio-dense tissue between the left and the right breasts. Increased "'fluctuating asymmetry" (FA), that is, increased anthropometrical asymmetry in paired features, is related to both fecundity and general health.[12–15] Furthermore,

breast FA appears to be related to many of the known reproductive breast cancer risk factors, such as parity, age at first birth and age at menopause.[16] Findings to date are consistent with breast volume (BV) asymmetry being associated with the presence of breast cancer[17–20] as well as with a higher risk of having a breast cancer diagnosed in the short- and medium-term.[21] There is also limited evidence that asymmetry in mammographic density might be associated with higher short-term,[22–24] and medium-term[25] risk of being diagnosed with breast cancer although the previous research used very specific bespoke algorithms to derive asymmetry scores based on comparing multiple bilateral mammographic density features.

Breast cancer subtypes, based on gene expression or receptor status, are clinically relevant because they are associated with differential treatment options and prognoses. Studies have shown that some hormonal risk factors associated with FA are also associated with particular breast cancer subtypes (*e.g.,* parity is inversely associated with FA[16] and with the risk of luminal-like breast tumours[26,27]).

The aim of this study is to investigate the association between left–right asymmetry in breast size and in the amount of radio-dense tissue, as ascertained by mammography, and the risk of being diagnosed with breast cancer (overall and by subtype) at the same or subsequent screens, or as an interval cancer, among a large population-based sample of 68,776 females who underwent mammographic screening in South West London, England, between March 2013 and June 2017.

METHODS

Study participants

The study participants were female residents in one of six London boroughs—Wandsworth, Merton, Croydon, Sutton, Richmond and Kingston—who underwent routine 3 yearly screening mammography as part of the NHSBSP at the South West London Breast Screening Service (SWLBSS) based in the St George's University Hospitals NHS Foundation Trust. The NHSBSP is an organised population-based mammographic screening programme, with a call–recall system, which targets females aged 50–70 years (with a trial for 50% of females aged 47–50 and 70–73) and has a coverage of approximately 75%.[28] Also included were small numbers of younger females who had been identified as having a higher risk of breast cancer and therefore invited for screening on an annual basis, plus any females over 73 years who had optionally contacted the service for a self-referred screening appointment. Participants were screened during the period 01 March 2013 to 20 June 2017.

Data on ethnicity were collected as part of the standard screening protocol via a self-completed screening questionnaire. Ethnicity was categorised according to the Census classification[29] and summarised as, "Asian" (Indian, Pakistani or Bangladeshi or other), "Black-African," "Black-British or Caribbean or other," "Chinese," "Mixed" (White and Black, White and Asian or any other mixed), "White" (British or Irish or other) and "Other." Data for other known breast cancer risk factors (*e.g.,* reproductive history, body mass index (BMI), family-history of breast cancer) are not collected in a systematic way across the NHSBSP screening programme and thus were unavailable. The type of screen (first (prevalent) versus subsequent (incident) screens) was recorded.

Exposure assessment

Each female underwent the NHSBSP standard, two-view (cranio-caudal (CC) and medio-lateral-oblique views (MLO)) mammography of each breast.[30] Raw digital mammographic images were processed via an automated algorithm, that is, Volpara® Density™ V.1.5.11, (Matakina Technology Limited, Wellington, New Zealand)[31] ; this algorithm provided fully automated estimates (in cm^3) of the volume of the BV and the volume of the radio-dense tissue (DV) separately for each of the four (left–right CC and MLO) images. The volume of non-dense volume (NDV) was calculated as the difference between BV and DV on the same image. The NHSBSP does not use mammographic density as a diagnostic aid, and participants are not informed on whether they have dense breasts.

For each participant, we estimated absolute measures of left–right asymmetry (in cm^3), that is, the unsigned difference between left BV (or DV) and right BV (or DV). Absolute asymmetry was estimated from the CC images because this view is likely to capture the whole of the breast while being less affected than the MLO view by the inclusion of variable amounts of retro-glandular fat tissue near the chest wall.

Subject eligibility

Screening events where exposure measurements (*i.e.,* breast asymmetry) and outcome ascertainment (screen-detected cancer) were done concurrently, were regarded as "contemporaneous screens" for the purposes of this study. In all, 93,416 contemporaneous screens took place during the study period. Screens were excluded from this analysis if: females had a previous history of breast cancer ($N = 2,068$); females had breast implants or where the standard set of four (*i.e.,* left–right CC and MLO) images was incomplete or exceeded ($N = 10,234$); and of these if one or both of the CC images was rejected by Volpara based on its internal consistency checks ($N = 1,383$). Thus, a total of 79,731 screens were eligible for inclusion in the analysis. Some females were screened more than once in the study period; 9,600 females had two screens; 221 females had three screens; 72 had four screens; and 3 females had five screens; all valid screens were included in the analysis. "Subsequent" screens were screens that took place, as a result of the next screening round invitation following on from a contemporaneous screen, at approximately 3 years after the contemporaneous screen. Approximately 20% of subsequent screens were also included in the contemporaneous screen study.

Cancer ascertainment

The images were double read with arbitration by consensus. In this study, cancers detected at the screen when breast asymmetry was estimated were called "contemporaneous screen detected cancers," cancers diagnosed symptomatically in the 3-year period following this measurement and prior to the next screening invitation were regarded as "interval cancers" and

breast cancers detected at the subsequent screen were considered as being "subsequent cancers.

Interval cancer case ascertainment was based on the sharing of data between the Screening Quality Assurance Service and Cancer Registries and via direct contact between the screening services and local treating NHS Trusts. Each NHSBSP screening service is responsible for recording and reviewing all reported interval cancers. We included all recorded interval cancers from the SWLBSS database as of 06 November 2019.

Contemporaneous screen cancers were categorised according to histological subtype and laterality (left-side, right-side, bilateral tumour). Tumour subtypes are routinely differentiated in the NHSBSP by immunohistochemical (IHC) analysis of the oestrogen (ER) and progesterone (PR) hormone receptors and the human epidermal growth factor (HER2) (using IHC plus *in situ* hybridisation (ISH) molecular analysis). These tests are carried out on diagnostic or surgical biopsies. In the NHSBSP, ER testing is required for all invasive tumours and guidelines are used to ensure standard reporting of results across the screening programme.[32] The results were used to approximately differentiate between the most clinically relevant subtypes based on the definitions proposed by Waks and Winer[33] as: Hormone+ (H+) cancers if ER +and/or PR+, HER2-; HER2 +cancers if ER ± PR+/-, HER2+; and triple negative cancers if ER-, PR- and HER2-. The size of tumours was estimated as maximum dimension of the whole tumour at surgical excision where such data were available. No data on receptor status or tumour size were available for interval cancers.

Study design

A cross-sectional screen-specific design was used to examine associations between left–right breast asymmetry and contemporaneous screen-detected cancers (Figure 1). Screens at which females were diagnosed with a first occurrence of breast cancer (n = 767) were defined as cases, and screens where no cancer was detected (n = 78,964) as non-cases. In all, 82 females had both a non-cancer contemporaneous screen and a later contemporaneous first screen-detected cancer; in the analysis, their non-cancer screens were included as being non-cancer while their screen-detected cancers were included as cases.

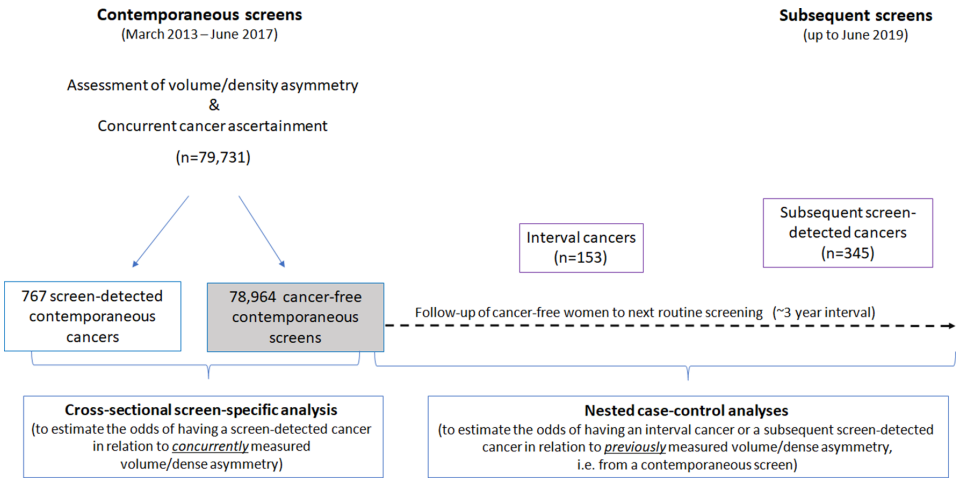
An incident-density-sampling (nested) case–control design was used to investigate the association between breast asymmetry and interval cancers (Figure 1). Cases were females who were diagnosed with an interval cancer after a normal contemporaneous screen. For each case, up to five controls were randomly selected among females who had a contemporaneous screen in the same year and month as the case and who had a verified "non-cancer" status (based on subsequent screening records) at the time that the case was diagnosed, matched to the case on age at contemporaneous screen (± 1 year). For cases aged >73 years at contemporaneous screen, controls were aged-matched within ± 5 years due to paucity of controls. A total of 153 interval cancer cases and 646 matched controls were identified corresponding to 87 cases with five controls each, 37 cases with four controls each, 14 cases with three controls each, seven cases with two controls each and seven cases with one control each; one case was excluded in the analysis because there were no valid matched controls.

A similar nested case–control approach was also used to assess the association between mammographic asymmetry and risk of being diagnosed with a breast cancer in a subsequent screen (Figure 1). This design was preferred to a cross-sectional analysis because subsequent screens had not yet been performed for around one-third of the study participants. Cases were females who had a normal contemporaneous screen but were diagnosed with breast cancer in the subsequent screening round (n = 345). Up to five age-matched controls per case were identified (a total of 1,438) using a similar approach to that outlined above for interval cancers, corresponding to 202 cases with five controls each, 58 cases with four controls each, 44 cases with three controls each, 25 cases with two controls each and 14 cases with one control each; two cases were excluded in the analysis because there were no eligible controls.

Statistical analyses

Tertiles of the distributions were used to categorise BV asymmetry and DV asymmetry into three equally sized categories (low,

Figure 1. Timing of mammography and cancer diagnosis



medium and high) based on the distributions in the non-cases/controls.

Logistic regression models were used to examine the strength of the associations between the exposures of interest, BV asymmetry and DV asymmetry, and the odds of being diagnosed with a contemporaneous screen-detected breast cancer (overall and by subtype). Robust standard errors (clustering by female) were used to account for the fact that some females had repeat screens over the 52-month study period. Similarly, separate conditional logistic regression models were used to examine the strength of the associations between BV asymmetry and DV asymmetry and the odds of an interval cancer and the odds of a subsequent screen-detected cancer.

All regression models were adjusted for a priori potential confounders: age at screening, ethnicity and mean mammographic NDV (a valid proxy for BMI when data for the latter are not available[34]) and additionally for mean BV (log transformed) in the BV asymmetry model and mean DV (log transformed) in the DV asymmetry model. DV was not added as a potential confounder in the BV asymmetry model because previous studies using this data showed that there was no association between DV and BV asymmetry.[35] Mean BV, NDV and DV values were calculated as averages of the corresponding fully automated readings obtained from each female's four contemporaneous CC and MLO images (all available image sets being used to derive the best estimate for these confounders). Trend tests for the association with the asymmetry measures were carried out fitting models with the ordinal values of each asymmetry measure and assessing their significance using Wald tests.

For the association between breast asymmetry and the odds of having a contemporaneous screen-detected cancer, further analysis included stratification by type of screen (prevalent *vs* incident) and reanalyses restricted to each tumour subtype. Adjustment for ethnicity was omitted for the latter due to sparsity of data.

Spearman rank correlation coefficients (r) were estimated to investigate whether the magnitude of the breast asymmetry in BV and DV among contemporaneous screen-detected breast cancer cases was correlated with the size of the tumour. The proportion of cancers detected in the larger breast was also calculated.

In all the analyses, we considered statistical significance (two-sided) at p-value < 0.05. All analyses were conducted in Stata (IC 14)[33].

RESULTS

Study participants

The characteristics of the participants, and of their screens, are shown in Table 1. The majority of the participants were White. The mean age at contemporaneous screening was 58.4 years when the screen did not lead to cancer detection and 60.4 years when it did. Mean time between contemporaneous screen and interval cancer diagnosis was 19.2 (range 0.14–36.0; SD = 9.1) months. Mean time between contemporaneous screen and subsequent screen diagnosis was 36.4 (range 9.6–70.8; SD = 8.2) months.

The median values for BV and DV asymmetry were higher for contemporaneous cases (65.4 cm^3 and 6.64 cm^3, respectively) than non-cases (60.3 cm^3 and 5.78 cm^3, respectively; Table 1). Median values for BV asymmetry and DV asymmetry were also higher for interval cancer cases (71.9 cm^3 and 8.90 cm^3, respectively) than their matched controls (57.5 cm^3 and 5.60 cm^3, respectively; Table 1). A similar pattern was observed for subsequent cancers but with smaller case–control differences in median BV asymmetry and DV asymmetry (Table 1) (Supplementary Table 1).

Tumour subtype was known for 88% of all contemporaneous screen-detected cases. Of these 84% were HR+, 11% HER2+ and 4.8% triple-negative tumours (Table 1). The median BV and DV asymmetry values for the latter were markedly higher (110.4 cm^3 and 11.66 cm^3, respectively) than for the other subtypes (average 65.4 cm^3 and 6.64 cm^3, respectively) (Supplementary Table 2).

Associations between BV and DV asymmetry and contemporaneous screen-detected breast cancer

There was a possible positive, but weak (p-for-linear-trend (Pt) = 0.105), log-linear association between BV asymmetry and the odds of being diagnosed with cancer at the contemporaneous screen. Relative to females in the bottom third of the BV asymmetry distribution (<36.4 cm^3), those in the top third (≥93.7 cm^3) appeared to have 1.17 times greater odds (OR 1.17; 95% CI 0.97, 1.44) of having a screen-detected cancer, in the fully adjusted models. There was stronger evidence that DV asymmetry was positively associated with the odds of being diagnosed with a cancer at the contemporaneous screen; (Pt = 0.018) with females in the top third of the DV asymmetry (≥9.04 cm^3) having 1.26 times greater odds (OR 1.26; 95% CI 1.04, 1.53) than those in the bottom third (<3.48 cm^3) (Figure 2).

In stratified analyses by type of screen, BV asymmetry was not associated with the odds of a contemporaneous screen-detected cancer in either group. DV asymmetry was however positively associated with the odds of a contemporaneous screen-detected breast cancer among females who had a prevalent screen (OR 1.56; 95% CI 1.07, 2.27; Pt = 0.012) but not among those who had an incident screen (OR 1.15 (0.92, 1.45); Pt = 0.21; Figure 3).

No clear associations were found between BV asymmetry and any specific tumour subtype. DV asymmetry however was positively associated with both the odds of having a contemporaneous screen-detected HR +breast cancer and the odds of having a triple negative breast cancer, but no association was found with HER2 +cancers. Relative to females in the bottom third of the DV asymmetry distribution those in the top third were 3.7 times more likely to have a triple negative cancer (OR 3.72; 95% CI 1.11, 12.45) and 1.3 times more likely to have a HR +cancer (1.28; 1.05, 1.58; Figure 4).

Table 1. Characteristics of the study participants, their mammographic screens and their breast cancer

	Contemporaneous screen-detected analysis		Subsequent interval cancer analysis		Subsequent screen-detected cancer analysis	
	Contemporaneous screen-detected cancer cases	Non-cancer at contemporaneous screen	Interval cancer cases	Controls	Subsequent screen-detected cancer cases	Controls
	n (%)	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**
No. females [a]	*n* = 767	*n* = 68,776 [a]	*n* = 153	*n* = 646	*n* = 345	*n* = 1,438
Age at contemporaneous screening	**Mean (SD)**	**Mean (SD)** [b]	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
	60.39 (7.66)	58.42 (7.17)	59.52 (7.29)	60.10 (7.38)	59.93 (6.21)	59.74 (6.23)
Ethnicity	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**
White [c]	526 (68.6%)	44,203 (64.3%)	114 (74.5%)	419 (64.9%)	252 (73.0%)	968 (67.3%)
Asian [d]	73 (9.5%)	6,422 (9.3%)	5 (3.3%)	78 (12.1%)	32 (9.3%)	132 (9.2%)
Black – Caribbean [e]	25 (3.3%)	3,265 (4.7%)	10 (6.5%)	25 (3.9%)	10 (2.9%)	78 (5.4%)
Black – African	22 (2.9%)	2,425 (3.5%)	3 (2.0%)	22 (3.4%)	1 (0.3%)	53 (3.7%)
Mixed	11 (1.4%)	1,311 (1.9%)	3 (2.0%)	16 (2.5%)	6 (1.7%)	24 (1.7%)
Chinese	8 (1.0%)	859 (1.2%)	1 (0.7%)	3 (0.5%)	7 (2.0%)	25 (1.7%)
Missing/not reported	102 (12.3%)	10,291 (15.0%)	17 (11.1%)	83 (12.9%)	37 (10.7%)	158 (11.0%)
No. Screens[f]	*n* = 767	*n* = 78,964	*n* = 153	*n* = 646	*n* = 345	*n* = 1,438
Type of screen						
Prevalent	205	21,242				
Incident	562	57,722				
Breast Volumetric measurements [g] Median BV, cm^3 Median DV, cm^3 Median NDV cm^3	**Median (IQR)** 778.8 (514.3–1149.9) 50.9 (39.5–69.6) 726.1 (463.4–1082.5)	**Median (IQR)** 756.7 (493.4–1112.9) 49.3 (37.2–67.2) 702.7 (443.1–1049.1)	**Median (IQR)** 920.2 (502.9–1225.9) 56.9 (43.4–86.2) 847.2 (413.6–1133.6)	**Median (IQR)** 780.1 (536.5–1133.1) 48.3 (36.1–65.5) 730.0 (484.1–1064.6)	**Median (IQR)** 794.8 (529.9–1156.4) 50.8 (39.1–70.5) 732.6 (485.3–1089.2)	**Median (IQR)** 748.3 (493.6–1116.5) 47.7 (35.9–65.1) 700.7 (438.5–1049.1)
Absolute Volumetric Asymmetry, cm^3[h] BV Bilateral Asymmetry DV Bilateral Asymmetry	**Median (IQR)** 65.4 (28.1–127.4) 6.64 (2.90–13.29)	**Median (IQR)** 60.3 (26.4–117.6) 5.78 (2.51–11.43)	**Median (IQR)** 71.9 (32.7–130.8) 8.9 (3.9–15.3)	**Median (IQR)** 57.5 (25.5–117.2) 5.6 (2.1–10.9)	**Median (IQR)** 69.8 (27.1–133.3) 6.24 (2.95–11.51)	**Median (IQR)** 61.0 (28.7–122.2) 5.40 (2.38–11.20)
Asymmetry laterality Right BV > Left BV Left BV > Right BV No difference[j] BV Right DV > Left DV Left DV > Right DV No difference[j] DV	**n (%)** 322 (41.98%) 445 (58.02%) 0 (0%) 348 (45.37%) 418 (54.50%) 1 (0.01%)	**n (%)** 31,169 (39.47%) 47,790 (60.52%) 5 (0.01%) 35,693 (45.13%) 43,288 (54.82%) 37 (0.05%)	**n/a**	**n/a**	**n/a**	**n/a**
No. Cancers	*n* = 767		*n* = 153		*n* = 345	

(Continued)

Table 1. (Continued)

Cancer Laterality Bilateral Left breast Right breast Unknown	**n (%)** 20 (2.6%) 380 (49.5%) 364 (47.5%) 3		**n (%)** 3 (2.0%) 42 (27.5%) 40 (26.1%) 68 (44.4%)		**n (%)** 0 165 (47.8%) 180 (52.2%) 0	
Tumour subtype HR+ (ER + and/or PR+, HER2-) HER2+ (ER+/-,PR+/-, HER2+) Triple negative (ER-, PR-, HER2-) Unknown or non-invasive	**n (%)** 566 (84.2%) 74 (11.0%) 32 (4.8%) 95 (12.4%)		**n/a**		**n/a**	
Tumour size mm[i] Left breast (n = 245) Right breast (n = 249)	**Median (IQR)** 19 (17 -21) 20 (18-22)		**n/a**		**n/a**	

BV, breast volume; DV, volume of radio-dense tissue on a mammogram; ER, oestrogen receptor; HER2, human epidermal growth factor; HR, hormone receptor; IQR, inter quartile range; NDV, volume of non-radio-dense (fatty) tissue on a mammogram; PR, progesterone receptor; SD, standard deviation; j, No measureable difference in volumes; n/a, not available.

[a]9,814 females have more than one contemporaneous screen.

[b]Where females have more than one contemporaneous screen the average of their age at screen is taken.

[c]White includes: British/Irish and other

[d]Asian includes: British Indian, Pakistani, Bangladeshi and other

[e]Black includes: British, Caribbean and other (non-African)

[f]Screens included must have at exactly four images taken, only screening images are included, excluded are images that were rejected as outliers by the Volpara algorithm. Screens for females known to have previous breast cancer were also excluded.

[g]Median values are calculated using all available images from the relevant view (MLO and CC from each side) at a contemporaneous screen

[h]Median value of maximum dimension in mm for unilateral cancers where size has been reported

[i]Calculated as absolute difference between left CC image and right CC image

[j]No measurable difference in volumes

Associations between BV and DV asymmetry and interval cancer

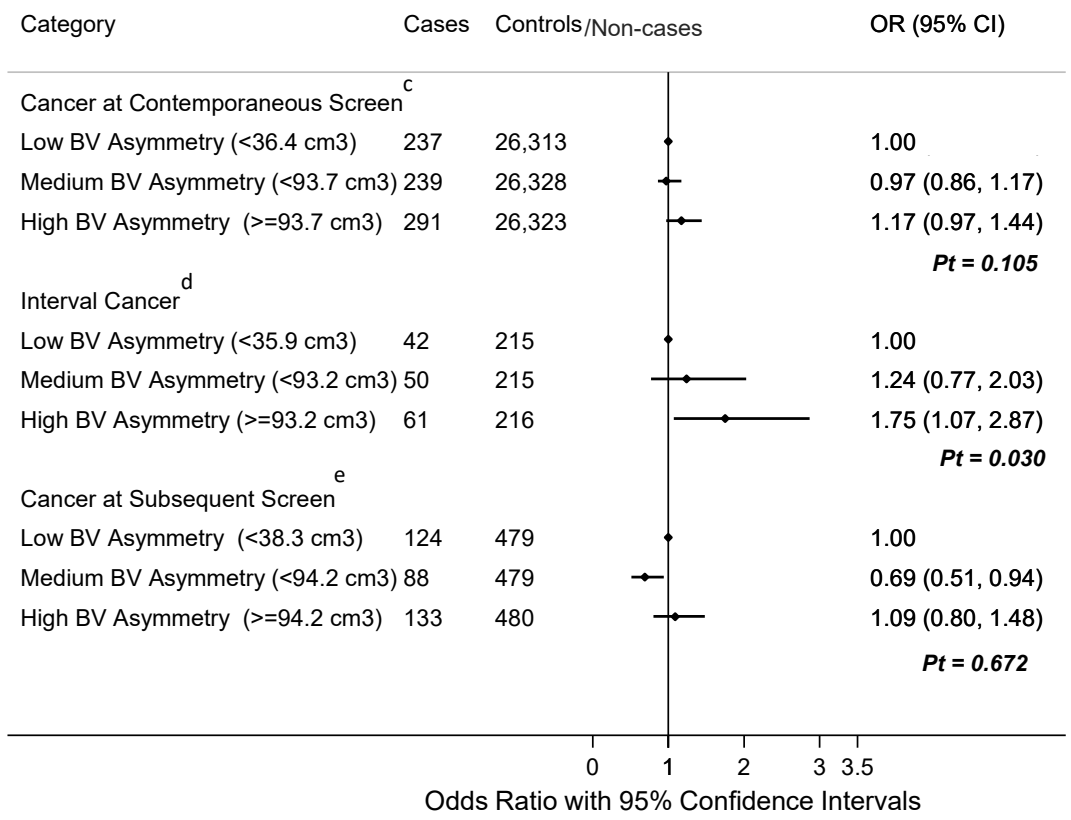
BV asymmetry was positively associated with the odds of having an interval cancer; relative to females in the bottom third of the BV asymmetry distribution those in the top third had significantly higher odds of being diagnosed with an interval cancer (adjusted OR 1.75; 95% CI 1.07, 2.87). Similarly, there was a positive, but weak (*P*-trend = 0.060), log-linear association between DV asymmetry and the odds of being diagnosed with a subsequent interval cancer (OR for females in the top third of the DV asymmetry distribution versus those in the bottom third: 1.68; 95% CI 0.97, 2.92).(Figure 2)

Associations between BV and DV asymmetry and a subsequent screen-detected cancer

There were no clear associations between BV or DV asymmetry and the odds of having a cancer detected at the next screening round.(Figure 2)

Cancer laterality

The cancer was detected in the breast with larger BV in approximately 52% of all cases and in the breast with larger DV in approximately 54% of cases. These proportions were similar irrespective of whether the cancer was detected at the contemporaneous screen (*i.e.,* from the same images that were used to measure BV/DV asymmetry) or whether it was an interval cancer or a cancer detected at a subsequent screen (Table 2).



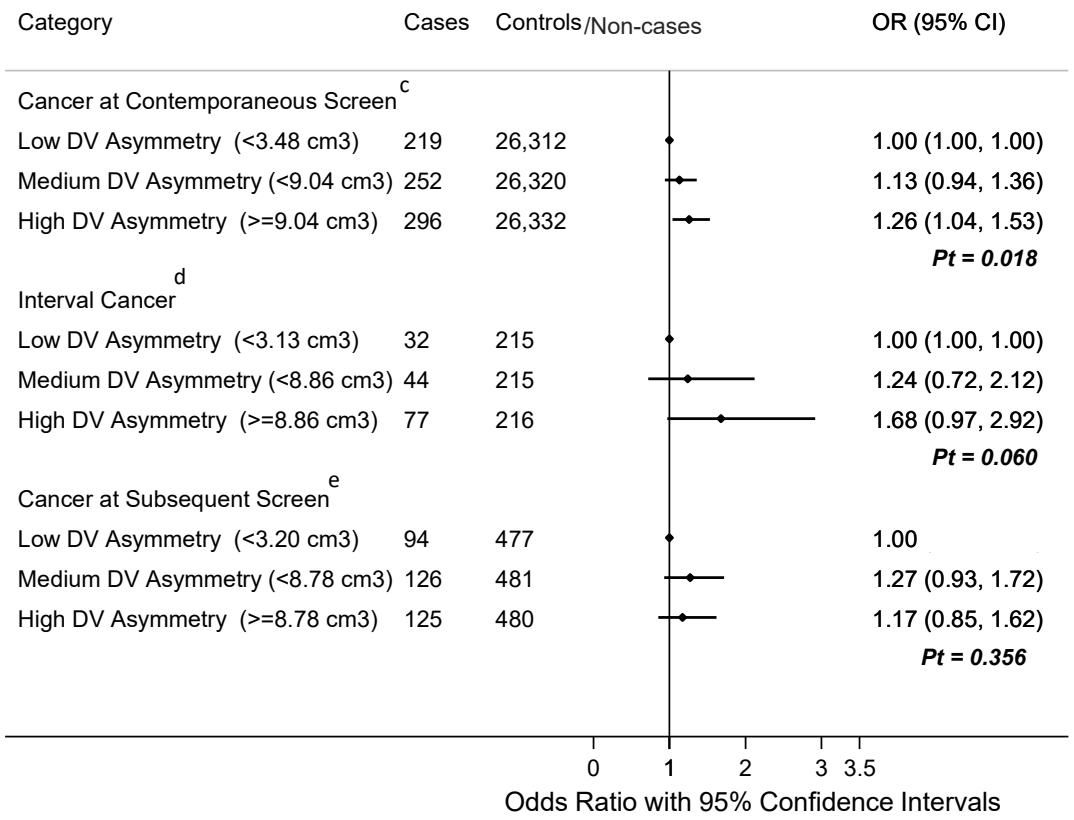
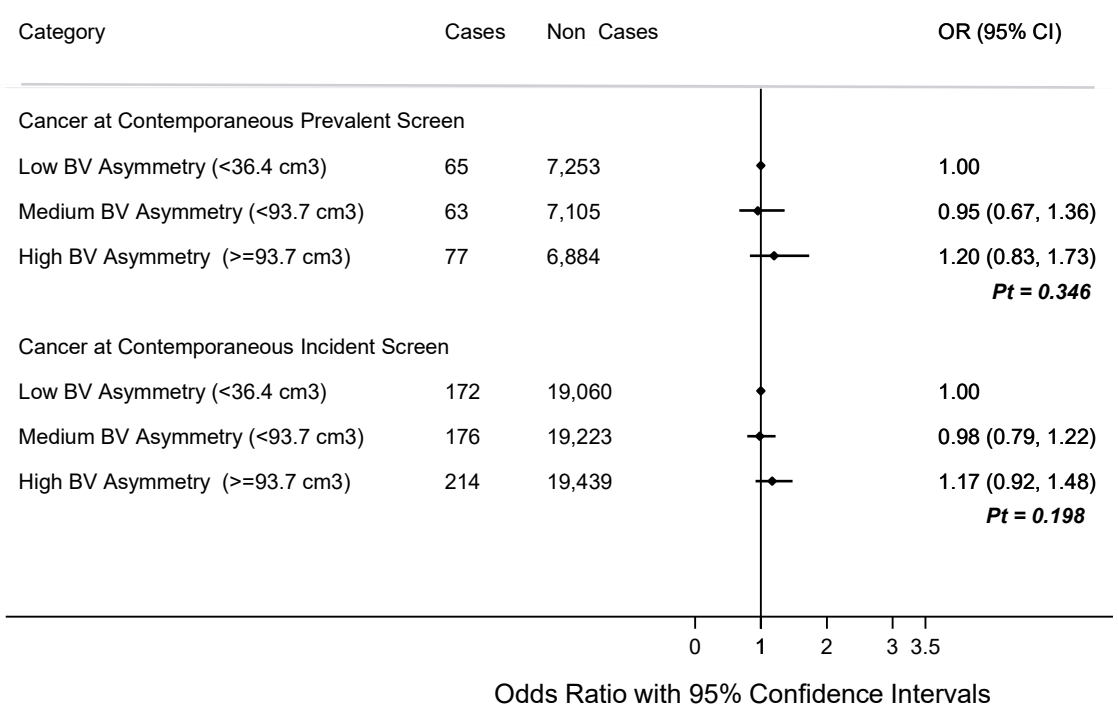


Figure 2. Associations[a] between breast volume (BV) asymmetry[b] and mammographic density volume (DV) asymmetry with the odds of having a breast cancer detected at the contemporaneous screen (a) Adjusted for age, ethnicity, NDV (volume of non- dense mammographic tissue, as a proxy for BMI), and log BV for BV asymmetry or log DV for DV asymmetry (b) Automated BV and DV asymmetry measures from the CC (cranio- caudal view) images categorised according thirds of the distribution in non- cases (c) Contemporaneous screen corresponds to the same screen whose images were used to estimate BV and DV asymmetry (d) Interval Cancers were cancers diagnosed within 3 years of the contemporaneous screen, but before a subsequent screened (e) Cancers at subsequent screen were those diagnosed at next the routine screening round after the contemporaneous screen



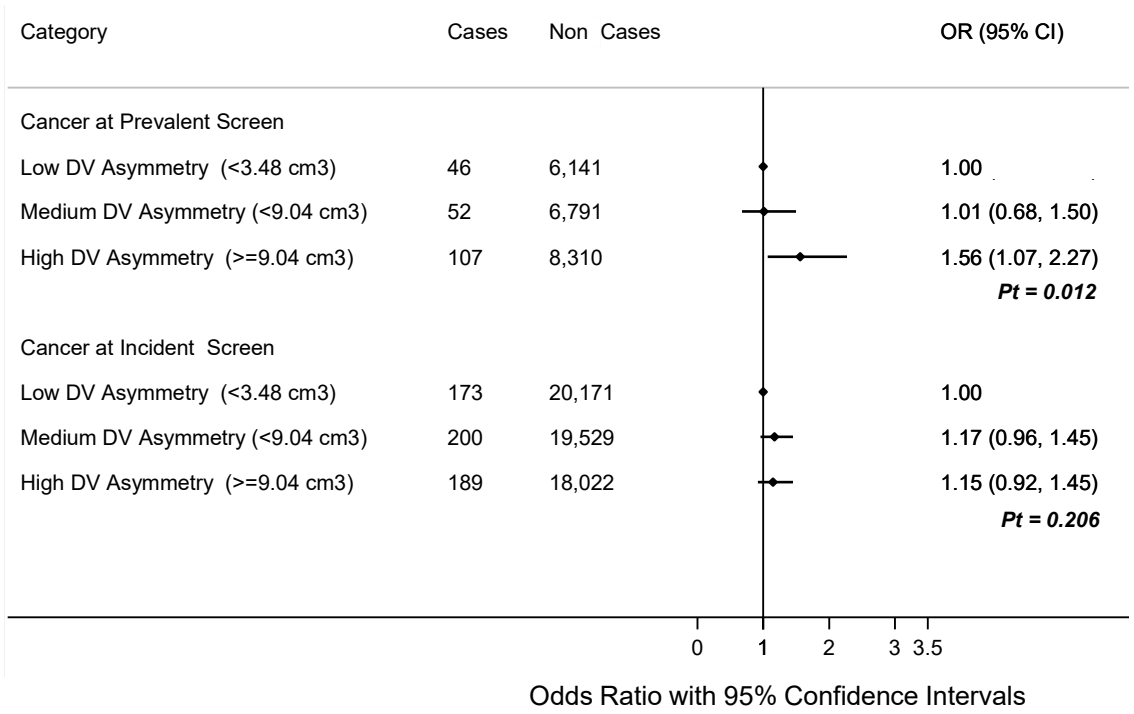
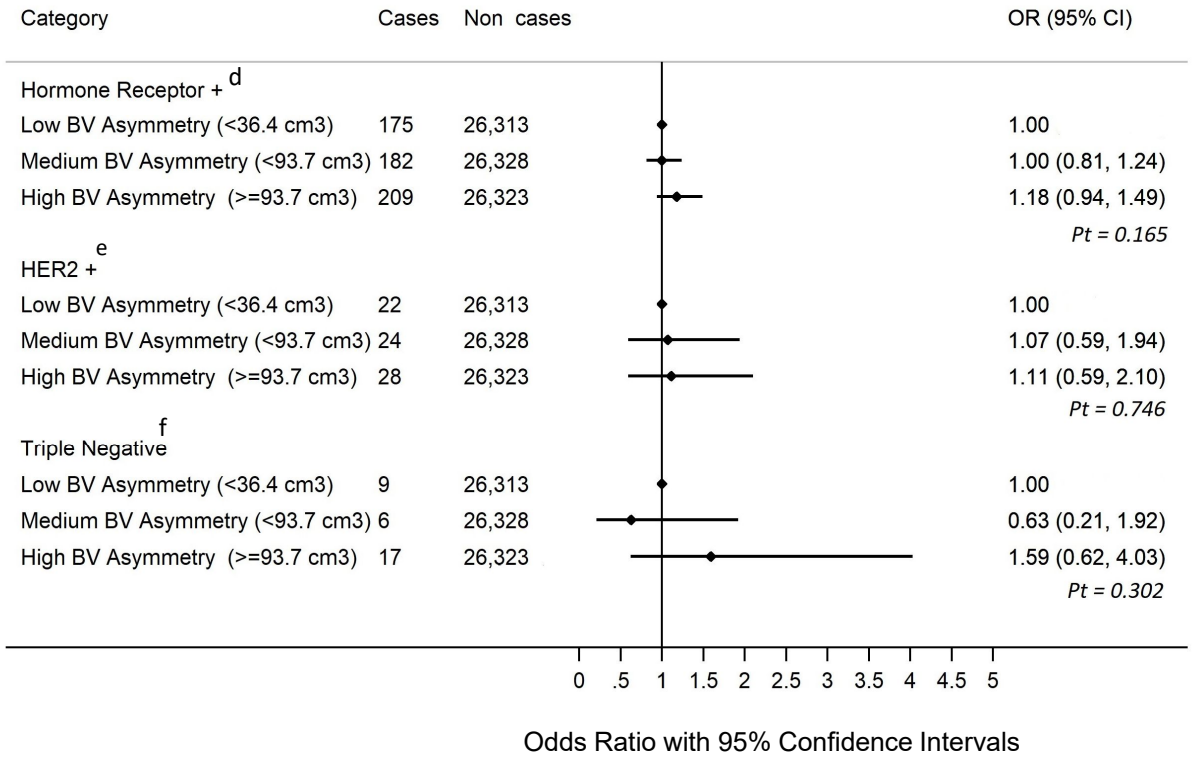


Figure 3. Associations[a] between breast volume (BV) asymmetry[b], and mammographic density volume (DV) asymmetry[b], and the odds of having a breast cancer detected at the contemporaneous screen[c], by type of screen[d]. (a) Adjusted for age, ethnicity, NDV (volume of non-dense mammographic tissue, as a proxy for BMI) and log BV for BV asymmetry or log DV for DV asymmetry. (b) Automated BV and DV asymmetry measures from the CC (cranio-caudal view) images categorised according thirds of the distribution in non- cases. (c) Contemporaneous screen corresponds to the same screen whose images were used to estimate BV and DV asymmetry. (d) Prevalent screen if the contemporaneous screen was the first screen a female had ever had; incident screen if the female had at least one screen prior to the contemporaneous screen.



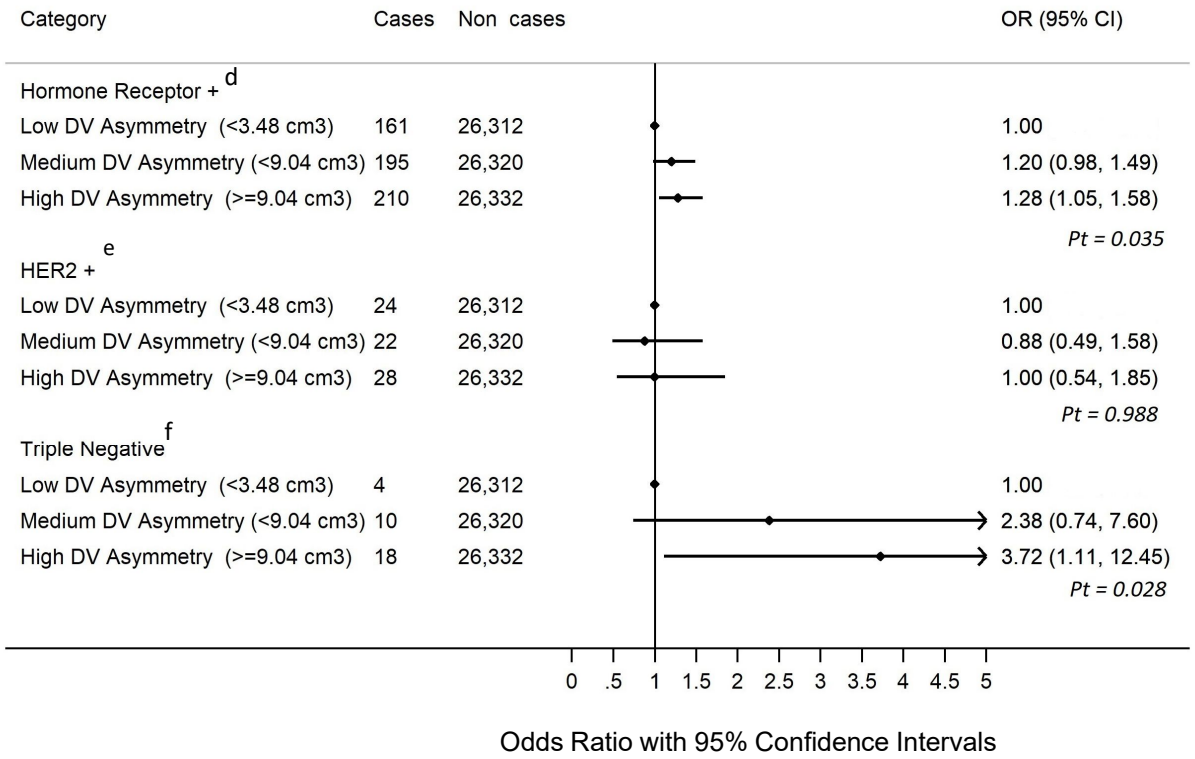


Figure 4. Associations[a] between breast volume (BV) asymmetry[b], and mammographic density volume (DV) asymmetry[b], and the odds of having a breast cancer detected at the contemporaneous screen, by tumour subtype[C]. (a) Adjusted for age, ethnicity, NDV (volume of non- dense mammographic tissue, as a proxy for BMI), and log BV for BV asymmetry or log DV for DV asymmetry (b) Automated BV and DV asymmetry measures from the CC (cranio-caudal view) images categorised according to thirds of the distribution in non-cases (c) 95 cases with unknown sub- type excluded from this figured (d) Hormone receptor positive (HR+) includes oestrogen receptor positive (ER+) and/or progesterone receptor positive (PR+) (e) HER2+ includes tumours with over- expression or amplification of human epidermal growth factor 2 (ER+/-, PR+/-) (f) Triple negative tumours correspond to those which were HER2-, PR- and ER-.

Correlations between contemporaneous screen tumour size and absolute asymmetry

Tumour size, as measured at surgical excision (available for 494 (64%) tumours), was not correlated with the degree of mammographic BV asymmetry (r = 0.01 (p = 0.82) and r = 0.03 (p = 0.60) for tumours located, respectively, in the left and right breasts). Similarly, there was only a very weak correlation between tumour size and DV asymmetry (r = 0.12 (p = 0.06)) for cancers located in the left breast and (r = −0.12 (p = 0.06)) for cancers located in the right breast).

The median percentage of BV occupied by the largest tumour was 0.57% (IQR; 0.14–2.15%). For DV, the median percentage of tumour size to overall breast DV was 8.04% (IQR; 1.9–29.48%).

Figure 5 shows that the distribution of tumour size versus signed difference in volume between the left and right breasts and tumour size was broadly similar irrespective of laterality of the tumour.

DISCUSSION

Main findings

The present study found positive associations between automatically estimated mammographic DV asymmetry and the odds of having a breast cancer diagnosed at a contemporaneous screen and as an interval cancer. Increasing BV asymmetry was also strongly associated with increasing odds of having an interval cancer but only weakly associated with higher odds of a contemporaneous screen cancer. Neither BV nor DV asymmetry were associated with the odds of a subsequent screen-detected cancer in our study.

Our findings are similar to previous smaller studies by Scutt et al who used visually assessed mammographic breast size (BV) asymmetry estimates (~250 cases;~250 age-matched controls) to show that absolute BV asymmetry was positively associated with contemporaneously detected cancer.[17] Our larger study using automated measurements also found a positive association between BV asymmetry and the odds of a breast cancer diagnosis at the contemporaneous screen (although with borderline significance). Unlike our study, Scutt et al also found an association between BV asymmetry and medium-term risk of breast cancer diagnosis (mean time to diagnosis 6.44 years) after adjustment for known-risk factors and absolute breast size.[21] As in our study, Scutt et al found no correlation between tumour size and BV asymmetry and they noted that approximately 50% of the tumours were found in the smaller breast by BV. Eltonsy et al used a computerised algorithm to estimate BV asymmetry from screening mammographic images (280 screen-detected cancer cases; 82 controls). They found that mean absolute BV

Table 2. Proportion of tumours occurring in the larger breast[a] [b]

Point of diagnosis	Cancer in larger breast by BV cm^3	Cancer in larger breast by DV cm^3
Contemporaneous screen cancers	377/744 (50.7%)	413/744 (55.5%)
Interval cancers	45/83 (54.2%)	46/83 (55.4%)
Subsequent screen cancers	181/345 (52.5%)	180/345 (52.2%)
All cancers	**603/1172 (51.5%)**	**639/1172 (54.5%)**

[a]Excludes bilateral cases
[b]Calculated as the signed difference in cm^3 between the BV (or DV) value from the left CC image and the BV (or DV) value from the right CC image.

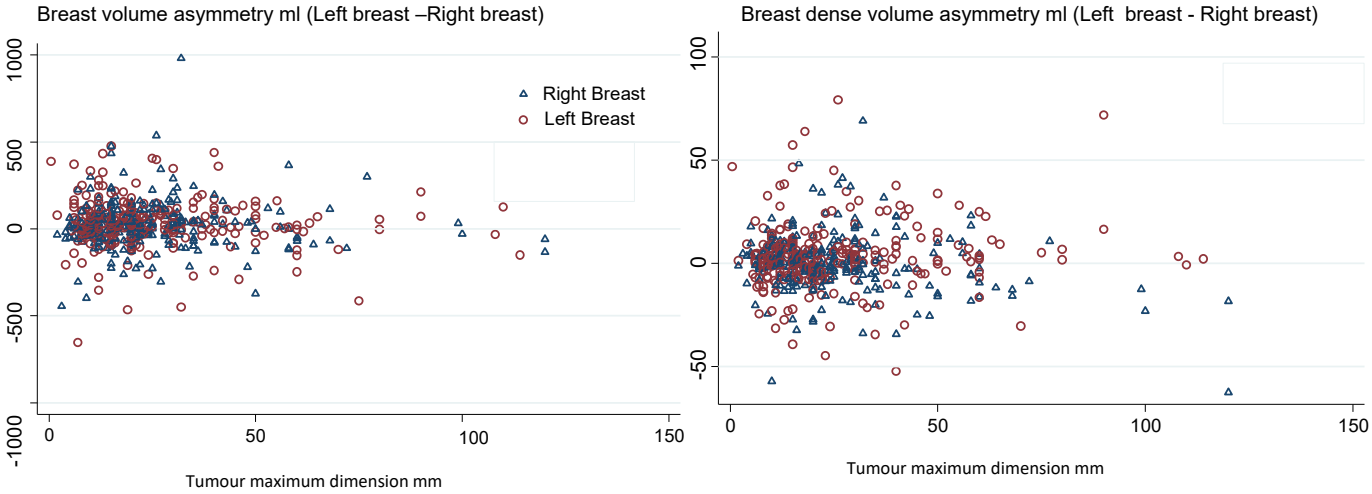


Figure 5. Tumour size and Left breast – Right breast asymmetry[a] by cancer laterality[b] (a) Calculated as the signed difference in cm3 between the BV (or DV) value from the left CC image and the BV (or DV) value from the right CC image. (b) Excludes bilateral cases

asymmetry, adjusting for BV, was significantly higher in cancer patients.[18] Only limited research has looked at the association between BV asymmetry and interval cancers. Kayar et al used physical breast measurements (251 cases; 466 controls) from a Turkish outpatient (non-screening) clinic, to identify a "pathological breast asymmetry ratio."[20] They found that approximately 50% of the tumours were located in the smaller breast but that left breast:right breast BV ratio of >±20% was associated with an increased risk of breast cancer being diagnosed within 1 year of the examination.[20] Similarly, we found a significant positive association between BV asymmetry at the original screening and interval cancer risk (mean time to diagnosis 1.6 years). Our findings are in line with Cheong et al who studied 87 breast cancer patients referred for breast reconstruction and found that only approximately 0.2% of the BV was occupied by the actual tumour. They found no association between tumour size and BV asymmetry.[36]

Our study also broadly agrees with the findings of a previous small case–control study, which used a bespoke algorithm for estimating mammographic density percentage (%MD) asymmetry in 230 cases found clear of cancer at the time the image was taken but who were subsequently diagnosed with breast cancer and 230 matched cancer-free controls.[22,25] Increasing %MD asymmetry was positively associated with the odds of cancer at the subsequent screen (1–3 years later) after adjusting for age and subjective breast density category (BIRADS),[22,25] in line with our findings.

DV asymmetry was more strongly associated with contemporaneous cancer detection in prevalent than in incident screens. Asymmetries might be more likely to be identified and investigated in the first (prevalent) screen when prior screening images are not normally available for comparison.[37]

To our knowledge, this is the first study to look at the association between breast asymmetry (DV or BV) and cancer subtypes. A systematic review by Antoni et al.[3838] showed that the density-breast cancer association did not differ by cancer subtype. Our analysis, albeit based on small numbers, suggests that the DV asymmetry association with breast cancer may be particularly strong for triple-negative cancers.

The findings that neither BV nor DV asymmetry were associated with the odds of a subsequent screen- detected cancer is possibly a result of the relatively small number of subsequent breast cancer cases; both BV and DV asymmetry showed positive associations with the odds of subsequent screen cancer but not at the 95% CI level. Larger studies will be required to investigate this fully.

The pathways through which asymmetry in BV and DV may affect the risk of being diagnosed with breast cancer (in the short or longer term) are poorly understood. If asymmetry is simply attributable to the presence of a tumour in the breast, then a higher correlation between tumour size and asymmetry would be expected together with a closer correspondence between tumour laterality and the breast with larger volume/density (in our study only ~55% of unilateral screen detected tumours were located in the breast with higher DV/BV) and previous studies found no evidence that the tumour was associated with the larger BV.[20,21,36] In our study, there was some evidence of a weak positive correlation between DV asymmetry and tumour size, but overall little of the observed asymmetry in our study can simply be explained by the presence of a tumour in the larger breast. We therefore conclude that asymmetry cannot be explained by the presence of a tumour alone but may be a biomarker of increased genetic/early life susceptibility to breast cancer.

Radiologists are able to identify abnormal signals from mammograms extremely quickly by extracting the "gist" of the image in fractions of a second[39] but they may also find it more difficult to read bilateral mammograms that display greater asymmetry between the breasts,[40,41] due to the "obfuscation" effect of increased asymmetry. The "masking effect" of DV has been recognised for some time[6,7] and this study suggests that the masking effect is enhanced where DV is asymmetrical. This association may however be subtle since Evans et al[42] found that, although asymmetry may be part of what signals an abnormal mammogram, there is still above-chance performance from clinicians when presented with artificial asymmetric conditions (e.g., where the contralateral breast was from a different female).

Strengths and limitations

Strengths of this study include its population-based design, large sample size, ethnic mix, and availability of information on receptor status. The images for both breasts were collected at the same point in time, and under similar technical conditions therefore within-woman left:right breast comparisons are unlikely to have been biased by anthropometric, reproductive and lifestyle characteristics or the equipment used. The study used an automated method to estimate BVs, therefore measurements were free from subject or observer biases.

The algorithm (Volpara Density) used gives reliable volumetric BV and DV estimates.[43–45] There is no published data specifically on the reliability of *asymmetry* measures derived from the Volpara volumetric measurements but examination of data from a subset of 464 females in our study, who had two sequential screens, using Bland Altman plots showed no systematic bias although the limits of agreement were large (unpublished).

A limitation of this study was the lack of data on potential reproductive confounders (*e.g.,* parity, age at menarche, menopausal status) which have been shown to be associated with breast asymmetry,[16] (these data are not routinely collected at screening in the UK). Our adjustment and matching for age (at least partly) dealt with the potential confounding effect of menopausal status; however, too little is known of the direction and strength of the cumulative effects of other reproductive variables to speculate on the direction of the potential residual confounding bias affecting the reported estimates. The number of interval cancers recorded was relatively small, partly reflecting the lag time between diagnosis and notification to the screening services. Information on tumour subtypes was limited and there was insufficient power to analyse asymmetry associations with the rarer tumour subtypes.

IMPLICATIONS

This study suggests that increasing left:right asymmetry in BV and DV may be of relevance when interpreting mammographic screening images as a signal of the likely presence of a cancer on a contemporaneous screen and the likelihood of being diagnosed with an interval cancer before the next screen. Further studies are needed to confirm these findings and, if confirmed, to assess how they may affect the performance of the screening programme. Nevertheless, the availability of automated algorithms, which allow volumetric assessment of BV and density in real-time from two-dimensional mammographic images, means that such studies can now be conducted on a large-scale as objective measurements of bilateral asymmetry can be easily obtained for all females screened.

AUTHORS' INFORMATION

LSW was previously Director of the South West London Breast Screening Service, SWLBSS, UK

AUTHORS' CONTRIBUTIONS

SMH and IdSS designed the study; LSW organised the collection of participants' data and provided clinical guidance on the design; SMH performed the statistical analysis with guidance from BDS; SMH wrote the first draft of the manuscript. All authors (SMH, LSW, BDS, IdSS) contributed to the interpretation of the results and critically reviewed the draft of the manuscript; they all read and approved the final version of the manuscript, and they all agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

ETHICAL APPROVAL

This retrospective study was carried out on fully anonymous, routinely collected data only, held in accordance with the National Health Service (NHS) Cancer Screening Programmes Confidentiality and Disclosure Policy 2011. The NHSBSP has section 251 support under the NHS Act 2006. The study was approved by all relevant ethics committees (Research Ethics Committees from St George's University Hospitals NHS Foundation Trust, and the London School of Hygiene and Tropical Medicine).

REFERENCES

1. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer* 2019; **144**: 1941–53. doi: https://doi.org/10.1002/ijc.31937
2. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012; **380**: 1778–86. doi: https://doi.org/10.1016/S0140-6736(12)61611-0
3. McCormack VA, dos Santos Silva I, Silva dosS I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2006; **15**: 1159–69. doi: https://doi.org/10.1158/1055-9965.EPI-06-0034
4. Boyd NF, Martin LJ, Bronskill M, Yaffe MJ, Duric N, Minkin S. Breast tissue composition and susceptibility to breast cancer. *J Natl Cancer Inst* 2010; **102**: 1224–37. doi: https://doi.org/10.1093/jnci/djq239
5. Huo CW, Chew GL, Britt KL, Ingman WV, Henderson MA, Hopper JL, et al. Mammographic density-a review on the current understanding of its association with breast cancer. *Breast Cancer Res Treat* 2014; **144**: 479–502. doi: https://doi.org/10.1007/s10549-014-2901-2
6. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 2007; **356**: 227–36. doi: https://doi.org/10.1056/NEJMoa062790
7. Pisano ED, Hendrick RE, Yaffe MJ, Baum JK, Acharyya S, Cormack JB, et al. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. *Radiology* 2008; **246**: 376–83. doi: https://doi.org/10.1148/radiol.2461070200

8. Destounis S, Johnston L, Highnam R, Arieno A, Morgan R, Chan A. Using volumetric breast density to quantify the potential masking risk of mammographic density. *AJR Am J Roentgenol* 2017; **208**: 222–7. doi: https://doi.org/10.2214/AJR.16.16489
9. Wanders JOP, Holland K, Karssemeijer N, Peeters PHM, Veldhuis WB, Mann RM, et al. The effect of volumetric breast density on the risk of screen-detected and interval breast cancers: a cohort study. *Breast Cancer Res* 2017; **19**: 67. doi: https://doi.org/10.1186/s13058-017-0859-9
10. Moshina N, Sebuødegård S, Lee CI, Akslen LA, Tsuruda KM, Elmore JG, et al. Automated volumetric analysis of mammographic density in a screening setting: worse outcomes for women with dense breasts. *Radiology* 2018; **288**: 343–52. doi: https://doi.org/10.1148/radiol.2018172972
11. Gastounioti A, Conant EF, Kontos D. Beyond breast density: a review on the advancing role of parenchymal texture analysis in breast cancer risk assessment. *Breast Cancer Research* 2016; **18**: 1–12. doi: https://doi.org/10.1186/s13058-016-0755-8
12. Moller A, Soler M, Thornhill R. Breast asymmetry, sexual selection, and human reproductive success. 1995;: 207–19p..
13. Thornhill R, Moller AP, stability D. Disease and medicine. *Biological Reviews of the Cambridge Philosophical Society* 1997; **72**: 497–548.
14. Jasienska G, Lipson SF, Ellison PT, Thune I, Ziomkiewicz A. Symmetrical women have higher potential fertility. *Evolution and Human Behavior* 2006; **27**: 390–400. doi: https://doi.org/10.1016/j.evolhumbehav.2006.01.001
15. Milne BJ, Belsky J, Poulton R, Thomson WM, Caspi A, Kieser J. Fluctuating asymmetry and physical health among young adults. *Evolution and Human Behavior* 2003; **24**: 53–63. doi: https://doi.org/10.1016/S1090-5138(02)00120-4
16. Manning JT, Scutt D, Whitehouse GH, Leinster SJ. Breast asymmetry and phenotypic quality in women. *Evolution and Human Behavior* 1997; **18**: 223–36. doi: https://doi.org/10.1016/S0162-3095(97)00002-0
17. Scutt D, Manning JT, Whitehouse GH, Leinster SJ, Massey CP. The relationship between breast asymmetry, breast size and the occurrence of breast cancer. *Br J Radiol* 1997; **70**(OCT): 1017–21. doi: https://doi.org/10.1259/bjr.70.838.9404205
18. Eltonsy HN, Elmaghraby A, Tourassi G. Bilateral breast volume asymmetry in screening mammograms as a potential marker of breast cancer: preliminary experience. 2007;V-5 p.
19. Williams AC, Hitt A, Voisin S, Tourassi G.editors.Automated assessment of bilateral breast volume asymmetry as a breast cancer biomarker during mammographic screening. *International Society for Optics and Photonics*. 2013;SPIE Medical Imaging; 2013.
20. Kayar R, Çilengiroğlu Özgül V. Breast volume asymmetry value, ratio, and cancer risk. *Breast Cancer* 2015; **9**: 87–92. doi: https://doi.org/10.4137/BCBCR.S32789
21. Scutt D, Lancaster GA, Manning JT. Breast asymmetry and predisposition to breast cancer. *Breast Cancer Res* 2006; **8**: R14. doi: https://doi.org/10.1186/bcr1388
22. Zheng B, Sumkin JH, Zuley ML, Wang X, Klym AH, Gur D. Bilateral mammographic density asymmetry and breast cancer risk: a preliminary assessment. *Eur J Radiol* 2012; **81**: 3222–8. doi: https://doi.org/10.1016/j.ejrad.2012.04.018
23. Tan M, Zheng B, Ramalingam P, Gur D. Prediction of near-term breast cancer risk based on bilateral mammographic feature asymmetry. *Acad Radiol* 2013; **20**: 1542–50. doi: https://doi.org/10.1016/j.acra.2013.08.020
24. Sun W, Zheng B, Lure F, Wu T, Zhang J, Wang BY, et al. Prediction of near-term risk of developing breast cancer using computerized features from bilateral mammograms. *Comput Med Imaging Graph* 2014; **38**: 348–57. doi: https://doi.org/10.1016/j.compmedimag.2014.03.001
25. Zheng B, Tan M, Ramalingam P, Gur D. Association between computed tissue density asymmetry in bilateral mammograms and near-term breast cancer risk. *Breast J* 2014; **20**: 249–57. doi: https://doi.org/10.1111/tbj.12255
26. Ellingjord-Dale M, Vos L, Tretli S, Hofvind S, Dos-Santos-Silva I, Ursin G, Parity UG. Parity, hormones and breast cancer subtypes - results from a large nested case-control study in a national screening program. *Breast Cancer Res* 2017; **19**: 10. doi: https://doi.org/10.1186/s13058-016-0798-x
27. Lambertini M, Santoro L, Del Mastro L, Nguyen B, Livraghi L, Ugolini D, et al. Reproductive behaviors and risk of developing breast cancer according to tumor subtype: a systematic review and meta-analysis of epidemiological studies. *Cancer Treat Rev* 2016; **49**: 65–76. doi: https://doi.org/10.1016/j.ctrv.2016.07.006
28. Health and Social Care Centre Breast screening programme England. ; 2018;4/4/20182016-2017. UK: NHS Digital.
29. Office for National Statistics (ONS)2011 Census Guidance and Methodology 2015 [Overview of methods and codes used for 2011 census].. Available from: https://www.ons.gov.uk/census/2011census/2011censusdata/2011censususerguide/variablesandclassifications.
30. The Royal College of RadiologistsGuidance *on screening and symptomatic breast imaging*2013London
31. Matakina Technology Ltd VolparaDensity™ User Manual Version 1.5.11. [User Manual Volpara Software. *In press* 2014;.
32. TRCo P. *Pathology reporting of breast disease in surgical excision specimens incorporating the dataset for histological reporting of breast cancer*. London: The Royal College of Pathologists; 2016.
33. Waks AG, Winer EP. Breast cancer treatment: a review. *JAMA* 2019; **321**: 288–300. doi: https://doi.org/10.1001/jama.2018.19323
34. Hudson S, Hjerkind KV, Vinnicombe S, Allen S, Trewin C, Ursin G, et al. Adjusting for BMI in analyses of volumetric mammographic density and breast cancer risk. 2018;.
35. Hudson SM, Wilkinson LS, Denholm R, De Stavola BL, Dos-Santos-Silva I. Ethnic and age differences in right-left breast asymmetry in a large population-based screening population. *Br J Radiol* 2020; **93**;:: 201903280(0). doi: https://doi.org/10.1259/bjr.20190328
36. Cheong AL, Liu J, Reece GP, Nicklaus KM, Catherine Bordes M, Hanson SE, et al. Natural breast symmetry in preoperative breast cancer patients. *Plast Reconstr Surg Glob Open* 2019; **7**: e2297: e2297: . doi: https://doi.org/10.1097/GOX.0000000000002297
37. Roelofs AAJ, Karssemeijer N, Wedekind N, Beck C, van Woudenberg S, Snoeren PR, et al. Importance of comparison of current and prior mammograms in breast cancer screening. *Radiology* 2007; **242**: 70–7. doi: https://doi.org/10.1148/radiol.2421050684
38. Antoni S, Sasco AJ, dos Santos Silva I, McCormack V. Is mammographic density differentially associated with breast cancer according to receptor status? A meta-analysis. *Breast Cancer Res Treat* 2013; **137**: 337–47. doi: https://doi.org/10.1007/s10549-012-2362-4
39. Evans KK, Georgian-Smith D, Tambouret R, Birdwell RL, Wolfe JM. The GIST of the abnormal: above-chance medical decision making in the blink of an eye. *Psychon Bull Rev* 2013; **20**: 1170–5. doi: https://doi.org/10.3758/s13423-013-0459-3
40. Lee HN, Sohn Y-M, Han KH. Comparison of mammographic density estimation by Volpara software with radiologists' visual assessment: analysis of clinical-radiologic

factors affecting discrepancy between them. *Acta Radiol* 2015; **56**: 1061–8. doi: https://doi.org/10.1177/0284185114554674

41. Onega T, Smith M, Miglioretti DL, Carney PA, Geller BA, Kerlikowske K, et al. Radiologist agreement for mammographic recall by case difficulty and finding type. *J Am Coll Radiol* 2012; **9**: 788–94. doi: https://doi.org/10.1016/j.jacr.2012.05.020
42. Evans KK, Haygood TM, Cooper J, Culpan A-M, Wolfe JM. A half-second glimpse often LETS radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast. *Proc Natl Acad Sci U S A* 2016; **113**: 10292–7. doi: https://doi.org/10.1073/pnas.1606187113
43. Brand JS, Czene K, Shepherd JA, Leifland K, Heddson B, Sundbom A, et al. Automated measurement of volumetric mammographic density: a tool for widespread breast cancer risk assessment. *Cancer Epidemiol Biomarkers Prev* 2014; **23**: 1764–72. doi: https://doi.org/10.1158/1055-9965.EPI-13-1219
44. Alonzo-Proulx O, Mawdsley GE, Patrie JT, Yaffe MJ, Harvey JA. Reliability of automated breast density measurements. *Radiology* 2015; **275**: 366–76. doi: https://doi.org/10.1148/radiol.15141686
45. Holland K, van Zelst J, den Heeten GJ, Imhof-Tas M, Mann RM, van Gils CH, et al. Consistency of breast density categories in serial screening mammograms: a comparison between automated and human assessment. *Breast* 2016; **29**: 49–54. doi: https://doi.org/10.1016/j.breast.2016.06.020

6.6 Summary Review

These studies suggest that natural breast fluctuating asymmetry, in both BV and DV, is observed across all socioeconomic and ethnic groups and all age bands within the screening age range. Unlike absolute BV and DV, which vary with age and ethnicity, the magnitude of relative breast asymmetry appears to be consistent across the ethnic and socioeconomic groups, (differences in absolute asymmetry largely driven by differences in volume). We observed marked between-woman variation in both BV and DV asymmetry in the cancer-free breast screening population with between breast differences of over 10% in DV not uncommon. Thus, DV and BV asymmetry offer the potential to be useful independent BC risk factors.

Both DV and BV asymmetry were positively associated with the odds of an interval cancer occurring in the period between routine screens, meaning that either a cancer was missed at screening (FN) or has developed to a symptomatic stage in the interval since the last screen. Possible explanations include the possibility that asymmetry between breasts is associated with faster growing tumours or that asymmetry between the images being read, results in obfuscation of an existing tumour. Further studies are needed to gain a better understanding of this.

DV asymmetry was also positively associated with the odds of cancer being detected at the contemporaneous screen. Although the logical explanation for this was that the tumour itself was responsible for the asymmetry, my findings suggest that the situation is more complex since only 55% of the tumours were located in the larger breast (by DV) and little of the observed DV asymmetry could be explained by the tumour size alone. Thus, it is possible that FA may be an independent risk factor for BC and mammographic breast L:R asymmetry provides a cue for film readers that a cancer may be present.

7 CHAPTER 7 Mammographic compression techniques, outcomes, and screening performance

7.1 Introduction

Chapters 5 and 6 have shown that automated mammographic image analysis tools, that provide comprehensive volumetric breast composition estimates, offer potential for improving our understanding and assessment of BC risk and screening performance beyond their original intended use, which was to estimate density of a woman's breast. In this chapter I look at whether these tools also offer potential for a better understanding of the compression techniques that we use to take mammographic images and their consequences for breast screening outcomes.

Mammographic practitioners agree that that the diagnostic quality of a mammographic image may be related to the compression techniques deployed during imaging and existing standards and guidelines partly reflect this. However, until recently empirical studies on this topic have been restricted to small data sets due to difficulties with collecting data in large volumes. The literature review in Chapter 2 (section 2.9) concluded that despite some recent research, there is a dearth of published empirical evidence that focusses on the association between mammographic compression technique and screening performance. Thus, at the time this thesis was undertaken there was little evidence to support the introduction of objectively measured mammography standards for compression, leaving screening mammographers to rely on comprehensive positioning guidelines and relatively vague, subjective guidelines for the amount of compression force to be used in mammography and with little or no guidance about how force should be adjusted to reflect the breast characteristics (such as size) of the woman being screened.

The introduction of full-field digital mammography (FFDM) has led to the development of automated algorithms such as Volpara, which allow collection of comprehensive image acquisition data as well as volumetric estimates of both breast size and mammographic density. Such automated methods make it feasible to conduct large-scale studies based on objective image acquisition parameters.

The potential for using mammographic acquisition data to improve breast screening performance, depends partly upon whether mammographers vary their compression technique and paddle usage to reflect the characteristics of women being screened, and whether these variations are associated with the compression outcomes that matter for diagnostic imaging, including thickness of the compressed breast. Therefore, I first conducted a descriptive study, (Study IV), of the variation in objectively measured imaging techniques and outcomes in a large, ethnically diverse, breast screening population.

Flexible compression paddles were introduced and promoted by manufacturers as a tool for allowing mammographers to alter the angle of the paddle to reflect the contours of a woman's breast, with the potential benefit that they may reduce pain during mammography. However, the longer-term impact of paddle tilt on screening performance is poorly understood and as far as I am aware this thesis includes the first large scale study to include paddle tilt measurements in a study of compression technique, outcomes, and screening performance.

The literature search found limited research to suggest that compression force and compression pressure may be associated with FNs at screening. Much of this only became available after my thesis had commenced. Limited research on flexible compression paddle usage suggested that the assumption that flexible paddles can be used without compromising image diagnostic quality should be challenged. Study V was therefore designed to address the research question: "Are objectively measurable breast compression techniques associated with risk of false negative outcomes at screening?". Studies IV and V together aim to further our empirical knowledge about objective image acquisition measurements and their association with BC detection in a screening setting.

7.2 Aims and Objectives

Aims:

i) to describe variations in objectively measured compression outcomes (compression thickness and dose) and to examine the extent to which they are related to mammographic technique (force, pressure, and paddle tilt), characteristics of the subject being screened (ethnicity, age) and differences in breast measurements (%BD, BV, DV) across an ethnically and socio-economically diverse screening population.

ii) to assess the strength of associations between these objectively measured mammographic imaging techniques and the odds of screen-detected cancers and the odds of FN at screening (as measured by the occurrence of interval breast cancers in the periods between routine screens).

Primary Objectives:

1) to describe Volpara imaging technique parameters (force, pressure, paddle tilt) and compression outcomes (thickness and dose) in a diverse UK screening population, by age, ethnicity, BV, and BD.

2) to examine the correlations between these compression outcomes (compression thickness and dose) and mammographic technique (force, pressure and paddle tilt), characteristics of the subject being screened (ethnicity, age) and differences in breast measurements (BV, %BD, DV).

3) to assess the strength of association between compression outcomes and compression technique after adjustment for age, ethnicity, BD and BV.

4) to investigate whether mammographic technique (force, pressure, tilt) is associated with increased odds of cancer detection at the current screen.

5) to investigate whether mammographic technique (force, pressure, tilt) is associated with increased odds of having an interval cancer.)

6) to investigate whether mammographic technique (force, pressure, tilt) is associated with increased odds of having a cancer detected at the next routine screen.

Secondary Objectives:

7) to describe how Volpara imaging technique parameters (force, pressure, paddle tilt) and compression outcomes (thickness and dose) vary between different mammographers with different levels of experience.

Objectives 1 – 3 were addressed in Research Paper IV. This paper is supplemented by unpublished results of the descriptive analyses (Section 7.4) that address objective 7.

Objectives 4 – 6 were addressed in Research Paper V. This paper is supplemented by unpublished results of the analyses that were not covered by the published paper due to word constraints in section 7.6.

7.3 Paper IV To what extent are objectively measured mammographic imaging techniques associated with compression outcomes (237).

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	12406683	**Title**	Ms
First Name(s)	Susan M		
Surname/Family Name	Hudson		
Thesis Title	Beyond breast density – Novel uses of automated mammographic analysis in breast cancer screening		
Primary Supervisor	Professor Isabel dos Santos Silva		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	The British Journal of Radiology		
When was the work published?	2023		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	n/a		
Have you retained the copyright for the work?*	**No**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I was responsible for the conceptualisation of the work, development of the methodology, setting up of the mammogram analysis tools in conjunction with Dr Wilkinson and data collection, cleansing and linking of data sources, data analysis and writing of the publication.

SECTION E

[redacted]

Student Signature	
Date	

Supervisor Signature	
Date	

BJR https://doi.org/10.1259/bjr.20230089

Received: 25 January 2023 | Revised: 13 March 2023 | Accepted: 14 March 2023

Cite this article as:
Hudson SM, Wilkinson LS, De Stavola BL, dos-Santos-Silva I. To what extent are objectively measured mammographic imaging techniques associated with compression outcomes. *Br J Radiol* (2023) 10.1259/bjr.20230089.

FULL PAPER

To what extent are objectively measured mammographic imaging techniques associated with compression outcomes

[1]SUE M HUDSON, BSc MSc, [2]LOUISE S WILKINSON, BA, BM, BCh, FRCR, BIANCA L DE STAVOLA, PhD and [1]ISABEL DOS-SANTOS-SILVA, MD MSc PhD

[1]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel, London, UK
[2]Oxford Breast Imaging Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

Address correspondence to: Ms Sue M Hudson
E-mail: *susan.hudson@lshtm.ac.uk;sue.hudson@pasconsulting.co.uk*

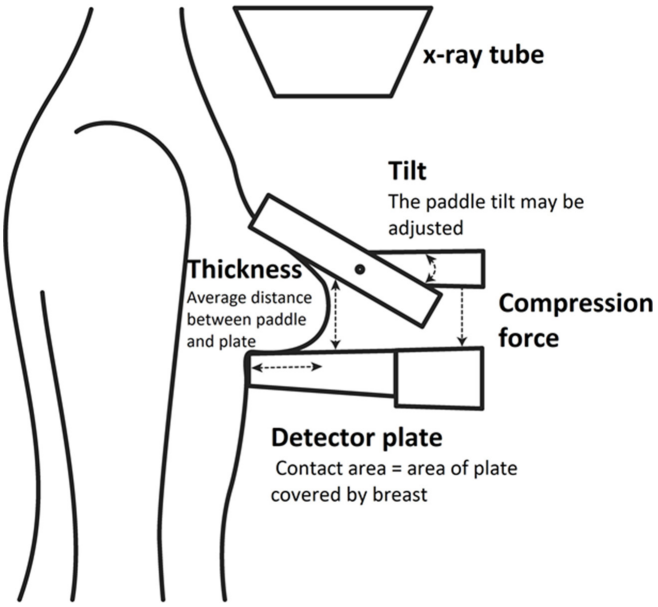
Objective: To describe the association between objectively measurable imaging techniques and the resulting compression thickness and dose.
Methods: The study included 80,495 routine screens from the South-West London Breast Screening Service between March 2013 and July 2017. Average compression force, paddle tilt and dose were calculated. The Volpara® DensityTM algorithm was used to estimate pressure, breast volume and density.
Linear regression models, using generalized estimating equations (GEEs) to account for clustering by practitioner, assessed the strength of the associations between the imaging compression outcomes, (thickness, dose) and imaging techniques (force, pressure and paddle tilt), adjusting for the subject's characteristics (age, ethnicity, breast volume and percent mammographic density).
Results: Fully adjusted linear regression models showed that compression thickness decreased by ~1mm (~2% of mean thickness) for every 1daN increase in force and decreased by ~0.8mm with an increase of 1kPa of pressure (at median pressure). Increasing pressure above 15kPa resulted in minimal reduction in thickness. Dose increased with increased force but decreased by ~1% of mean dose with every increase in 1kPa of pressure. For 1° increase in paddle tilt, the compression thickness increased by ~1.5mm (~2.5%) and dose increased by ~2.5%, (Pt <0.001 in all cases).
Conclusion: Differences in imaging technique are associated with imaging outcome measures (thickness and dose). A better understanding of the association between objective image acquisition parameters and tumour conspicuity could lead to clearer guidelines for practitioners.
Advances in knowledge: Increased paddle tilt is associated with increased compression thickness and increased dose after adjustment for breast volume and force applied.

INTRODUCTION

Breast cancer screening, through mammography, offers one approach to reducing mortality from breast cancer through early diagnosis of non-palpable tumours followed by early treatment. In England and Wales, ~3 million women, aged 50–70 years are invited once every 3 years to undergo standard 2-view mammography of each breast as part of the National Health Breast Screening Programme (NHSBSP).[1] Although the radiation risks associated with screening mammography are relatively low, the regular exposure of well-women to potentially harmful X-ray exposures should be kept to the minimum required to obtain an adequate breast image.[2] Not all mammographic images however are high quality and any factor that leads to a reduction in image quality could be detrimental to cancer detection. A key factor that influences both absorbed dose and image quality, is breast compression. Compression is required to reduce movement, separate superimposed tissue and to reduce tissue thickness which is associated with both increased tumour conspicuity[3,4] and reduction in radiation load.[5,6] At the same time, breast compression is associated with pain which may deter females from attending a routine screen.[7,8] There is evidence that 'too much' compression can lead to unnecessary pain for no additional improvement in image quality,[9,10] however, research remains very limited in this area.

Despite the importance of compression technique, guidelines remain largely subjective and the most recent NHSBSP recommendations, which stipulate regular image quality

Figure 1. Compression of the breast during CC image acquisition schematic. CC, craniocaudal.



audits, are restricted to subjective checks for image sharpness and for 'adequate compression to hold breast firmly/no movement'.[6] Earlier NHSBSP guidelines, in force at the time images in this study were taken, also included the recommendation that force should not exceed 20 daN.[11]

Compression is achieved by pressing the breast against the top of the detector using a transparent plastic compression paddle (Figure 1). The force applied can be monitored by the practitioner during the process however 'pressure', (force divided by the area of contact between breast and detector plate) may be better measure of compression, because it takes account of the size of the breast as well as the force applied.[12] Studies show that screening performance, including cancer detection, may be associated with compression pressure.[13,14] With a rigid paddle, the paddle remains parallel to the detector during imaging. The optional, flexible paddle, was introduced by equipment manufacturers to make the process of mammography more comfortable, allowing the practitioner to 'tilt' the paddle using a hinging mechanism to accommodate the shape and size of the breast, although the effectiveness of these flexible paddles for pain reduction has been queried[15,16] as has their effect on image quality and dose.[15,17]

Studies have found considerable variation in breast compression force and pressure between practitioners[18–20] both within screening organisations[21,22] and also across different countries.[23,24] Studies have also shown that the breast characteristics such as mammographic density are correlated with compression force, pressure and thickness.[25] Less is known about the variation in flexible-paddle tilt, used during image acquisition and its association with compression measurements. It is also possible that the ethnicity of the subject being screened could modify any association between compression parameters and compression outcomes, if, *e.g.* slight physiological differences between ethnic groups influenced the way that mammography is conducted, likewise the actual mammography machine may modify these associations due to physical location and constraints.

Mammographic screening in the UK utilises full-field digital mammography and the resulting images may be processed using automated algorithms which provide comprehensive image acquisition data as well as volumetric estimates of both breast size and mammographic density. Therefore, it has become feasible to carry out large-scale studies based on objective image acquisition parameters.

This study aims to describe the variation in the image acquisition parameters that are controlled by the practitioner during the imaging process (*i.e.* force, pressure and paddle tilt) using a sample of over 80,000 examinations of females who participated in a population-based breast screening programme. We believe that this will be the first large-scale study to include the variation in use of the flexible paddle. Secondly, the study aims to describe the association between imaging technique (force, pressure and paddle tilt) and the resulting compression thickness and dose (after adjusting for the screening subject's breast volume and density) since research suggests that these factors are likely to be key to successful diagnostic imaging in a screening programme.

METHODS

Study participants

The study includes mammographic examinations undertaken in the period 01 March 2013 to 20 June 2017 as part of the NHSBSP routine 3-yearly screening programme at the South-West London Breast Screening Service (SWLBSS) based in the St George's University Hospitals National Health Service (NHS) Foundation Trust. Females screened were resident in one of six London boroughs—Wandsworth, Merton, Croydon, Sutton, Richmond and Kingston and all were asymptomatic for breast cancer at the time of screening. Data on ethnicity were recorded according to the Census classification[26] as part of standard practice via a self-completed screening questionnaire. The practitioner responsible for each screening examination (*i.e.* each set of four images) was recorded on the screening administrative system.

Each screening examination consisted of the NHSBSP standard 2-view [craniocaudal (CC) and mediolateral oblique (MLO) views] mammography of each breast.[27] The raw anonymised digital mammographic images were processed using the automated algorithm, Volpara® Density™ v. 1.5.11 (Matakina Technology Limited, Wellington, New Zealand)[28] to generate automated estimates (in cm^3) of the volume of the breast (BV) and the volume of the radiodense tissue (DV) for each image. Mammographic density (%MD) was estimated as DV/BV×100. Data on the imaging technique were also collected; compression force (decaNewton, daN), compressed breast thickness (mm) and compression paddle tilt (degrees from horizontal) which were available from the Digital Imaging and Communications in Medicine (DICOM) image header. The Volpara® Density™ algorithm estimated contact area (cm^2) between breast and plate for each image and the resulting pressure (kiloPascals,kPa) from force*10/contact area. The mean glandular dose (MGD)(in milli

Gray (mGy)) as calculated by the machine manufacturer and the identification of the mammography machine 'detector' were available from the DICOM header.

Exclusions

In all, 94,408 screening examinations were carried out during the study period. We excluded examinations for which no reason was specified (*i.e.* screening episode type missing) (n = 992). Examinations were also excluded if there were not exactly four images taken, because the automated algorithm is not designed to make estimates in these circumstances.[28] Thus, we excluded examinations of females who have exceptionally large breasts requiring additional (mosaic) images, examinations that were repeated for technical reasons, and examinations where fewer images were taken because of mastectomy or lack of tolerance of the procedure (n = 10,882; Supplementary Table 1). Because of potential differences between manufacturers, examinations using non-Hologic systems were also excluded (n = 626) and we also excluded examinations on subjects known to have previous breast cancer (n = 1413) because this may influence the imaging technique, leaving a total of 80,495 examinations (321,980 compressions) eligible for inclusion in the analysis (Supplementary Figure 1).

Ethical approval

This retrospective study was carried out on fully anonymous, routinely collected data only, held in accordance with the NHS Cancer Screening Programmes Confidentiality and Disclosure Policy 2011. The NHSBSP has section 251 support under the NHS Act 2006. The study was approved by all relevant ethics committees (Research Ethics Committees from St George's University Hospitals NHS Foundation Trust, and the London School of Hygiene and Tropical Medicine).

Statistical methods

The distributions of the imaging parameters (force, pressure and paddle tilt) and the imaging compression outcomes (dose and compressed breast thickness) were examined. Scatter plots were created to examine the distribution and Spearman correlations between the outcomes and imaging parameters. Similarly, scatter plots and Spearman correlation coefficients were used to examine the correlations between imaging compression outcomes (thickness, dose) and the characteristics of the imaging subject (age, BV, %MD). A line of best fit was calculated for each plot using a locally weighted scatterplot smoothing (Lowess) function.

Linear regression models were used to examine the strength of the associations between the imaging outcomes, (thickness and dose) and the three imaging parameters (force, pressure and paddle tilt, treated as continuous variables), after adjusting for the subject's characteristics (age, BV and %MD). Spearman rank correlation coefficients were used to identify potential collinearity between predictors in the proposed models. Models for each exposure were further adjusted for the other two compression parameters where collinearity was not an issue. General estimating equations (GEEs) and robust standard errors (clustering by practitioner) were used to account for the fact that each practitioner carried out multiple examinations in the study period and practitioners may have their own imaging technique. Tests for departure from linear trend were conducted by including quadratic terms for each exposure variable and plotting the estimated exposure response curves. In all the analyses, regression coefficients represent the change in per one-unit change in the exposure variable.

We categorised compression parameters (force, pressure and tilt) into high and low categories and used linear regression models to test for effect modification by ethnicity or mammography machine (detector) on the association between compression parameters and outcomes (thickness and dose).

We considered statistical significance (two-sided) at p-value < 0.05. All analyses were conducted in Stata (IC 14).[29]

RESULTS

Characteristics of screening examinations

The characteristics of the 80,495 screening examinations are shown in Table 1. The majority (~86%) of examinations were on females who were within the ages of 50–70 years, the main age-group targeted by the NHSBSP. Among the 86% of the subjects who reported their ethnicity, ~76% were White but there were also high numbers of females of Black and Asian ethnicity. In all, 87 different practitioners carried out examinations during the study.

The mean force applied during a single MLO compression was higher than that for a CC compression (9.12 and 7.74 daN respectively), likewise, the mean paddle tilt was greater during an MLO compression than a CC compression (2.87° and 2.43° from horizontal respectively). In contrast, the mean pressure was higher for CC views than for MLO views (9.97 and 7.36 kPa respectively). The mean of MLO and CC values were used for this study unless otherwise stated.

The distributions of the imaging parameters and the outcomes were approximately normal (Supplementary Figure 2).

Correlations between characteristics of screening subject and compression outcomes

There was a strong/moderate correlation between the compression outcomes and the subject's BV with larger BV associated with increased thickness and dose (Spearman correlation coefficient (ρ): 0.83 and 0.56, respectively; p < 0.001 for both). Both compression thickness and dose were negatively correlated with %MD (ρ = −0.63 and −0.11, respectively; p < 0.001 for both). The compression outcomes (thickness and dose) both decline somewhat with age of subject, but the correlations were very weak (Figure 2).

Associations between imaging parameters and compression outcomes

Figure 3 shows weak positive Spearman's correlations between force and the outcome measures (thickness ρ = 0.14; and dose ρ = 0.28; p < 0.001 in both cases). In contrast the correlation between pressure and thickness is moderate but is negative (ρ = −0.44, p < 0.0001) and the relationship appears to be non-linear.

Table 1. Technical and subject characteristics associated with screening examinations

	Frequency	Percent %
All standard screening examinations[a]	80,495	
Age at screening, years		
<45-	551	0.70%
45–49	5,928	7.40%
50–54	22,082	27.40%
55–59	18,175	22.60%
60–64	15,080	18.70%
65–69	13,633	16.90%
70+	5,046	6.30%
Missing	0	0.00%
Ethnicity (of subject screened[b])		
White—British or Irish or other	52,461	65.20%
Asian—British Indian or Pakistani or Bangladeshi or other	7,611	9.50%
Black—British or Caribbean or other	3,745	4.70%
Black—African	2,725	3.40%
Mixed White and Black, White and Asian or any other mixed	1,526	1.90%
Chinese	1,062	1.30%
Missing or not reported	11,365	14.10%
Breast volumetric measurements[c]	**Median**	**IQR**
Breast volume, cm^3	753	489–1,110
Breast dense volume, cm^3	49.4	37.2–67.3
Mammographic density, %	6.40%	4.6–10.2%
Imaging acquisition parameters average across MLO and CC views[c]	**Mean**	**SD**
Mean compression force applied, daN	8.27	2.09
Mean paddle tilt angle, degrees positive from horizontal	2.69	1.06
Mean pressure, kPa	8.6	3.53
Imaging outcome estimates average across MLO and CC views[c]		
Manufacturers mean glandular dose, mGy[d]	1.32	0.36
Mean breast thickness, mm	56	12.4
Imaging acquisition parameters for MLO[e]	**Mean**	**SD**
Mean compression force applied, daN	9.12	2.59
Mean paddle tilt angle, degrees positive from horizontal	2.87	1.28
Mean pressure, kPa	7.36	2.51
Imaging outcome estimates for MLO[e]		
Mean glandular dose, mGy[d]	1.39	0.41
Mean breast thickness, mm	57.8	13.67
Imaging acquisition parameters for CC[e]	**Mean**	**SD**
Mean compression force applied, daN	7.74	1.95
Mean paddle tilt angle, degrees positive from horizontal	2.43	1.16
Mean pressure, kPa	9.97	4.96

(Continued)

Table 1. (Continued)

	Frequency	Percent %
Imaging outcome estimates for CC[e]		
Mean glandular dose, mGy[d]	1.25	0.34
Mean breast thickness, mm	54.08	11.72

CC, craniocaudal;DICOM, Digital Imaging and Communications in Medicine; IQR, interquartile range; MLO, mediolateral oblique; SD, standard deviation.

[a]A screening examination was included if it had exactly four images taken, only screening appointments were included. We excluded images taken on non-Hologic systems and screens where females were known to have previous cancer. Total number of images (compressions) was 321,980

[b]Data on ethnicity were collected as part of standard screening protocol via a self-completed screening questionnaire and recorded according to the Census classificationand summarised as, "Asian" (Indian, Pakistani or Bangladeshi or other), "Black-African", "Black-British or Caribbean or other", "Chinese", "Mixed" (White and Black, White and Asian or any other mixed), "White" (British or Irish or other) and "Other". Count per screening examination (subjects may have more than one examination over the study period).

[c]Breast volumetric measures were calculated from the mean value from the four images: left CC image, right CC image, left MLO image, right MLO image.

[d]Manufacturers mean glandular dose as recorded in DICOM header.

[e]Calculated from the average value from the two relevant images left and right sides.

The adjusted fitted linear regression model shows that compression thickness decreased with increasing force (P for trend (Pt) <0.001; Table 2), after a simple adjustment for BV alone. Further adjustment for %MD, age and paddle tilt did not change the direction of the association, but the regression coefficients were strengthened somewhat (β = −0.81 and −1.07, respectively; p < 0.001 for both). There was little evidence for departure from linear trend in the exposure response plots (Supplementary Figure 3).

Thickness also decreased with increasing pressure, (Pt <0.001; Table 2) when controlling for %MD, age and paddle tilt. BV adjustment was omitted from the pressure model due to collinearity because it is strongly negatively correlated with pressure (ρ = −0.73 see Supplementary Table 2). The exposure response curve shows that, at increased pressures, there was no longer a reduction of breast thickness suggesting a diminishing return from additional pressure above an optimal point, after controlling for %MD, age and paddle tilt (Supplementary Figure 3).

Dose increased with increasing force (Pt<0.001 in all models; Table 2) although this was attenuated after adjustment for BV. In contrast dose decreased slightly with increasing pressure (Pt<0.001) after adjustment for %MD, age and paddle tilt.

Figure 2. Scatter plots and heat maps of compression outcomes (thickness and dose) against characteristics of imaging subject (lowess smoothing). All measurements derived from average of four mammographic views. Left breast CC, Right breast CC, Left breast MLO, Right breast MLO. Dose is manufacturer's recorded mean glandular dose in milligray (mGy). Showing Spearman's correlation coefficient (ρ). CC, craniocaudal; MLO, mediolateral oblique.

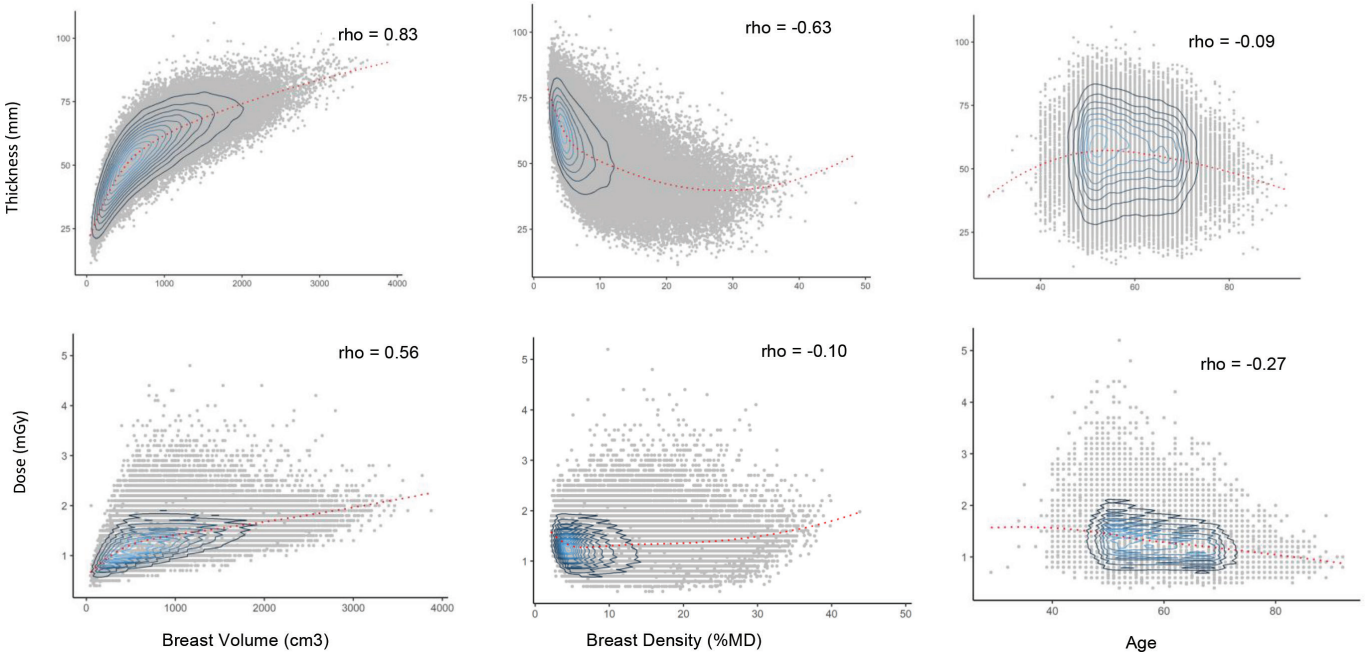


Figure 3. Scatter plots of compression outcomes (thickness and dose) with image acquisition parameters (force, pressure and paddle tilt) lowess smoothing. All measurements derived from average of four mammographic views. Left breast CC, Right breast CC, Left breast MLO, Right breast MLO. Dose is manufacturer's recorded mean glandular dose in milligray (mGy). Showing Spearman's ρ correlation coefficient. Tilt is positive tilt in degrees from horizontal. CC, craniocaudal; MLO, mediolateral oblique.

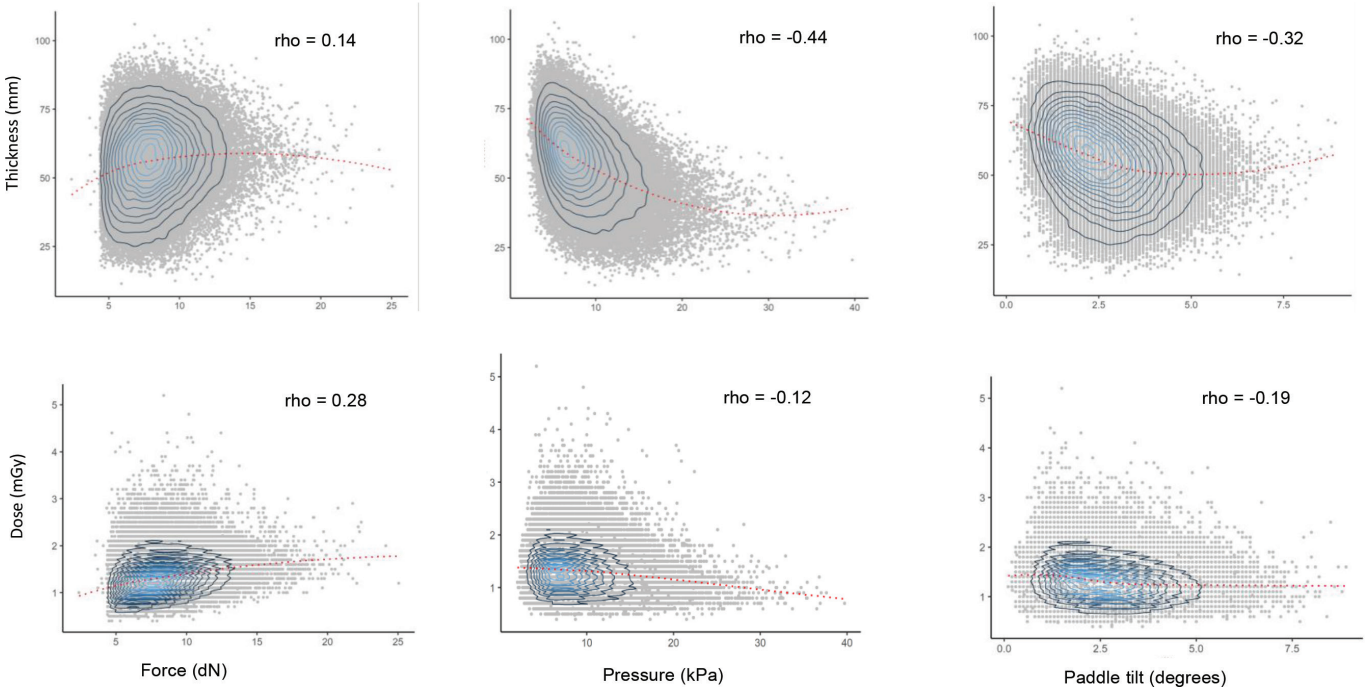


Table 2. Linear regression analysis of associations between compression parameters (force, pressure and tilt) and compression outcomes (thickness and dose) crude and after adjustment [a]

	Compression thickness mm(n = 79,476)			
	Crude	**Adjusted for breast volume**	**Fully adjusted[a]**	**Mutually adjusted[b]**
	Change per unit (95% CI)	**Change per unit (95% CI)**	**Change per unit (95% CI)**	**Change per unit (95% CI)**
Force daN	1.89 (1.0, 2.19)	−0.81 (-0.97,−0.65)	−0.92 (-1.07,−0.77)	−1.07 (-1.21,−0.92)
Pressure kPa	−1.52 (-1.60,−1.43)	N/A[c]	−0.79 (-0.85,−0.73)	−0.59 (-0.65,−0.53)
Paddle tilt °	−3.25 (-3.43,−3.07)	1.48 (1.36, 1.60)	1.27 (1.15, 1.39)	1.48 (1.36, 1.59)
	Mean glandular dose milliGray[d] (n = 69,924)			
	Crude	**Adjusted for breast volume**	**Fully adjusted[a]**	**Mutually adjusted[b]**
	Change per unit (95% CI)	**Change per unit (95% CI)**	**Change per unit (95% CI)**	**Change per unit (95% CI)**
Force daN	0.061 (0.056, 0.067)	0.017 (0.013, 0.021)	0.023 (0.019, 0.028)	0.020 (0.016, 0.024)
Pressure kPa	−0.015 (-0.016,−0.013)	N/A[c]	−0.019 (-0.021,−0.017)	−0.011 (-0.131,−0.009)
Paddle Tilt °	−0.060 (-0.064,−0.057)	0.029 (0.021, 0.035)	0.041 (0.036, 0.046)	0.037 (0.032, 0.042)

BV, breast volume ; CI, confidence interval; GEE, generalized estimating equation.
P for linear trend <0.001 in all cases
[a]Adjusted for: %MD, Age with GEE and robust standard errors to account for mammographer clusters. Force and Tilt models additionally adjusted for BV, which was omitted in the Pressure model due to collinearity.
[b]Mutually adjusted, *i.e* force models additionally adjusted for tilt;pressure models additionally adjusted for tilt; tilt models additionally adjusted for force. Pressure omitted from force and tilt models due to collinearity.
[c]Pressure models are not additionally adjusted for breast volume due to collinearity.
[d]Manufacturer's estimatedmean glandular dose.

Both thickness and dose are weakly negatively correlated with paddle tilt (ρ = −0.32 and −0.19, respectively; $p < 0.001$ for both), suggesting that, in a model where there is no adjustment for breast volume, both dose and thickness decline with increasing paddle tilt. However, after controlling for BV, %MD, age and force, compression thickness increased with increasing paddle tilt and there was no evidence of departure from linear trend(Pt<0.001; Table 2)). For each 1° increase in paddle tilt, the compression thickness increased by ~1.5 mm (~2.5%). Likewise, after adjustment for BV and force, we found a positive association between dose and paddle tilt. For each 1° increase in paddle tilt, dose increased by ~2.5%, (P for trend (Pt) <0.001; Table 2) in the fully adjusted model.

There was evidence for an interaction between ethnicity and force applied ($p < 0.001$; results not shown) after adjustment for BV, %MD and age. However, the differences in coefficients between the different ethnicities represented small differences in thickness (~1 mm).There was no evidence of interaction between detector plate and explanatory variables in any of the models.

DISCUSSION

Main findings

Our study of over 80,000 screening examinations shows that there is large variation in imaging technique, as measured by the compression parameters; force, paddle tilt and pressure. Although the strongest correlate of compression outcome measures (thickness and dose) is BV these outcomes were also associated with the technique applied after adjustment for BV. Compression thickness decreased by ~1 mm (~2% of mean thickness) for every 1 daN increase in force after adjusting for the imaging and subject dependent confounders (tilt, BV, %MD, and age). Thickness decreased by ~0.8 mm (~1.5% of mean thickness) with an increase of 1 kPa of pressure after adjustment for tilt, %MD and age (quadratic model at median pressure value). Dose increased by 1.5% of mean dose with 1 daN increased force (after adjustment), whereas dose decreased by ~0.8% of MGD with every increase in 1 kPa of pressure after adjustment. Outcome measures were also associated with the degree of paddle tilt employed, after full adjustment for subject-dependent confounders. For every 1° increase in paddle tilt, the compression thickness increased by ~1.5 mm (~2.5%) and dose increased by 0.037 mGy (~2.5%). This supports findings in a Dutch study, which found that mean radiation dose was 4.5% lower when rigid (horizontal) paddles were used rather than tilting paddles.[15]

Whether these changes are of clinical relevance is uncertain, although Salvagnini found that lesion detectability decreased from 70 to 37% as thickness increased from the lowest thickness quartile group (<29 mm) to the greatest thickness quartile (>70 mm), in a study of simulated lesions in real breast images.[3]

As expected, compression outcomes were correlated with the characteristics of the subjects being screened; a larger BV was strongly correlated with increased thickness and moderately correlated with higher dose. Thickness was negatively correlated with %MD (possibly because higher %MD is associated with smaller BV); a finding that is similar to Khan-Perez et al in a study of 211 UK females[30] and Waade from a study of ~11,000 women in Norway.[31] We found a low correlation between dose and %MD, a finding also supported by Khan-Perez et al.[30] Ng et al analysed images from 17 different counties and concluded that beyond the practitioner's breast compression choices, the subject's age, and breast composition (BV and %MD), there are other factors influencing compression.[24]

Overall mean force and pressure were low in our study in comparison with a study of ~37,000, similar-aged, Dutch females, using the same analytical algorithm[23]; mean MLO force in the Netherlands was 13.8 (SD 2.7) daN compared to a mean of 9.12daN (SD 2.59) in our study. The practitioners in the Netherlands used protocols instructing them to compress to at least 12 daN but at the time that our study data were collected, the UK NHSBSP guidelines did not specify a minimum compression force and were limited to the guidance that force should not exceed 20 daN.[11] Mean MLO pressure (Netherlands) was 13.7 kPa (SD 5.9) compared to 7.36 kPa (SD 2.51) in our study. Comparative results from the USA[23] were 7.4 daN (SD 3.1) for force and 8.1 kPa (SD 4.1) for pressure, similar to our study. A Norwegian study on ~18,000 examinations found large variation between centres and reported a mean force (average of CC and MLO compressions) of 11.6 daN; higher than our mean of 8.27 daN.[32] This suggests that even across European screening programmes where guidelines[5] have been shared, there is a large variation between programmes. In our study, mean tilt was 2.69° (SD = 1.06; range 2.29 to 3.15), somewhat lower than the 3.73 ° (SD = 2.18) reported by Kallenburg et al. from a sample of 287 examinations in Netherlands that used flexible paddles.[33] Note, however, that the mean resulting thickness achieved in our study (57.80 mm) was lower than, but very similar to both the Netherlands and US studies (60.7 mm and 59.9 mm respectively) suggesting that the direct comparisons of compression force across different screening populations are not straightforward.

Strengths and limitations

Strengths of this study include its population-based design and large sample size. We believe that this is the first large study to look at the association of paddle tilt with compression outcomes in a large screening population.

A limitation of this study was that most females were postmenopausal, and average breast density is likely to be lower than a sample that includes younger females, therefore findings are only applicable to this age range. A large number (10,212) of females undergoing routine screening were excluded from the study because there were not exactly four images taken. Whilst the demographic characteristics of the excluded group are not very different from the main study group, it would still be informative to study this group using other technology. We only included examinations from a single breast screening unit where we might expect some consistency due to local quality assurance and supervision. We used one specific algorithm for estimating breast measurements, however this algorithm has been found to produce reliable and repeatable results.[34–36] The study also uses the X-ray machine manufacturers' own estimate of MGD, which has been shown to be rather a crude estimate[37] and not

specifically adapted to incorporate tilting paddles. None-the-less despite these uncertainties, the general findings related to dose are likely to be of interest.

Implications

Compression outcomes are important because Salvagnini et al[3] have shown, using simulated breast lesions, that tumour detectability increases with reduced compressed breast thickness. Furthermore. recent studies in both the Netherlands[14,38] and Norway have found that cancer detection is associated with force and pressure used in the image acquisition process. Alongside these considerations, the breast is a radiation-sensitive organ and it is important restrict the glandular dose as much as possible without compromising image quality and cancer detection.

However, these relationships are complex and in our study, we found that increasing pressure beyond 15 kPa, had substantially diminishing returns in terms of decreased thickness, supporting the suggestion from Hogg et al. that, above a certain level (~13 daN in their UK study on younger, symptomatic females), increased force does not reduce thickness considerably and could be avoided,[39] although their study is not directly comparable.

Flexible (tilting) paddles were introduced as a way of reducing pain during mammography but Broeders et al[15] and Moshina et al[16] found no pain reduction during use. Because increased paddle tilt is associated with increased compression thickness and dose after adjustment for BV, it is possible that flexible paddle use has a detrimental effect on screening performance without any reduction in pain. Further studies are required to examine the association between flexible paddle use and breast cancer detection.

Our findings suggests that compression force, pressure and tilt are not systematically adjusted in accordance with subjective breast characteristics and consequently, there is inconsistency in technique and outcome. In particular, force is not systematically adjusted to reflect BV, resulting in variation in pressure, with larger females being compressed using lower pressures and smaller females experiencing higher pressures. Our study further suggests that ethnicity may play a role in the imaging process, but further research is required in this area.[A]

Studies by de Groot et al. in the Netherlands also found that females with smaller BV experienced severe pain more commonly than other subjects[40] suggesting that protocols are not always appropriate for females of smaller BV. They proposed that pressure-based guidelines could be better than force-based guidelines in mammography. Our study suggests that force-based guidelines could be appropriate but only if controlled for breast volume. However, under real-time conditions, objective measures of BV are not available and therefore pressure guidance may provide a practical alternative. A recent systematic review by Serwan et al, looked at the relative merits of introducing a pressure-standardised protocol in place of force standardisation and concluded that pressure-standardised protocols could be implemented to reduce pain levels without compromising image quality.[41] Until recently, real-time estimates of detector plate contact area were not readily available, which made real-time estimation of pressure impossible and hence implementation of pressure protocols was impractical in a screening setting. However, recent technological developments are becoming available to support the introduction of such protocols.

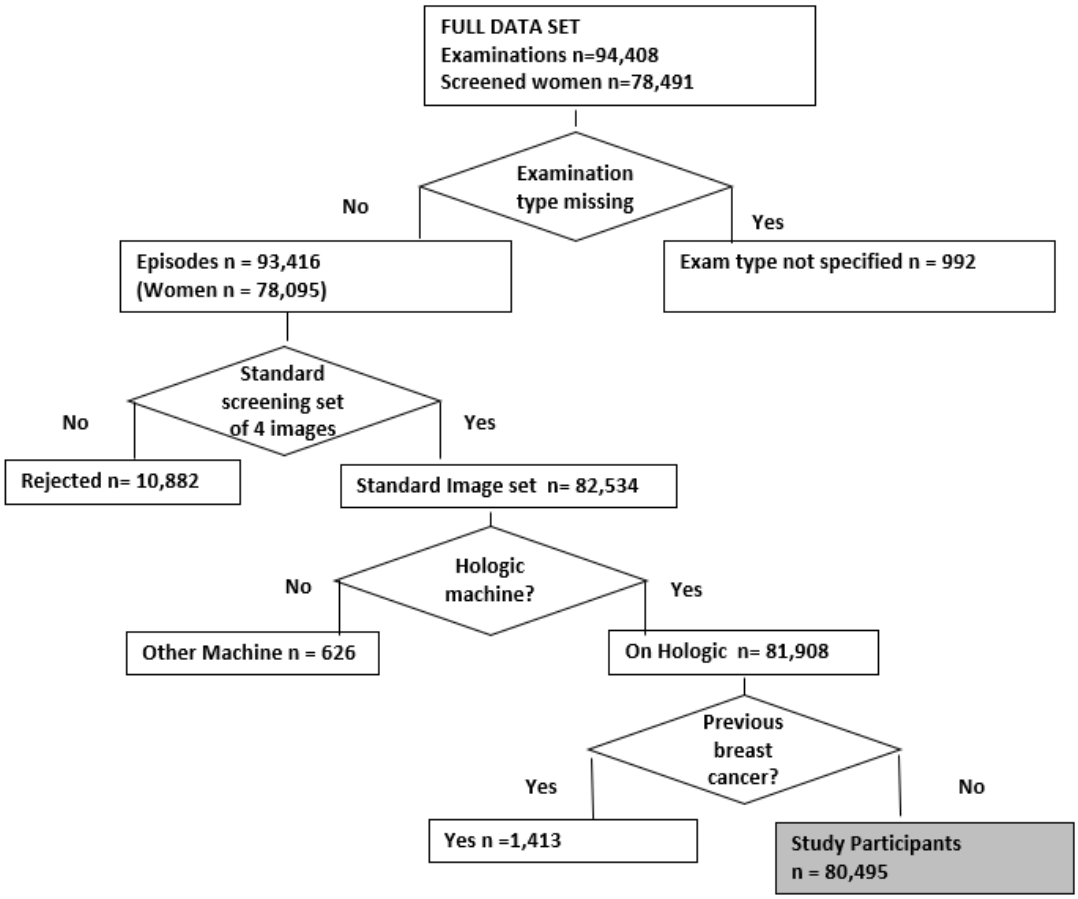
Mammography involves consideration of both objective and subjective parameters and an 'appropriate pressure' level is achieved using judgement about size, density and elasticity of the subject's breast as well as the subject's pain tolerance.[B] It is possible that a better understanding of the association between directly measurable image acquisition parameters and tumour conspicuity could add to this judgement and inform new guidelines, potentially improving overall screening performance through the provision of more objective imaging guidelines.

REFERENCES

1. (NHS Digital.). Breast screening programme. England - 2013-14. Available from: http://www.hscic.gov.uk/catalogue/PUB16803[
2. Yaffe MJ, Mainprize JG. Risk of radiation-induced breast cancer from mammographic screening. *Radiology* 2011; **258**: 98–105. https://doi.org/10.1148/radiol.10100655
3. Salvagnini E, Bosmans H, Van Ongeval C, Van Steen A, Michielsen K, Cockmartin L, et al. Impact of compressed breast thickness and dose on lesion detectability in digital mammography: FROC study with simulated lesions in real mammograms. *Med Phys* 2016; **43**: 5104: 5104–16: . https://doi.org/10.1118/1.4960630
4. Saunders RS Jr, Samei E. The effect of breast compression on mass conspicuity in digital mammography. *Med Phys* 2008; **35**: 4464–73. https://doi.org/10.1118/1.2977600
5. Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L. European guidelines for quality assurance in breast cancer screening and diagnosis. fourth edition -- summary document. *Ann Oncol* 2008; **19**: 614–22. https://doi.org/10.1093/annonc/mdm481
6. NHS Breast Screening Programme: NHS Breast Screening Programme Guidance for breast screening mammographers. In. London, UK: Public Health England; 2017.
7. Whelehan P, Evans A, Wells M, Macgillivray S. The effect of mammography pain on repeat participation in breast cancer screening: A systematic review. *Breast* 2013; **22**: 389–94. https://doi.org/10.1016/j.breast.2013.03.003
8. Meyer J, Maxwell A, Harkness E, Astley S, Mercer C, Wilson M, et al. PB.22. does mammographic compression force at breast screening influence the likelihood of subsequent screening attendance? *Breast Cancer Res* 2014; **16**(S1). https://doi.org/10.1186/bcr3710
9. Lau S, Abdul Aziz YF, Ng KH. Mammographic compression in Asian women. *PLoS ONE* 2017; **12**(4): e0175781. https://doi.org/10.1371/journal.pone.0175781
10. de Groot JE, Hopman IGM, van Lier M, Branderhorst W, Grimbergen CA,

[A] Section 7.4 includes additional unpublished analysis on ethnic variations in compression parameters
[B] Section 7.4 includes additional unpublished analysis on the variation in compression parameters by mammographer and experience

den Heeten GJ. Pressure-standardised mammography does not affect visibility, contrast and sharpness of stable lesions. *Eur J Radiol* 2017; **86**: 289–95. https://doi.org/10.1016/j.ejrad.2016.11.030
11. NHS Cancer Screening Programmes: QUALITY ASSURANCE GUIDELINES FOR MAMMOGRAPHY Including Radiographic Quality Control. In. Sheffield: NHS Cancer Screening Programmes; 2006.
12. De Groot JE: Pressure-standardized breast compression in mammography. Amsterdam: Amsterdam; 2015.
13. Moshina N, Sebuødegård S, Hofvind S. Is breast compression associated with breast cancer detection and other early performance measures in a population-based breast cancer screening program? *Breast Cancer Res Treat* 2017; **163**: 605–13. https://doi.org/10.1007/s10549-017-4214-8
14. Holland K, Sechopoulos I, Mann RM, den Heeten GJ, van Gils CH, Karssemeijer N. Influence of breast compression pressure on the performance of population-based mammography screening. *Breast Cancer Res* 2017; **19**(1): 126. https://doi.org/10.1186/s13058-017-0917-3
15. Broeders MJM, Ten Voorde M, Veldkamp WJH, van Engen RE, van Landsveld-Verhoeven C, 't Jong-Gunneman MNL, et al. Comparison of a flexible versus a rigid breast compression paddle: Pain experience, projected breast area, radiation dose and technical image quality. *Eur Radiol* 2015; **25**: 821–29. https://doi.org/10.1007/s00330-014-3422-4
16. Moshina N, Sebuødegård S, Evensen KT, Hantho C, Iden KA, Hofvind S. Breast compression and experienced pain during mammography by use of three different compression paddles. *Eur J Radiol* 2019; **115**: 59–65. https://doi.org/10.1016/j.ejrad.2019.04.006
17. Ma WK, Brettle D, Howard D, Kelly J, Millington S, Hogg P. Extra patient movement during mammographic imaging: An experimental study. *Br J Radiol* 2014; **87**(1044). https://doi.org/10.1259/bjr.20140241
18. Mercer CE, Hogg P, Lawson R, Diffey J, Denton ERE. Practitioner compression force variability in mammography: A preliminary study. *Br J Radiol* 2013; **86**(1022). https://doi.org/10.1259/bjr.20110596
19. Mercer CE, Hogg P, Szczepura K, Denton ERE. Practitioner compression force variation in mammography: A 6-year study. *Radiography* 2013; **19**: 200–206. https://doi.org/10.1016/j.radi.2013.06.001
20. Waade GG, Sebuødegård S, Hogg P, Hofvind S. Breast compression across consecutive examinations among females participating in breastscreen norway. *Br J Radiol* 2018; **91**(1090). https://doi.org/10.1259/bjr.20180209
21. Mercer CE, Szczepura K, Kelly J, Millington SR, Denton ERE, Borgen R, et al. A 6-year study of mammographic compression force: Practitioner variability within and between screening sites. *Radiography* 2015; **21**: 68–73. https://doi.org/10.1016/j.radi.2014.07.004
22. Waade GG, Sanderud A, Hofvind S. Compression force and radiation dose in the norwegian breast cancer screening program. *Eur J Radiol* 2017; **88**: 41–46. https://doi.org/10.1016/j.ejrad.2016.12.025
23. Branderhorst W, de Groot JE, Highnam R, Chan A, Böhm-Vélez M, Broeders MJM, et al. Mammographic compression--a need for mechanical standardization. *Eur J Radiol* 2015; **84**: 596–602. https://doi.org/10.1016/j.ejrad.2014.12.012
24. Ng K, Hill M, Johnston L, Highnam R, Tomal A: Large variation in mammography compression internationallyIn: European Congress of RadiologyVienna: European Society of Radiology; 2017.
25. Moshina N, Roman M, Waade GG, Sebuødegård S, Ursin G, Hofvind S. Breast compression parameters and mammographic density in the Norwegian breast cancer screening programme. *Eur Radiol* 2018; **28**: 1662–72. https://doi.org/10.1007/s00330-017-5104-5
26. 2011 Census Guidance and Methodology [https://www.ons.gov.uk/census/2011census/2011censusdata/2011censususerguide/variablesandclassifications]
27. The Royal College of Radiologists: Guidance on screening and symptomatic breast imaging. ., vol. BFCR(13)5,Third edition edn. London;2013.
28. Matakina Technology Ltd.: VolparaDensity™ User Manual Version 1.5.11. In.; 2014.
29. StataCorp: Stata Statistical Software: Release 14. College Station, TX: StataCorp LP. In., 14 edn; 2015.
30. Khan-Perez J, Harkness E, Mercer C, Bydder M, Sergeant J, Morris J, Maxwell A, Rylance C, Astley SM: Volumetric Breast Density and Radiographic Parameters. In: 2014; Cham: Springer International Publishing; 2014:265-272.
31. Waade GG, Holen Å, Sebuødegård S, Aase H, Pedersen K, Hanestad B, et al. Breast compression parameters among women screened with standard digital mammography and digital breast tomosynthesis in a randomized controlled trial. *Acta Radiol* 2020; **61**: 321–30. https://doi.org/10.1177/0284185119863989
32. Waade GG, Moshina N, Sebuødegård S, Hogg P, Hofvind S. Compression forces used in the norwegian breast cancer screening program. *Br J Radiol* 2017; **90**(1071). https://doi.org/10.1259/bjr.20160770
33. Kallenberg MGJ, van Gils CH, Lokate M, den Heeten GJ, Karssemeijer N. Effect of compression paddle tilt correction on volumetric breast density estimation. *Phys Med Biol* 2012; **57**: 5155–68. https://doi.org/10.1088/0031-9155/57/16/5155
34. Brand JS, Czene K, Shepherd JA, Leifland K, Heddson B, Sundbom A, et al. Automated measurement of volumetric mammographic density: A tool for widespread breast cancer risk assessment. *Cancer Epidemiol Biomarkers Prev* 2014; **23**: 1764–72. https://doi.org/10.1158/1055-9965.EPI-13-1219
35. Alonzo-Proulx O, Mawdsley GE, Patrie JT, Yaffe MJ, Harvey JA. Reliability of automated breast density measurements. *Radiology* 2015; **275**: 366–76. https://doi.org/10.1148/radiol.15141686
36. Holland K, van Zelst J, den Heeten GJ, Imhof-Tas M, Mann RM, van Gils CH, et al. Consistency of breast density categories in serial screening mammograms: A comparison between automated and human assessment. *Breast* 2016; **29**: 49–54. https://doi.org/10.1016/j.breast.2016.06.020
37. Highnam R: Patient-Specific Radiation Dose Estimation in Breast Cancer Screening. In: Volpara White Papers. New Zealand;2014.
38. Holland K, Sechopoulos I, Den Heeten GJ, Mann RM, Karssemeijer N: Performance of Breast Cancer Screening Depends on Mammographic CompressionIn: Breast Imaging: 13th International Workshop, IWDM 2016Malmö, Sweden, Jun e 19-22, 2016, Proceedings Springer International Publishing; 2016:-183-189.
39. Hogg P, Taylor M, Szczepura K, Mercer C, Denton E. Pressure and breast thickness in mammography -- an exploratory calibration study. *Br J Radiol* 2013; **86**(1021): 20120222. https://doi.org/10.1259/bjr.20120222
40. de Groot JE, Broeders MJM, Branderhorst W, den Heeten GJ, Grimbergen CA. A novel approach to mammographic breast compression: Improved standardization and reduced discomfort by controlling pressure instead of force. *Med Phys* 2013; **40**(8): 081901. https://doi.org/10.1118/1.4812418
41. Serwan E, Matthews D, Davies J, Chau M. Mammographic compression practices of force- and pressure-standardisation protocol: a scoping review. *J Med Radiat Sci* 2020; **67**: 233–42. https://doi.org/10.1002/jmrs.400



Paper IV Supplementary Figure 1 Tree of exclusions from study

Paper IV Supplementary Table 1 Characteristics associated with screening examinations in non-standard image sets

	Frequency	Percent %
All non-standard screening examinations [a]	10,212	
No of images taken		
<2	50	0.49%
2	6,299	61.68%
3	362	3.54%
5	40	0.39%
6	652	6.38%
7	9	0.09%
8	2,396	23.46%
>8	404	3.96%
Age at screening		
<45-	81	0.79%
45-49	404	3.96%
50-54	2,758	27.01%
55-59	2,464	24.13%
60-64	2,051	20.08%
65-69	1,944	19.04%
70+	509	4.98%
Missing	1	0.01%
Ethnicity (of subject screened[b])		
White – British or Irish or other	6,455	63.21%
Asian – British Indian or Pakistani or Bangladeshi or other	643	6.30%
Black – British or Caribbean or other	892	8.73%
Black – African	189	1.85%
Mixed White and Black, White and Asian or any other mixed	119	1.17%
Chinese	319	3.12%
Missing or not reported	1,595	15.62%
Breast Volumetric measurements[c]	**Median**	**IQR**
Median Breast volume, cm^3	758	479-1,145
Median Dense volume cm^3	49.9	36.7-69.8
Median %Mammographic Density	6.4%	4.5%-10.2%
Imaging acquisition parameters average across MLO and CC views[c]	**Mean**	**SD**
Mean compression force applied, N	8.26	2.11
Mean paddle tilt angle, degrees positive from horizontal [d]	2.52	0.99
Mean pressure, kPa	8.48	3.52
Imaging outcome estimates average across MLO and CC views		
Manufacturers mean glandular dose, mGy [e]	1.33	0.36
Mean breast thickness, mm	55.8	12.4

Footnotes:

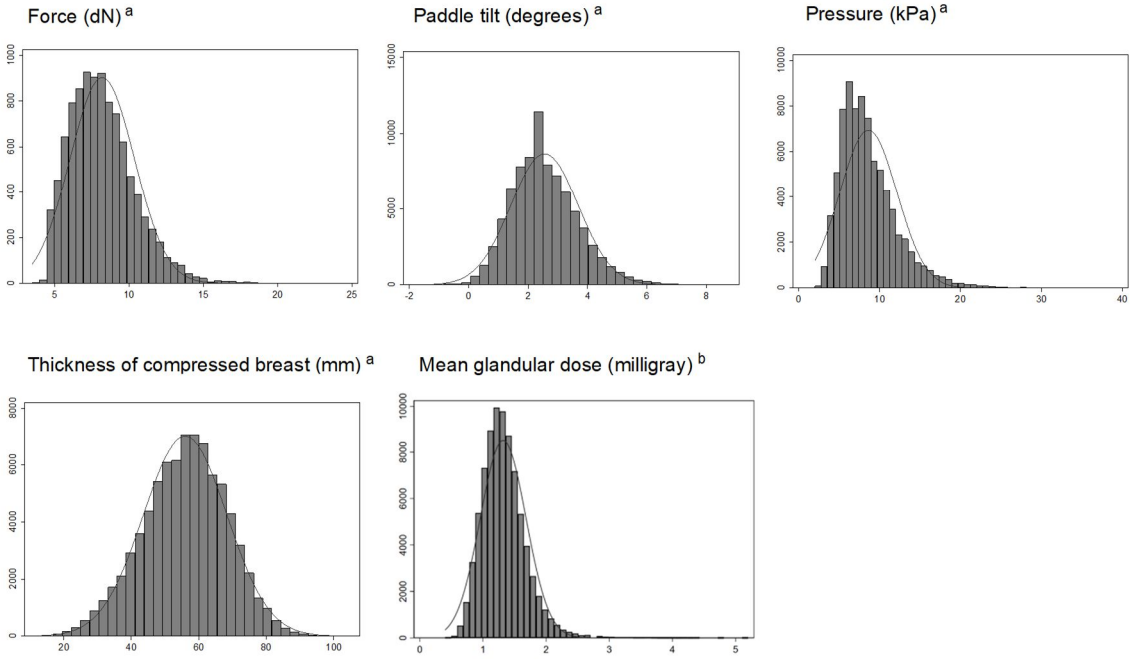
[a] A non-standard screening examination had <4 or >4 images taken, only screening appointments are included, excluded 22 images taken on non-Hologic systems and 648 screens excluded where women were known to have previous cancer.

[b] Count for each screening examination (subjects may have more than one examination over the study period).

[c] Calculated from the average value from the images available. Where > 4 images taken each image may only include part of the breast and the automated estimating algorithm is not able to make reliable overall volumetric estimates in these conditions.

[d] Mean paddle tilt from horizontal (where paddle tilt >=0)

[e] Manufacturers mean glandular dose as recorded in DICOM header



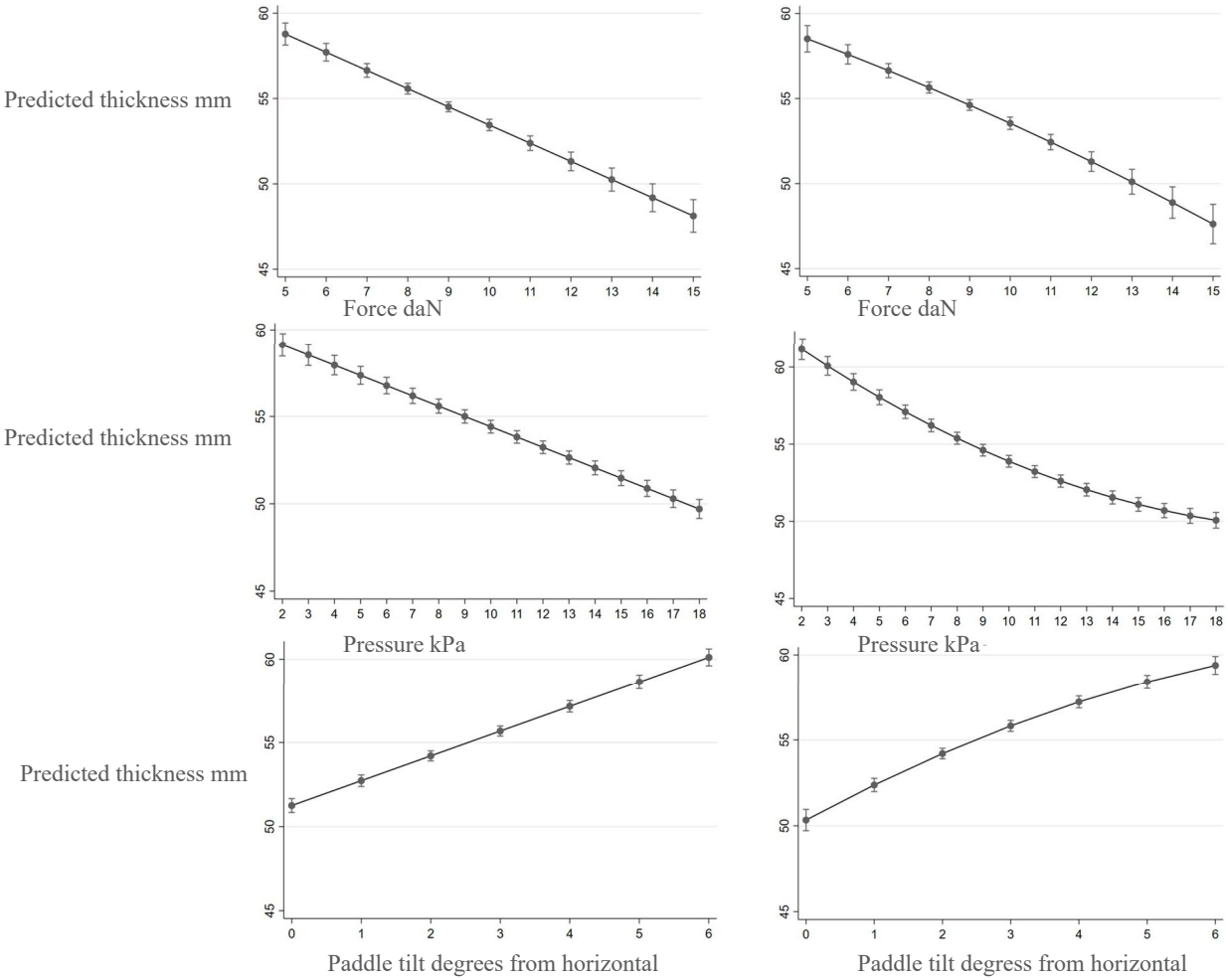
Paper IV Supplementary Figure 2 Distribution of mammogram acquisition and outcome measurements

Footnotes :
[a] Derived from average of 4 mammographic views Left breast CC, Right breast CC, Left breast MLO, Right breast MLO
[b] The mean glandular dose estimation in mGy as provided by the manufacturer of the mammography equipment. Derived from average of 4 mammographic views Left breast CC, Right breast CC, Left breast MLO, Right breast MLO. Represents the estimated average absorbed radiation dose per unit of glandular tissue.

Paper IV Supplementary Table 2 Spearman's rank correlation coefficients (rho) between compression parameters and breast characteristics

Variables	Thickness	MGD	Age	BV	%MD	Force	Pressure	Tilt
Thickness	1.00							
Mean glandular dose (MGD)	0.73	1.00						
Age	-0.09	-0.27	1.00					
Breast Volume (BV)	0.83	0.56	0.01	1.00				
Mammographic density (%MD)	-0.63	-0.10	-0.19	-0.70	1.00			
Force	0.14	0.28	0.02	0.37	-0.26	1.00		
Pressure	-0.44	-0.12	-0.09	-0.73	0.48	0.07	1.00	
Paddle tilt	-0.32	-0.19	0.03	-0.52	0.21	-0.06	0.57	1.00

p<0.0001 in all cases



Paper IV Supplementary Figure 3 Estimated exposure response curves[1] for compression thickness in mutually adjusted[2,3] linear and quadratic models

Footnotes:
[1] Plots show predicted margins with 95% CIs.
[2] Adjusted for Age, %mammographic density, Breast Volume (omitted in pressure models due to collinearity).
[3] Pressure models additionally adjusted for tilt. Force model additionally adjusted for tilt and vice versa.

7.4 Further descriptive analyses

Introduction to supplementary descriptive analyses

This section describes additional descriptive analyses that were carried out as an adjunct to the published Paper IV, using the same study population, data set and exposure and outcome measurements.

Variation of imaging technique in different practitioners of different experience levels

The literature review of technical imaging techniques (Chapter 2, section 2.9) concluded that there is poor consistency (in measured compression force and thickness) between mammographers irrespective of experience and screening setting. In the NHSBSP mammographers undergo comprehensive training and their images are reviewed. The convention for assessing images during the period of my research studies (prior to December 2017) was based on rating a sample of screening images on the PGMI (Perfect, Good, Moderate, Inadequate) but despite a number of more obvious image flaws, such as incomplete pectoralis muscle the scoring system is naturally subjective and inter-rater agreement tends to be poor (168). Therefore, it is possible that there will be variations in practice and outcomes between individual practitioners within any given screening service and these variations may or may not be related to their degree of experience. The first objective of this supplementary analysis was therefore to describe the variation in image acquisition parameters (force pressure and paddle tilt) and objective compression outcomes (thickness and dose) in mammographers of different levels of experience using the same sample of ~80,000 images described in Paper IV taken by over 80 different practitioners of varying levels of experience.

The practitioner (mammographer, advanced practitioner or assistant practitioner) responsible for each screening examination was retrieved from the NBSS breast screening administrative system. The superintendent mammographer allocated each practitioner to one of 3 groups depending upon their years of experience at the time of their first recorded examination (Low <3 years, Medium 3 to <7 years and High 7+ years). Because SWLBSS hosts a national mammography training centre, a number of visiting practitioners (n=49) took a small number of images; their level of experience was unobtainable, therefore only practitioners who were responsible for 700 or more studies (n=39) during the period of the study were categorised into the experience groups. Training examinations, where the practitioner was supervised, were excluded from the analysis.

Medians, 25th and 75th centiles of the force and paddle tilt and thickness, dose and pressure distributions were calculated and plotted by practitioner. Similarly, for each of the three levels of experience, violin plots were created to show the medians, 25th and 75th centiles and the distribution for each of the imaging and outcome variables. The corresponding means and SD for each

practitioner were also calculated and the means and SDs for each practitioner experience group. Analysis of variance (ANOVA) was used to test whether the average imaging parameters (force pressure and paddle tilt) and outcome metrics (thickness and dose) varied between practitioners and across the practitioner experience groups.

Findings - Differences between practitioners

87 different practitioners carried out examinations during the study. 71,099 examinations (88.3%) were carried out by practitioners who individually carried out at least 700 examinations during the study period. 8,957 (11.1%) were taken by practitioners who had taken fewer than 700 examinations (or were training during the period). A further 439 examinations (0.6%) were taken by unidentified practitioners.

Table 7.1 Examination volumes and experience level of practitioners

	Frequency	Percent %
All standard screening events[a]	80,495	
Practitioner characteristics		
Studies taken by practitioners taking over 700 studies in period (n=39)	71,099	88.33%
Studies taken by practitioners taking less than 700 studies in period (n=48)	8,957	11.13%
Studies taken by unidentified practitioner	439	0.55%
Experience of practitioners [b]		
<3 years (Low) (n=9)	13,943	17.32%
3-6 years (Medium) (n=9)	16,663	20.70%
7+ years (High) (n=21)	40,493	50.30%
Unidentified practitioner or <700 images in period	9,396	11.67%

Footnotes:

[a] Examination excluded if: not exactly 4 images taken, not a routine screening exam, images taken on non-Hologic systems and exams where women were known to have previous cancer.

[b] Practitioners only included individually if they have taken over 700 images (median 1,335; range 712 to 4210) that were analysed by Volpara software in the study period March 2013 to June 2017, training mammograms are not included.

There was variation between practitioners in the exposure variables (force, pressure and paddle tilt) employed during the compression process and the compression outcome measures (thickness and dose) (Figure 7.1).

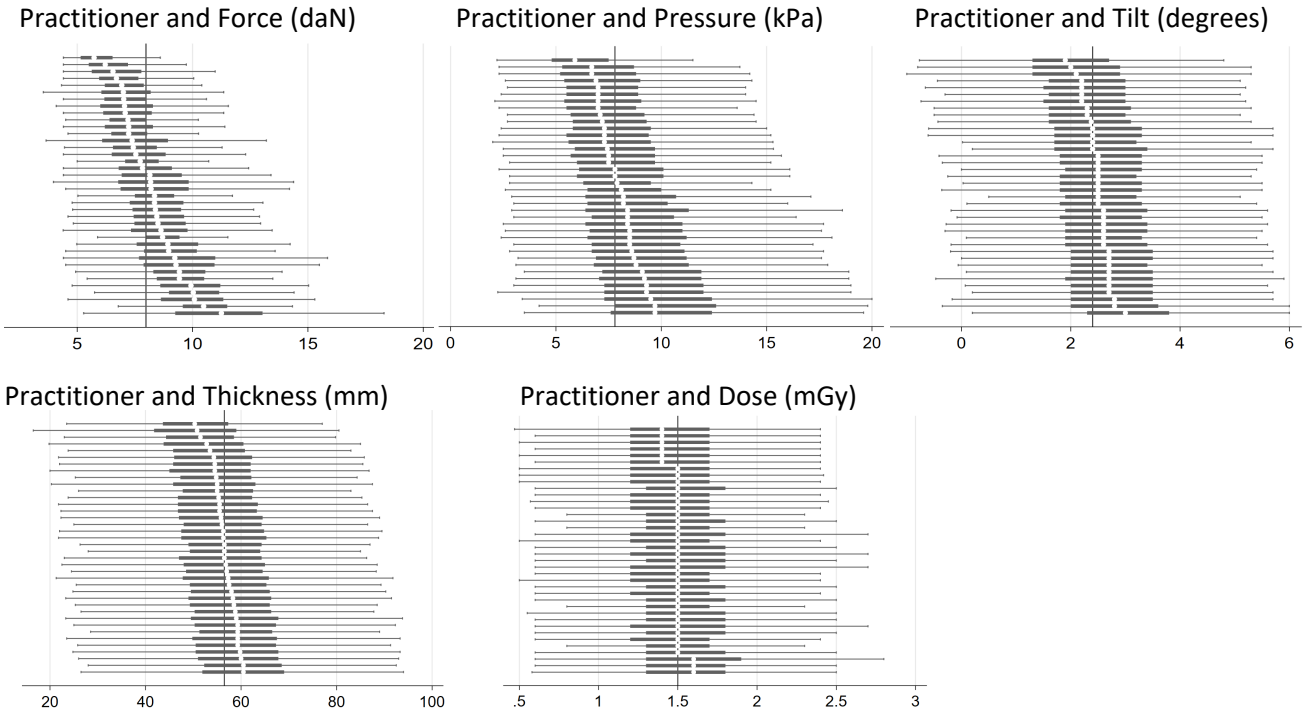


Figure 7.1 Median compression technique measurements and outcome estimates[a,b] by practitioner [c]

Footnotes :

[a]All measurements derived from average of 4 mammographic views Left breast CC, Right breast CC, Left breast MLO, Right breast MLO. Median value for all images shown as vertical line.

[b] Manufacturer's dose calculation average over 4 images.

[c] Practitioners listed in order of increasing median value for measure being shown, excludes practitioners taking less than 700 image sets during the study period

There was significant variation between the practitioners (ANOVA $p<0.001$ for all three acquisition parameters). The mean force for an individual practitioner (who had completed at least 700 examinations) ranged from ~6.8daN in the practitioner employing least force to 11.4daN in the practitioner employing most force, an increase of 67%. The mean compression pressure for an individual practitioner ranged from 7.29 kPa to 10.59 kPa, an increase of ~45%. The mean paddle tilt for an individual practitioner ranged from 2.29 ^{o} to 3.15^{o}, an increase of ~38%, (data not shown).

There was also significant variation between the practitioners in compression outcomes (thickness and dose) (ANOVA $p<0.001$ for both outcomes). The mean compression thickness for an individual practitioner ranged from 50.14mm in the practitioner with lowest compression thickness to 59.26mm in the practitioner with highest thickness, an increase of ~18%, (data not shown). Violin plots of distributions by practitioner experience show that variations between experience groups are small and the distributions are similar (Fig 7.2).

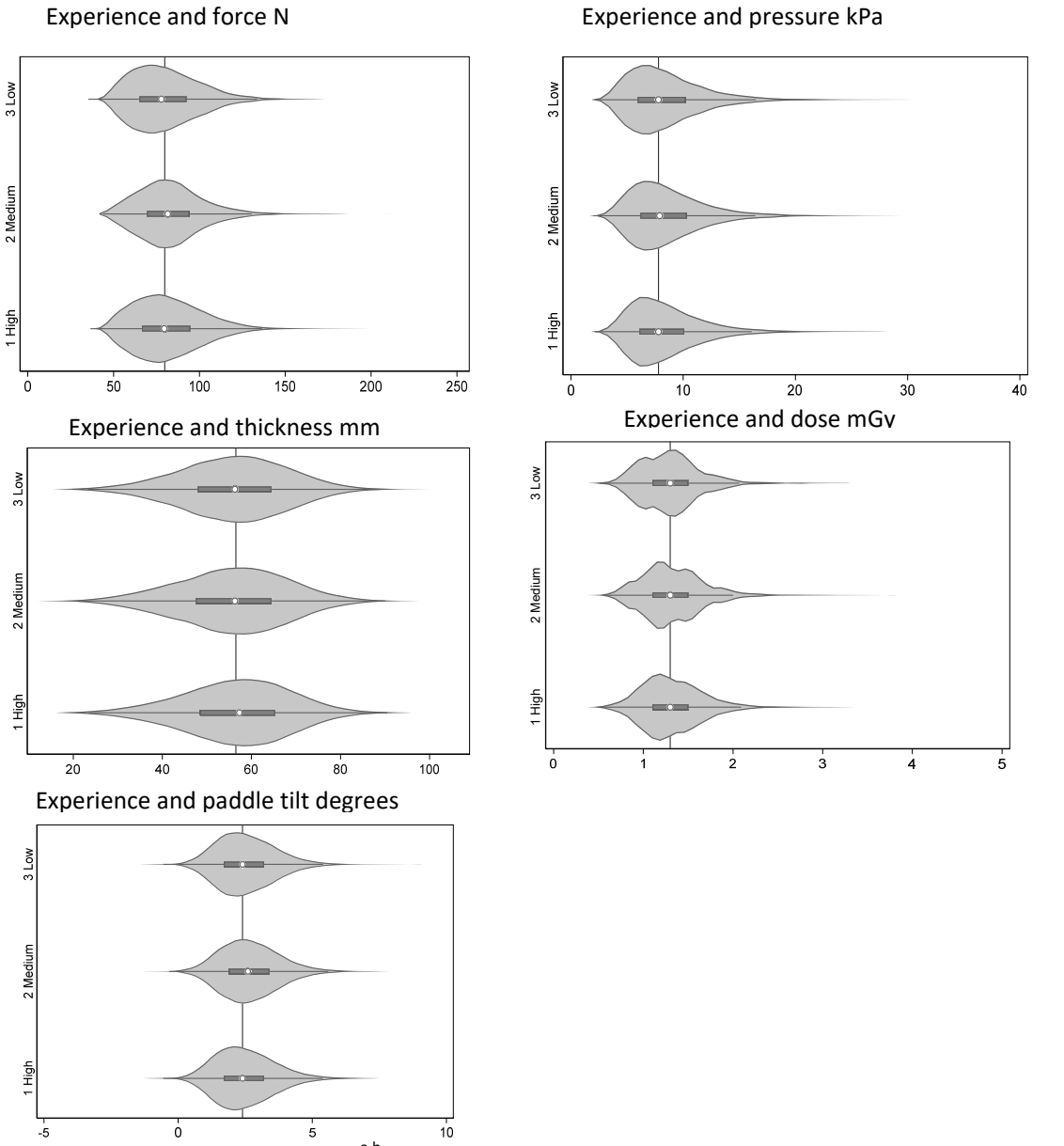


Figure 7.2 Distribution of imaging[a,b] and compression outcomes across practitioner experience groups

Footnotes : [a]All measurements derived from average of 4 mammographic views Left breast CC, Right breast CC, Left breast MLO, Right breast MLO. Median value for all images shown as vertical line.[b] Manufacturer's dose calculation average over 4 images.

Analyses of variance (ANOVA) show that the variations in mean force, pressure and paddle tilt between the experience groups are significant but not large, likewise the variations in mean compression outcomes (mean thickness and dose) between groups of different experience are significant but not large (Table 7.2). Regression models that used practitioners' experience as an exposure category found no significant differences between the categories for any of the compression outcomes (thickness, dose and pressure) after full adjustment (for age, BV, %MD and ethnicity) (data not shown).

Table 7.2 Mean (95% CI)[a] observed compression force, pressure, paddle tilt, and resulting thickness and dose grouped by practitioners level of experience

	High Experience (7+ years) (n=40,493 examinations) [b]	Medium Experience (3-6 years) (n=16,663 examinations) [b]		Low Experience (0-2 years) (n=13,943 examinations) [b]	
Imaging parameters [a]	**Mean (95% CI)**	**Mean (95% CI)**	**Difference in means (%)** [d]	**Mean (95% CI)**	**Difference in means (%)** [d]
Force, daN	8.25 (8.23, 8.27)	8.50 (8.47, 8.53)	+0.25 (+3.01%)	8.03 (8.00, 8.06)	-0.22 (-2.69%)
Paddle tilt, degrees	2.66 (2.65, 2.67)	2.79 (2.77, 2.80)	+0.13 (+4.90%)	2.66 (2.65, 2.68)	+0.00 (0.00%)
Pressure, kPa	8.57 (8.53, 8.60)	8.75 (8.70, 8.80)	+0.18 (+2.10%)	8.49 (8.43, 8.54)	-0.08 (-0.93%)
Compression outcomes[a]					
Thickness, mm	56.34 (56.23, 56.46)	55.22 (55.05, 55.39)	-1.12 (-1.99%)	55.90 (55.70, 56.10)	-0.44 (-0.78%)
Dose, mGy [c]	1.52 (1.52, 1.52)	1.54 (1.53, 1.54)	+0.02 (+1.32%)	1.50 (1.50, 1.51)	-0.02 (-1.32%)

Footnotes

[a] Mean values per examination are calculated using all available images from the relevant view (MLO and CC from each side).

[b] For a screening examination to be included it must have with exactly 4 images taken, only screening appointments are included, excluded images taken on non-Hologic systems and screens excluded where women were known to have previous cancer. Observations where the practitioner took <700 images per year were excluded (8,957) and in addition, observations where the identification of the practitioner was not known were excluded (n=439 examinations)

[c] Volpara calculated mean glandular dose in milligray

[d] Difference between mean value for the Highly experienced group versus the Medium/Low group means. The % difference is the difference/high experience mean* 100

Ethnicity and image acquisition parameters

Ng et al analysed images from 17 different counties and concluded that beyond breast compression behaviour, age, and breast composition (BV and %MD) there are other factors influencing compression (182). The NHSBSP invites over 3 million women a year from a wide range of different ethnicities and therefore it is possible that there are variations in imaging parameters that are dependent upon the screening client's ethnicity. The objective of this analysis was therefore to describe the variation in the image acquisition parameters (force pressure and paddle tilt) and measured compression outcomes (thickness and dose) between women of different ethnic groups in the same study population.

Findings - Ethnicity and image acquisition parameters

Figure 7.3 shows that subjects of Chinese and Asian ancestry experienced higher pressure during compression than White and Black subjects with the same BV. Chinese subjects experienced on average 2.6 kPa greater pressure than White subjects with the same age BV and %MD (data not shown). Despite greater pressure, Chinese subjects had greater compression thickness and dose than other ethnic groups for subjects with the same BV. Greater paddle tilt was used for Asian and Chinese subjects in the lower BV categories. Previous published research (on a subset of these

women) also found that women of Chinese ethnicity had greater %MD than other ethnicities of a similar age (238). Asian subjects also experienced higher pressure during compression, (on average 0.34 kPa higher pressure than Whites after further adjustment for BV, age and %MD (data not shown)).

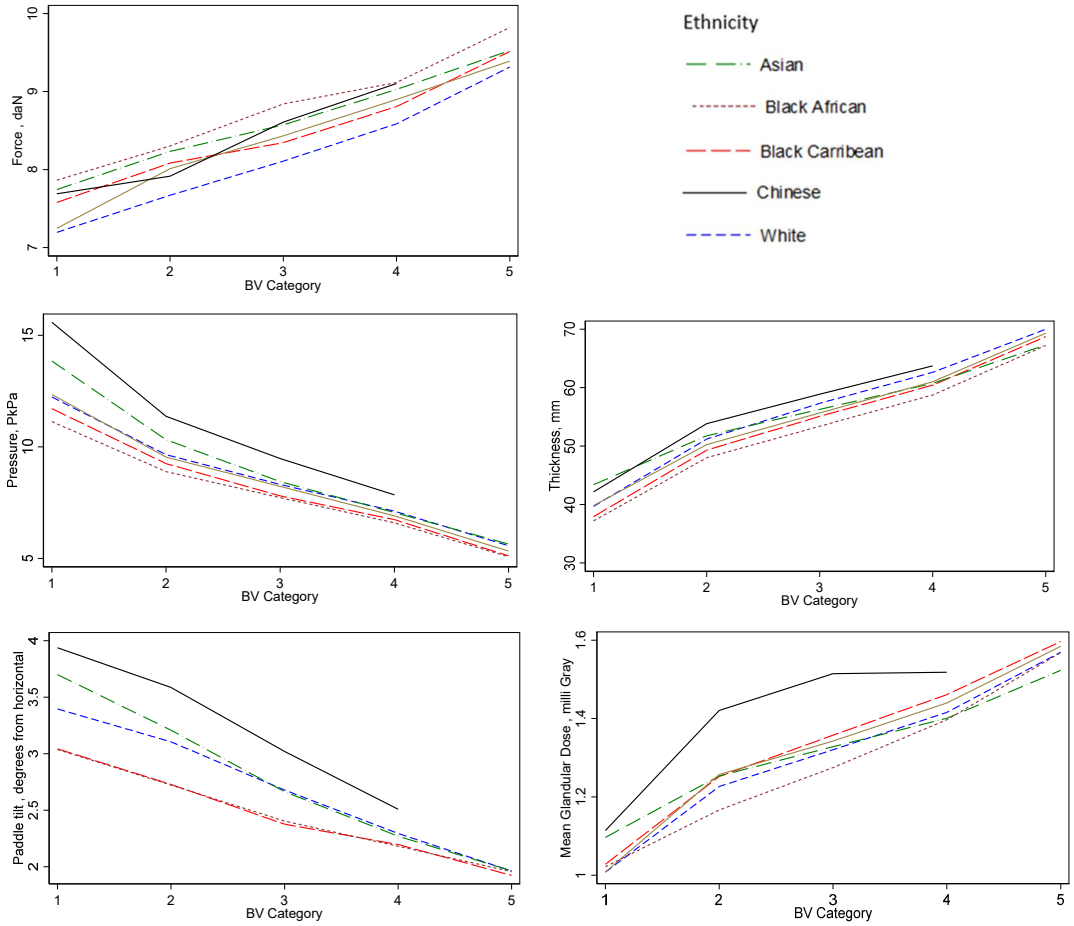


Figure 7.3 Mean force, pressure, thickness, tilt, and dose, by breast volume category plotted by ethnic groups (lowess smoothing)

[b] Data on ethnicity were collected as part of standard screening protocol via a self-completed screening questionnaire and recorded according to the Census classification and summarised as, “Asian” (Indian, Pakistani or Bangladeshi or other), “Black-African”, “Black-British or Caribbean or other”, “Chinese”, “Mixed” (White and Black, White and Asian or any other mixed), “White” (British or Irish or other) and “Other”.

Conclusions - supplementary analyses

The supplementary descriptive analyses found considerable variation in imaging technique between practitioners, as measured by compression force, pressure and paddle tilt choices. For example, force was almost twice as high for the practitioner using most force in comparison to the practitioner using lowest mean force. The variation between practitioners was not strongly related to their degree of experience.

Ethnicity of the screening subject may play a role in the imaging process; subjects of Chinese ancestry experienced higher pressure during compression but unexpectedly had greater compression thickness and dose than White subjects with the same BV. This finding could also be a result of the greater degree of paddle tilt used for subjects of Chinese ethnicity. Asian subjects also experienced higher pressure during compression than Whites after adjustment for BV age and %BD therefore it is possible that differences may be associated with some ethnic differences in anatomy and in the stretching and flattening properties of breast tissue, which results in relatively low detector plate contact area.

7.5 Paper V Are mammography image acquisition factors, compression and paddle tilt, associated with breast cancer detection in screening?(239)

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	12406683	**Title**	Ms
First Name(s)	Susan M		
Surname/Family Name	Hudson		
Thesis Title	Beyond breast density – Novel uses of automated mammographic analysis in breast cancer screening		
Primary Supervisor	Professor Isabel dos Santos Silva		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	The British Journal of Radiology		
When was the work published?	2023		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	n/a		
Have you retained the copyright for the work?*	**No**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I was responsible for the conceptualisation of the work, development of the methodology, setting up of the mammogram analysis tools in conjunction with Dr Wilkinson and data collection, cleansing and linking of data sources, data analysis and writing of the publication.

SECTION E

Student Signature	[redacted]
Date	

Supervisor Signature	
Date	

BJR https://doi.org/10.1259/bjr.20230085

Received: 25 January 2023 | Revised: 19 April 2023 | Accepted: 28 April 2023

Cite this article as:
Hudson SM, Wilkinson LS, De Stavola BL, dos-Santos-Silva I. Are mammography image acquisition factors, compression pressure and paddle tilt, associated with breast cancer detection in screening?. *Br J Radiol* (2023) 10.1259/bjr.20230085.

FULL PAPER

Are mammography image acquisition factors, compression pressure and paddle tilt, associated with breast cancer detection in screening?

[1]SUE M HUDSON, BSc, MSc, [2]LOUISE S WILKINSON, BA, BM, BCh, FRCR, [3]BIANCA L DE STAVOLA, PhD and [1]ISABEL DOS-SANTOS-SILVA, MD, MSc, PhD

[1]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom
[2]Oxford Breast Imaging Centre, Churchill Hospital,Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom
[3]Faculty of Pop Health Sciences, Institute of Child Health, University College London, London, United Kingdom

Address correspondence to: Sue M Hudson
E-mail: *susan.hudson@lshtm.ac.uk;sue.hudson@pasconsulting.co.uk*

Objectives: To assess the associations between objectively measured mammographic compression pressure and paddle tilt and breast cancer (BC) detected at the same ("contemporaneous") screen, subsequent screens, or in-between screens (interval cancers).
Methods: Automated pressure and paddle tilt estimates were derived for 80,495 mammographic examinations in a UK population-based screening programme. Adjusted logistic regression models were fitted to estimate the associations of compression parameters with BC detected at contemporaneous screen (777 cases). Nested case-control designs were used to estimate associations of pressure and tilt with: (a) interval cancer (148 cases/625 age-matched controls) and (b) subsequent screen-detected cancer (344/1436), via conditional logistic regression.
Results: Compression pressure was negatively associated with odds of BC at contemporaneous screen (odds ratio (OR) for top versus bottom third of the pressure distribution: 0.74; 95% CI 0.60, 0.92; P-for-linear-trend (Pt) = 0.007). There was weak evidence that moderate pressure at screening was associated with lower odds of interval cancer (OR for middle versus bottom third: 0.63; 95% CI 0.38, 1.05; p = 0.079), but no association was found between pressure and the odds of BC at subsequent screen. There was no evidence that paddle tilt was associated with the odds of contemporaneous, subsequent screen or interval cancer detection.
Conclusions: Findings are consistent with compression pressure, but not paddle tilt, affecting the performance of mammographic screening by interfering with its ability to detect cancers.
Advances in knowledge: Inadequate or excessive compression pressure at screening may contribute to a reduced ability to detect cancers, resulting in a greater number of interval cancer cases.

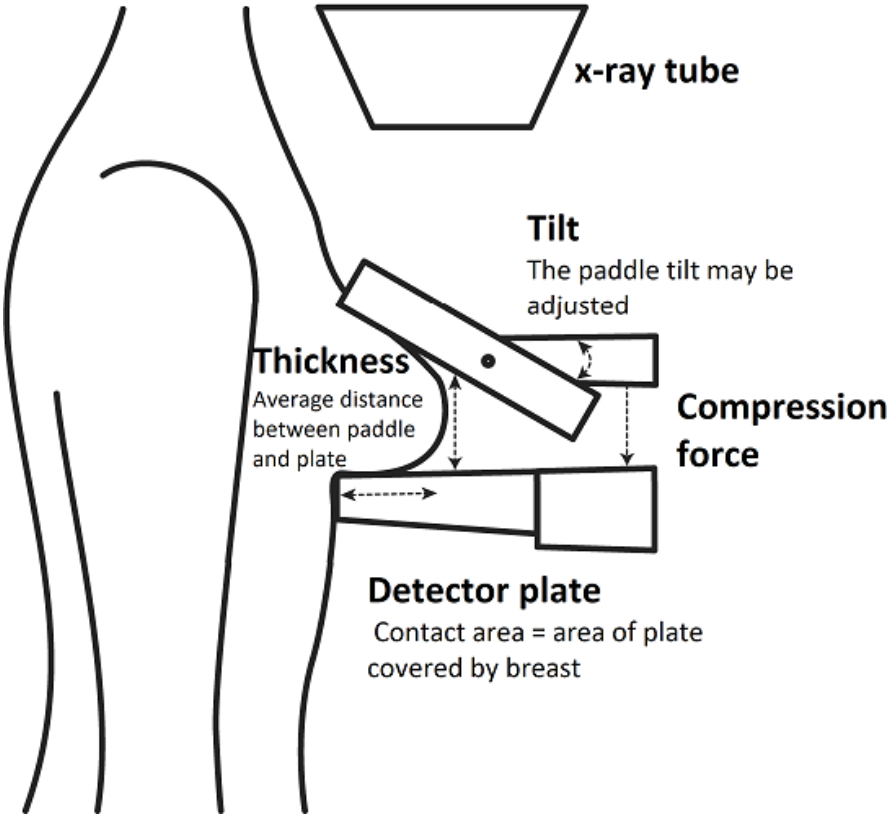
INTRODUCTION

Population-based mammographic screening programmes, such as the England and Wales Breast Screening Programme (NHSBSP), have been found to reduce mortality through detection of asymptomatic cases coupled with early treatment.[1] However, such programmes rely on the quality of the mammographic images to enable radiological detection of suspicious features in the breast.

Mammography involves compressing the breast between a detector plate and a transparent paddle such that the breast is immobilised, and the thickness of tissue minimised without causing unnecessary pain. The force applied to achieve this compression can be monitored by the practitioner. A tilting or hinging paddle may optionally be used to adjust the angle of the top paddle away from parallel to reflect the natural shape of the compressed breast (Figure 1).

Breast compression is thought to be a key factor in the production of high-quality images because it helps to reduce movement (blur), separate overlying tissues and also reduce thickness, thereby improving tumour conspicuity.[2,3] Furthermore compression reduces the absorbed radiation dose during the screening procedure.[4] Tilting paddles were introduced with the aim of reducing pain during mammography, but a previous study, conducted within the same study population as the present investigation found that increased paddle tilt was associated with increased compression thickness; therefore, it is possible that tilting paddle use also affects cancer conspicuity.[5]

Figure 1. Compression of the breast during CC image acquisition schematic



Internationally, a wide variation in objectively measurable imaging parameters has been observed.[6–9] In the UK, although regular image audits take place, objective guidelines on optimal breast compression are currently limited to the guidance that force should not exceed 20daN.[10,11] Our previous study on the same population found that compression pressure and paddle tilt are not systematically adjusted in accordance with objective breast characteristics and consequently there is inconsistency in technique and compression outcomes.[5]

A limited number of studies have used objective mammographic acquisition measures to show that screening performance is associated with the degree of compression force and pressure used during image acquisition.[12,13] However, little is known, as yet, about the association between paddle tilt and cancer detection in screening programmes.

BC risk is increased in females with denser breasts, with females in the densest category having 4.6 times the risk of females in the fatty category.[14] Furthermore, breast density affects the effectiveness of mammographic screening because fibro-glandular tissue, that makes up breast density, can mask cancers, resulting in reduced sensitivity and a higher risk of interval cancers in denser breasts.[15–18] Studies have shown that mammographic acquisition measures are correlated with breast density.[19]

The aim of this study is to investigate the association between image acquisition pressure and paddle tilt, and the risk of being diagnosed with BC at the same or subsequent screens, or as an interval cancer between screens, among a large population-based sample of 94,408 examinations taken amongst 68,776 women who underwent mammographic screening on one or more occasion in South-West London, England, between March 2013 and June 2017.

METHODS

Study participants

Study participants underwent routine 3-yearly screening mammography at the South-West London Breast Screening Service (SWLBSS) based in the St George's University Hospitals NHS Foundation Trust. SWLBSS is a part of the NHSBSP, an organised population-based mammographic screening programme, which targets females aged 50–70. We also included females aged 47–49 and 71–73 screened as part of a national trial[20] plus any females over 73 years who had contacted the service for a self-referred screening appointment. A small number of females who are invited to annual screening due to higher familial risk, were also included. Participants were screened during the period 01/03/2013 to 20/06/2017. The subject's age at the time of screening was recorded. A self-completed questionnaire is routinely used at SWLBSS to gather ethnicity data according to the Census classification[21] and these data were further categorised into, "Asian" (Indian, Pakistani or Bangladeshi or other), "Black-African", "Black-British or Caribbean or other", "Chinese", "Mixed" (White and Black, White and Asian or any other mixed), "White" (British or Irish or other) and "Other". The NHSBSP does not systematically record data on

other known BC risk factors and thus we were unable to collect data on factors such as reproductive history, body mass index (BMI), family-history of breast cancer and menopause hormone therapy usage.

Exposure assessment

Each female was screened according to the NHSBSP standard, 2-view (cranio-caudal (CC) and medio-lateral-oblique view (MLO)) mammography of each breast.[22] Raw digital mammographic images were processed using an automated algorithm, i.e. Volpara® Density™ version 1.5.11 (Volpara), (Matakina Technology Limited, Wellington, New Zealand),[23] which provided estimates (in cm^3) of the volume of the breast (BV) and the volume of the radio-dense tissue (DV) plus an estimate contact area (cm^2) between breast and detector plate. Estimates were provided separately for each of the four (left/right CC and MLO) images and as an average across all four images. The algorithm also yielded estimates, separately for each image and also as an overall average, of non-dense volume (NDV) as BV-DV and of % dense volume as the ratio of DV to BV expressed as a percentage. In addition, the algorithm provided a density grade (DG) score of 1 to 4, corresponding to the BI-RADS (Breast Imaging Reporting and Data System fourth Edition) classification for mammographic density *i.e.* A: almost entirely fatty, B: scattered areas of fibroglandular density, C: heterogeneously dense, and D: extremely dense.

The Digital Imaging and Communications in Medicine (DICOM) image header provided additional data on compression force (in decaNewton, daN) and compression paddle tilt (in degrees from horizontal). The resulting pressure (in kiloPascals, kPa) was calculated by the algorithm from force*10/contact area. The anonymised identifiers of the mammographer taking the image and the type of screen (first (prevalent) versus subsequent (incident) screens) were also recorded.

Examination eligibility

Screening examinations where exposure measurements (*i.e.,* pressure and paddle tilt) and outcome ascertainment (screen-detected cancer) were collected concurrently, were regarded as "contemporaneous screens for the purposes of this study. A total of 94,408 contemporaneous screening examinations took place during the study period. Examinations were excluded from the analysis if: (i) the reason for the examination was not known (*i.e.,* screening episode type was missing) (n = 992); (ii) the females had a previous BC (n = 2,068) because this might have influenced the physical nature and compressibility of the breast; (iii) examinations were performed using non-Hologic systems (n = 836) because of potential differences between manufacturers (iv) if more, or less, than the four standard images were taken, because the automated algorithm is not designed to make estimates where non-standard imaging sets are taken[23] (n = 10,234). Thus, a total of 80,495 examinations (321,980 compressions) were eligible for inclusion in the analyses.

Some examinations were on females who were screened more than once in the study period; 13,489 women had two examinations, 439 women had three examinations and 157 women had four examinations or more. All valid screening examinations were included in the analysis.

"Subsequent" screens were screens that took place at ~3 years after the contemporaneous screen *i.e.* at the next screening invitation following on from a contemporaneous screen. ~20% of subsequent screens were examinations that were taken in the period 2013 to 2017 and were therefore also eligible for inclusion in the contemporaneous screen analysis.

Cancer ascertainment

For the purpose of this study "contemporaneous" screen detected cancers were classified as cancers detected at the same time that the compression exposures (*i.e.,* pressure and tilt) were estimated (Figure 2). "Interval cancers" were those diagnosed symptomatically in the 3-year period following the initial screen and exposure measurement but before the next screening invitation. Any cancers detected at the subsequent screen, were classified as "subsequent cancers". All screens in this study were double read. A third, arbitration read, was conducted and a consensus agreed for all abnormal reads.

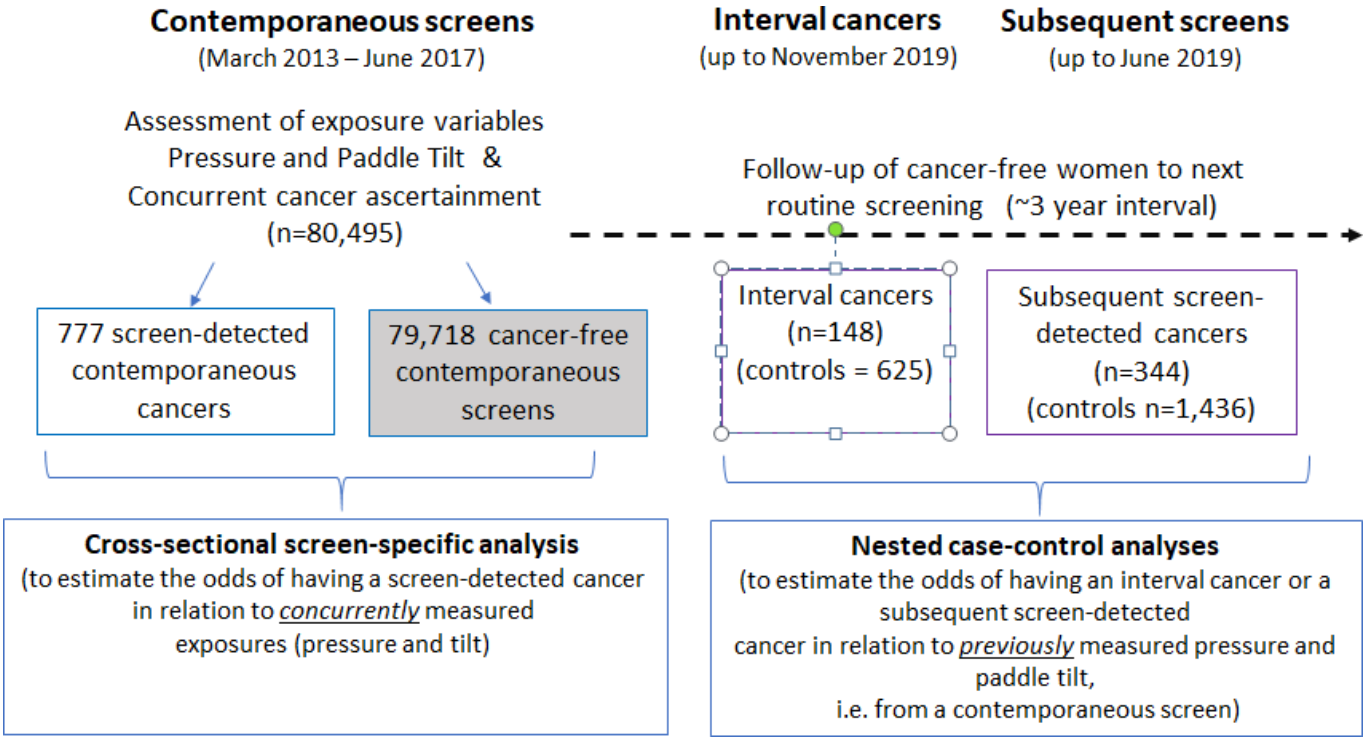
Screen detected cancers (contemporaneous or subsequent screen) were routinely recorded by the SWLBSS at the time of the relevant screening. Interval cancer cases were notified to SWLBSS through sharing of data between the Screening Quality Assurance Service and Cancer Registries and via direct contact between the screening services and local treating NHS Trusts and then recorded in the SWLBSS screening database. We included all subsequent screens up to June 2019 and all recorded interval cancers from the SWLBSS database as of 6/11/2019.

Study design

A cross-sectional screen-specific design was used to examine associations between the pressure and tilt used in the mammography examination and contemporaneous screen-detected cancers (Figure 2). Examinations at which females were diagnosed with BC (n = 777) were defined as cases, and examinations where no cancer was detected (n = 79,718) as non-cases.

An incident-density-sampling (nested) case-control design was used to investigate the association between mammography technique (pressure and tilt) and interval cancers (Figure 2). Cases were examinations where females were diagnosed with an interval cancer after a previous negative contemporaneous screen. Up to five matching controls were randomly selected for each case from females of the same age (± 1 year) who had a contemporaneous screen in the same year and month as the case with a verified 'non-cancer' status at the time that the case was diagnosed (based on subsequent screening records). For cases aged >73 years at contemporaneous screen controls were aged-matched within ± 5 years due to lack of qualifying controls. A total of 148 interval cancer cases and 625 matched controls were identified, corresponding to 86 cases with five controls each, 29 cases with four controls each, 20 cases with three controls

Figure 2. Timing of mammography and cancer diagnosis



each, 6 cases with two controls each and 7 cases with one control each; one case was excluded in the analysis because there were no valid matched controls.

A nested case-control approach was also used to assess the association between mammographic technique and risk of being diagnosed with a BC at a subsequent screen (Figure 2). This design was preferred to a cross-sectional analysis because at the time that the data was available, subsequent screens had only been performed for around 65% of study participants. Cases were examinations where females had a negative contemporaneous screen and no interval cancer diagnosis but were then diagnosed with breast cancer in the subsequent screening round (n = 344). Up to five age-matched controls per case were identified (a total of 1,436) using a similar approach to that outlined above for interval cancers.

Ethical approval

This retrospective study was carried out on fully anonymous, routinely collected data only, held in accordance with the National Health Service (NHS) Cancer Screening Programmes Confidentiality and Disclosure Policy 2011. The NHSBSP has section 251 support under the NHS Act 2006. The study was approved by all relevant ethics committees (Research Ethics Committees from St George's University Hospitals NHS Foundation Trust, and the London School of Hygiene and Tropical Medicine).

Statistical analyses

Mean compression pressure and mean paddle tilt used for an examination were calculated using all compressions from both views (MLO, CC) and each side; the distributions of these variables were approximately normal, and we further categorised them into thirds (low, medium and high) using as cut-off points the tertiles of the distributions in the non-cases/controls. The mean acquisition pressure and tilt for MLO and CC views separately were also calculated and thirds defined using tertiles as above.

Separate logistic regression models were used to examine the strength of the associations between the categorical exposures of interest, pressure and tilt, and the odds of being diagnosed with a contemporaneous screen-detected BC. Robust standard errors (clustering by female screened) were used to account for the fact that some females may have been screened more than once during the period. Similarly, separate conditional logistic regression models were used to examine the strength of the associations between pressure and tilt and the odds of an interval cancer and the odds of a subsequent screen-detected cancer.

All regression models were adjusted for a priori potential confounders: age at screening, ethnicity, DG (as estimated by the Volpara algorithmn) and additionally, in the tilt models only, for mammographic NDV (a valid proxy for BMI when data for the latter are not available[24]). NDV was not included as a potential confounder in the pressure model because of collinearity between pressure and NDV (data not shown). Linear trend tests for the association with the exposures of interest were carried out fitting models with the ordinal values of each categorical measure, assessing their significance using Wald tests. To allow comparison to other studies an alternative pressure model was fitted to replicate Moshina's Norwegian model,[12] adjusting for absolute DV rather than the relative density measure DG.

In all contemporaneous screening models we additionally adjusted for type of screen, (incident or prevalent) since a female undergoing her first (prevalent) screen is more likely to be recalled for additional tests and a higher cancer detection rate is normally observed.[25]

In all the analyses, we considered statistical significance (2-sided) at p-value < 0.05. All analyses were conducted in Stata (IC 14) [33].

RESULTS

Study participants

The characteristics of the participants, and of their screens, are shown in Table 1. The majority of the participants were White. The mean age, at contemporaneous screening, was 58.4 years in non-cases and 60.4 years in BC cases. Mean time between contemporaneous screen and interval cancer diagnosis was 19.2 (range 1.7–36.0; SD = 9.1) months. Mean time between contemporaneous screen and subsequent screen diagnosis was 36.4 (range 9.6–70.8; SD = 8.2) months by design, since the screening programme aims to invite females at ~36 monthly intervals

The median values for pressure and tilt were lower for contemporaneous cases (8.41 kPa and 2.59 degrees, respectively) than non-cases (8.65 kPa and 2.69 degrees, respectively; Table 1). In contrast median value for exposure pressure used in the original mammogram was higher for interval and subsequent screen cancer cases (8.54 kPa and 8.55 kpa, respectively) than their matched controls (8.28 kPa and 8.50 kpa, respectively; Table 1). This difference mainly reflects differences in pressure used during the CC compressions with smaller case-control differences observed in the MLO view (Table 1).

In each category (contemporaneous, interval and subsequent round screens) the DV, DG and NDV were higher in cases than non-cases/controls, (Table 1).

Associations between image acquisition pressure and tilt and contemporaneous screen-detected breast cancer

There was a negative association between compression pressure and the odds of being diagnosed with BC at the contemporaneous screen (Figure 3) (p-for-linear-trend (Pt) = 0.007). Relative to females in the bottom third of the pressure distribution (<6.7 kPa), those in the top third (>9.3 kPa) had 26% lower odds (OR 0.74; 95% CI 0.60, 0.92) of having a screen detected cancer in the fully-adjusted models (Figure 3 and Supplementary Table 1). There was a possible negative association between paddle tilt and odds of breast cancer detected at contemporaneous screen, but trends were non-significant (Pt = 0.119), (Figure 3). In all models age and breast density were strongly positively associated with increased risk of BC (Supplementary Tables 1 and 2).

Associations between image acquisition pressure and tilt and interval cancer

After adjustment for relative breast density, age and ethnicity, compression pressure was weakly negatively associated with the odds of having an interval cancer; females in the top third of the pressure distribution had odds, similar to, but somewhat lower than, those in the bottom third (adjusted OR 0.87; 95% CI 0.53, 1.43; Figure 3). However, females in the middle third had lowest odds of being diagnosed with an interval cancer relative to those in the lowest third of the pressure distribution (adjusted OR 0.63; 95% CI 0.38, 1.05; Figure 3). These results were however of borderline significance p = 0.079 (Supplementary Table 1). This association was stronger but also non-significant, in the CC compressions (adjusted OR 0.64; 95% CI 0.36, 1.14) than the MLO (adjusted OR 0.88; 95% CI 0.49, 1.56) compressions (Supplementary Table 3).

The odds of being diagnosed with an interval cancer were higher for greater degrees of paddle tilt but these estimates were very imprecise as reflected by the wider confidence intervals (Figure 3).

Associations between image acquisition pressure and tilt and a subsequent screen-detected cancer

There were no clear associations between pressure and the odds of being diagnosed with cancer at the subsequent screen (Figure 3). Nor were there associations between paddle tilt and the odds of having a cancer detected at the next screening round (Figure 3).

DISCUSSION

Main findings

Females who received the highest-pressure compressions were less likely to have a contemporaneous screen detected cancer. The findings for interval cancers show no clear trend but females in the middle third of the pressure distribution had lower odds of an interval cancer diagnosis than females in the lowest and highest pressure thirds of the distribution (but with borderline significance). We found no evidence of an association between pressure and the odds of a BC diagnosis at the subsequent routine screen. Increasing compression paddle tilt was not strongly associated with increasing odds of having an interval cancer or a subsequent routine-screen cancer in our study.

Our findings on pressure partly support those from a similar study by Holland et al who used the same computer algorithm and controlled for similar confounders, but used MLO views only, from over 100,000 women invited for screening in the Netherlands breast screening programme.[13] Mean BV was higher in the Dutch study than in our study (974 cm^3 and 850 cm^3 respectively) and average pressure for the MLO view was also higher than in our study (10.5 kPa and 7.4 kPa respectively). Holland et al found that screening sensitivity (based on interval cancers) was significantly lower in the highest pressure compression quintiles but higher in the middle pressure quintile of the distribution. We also found that odds of interval cancer were lowest in the middle pressure third of the pressure distribution. In our study the association between pressure and interval cancer was stronger for CC compressions than for MLO compressions. This may be related to the higher mean compression pressures that are used for CC views, which only include breast tissue and are not limited by inclusion of the pectoral muscle.

Table 1. Characteristics of the study participants, their mammographic examinations and cancers detected

	Contemporaneous screen-detected analysis		Subsequent interval cancer analysis		Subsequent screen-detected cancer analysis	
	Contemporaneous screen-detected cancer cases	Non-cancer at contemporaneous screen	Interval cancer cases	Controls	Subsequent screen-detected cancer cases	Controls
No. examinations[a]	*n* = 777	*n* = 79,718 a	*n* = 148	*n* = 625	*n* = 344	*n* = 1,436
Age at contemporaneous screening	**Mean (SD)**	**Mean (SD)** [b]	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
	60.35 (7.69)	58.40 (7.18)	59.19 (7.16)	59.59 (7.05)	59.90 (6.18)	59.74 (6.21)
Ethnicity	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**
White [c]	530 (68.2%)	51,931 (65.1%)	113 (76.4%)	405 (64.8%)	252 (73.3%)	965 (67.2%)
Asian [d]	75 (9.7%)	7,536 (9.5%)	5 (3.4%)	75 (12.0%)	32 (9.3%)	135 (9.4%)
Black – Caribbean [e]	25 (3.2%)	3,720 (4.7%)	9 (6.1%)	26 (4.2%)	10 (2.9%)	75 (5.2%)
Black – African	22 (2.8%)	2,703 (3.4%)	2 (1.4%)	20 (3.2%)	1 (0.3%)	53 (3.7%)
Mixed	11 (1.4%)	1,515 (1.9%)	3 (2.0%)	16 (2.6%)	6 (1.7%)	24 (1.7%)
Chinese	10 (1.3%)	1,052 (1.3%)	1 (0.7%)	3 (0.5%)	7 (2.0%)	24 (1.7%)
Missing/not reported	104 (13.4%)	11,261 (14.1%)	15 (10.1%)	80 (12.8%)	36 (10.5%)	160 (11.1%)
Type of screen	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**
Prevalent	212 (27.3%)	21,558 (27.0%)	36 (24.3%)	130 (20.8%)	61 (17.7%)	262 (18.2%)
Incident	565 (72.7%)	58,160 (73.0%)	112 (75.7%)	495 (79.2%)	283 (82.3%)	1,174 (81.8%)
Imaging parameters average for MLO and CC views [g] Mean pressure, kPa Mean positive paddle tilt, degrees	**Mean (SD)** 8.41 (3.47) 2.59 (1.01)	**Mean (SD)** 8.65 (3.54) 2.69 (1.06)	**Mean (SD)** 8.54 (3.29) 2.47 (1.01)	**Mean (SD)** 8.28 (3.16) 2.59 (1.01)	**Mean (SD)** 8.55 (3.28) 2.53 (0.96)	**Mean (SD)** 8.50 (3.33) 2.62 (1.04)
Imaging parameters for MLO [g] Mean pressure, kPa Mean positive paddle tilt, degrees	7.24 (2.43) 2.77 (1.23)	7.36 (2.51) 2.87 (1.28)	7.24 (2.35) 2.59 (1.25)	7.24 (2.48) 2.79 (1.21)	7.45 (2.37) 2.68 (1.16)	7.31 (2.38) 2.80 (1.26)
Imaging parameters for CC [g] Mean pressure, kPa Mean positive paddle tilt, degrees	9.63 (4.93) 2.34 (1.12)	9.98 (4.96) 2.43 (1.16)	9.95 (4.72) 2.28 (1.10)	9.44 (4.38) 2.32 (1.10)	9.92 (4.74) 2.29 (1.07)	9.82 (4.79) 2.35 (1.16)

(Continued)

Table 1. (Continued)

	Contemporaneous screen-detected analysis		Subsequent interval cancer analysis		Subsequent screen-detected cancer analysis	
Breast Volumetric measurements[h]	**Median**	**Median**	**Median**	**Median**	**Median**	**Median**
Breast dense volume,[i] cm^3	51.1 (39.9–69.6)	49.3 (37.2–67.3)	57.1 (43.0–89.4)	48.1 (36.0–65.5)	50.3 (43.0–89.4)	47.7 (36.0–65.1)
Breast non-dense volume, cm^3	725.7 (462.3–1075.5)	698.8 (438.0–1046.4)	838.7 (386.0–1144.3)	728.7 (484.1–1074.5)	727.65 (447.5–1085.4)	697.8 (437.5–1048.3)
Volpara Density Grade (n, %)[j]	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**	**n (%)**
Category 1	132 (16.99%)	17,512 (21.97%)	20 (13.51%)	146 (23.36%)	70 (20.35%)	356 (24.79%)
Category 2	303 (39.00%)	29,771 (37.35%)	47 (31.76%)	258 (41.28%)	122 (35.47%)	531 (36.98%)
Category 3	278 (35.78%)	23,270 (29.19%)	55 (37.16%)	169 (27.04%)	121 (35.17%)	418 (29.11%)
Category 4	64 (8.24%)	9,165 (11.50%)	26 (17.57%)	52 (8.32%)	31 (9.01%)	131 (9.12%)

SD, standard deviation ; n/a, not available.
[a]14,085 of the screens were on females who had two or more contemporaneous screen. Each examination is included independently.
[b]Where females have more than one contemporaneous screen the average of their age at screen is taken.
[c]White includes: British/Irish and other
[d]Asian includes: British Indian, Pakistani, Bangladeshi and other
[e]Black includes: British, Caribbean and other (non-African)
[f]Mean values are calculated using all available images from the relevant view (MLO and CC from each breast side) at a contemporaneous screen
[g]Mean and Median values are calculated using all available images (MLO and CC from each side) at a contemporaneous screen. Median (with 25th and 75th centile values) are shown.
[h]Screens included must have exactly four images taken, only screening images are included. Screens for females known to have previous breast cancer were also excluded.
[i]Mean (SD) Breast Volume for all contemporaneous screens was 850.1cm^3 (489.6), median 753.2 cm^3 (488.8–1110.1). Data included for comparative purposes with other studies.
[j]Density grade as calculated by the Volpara algorithm using relative % of dense and non-dense areas on the image. Estimates correspond to BIRADS fourth Edition. Images taken at the contemporaneous screen

A similar Norwegian study by Moshina et al using pressure estimates based on averaged MLO and CC views, yielded by the same algorithm (~339 interval cases; ~83,000 non-cases), found that compression pressure was positively associated with interval cancer.[12] The Norwegian screening programme participants were of similar average age as the participants in our study, but their median breast volume was somewhat greater (814.7 cm^3 and 776.8 cm^3 respectively). A key difference in the Norwegian study was that it controlled for absolute breast DV whilst, like Holland et al,[13] we adjusted for a relative measure of density to reflect breast composition and compressibility. When we replicated the Norwegian model by adjusting for DV rather than DG we found the adjusted ORs (high pressure third versus low pressure third) at interval cancer to be rather similar (OR 1.86 (95% CI 1.41, 2.45) in the Norwegian study, versus OR 2.03 (95% CI 1.21, 3.37) our study (Supplementary Table 4). Controlling for an absolute measure of breast dense volume, as in the Norwegian model, increased the magnitude of our findings, possibly because in our study population compression force was not altered adequately for breast size during mammography and hence smaller, denser breasts received higher average compression pressure (see previous study on the same study population[5]). On the other hand, it is possible that controlling for a relative measure of density attenuates the associations with pressure because relative density is relative to breast volume. Despite these difficulties it is clear from our study that the association between pressure and odds of interval cancer is not linear, and it is possible that moderate levels of pressure are associated with lower risk of interval cancer.

A recent UK study by Hill et al, which used a different design, appears to contradict these findings. They compared interval cancers with age and Volpara density grade matched screen detected cancer controls and found that pressure measured at initial screen was a significant predictor of interval versus screen detected cancers, with higher pressure being associated with a lower risk of interval cancers.[26] The results of Hill's study are not directly comparable to ours but suggest that the exact nature of the relationship between pressure and cancer detection is still not clear.

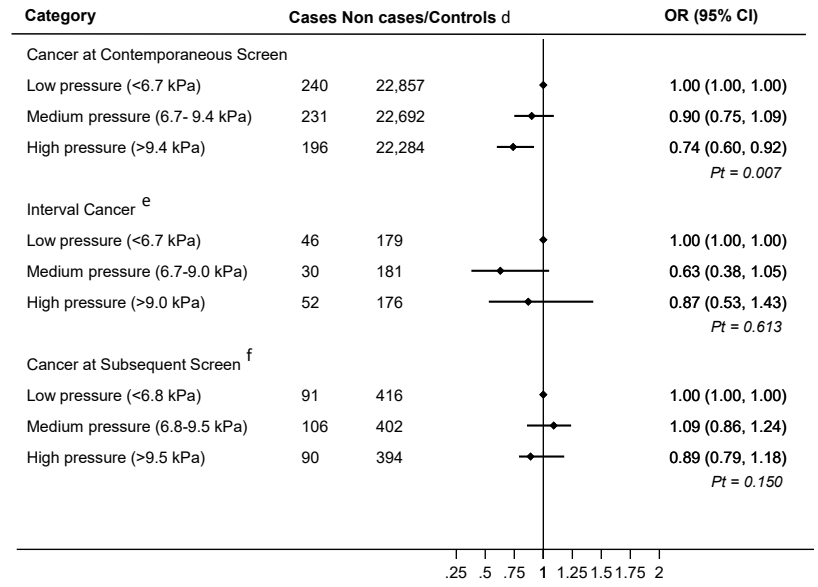
To our knowledge ours is the first study to look at the association between paddle tilt and cancer detection. Our findings, albeit non-significant at the 5% significance level suggest that fewer interval cancers may be associated with the lowest paddle tilt, but further studies are required to clarify this association.

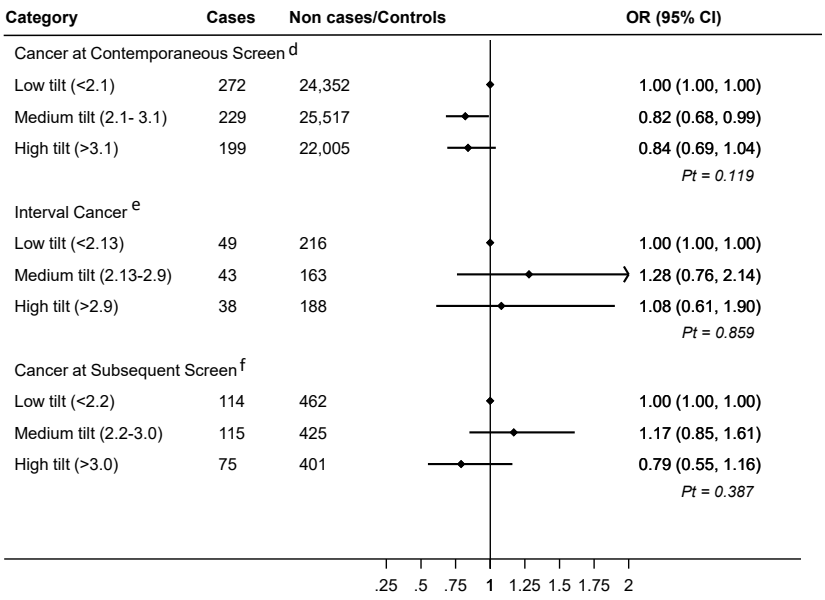
The pathways through which variations in pressure and tilt applied at imaging may influence cancer conspicuity and hence the likelihood of an interval cancer are poorly understood. Our findings also suggest that the association between compression thickness and tumour conspicuity may be more subtle in real-life than when simulated lesions in breast phantoms are used.[2]

Our study suggests that applying a moderate level of pressure may reduce the odds of cancers being missed at screening (albeit with non-significant findings). Under compression is likely to lead to increased thickness and more possibility of image blur

Figure 3. Regression analysis, fully adjusted [a, b], of associations between pressure [c] and tilt [c] measured at the contemporaneous screen and breast cancer. Footnotes a Adjusted associations: Age, Ethnicity and Volpara Density Grade (4th Edition) which corresponds to the breast imaging reporting & data system (BI-RADS) density category. All exposures measured at the contemporaneous screen. b Tilt model additionally adjusted for NDV (as a proxy for BMI), which was omitted in the pressure model due to collinearity; NDV was strongly negatively correlated with compression pressure (Pearson correlation coefficient <-0.70). c Automated pressure and tilt measures from the mean values from CC (cranio-caudal view) and MLO (medio-lateral oblique) images categorised according thirds of the distribution in non-cases. d Contemporaneous models additionally adjusted for incident or prevalent screening. e Interval Cancers – diagnosed during 3-year period since contemporaneous screen but before a subsequent screen. f Cancers at subsequent screen - diagnosed at next routine screening event after contemporaneous screen.

Fig 3 Regression analysis, fully adjusted [a, b], of associations between pressure [c] and tilt [c] measured at the contemporaneous screen and breast cancer





associated with movement, but higher pressure than strictly necessary may also be detrimental to the screening process. In practice film readers suggest that conspicuity depends on the relative density of fat and lesion and whereas fat is compressible and displaced from the image, the fibroglandular tissue and lesion are less compressible. At high levels of compression therefore the relative difference between fat and dense tissue may be reduced, hence reducing tumour conspicuity. It has also been suggested that reduced conspicuity may be the result of high compression pressure spreading tumour tissue and thereby diminishing the contrast required to identify the lesion.[27] An alternative explanation for possible reduced sensitivity at higher levels of compression pressure, was proposed by Hauge et al who found that the paddle moved for a significant period after the mammographer stopped increasing the compression force[28] and Ma et al also noted that the settling period was longer when higher compression force was used.[29] It is possible therefore, that at higher pressures, blurring can occur if the image is taken too soon after compression ceases *i.e.* whilst the breast is still undergoing settling movement. Others have suggested that the fact that fluids, including blood, are forced out of the breast during compression, whilst necessary for exposing some tumours may diminish the increased blood flow into a mass that can be a clue to identifying invasive cancers.[30]

The term "pressure" to describe force/contact area is not strictly correct since fluids, such as breast tissues, cannot be compressed, nevertheless it is a useful shorthand to describe the stretching of the breast. It is possible that models based on compression force adjusted for BV, may be better for understanding the association between relative force and cancer conspicuity because they take into account the entirety of the breast tissue undergoing compression and it is easier to control for breast density; however, unlike the compression pressure, it cannot easily be estimated in real-time and therefore has more limited use in practical settings.

Our study is inconclusive with respect to the association of paddle tilt and cancer detection although there is a possibility that lower tilt is associated with better screening outcomes. This could be related to the finding, in qualitative studies, that images taken with tilting paddles tend to show less tissue and have reduced contrast compared to rigid paddles.[31]

Strengths and limitations

Strengths of this study include its population-based design, large sample size, ethnic mix, and unbiased exposure measurements.

The algorithm (Volpara Density) used, also gives objective and reliable volumetric BV, plate contact area and DV estimates[32–34] which were used to calculate the exposure measures of interest as well as potential confounders. Force and tilt measurements are calibrated by the machine manufacturers and therefore the raw objective exposure variables are reliable and unbiased as they are independent of the outcome status of the participants (*i.e.,* their current or subsequent cancer status).

A limitation of this study was its low power to detect true associations, resulting from relatively few interval cancers being recorded, partly because of the lag time between diagnosis and notification to the screening services. Similarly, the number of subsequent round screen detected cancers was relatively low after excluding all the image sets that did not meet our inclusion criteria.

Implications

This study suggests that breast screening mammography technique, reflected in mammographer's discretionary decision making about positioning, force and paddle tilt, although poorly understood, has an impact on screening programme outcomes. Mammography is not a perfect screening tool and although cancer is successfully detected in almost 0.9% of females screened in the UK,[35] around 0.3% of females screened, actually present as interval cancers. Interval cancers tend to have a poorer prognosis than screen-detected cancers[36] therefore any improvements that increase the proportion of cancers that are detected at screening, will potentially save lives. Simple guidelines such as 'higher pressure is better' are unlikely to be helpful since there is evidence to suggest that outcomes (in terms of interval cancers) at medium-pressure levels may be better. As Ekpo et al pointed out 'errors in mammography cannot be solved through technology alone'[37] however by further improving our knowledge, and by challenging current assumptions incremental improvements may be made. The availability of automated image analysis could be used to increase the scope of routine image audits and enable more objective measures such as pressure or relative force to be incorporated into the audit process.

Further studies are required to compare outcomes where mammographers rely on their own discretion, with those where stricter pressure or force guidelines are adhered to. Research is also required to investigate whether the use of tilting or flexible paddles is associated with better or worse screening outcomes.

REFERENCES

1. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012; **380**: 1778–86. https://doi.org/10.1016/S0140-6736(12)61611-0
2. Salvagnini E, Bosmans H, Van Ongeval C, Van Steen A, Michielsen K, Cockmartin L, et al. Impact of compressed breast thickness and dose on lesion Detectability in Digital Mammography: FROC study with simulated lesions in real mammograms. *Med Phys* 2016; **43**: 5104–16. https://doi.org/10.1118/1.4960630
3. Saunders RS, Samei E. The effect of breast compression on mass Conspicuity in Digital Mammography. *Med Phys* 2008; **35**: 4464–73. https://doi.org/10.1118/1.2977600
4. Yaffe MJ, Mainprize JG. Risk of radiation-induced breast cancer from Mammographic screening. *Radiology* 2011; **258**: 98–105. https://doi.org/10.1148/radiol.10100655
5. Hudson SM, Wilkinson LS, De Stavola BL, Dos-Santos-Silva I. To what extent are objectively measured Mammographic imaging techniques associated with compression outcomes. *Br J Radiol* 2023; **96**(1146): 20230089. https://doi.org/10.1259/bjr.20230089
6. Mercer CE, Hogg P, Szczepura K, Denton ERE. Practitioner compression force variation in Mammography: A 6-year study. *Radiography* 2013; **19**: 200–206. https://doi.org/10.1016/j.radi.2013.06.001
7. Branderhorst W, de Groot JE, Highnam R, Chan A, Böhm-Vélez M, Broeders MJM, et al. Mammographic compression--a need for mechanical standardization. *Eur J Radiol* 2015; **84**: 596–602. https://doi.org/10.1016/j.ejrad.2014.12.012
8. Lau S, Abdul Aziz YF, Ng KH. Mammographic compression in Asian women. *PLoS ONE* 2017; **12**(4): e0175781. https://doi.org/10.1371/journal.pone.0175781
9. Waade GG, Sanderud A, Hofvind S. Compression force and radiation dose in the Norwegian breast cancer screening program. *Eur J Radiol* 2017; **88**: 41–46. https://doi.org/10.1016/j.ejrad.2016.12.025
10. NHS Breast Screening Programme. NHS Breast Screening Programme Guidance for breast screening mammographers. London, UK: Public Health England; 2017.
11. Programmes NCS. Quality assurance guidelines for mammography including radiographic quality control. [Contract No.: NHSBSP Publication No 63]. 2006.
12. Moshina N, Sebuødegård S, Hofvind S. Is breast compression associated with breast cancer detection and other early performance measures in a population-based breast cancer screening program *Breast Cancer Res Treat* 2017; **163**: 605–13. https://doi.org/10.1007/s10549-017-4214-8
13. Holland K, Sechopoulos I, Mann RM, den Heeten GJ, van Gils CH, Karssemeijer N. Influence of breast compression pressure on the performance of population-based Mammography screening. *Breast Cancer Res* 2017; **19**(1): 126. https://doi.org/10.1186/s13058-017-0917-3
14. McCormack VA, dos Santos Silva I. Breast density and Parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2006; **15**: 1159–69. https://doi.org/10.1158/1055-9965.EPI-06-0034
15. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 2007; **356**: 227–36. https://doi.org/10.1056/NEJMoa062790
16. Pisano ED, Hendrick RE, Yaffe MJ, Baum JK, Acharyya S, Cormack JB, et al. Diagnostic accuracy of Digital versus film Mammography: exploratory analysis of

selected population subgroups in DMIST. *Radiology* 2008; **246**: 376–83. https://doi.org/10.1148/radiol.2461070200
17. Wanders JOP, Holland K, Karssemeijer N, Peeters PHM, Veldhuis WB, Mann RM, et al. The effect of volumetric breast density on the risk of screen-detected and interval breast cancers: a cohort study. *Breast Cancer Res* 2017; **19**(1): 67. https://doi.org/10.1186/s13058-017-0859-9
18. Burnside ES, Warren LM, Myles J, Wilkinson LS, Wallis MG, Patel M, et al. Quantitative breast density analysis to predict interval and node-positive cancers in pursuit of improved screening protocols: a case-control study. *Br J Cancer* 2021; **125**: 884–92. https://doi.org/10.1038/s41416-021-01466-y
19. Moshina N, Roman M, Waade GG, Sebuødegård S, Ursin G, Hofvind S. Breast compression parameters and Mammographic density in the Norwegian breast cancer screening programme. *Eur Radiol* 2018; **28**: 1662–72. https://doi.org/10.1007/s00330-017-5104-5
20. Oxford Uo. AgeX trial. Oxford: University of Oxford. 2020. Available from: http://www.agex.uk/
21. Office for National Statistics (ONS). Census Guidance and Methodology 2015 [Overview of methods and codes used for 2011 census. 2011. Available from: https://www.ons.gov.uk/census/2011census/2011censusdata/2011censususerguide/variablesandclassifications
22. The Royal College of Radiologists. Guidance on screening and symptomatic breast imaging. 2013.
23. Matakina Technology Ltd. VolparaDensity™ User Manual Version 1.5.11. [User Manual Volpara Software]. In press 2014.
24. Hudson S, Vik Hjerkind K, Vinnicombe S, Allen S, Trewin C, Ursin G, et al. Adjusting for BMI in analyses of volumetric Mammographic density and breast cancer risk. *Breast Cancer Res* 2018; **20**(1): 156. https://doi.org/10.1186/s13058-018-1078-8
25. Health and Social Care Centre. UK :NHS Digital. *Breast Screening programme England 2016-2017*2018.
26. Hill ML, Martis L, Halling-Brown M, Highnam RP, Chan A, Bosmans H, et al. Mammographic compression pressure as a predictor of interval cancer. Sixteenth International Workshop on Breast Imaging; Leuven, Belgium; 2022. https://doi.org/10.1117/12.2625460
27. Tingberg A, Lång K, Timberg P. Breast Imaging. In: Karssemeijer N, ed. Performance of Breast Cancer Screening Depends on Mammographic Compression. Breast Imaging: 13th International Workshop, IWDM 2016. Cham: Springer International Publishing; 2016. pp. 19–22. https://doi.org/10.1007/978-3-319-41546-8
28. Hauge IHR, Hogg P, Szczepura K, Connolly P, McGill G, Mercer C. The Readout thickness versus the measured thickness for a range of screen film Mammography and full-field Digital Mammography units. *Med Phys* 2012; **39**: 263–71. https://doi.org/10.1118/1.3663579
29. Ma WK, Brettle D, Howard D, Kelly J, Millington S, Hogg P. Extra patient movement during Mammographic imaging: an experimental study. *Br J Radiol* 2014; **87**: 20140241. https://doi.org/10.1259/bjr.20140241
30. Highnam R, Brady JM. Mammographic Image Analysis: Springer Netherlands; 2012.
31. Broeders MJM, ten Voorde M, Veldkamp WJH, van Engen RE, van Landsveld – Verhoeven C, 't Jong – Gunneman MNL, et al. Comparison of a flexible versus a rigid breast compression paddle: pain experience, projected breast area, radiation dose and technical image quality. *Eur Radiol* 2015; **25**: 821–29. https://doi.org/10.1007/s00330-014-3422-4
32. Brand JS, Czene K, Shepherd JA, Leifland K, Heddson B, Sundbom A, et al. Automated measurement of volumetric Mammographic density: A tool for widespread breast cancer risk assessment. *Cancer Epidemiology, Biomarkers & Prevention* 2014; **23**: 1764–72. https://doi.org/10.1158/1055-9965.EPI-13-1219
33. Alonzo-Proulx O, Mawdsley GE, Patrie JT, Yaffe MJ, Harvey JA. Reliability of automated breast density measurements. *Radiology* 2015; **275**: 366–76. https://doi.org/10.1148/radiol.15141686
34. Holland K, van Zelst J, den Heeten GJ, Imhof-Tas M, Mann RM, van Gils CH, et al. Consistency of breast density categories in serial screening mammograms: A comparison between automated and human assessment. *Breast* 2016; **29**: 49–54. https://doi.org/10.1016/j.breast.2016.06.020
35. Public Health England. Breast Screening Programme, England Statistics for 2014-15 2016. 2016. Available from: http://www.hscic.gov.uk/pubs/brstscreen1415
36. Bennett RL, Sellars SJ, Moss SM. Interval cancers in the NHS breast cancer screening programme in England. *Br J Cancer* 2011; **104**: 571–77. https://doi.org/10.1038/bjc.2011.3
37. Ekpo EU, Alakhras M, Brennan P. Errors in Mammography cannot be solved through technology alone. *Asian Pac J Cancer Prev* 2018; **19**: 291–301. https://doi.org/10.22034/APJCP.2018.19.2.291

Paper V Supplementary Table 1. Odds ratios of contemporaneous screen-detected, interval and next round screen detected cancers associated with compression pressure[a], before and after adjustment for relative breast density[b]

	Crude OR	P		Fully Adjusted using Volpara Density (BIRADS) OR	P
	Contemporaneous Screen Detected Cancers [c]				
Pressure kPa					
Low (<6.7)	ref			ref	
Medium (6.7- 9.4)	0.97 (0.81, 1.16)	*0.738*		0.90 (0.75, 1.09)	*0.290*
High (>9.4)	0.84 (0.69, 1.01)	*0.067*		0.74 (0.60, 0.92)	*0.007*
	Pt=0.068			Pt=0.007	
Age at screen	1.04 (1.03, 1.05)	*<0.001*		1.06 (1.05, 1.07)	*<0.001*
Dense Volume cm^3	1.00 (1.00, 1.00)	*0.015*			
Density Grade	1.06 (0.99, 1.13)	*0.101*		1.18 (1.08, 1.28)	*<0.001*
	Interval Cancers				
Pressure kPa					
Low (<6.7)	ref			ref	
Medium (6.7-9.0)	0.72 (0.45, 1.16)	*0.174*		0.63 (0.38, 1.05)	*0.079*
High (>9.0)	1.22 (0.79, 1.88)	*0.365*		0.87 (0.53, 1.43)	*0.581*
	Pt=0.356			Pt=0.613	
Age at screen	0.56 (0.44, 0.70)	*<0.001*		0.56 (0.44, 0.71)	*<0.001*
Dense Volume cm^3	1.01 (1.01, 1.02)	*<0.001*			
Density Grade	1.62 (1.32, 1.99)	*<0.001*		1.60 (1.27, 2.01)	*<0.001*
	Subsequent Screen Detected Cancers				
Pressure kPa					
Low (<6.8)	ref			ref	
Medium (6.8-9.5)	1.19 (0.86, 1.63)	*0.287*		1.09 (0.86, 1.24)	*0.629*
High (>9.5)	1.10 (0.79, 1.55)	*0.563*		0.89 (0.79, 1.18)	*0.545*
	Pt=0.545			Pt=0.556	
Age at screen	1.02 (0.88, 1.19)	*0.759*		1.05 (0.88, 1.26)	*0.593*
Dense Volume cm^3	1.01 (1.00, 1.01)	*0.003*			
Density Grade	1.15 (1.00, 1.31)	*0.044*		1.19 (1.00, 1.41)	*0.045*

Footnotes:

[a] Automated pressure calculated from the mean values from CC (cranio-caudal view) and MLO (medio-lateral oblique) images categorised according thirds of the distribution in non-cases

[b] Adjusted associations: Age, Ethnicity and Volpara Density Grade (4th Edition) (DG) which corresponds to the breast imaging reporting & data system (BI-RADS) density category calculated by the Volpara algorithm using relative % of dense and non-dense areas on the image. All exposures measured at the contemporaneous screen.

[c] Contemporaneous model was additionally adjusted for incident or prevalent status. Robust standard errors, clustered by woman screened, were used in contemporaneous model to account for possibility of multiple screens per woman

Paper V Supplementary Table 2. Odds ratios of contemporaneous screen-detected, interval and next round screen detected cancers associated with compression tilt[a] before and after adjustment [b]

	Crude OR	P		Fully Adjusted using Volpara Density (BIRADS) OR	P
	Contemporaneous Screen Detected Cancers [c]				
Paddle Tilt °					
Low (<2.1)	ref			ref	
Medium (2.1- 3.1)	0.80 (0.67, 0.96)	*0.015*		0.82 (0.68, 0.99)	*0.035*
High (>3.1)	0.81 (0.67, 0.97)	*0.024*		0.84 (0.69, 1.04)	*0.111*
	Pt = 0.021			Pt=0.105	
Age at screen	1.04 (1.03, 1.05)	*<0.001*		1.06 (1.05, 1.07)	*<0.001*
Dense Volume cm^3	1.00 (1.00, 1.00)	*0.015*			
Density Grade [e]	1.06 (0.99, 1.13)	*0.101*		1.26 (1.15, 1.39)	*<0.001*
	Interval Cancers				
Paddle Tilt °					
Low (<2.13)	ref			ref	
Medium (2.13-2.9)	1.16 (0.73, 1.85)	*0.528*		1.28 (0.76, 2.14)	*0.357*
High (>2.9)	0.85 (0.53, 1.37)	*0.511*		1.08 (0.61, 1.90)	*0.790*
	Pt = 0.559			Pt=0.739	
Age at screen	0.61 (0.48, 0.71)	*<0.001*		0.62 (0.49, 0.80)	*<0.001*
Dense Volume cm^3	1.01 (1.01, 1.02)	*<0.001*			
Density Grade [e]	1.54 (1.24, 1.92)	*<0.001*		2.01 (1.46, 2.76)	*<0.001*
	Subsequent Screen Detected Cancers				
Paddle Tilt °					
Low (<2.2)	ref			ref	
Medium (2.2-3.0)	1.23 (0.91, 1.67)	*0.173*		1.17 (0.85, 1.61)	*0.330*
High (>3.0)	0.86 (0.62, 1.19)	*0.360*		0.79 (0.55, 1.16)	*0.232*
	Pt = 0.457			Pt=0.272	
Age at screen	1.02 (0.88, 1.19)	*0.759*		1.04 (0.88, 1.22)	*0.634*
Dense Volume cm^3	1.01 (1.00, 1.01)	*0.003*			
Density Grade [e]	1.15 (1.00, 1.31)	*0.044*		1.19 (0.98, 1.45)	*0.073*

Footnotes:

[a] Automated tilt calculated from the mean values from CC (cranio-caudal view) and MLO (medio-lateral oblique) images categorised according thirds of the distribution in non-cases

[b] Adjusted associations: Age, Ethnicity, NDV and Volpara Density Grade (4th Edition) (DG) which corresponds to the breast imaging reporting & data system (BI-RADS) density category calculated by the Volpara algorithm using relative % of dense and non-dense areas on the image. All exposures measured at the contemporaneous screen.

[c] Contemporaneous model was additionally adjusted for incident or prevalent status. Robust standard errors, clustered by woman screened, were used in contemporaneous model to account for possibility of multiple screens per woman

Paper V Supplementary Table 3. Adjusted odds ratios of contemporaneous screen-detected, interval and next round screen detected cancers associated with compression pressure, by imaging view (MLO or CC)

	Adjusted OR [a] MLO view	**P**			**Adjusted OR[a] CC view**	**P**
	Contemporaneous Screen Detected Cancers[b, c]					
Pressure MLO kPa				**Pressure CC KPa**		
Low (<6.03)	ref			Low (<7.24)	ref	
Medium (6.03-8.00)	0.93 (0.78, 1.11)	*0.440*		Medium (7.24-10.90)	0.95 (0.79, 1.14)	*0.557*
High (>8.00)	0.80 (0.66, 0.98)	*0.028*		High (>10.90)	0.77 (0.62, 0.95)	*0.015*
	Pt=0.028				Pt=0.015	
	Interval Cancers					
Pressure MLO kPa				**Pressure CC KPa**		
Low (<5.92)	ref			Low (<7.00)	ref	
Medium (5.92-7.91)	0.89 (0.53, 1.49)	*0.663*		Medium (7.00-10.25)	0.64 (0.36, 1.14)	*0.130*
High (>7.91)	1.04 (0.62, 1.75)	*0.874*		High (>10.25)	0.88 (0.49, 1.56)	*0.679*
	Pt=0.858				Pt=0.755	
	Subsequent Screen Detected Cancers					
Pressure MLO kPa				**Pressure CC KPa**		
Low (<6.05)	ref			Low (<7.13)	ref	
Medium (6.05-7.98)	1.02 (0.74, 1.39)	*0.910*		Medium (7.13-11.02)	1.12 (0.80, 1.57)	*0.493*
High (>7.98)	0.96 (0.8, 1.35)	*0.820*		High (>11.02)	0.83 (0.57, 1.22)	*0.339*
	Pt=0.814				Pt=0.325	

Footnotes:
[a] Adjusted for: Volpara Breast density grade (BI-RADS 4th edition) , age, ethnicity. Not adjusted for NDV, which was omitted in the pressure model due to collinearity. All exposures measured at contemporaneous screen.
[b] Contemporaneous model was additionally adjusted for adjusted for incident or prevalent status.
[c] Robust standard errors, clustered by woman screened, were used contemporaneous model to account for possibility of multiple screens per woman

Paper V Supplementary Table 4. Odds ratios of contemporaneous screen-detected, interval and next round screen detected cancers associated with compression pressure, before and after adjustment for absolute breast density.

	Crude OR	**P**		**Model Adjusted for Breast Dense Volume OR [a]**	**P**		**Fully Adjusted OR [b c]**	**P**	
	Contemporaneous Screen Detected Cancers								
Pressure kPa									
Low (<6.7)	ref			ref			ref		
Med (6.7- 9.4)	0.97 (0.81, 1.16)	*0.738*		0.99 (0.83, 1.20)	*0.953*		1.03 (0.86, 1.24)	*0.758*	
High (>9.4)	0.84 (0.69, 1.01)	*0.067*		0.88 (0.72, 1.07)	*0.192*		0.97 (0.79, 1.18)	*0.750*	
	Pt=0.068			Pt=0.197			Pt=0.770		
Age at screen[d]	1.04 (1.03, 1.05)	*<0.001*					1.06 1.05, 1.08)	*<0.001*	
Dense Vol cm^3	1.00 (1.00, 1.00)	*0.015*					1.01 (1.00, 1.01)	*<0.001*	
	Interval Cancers								
Pressure kPa									
Low (<6.7)	ref			ref			ref		
Med (6.7-9.0)	0.72 (0.45, 1.16)	*0.174*		0.93 (0.56, 1.54)	*0.776*		0.96 (0.57, 1.62)	*0.880*	
High (>9.0)	1.22 (0.79, 1.88)	*0.365*		2.01 (1.24, 3.26)	*0.005*		2.03 (1.21, 3.37)	*0.007*	
	Pt=0.356			Pt=0.005			Pt=0.007		
Age at screen[d]	0.56 (0.44, 0.70)	*<0.001*					0.59 (0.46, 0.74)	*<0.001*	
Dense Vol cm^3	1.01 (1.01, 1.02)	*<0.001*					1.01 (1.01, 1.02)	*<0.001*	
	Subsequent Screen Detected Cancers								
Pressure kPa									
Low (<6.8)	ref			ref			ref		
Med (6.8-9.5)	1.19 (0.86, 1.63)	*0.287*		1.36 (0.98, 1.89)	*0.068*		1.32 (0.94, 1.85)	*0.105*	
High (>9.5)	1.10 (0.79, 1.55)	*0.563*		1.39 (0.97, 2.01)	*0.076*		1.31 (0.90, 1.90)	*0.155*	
	Pt=0.545			Pt=0.073			Pt=0.150		
Age at screen[d]	1.02 (0.88, 1.19)	*0.759*					1.06 (0.88, 1.27)	*0.542*	
Dense Vol cm^3	1.01 (1.00, 1.01)	*0.003*					1.01 (1.00, 1.01)	*0.001*	

Footnotes:

[a] Adjusted for: Breast Dense Volume (DV) cm^3

[b] Additionally adjusted for age, ethnicity. NB Not adjusted for NDV, which was omitted in the pressure model due to collinearity. NDV was strongly negatively correlated with compression pressure (Pearson correlation coefficient =-0.64).

[c] Contemporaneous model was additionally adjusted for adjusted for incident or prevalent status. Robust standard errors, clustered by woman screened, were used in contemporaneous model to account for possibility of multiple screens per woman

[d] Age at the contemporaneous screen

7.6 Further analyses

As noted in the discussion section of paper V, 'pressure' is a misnomer in the way that it is used to describe the relationship between force applied to the breast and the breast volume. The breast itself does not change in volume as a result of the compression applied, rather with increasing compression (to a limiting degree as show in paper IV) the tissue is spread and stretched across the imaging plate. In our studies the pressure models are somewhat difficult to interpret due to the degree of collinearity between the explanatory variable of interest (pressure) and the variables (BV, BD) that we wish to control for in the models since BV is strongly correlated with pressure and moderately correlated with BD, (see Paper IV Supplementary Table 2 for raw correlation coefficients).

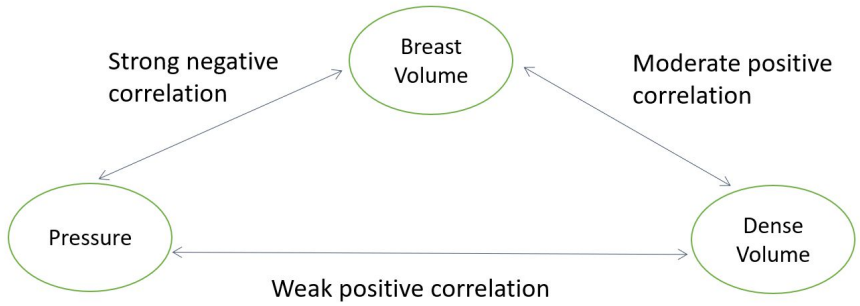


Figure 7.4 Unadjusted correlations between compression pressure and breast volumetric estimates

Footnote:
For the purposes of this figure the following interpretation of the strength of coefficients is used: absolute values of r 0-0.19 is regarded as very weak, 0.2-0.39 as weak, 0.40-0.59 as moderate, 0.6-0.79 as strong and 0.8-1 as very strong correlation.

Therefore, additional analyses were carried out to clarify the association between force applied to the breast and BC. These models are less subject to collinearity (see Fig 7.4) and therefore easier to interpret in the context of my data.

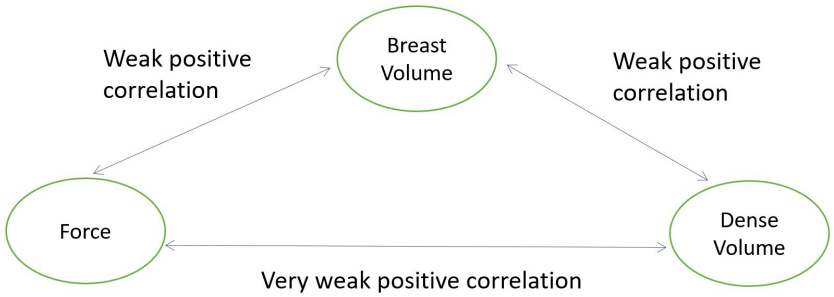


Figure 7.5 Unadjusted correlations between compression force and breast volumetric estimates

The same study participants and study designs were used as described in paper V. Additional adjusted regression models were fitted to estimate the associations of compression force with BC. The results are summarised in Table 7.3.

Findings – Associations between compression force and BC

Unadjusted odds of interval cancers are higher in the women who experienced more force, (OR for top versus bottom third of the force distribution 1.18; 95%CI 0.76 ,1.81) (Table 7.3) and the odds were lower in the medium force category (as with the pressure models). Adjusting for breast size had little effect upon the magnitude or significance of the ORs. Further adjustment for absolute BD further strengthened the association between high force and interval cancer (OR for top versus bottom third of the force distribution 1.32; 95%CI 0.80 ,2.17 in women of the same BV and BD and age), although this was not significant at the 5% CI level. There was a similar finding for subsequent screening round BC with women in the top third of the force distribution having greater odds of a cancer being detected in the next screening round, (OR for top versus bottom third of the force distribution 1.43; 95%CI 1.04 ,1.97) in the fully adjusted model.

Table 7.3 Odds ratios of contemporaneous screen-detected, interval and next round screen detected cancers associated with compression force, before and after adjustment

	Crude OR	**P**		**Model Adjusted for Non-Dense Volume OR** [a]	**P**		**Fully Adjusted OR** [b c]	**P**	
	Contemporaneous Screen Detected Cancers								
Force N									
Low (<72.1)	ref			ref			ref		
Med (72.1-89.3)	1.16 (0.97, 1.38)	*0.108*		1.15 (0.96, 1.38)	*0.124*		1.15 (0.96, 1.38)	*0.124*	
High (>89.3)	1.19 (1.00, 1.41)	*0.056*		1.17 (0.98, 1.41)	*0.089*		1.17 (0.97, 1.41)	*0.093*	
	Pt=0.056			Pt<0.089			Pt=0.093		
	Interval Cancers								
Force N									
Low (<72.3)	ref			ref			ref		
Med (72.3-89)	0.90 (0.58, 1.40)	*0.643*		0.89 (0.57, 1.40)	*0.625*		0.97 (0.60, 1.5)	*0.905*	
High (>89)	1.18 (0.76, 1.81)	*0.458*		1.16 (0.74, 1.83)	*0.517*		1.32 (0.80, 2.17)	*0.247*	
	Pt=0.480			Pt=0.534			Pt=0.287		
	Subsequent Screen Detected Cancers								
Force N									
Low (<71.7)	ref			ref			ref		
Med (71.7-87.8)	1.08 (0.80, 1.45)	*0.636*		1.06 (0.79, 1.44)	*0.688*		1.15 (0.84, 1.56)	*0.381*	
High (>87.8)	1.32 (0.97, 1.78)	*0.074*		1.24 (0.92, 1.69)	*0.160*		1.43 (1.04, 1.97)	*0.029*	
	Pt=0.074			Pt=0.161			Pt=0.029		

Footnotes:

[a] Adjusted for: Breast Non Breast Dense Volume (a proxy for BV)

[b] Additionally adjusted for age, ethnicity and DV

[c] Contemporaneous model was additionally adjusted for adjusted for incident or prevalent status. Robust standard errors, clustered by woman screened were used in contemporaneous model to account for possibility of multiple screens per woman.

These findings add weight to the findings of Paper V, that suggest that the use of 'too much' compression force could be detrimental to screening performance.

7.7 Summary Review

These studies found that aspects of mammographic technique, which can be objectively measured and recorded using automated image analysis tools are associated with the technical outcomes of compression (thickness and dose), but that these associations are not always linear with for example a decreasing effect of pressure on reduction of breast compression thickness above a certain pressure level. Women with smaller breasts received significantly higher compression pressure than those with larger breasts. There was some evidence that women of different ethnicities with the same breast measurements, experienced different levels of compression pressure. Marked between mammographer differences were observed in force, pressure and paddle tilt deployed which suggests that mammographers have their own preferred compression techniques, irrespective of their level of experience even within the setting of a single screening service.

The importance of these factors for cancer detection and screening performance was clarified and findings were consistent with the view that extremes of compression (either inadequate or excessive) as measured in terms of pressure or force may be detrimental to screening performance and result in more FNs at screening. The findings on paddle tilt were not conclusive but there is a possibility that lower paddle tilt is associated with better screening performance.

8 CHAPTER 8 DISCUSSION AND CONCLUSIONS

8.1 Introduction

The global burden of breast cancer is increasing, and it has become the most common cause of cancer death for women worldwide. In the UK the incidence, in absolute numbers, is set to rise as the population ages, with the number of new cases per year estimated to reach 70,000 by 2040 (up 13% on the current levels)(240). The risk factors for breast cancer, although generally well understood, are multifactorial and many of these risks accumulate over a woman's lifetime. Primary prevention is thus hard to address because many risk factors are not possible, amenable, or even desirable to change. Therefore, in HIC, secondary prevention through population based mammographic screening, has become a valuable tool for reducing mortality from breast cancer, allowing cancers to be treated before they become invasive or whilst they are small, non-palpable and more amenable to treatment.

Whilst breast cancer mortality risk is estimated to be reduced by ~40% in women who attend screening (75-77), mammographic screening comes with drawbacks and there is an ongoing debate about the balance of benefits and harms accruing to breast cancer screening. The main harm is perceived to be the overdiagnosis and overtreatment of women who have their breast cancer diagnosed through screening, but whose cancer would otherwise never have been diagnosed and would never have harmed them. The current consensus is that breast cancer screening in the UK is beneficial but there is a need to be constantly aware of the potential harms and to find new ways of improving the benefit/harm balance.

Breast cancer screening in the UK has continued largely unchanged since 1988 when the programme was instigated (apart from the introduction of double reading of mammograms and FFDM plus a tailored approach for the very small number of women with higher familial risk). It has been generally accepted that screening at a fixed interval with one modality is the best that can be offered at the moment[3]. The improved understanding of risk factors, coupled with recent technological advances that allow us to estimate some of those risk factors more easily, suggest that it may be time to challenge the idea that 'one-size-fits-all'.

Recent technological developments in automated mammographic image analysis offer new possibilities for estimating empirical data on breast cancer risk factors such as breast tissue composition (e.g. as assessed by density) and also on the image acquisition technique deployed for

[3] There is currently an ongoing RCT (PROSPECTS trial), recruiting up to 100,000 screening women in the UK to assess the cost-effectiveness of tomosynthesis plus standard 2D mammography versus standard mammography for breast cancer screening, which may in future affect the choice of modality in screening, but the protocols and procedures for screening acquisition remain substantially unchanged.

every woman screened. How this opportunity can be harnessed for the improvement of breast cancer screening performance is however less clear. The use of breast density assessments in screening is relatively well researched, therefore in my thesis I set out to gain a better understanding of some of the more novel features of automated mammogram analysis tools, in the hope that they could be relevant to the continual challenge of breast screening performance improvement. An overview of my research structure and thesis findings is shown in Figure 8.1.

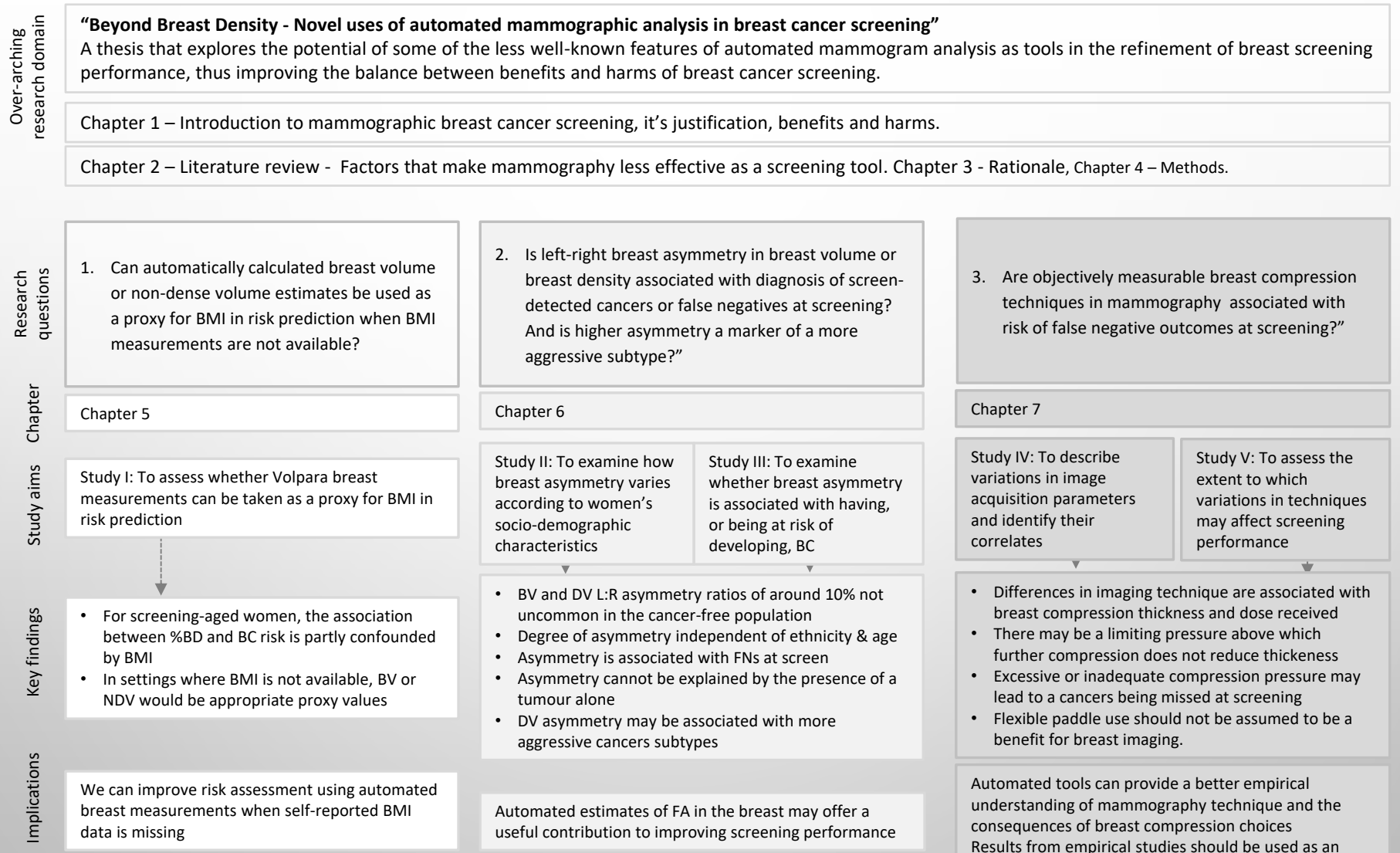


Figure 8.1. Overview of research structure and thesis findings

8.2 Summary of findings

My thesis addressed three main research questions as laid out below with key findings.

1) *"Can automatically calculated breast volume or non-dense volume estimates be used as a proxy for BMI in settings where BMI measurements are not available?"*

A major key limitation to the use of breast density as an indicator of a woman's subsequent risk of developing breast cancer in the UK, as in many other similar settings, is the fact that information on BMI is not routinely collected by the screening program. In Paper I (Chapter 5) a study was conducted using pooled data from two different countries (UK and Norway) to determine whether automatically estimated breast volume (which largely comprises fatty tissue), or its correlate non-dense volume, can be used as a proxy for BMI. BMI is both a breast cancer risk factor and an important confounder in the relationship between breast density and breast cancer risk. This study offers a possible solution for future situations that wish to control for BMI appropriately but where these data are unavailable.

Paper I confirmed previous findings that, for screening-aged women, the association between absolute volume of breast density and breast cancer risk is not confounded by BMI but that the association between %breast density and breast cancer risk is partly confounded by levels of adiposity (as measured by BMI), leading to an underestimate of the breast cancer risk associated with %breast density in models where BMI is not included. We repeated the modelling using breast volume and non-dense volume as potential proxy values for BMI and found that either proxy leads to very similar estimates of the magnitude of the adjusted association. We therefore concluded that in settings where BMI is not available, either of the mammographic measures i.e. breast volume or non-dense volume would be appropriate proxy values. This may be of importance in large scale studies of screening programmes (such as in the UK) where BMI is not routinely recorded for women being screened but is none-the-less both an important risk factor itself and a confounder in %breast density-breast cancer risk assessments. Furthermore, if there is a move towards the stratification of screening based on a woman's multifactorial breast cancer risk assessment then the unavailability of BMI might be a barrier. In circumstances where automated mammographic analysis tools are available then breast volume (or its close correlate non-dense volume) would provide an adequate surrogate for the BMI risk factor to be used in a risk assessment tool. Although existing published research had used breast size measurements previously as a tacitly assumed valid surrogate for BMI, Paper I was the first published formal empirical test of this proxy in a screening setting.

2) *"Is left-right breast asymmetry in breast volume or breast density associated with diagnosis of screen-detected cancers or false negatives at screening and is higher asymmetry a cue that the cancer may be of a more aggressive subtype?"*

Studies II and III (Chapter 6) sought to address gaps in our knowledge about lateral breast volume and breast density asymmetry and how this might be associated with breast cancer risk and tumour masking. The literature search uncovered a plausible link between fluctuating asymmetry (FA) in paired features (including breast size left:right lateral asymmetry) and breast cancer risk but there was very limited previous existing research into breast FA, and no previous studies into the prevalence of, or breast cancer risk associated with, dense volume left:right asymmetry in particular.

I first undertook an observational descriptive study using automated mammographic image analysis from over 50,000 screening images in cancer-free women who attended the S.W. London Breast Screening Programme (Study II). I found that breast fluctuating asymmetry (in both dense volume and breast volume) is prevalent across all ages, socio-economic and ethnic groups of the screening age population and positively correlated with the absolute underlying dimension, e.g. larger absolute asymmetry being observed in larger breasts. There were significant differences in mean absolute breast volume (and dense volume) between women of different age and ethnic groups but, critically for an exposure to be considered as a candidate risk factor for breast cancer, between-woman variation in lateral asymmetry was also observed within all ethnic and age groups and irrespective of absolute breast volume (or dense volume), with relative asymmetry values of around 10% not uncommon in the cancer-free population. This between-woman heterogeneity suggests that breast left:right asymmetry (if found to be associated with breast cancer) therefore has the potential to be an independent breast cancer risk factor.

Study III was therefore designed to determine whether women with increased FA in breast volume or dense volume estimates were more likely to have a BC diagnosed at screening and also whether they were more likely to have their cancer missed at screening. This is plausible not only because the asymmetry may reflect the presence of a tumour in the breast, but also because FA itself may be associated with an inherently higher risk of developing breast cancer. I found that dense volume asymmetry was positively associated with the risk of breast cancer being detected contemporaneously at screening, with women with highest dense volume asymmetry being 26% more likely to be diagnosed with a cancer at screening than those in the bottom third of the distribution. Both breast volume and dense volume asymmetry were positively associated with false negatives, i.e. cancers being missed at screening or developing in the period between screens (the latter would be consistent with an inherent higher risk). The tumour was located in the larger breast

in only 52% of all cases and there was no correlation between magnitude of the dense volume asymmetry and the tumour size, therefore I concluded that the observed asymmetry in breast tissue volumes in cancer cases was not purely a result of the presence of a tumour in the larger breast but may also reflect an inherent increased susceptibility to developing breast cancer.

A secondary aim of study III was to investigate whether breast FA was associated with particular subtypes of breast cancer. This is biologically plausible given the associations between reproductive and hormonal risk factors and both FA and breast cancer of different sub-types, although I found no previous research that had examined this potential association. The findings from my study showed that dense volume asymmetry (but not breast volume asymmetry) was most strongly associated with triple negative breast cancer. Relative to women in the bottom third of the dense volume asymmetry distribution, in women with breast cancer, those with the greatest dense volume asymmetry were over 3 times as likely to have a triple negative cancer. This finding should be caveated by noting that only 5% (n=32) of the cancers detected were of the triple negative subtype but it is nonetheless a potentially interesting finding with clinical implications since this cancer subtype tends to be both more aggressive and more difficult to treat than hormone receptor positive breast cancers making timely diagnosis of prime importance.

3) *"Are objectively measurable breast compression techniques associated with risk of false negative outcomes at screening?"*

The diagnostic quality of a mammographic image may be related to the compression techniques deployed during imaging and existing standards and guidelines partly reflect this. However, the literature review (section 2.9) concluded that despite some recent research, there is a dearth of published empirical evidence that focusses on the association between mammographic compression technique and screening performance. Publications IV and V sought to address some of the missing knowledge in this area.

Study IV, which included over 80,000 routine mammographic screens, set out to describe the distributions of, and associations between, objectively measured compression techniques and outcomes. Using fully adjusted linear regression models I found that differences in imaging technique are associated with breast compression thickness and dose received. The negative association between compression pressure and thickness was not linear and above a certain level of pressure (15kPa in our study participants) an increase in pressure did not lead to substantial reduction in thickness. Increased paddle tilt was associated with increased compression thickness

even after adjusting for subject-specific confounders suggesting that mammographers preferred technique in the use of compression paddles may be of importance to screening outcomes. Potentially harmful radiation dose was reduced slightly with increased compression pressure but increased with increasing paddle tilt; for every 1^{o} increase in paddle tilt mean glandular dose increased by 2.8% (95% CI 2.4%, 3.2%). The clinical significance of these small relative differences, however, is not known.

Study V found that compression pressure was negatively associated with the odds of a cancer being detected at the contemporaneous screen. There was evidence (albeit borderline at the 5% significance level) that moderate compression pressure was associated with around 37% lower risk of interval cancers suggesting that either excessive or inadequate compression pressure can lead to cancers being missed at screening. The associations were stronger in the CC view than the MLO view and it is possible that these associations are somewhat imaging view dependent. There was no strong evidence that paddle tilt affected the performance of mammographic screening although there are tentative indications that reduced paddle tilt may be associated with better compression outcomes and screening performance.

8.3 Strengths and limitations of study designs

The main strengths of my studies are that they used a population-based design with large sample sizes and a relatively wide ethnic and socioeconomic mix. However, because of the screening service setting, they are limited to women of screening age and care should be taken when attempting to generalize from these findings. In particular, the findings of Study I are not generalizable to younger (pre-menopausal) women, because the association between BMI and breast cancer risk is of a different magnitude and direction to that in post-menopausal women. Study I tested breast volume measurements as proxies for BMI when assessing near term (intrinsic) breast cancer risk but was not tested in models that predict the masking effect of breast density, where the role of BMI is less well understood.

There were two main study designs used in Studies II to V (Chapters 6 and 7) of this thesis, firstly cross-sectional designs, used to carry out descriptive analyses and to examine the associations between different exposures and contemporaneous screen detected cancers, secondly nested case-control designs were used to examine the associations between exposures and breast cancer (subsequent screen detected or interval cancers) risk. With a longer follow up period it might have been possible to conduct a cohort study design which would have allowed a more comprehensive analysis of screening performance statistics such as interval cancer rates and estimates of disease rates in the different exposure groups. Given the available time frame and knowing that loss to

follow up was likely to be high and considering that the outcome of interest is rare it was decided that nested-case control designs were a reasonable compromise.

A strength of the design was that I was able to collate several different exposures of interest and the potential confounders using objective methods. I found no evidence of bias in the way that Volpara estimates the exposure variables (Appendix D,E). There is a possibility of misclassification of asymmetry but because this is likely to be non-differential, the impact on my estimates of association with breast cancer is likely to have been a bias towards the null. Unfortunately, the Volpara algorithm was only designed for standard, 4 view, mammography and therefore it was necessary to exclude women who had fewer, or more than 4 images taken at screening. Even though this is a relatively small number of women (see Chapter 6 and 7 for specific exclusions), there is a bias towards excluding women with very large breasts because they are more likely to require more than 4 images and women who have had a previous mastectomy. Other exposure variables are taken from NHS registers (e.g., age, postcode, previous screening status). The only self-reported exposure was ethnicity which is collected routinely by the NHSBSP (self-reporting being considered the 'gold-standard for ethnic data collection (241)).

The cross-sectional studies could be subject to selection bias because, only women who actually attended screening had their exposures and outcomes measured. It was not possible to carry out a non-responder analysis, but other research has shown that women from areas of higher socio-economic deprivation and women who are immigrants are slightly less likely to attend screening than those who live in less deprived areas and those who were born in the UK (242-245). This would not have affected the internal validity of the findings but may, to some extent, make the findings less generalisable.

Recognising that a weakness in research to date has been its bias towards White European populations, this research was designed be conducted on a large multi-ethnic screening population. Cancer subtype may also be associated with ethnicity. Studies in the UK suggest that Black women develop cancer at a younger age and have a higher risk of triple negative breast cancer (37, 38). Breast cancer is a rare outcome in screening (<1% of screens have a cancer outcome) therefore despite the multi-ethnic population the volumes of women in the different ethnic groups were unfortunately not sufficient to make sub-group analysis by cancer subtype feasible for the different ethnic groups.

Interval cancer ascertainment is not available for cancers that arise in women who have moved abroad since screening and are hence not recorded in the UK National Cancer Registry. The number

of such cases is likely to have been small, but the possibility is acknowledged and again makes findings somewhat less generalizable. In addition, there is a delay in reporting of interval cancers which means it is highly unlikely that all were recorded before the end of the study period reducing the potential sample size.

Interval cancers arise for different reasons (see section 1.4) but some develop in the interval between screens and would not have been present or detectable at screening. It is not easy to differentiate between these cancers and the cancers that are missed because of masking. Interval cancer review does take place at the screening services and aims to classify these interval cancers according to the reason they arose, but these data were unfortunately not available for this study due to timing. This is not a problem that is unique to the screening service where my studies were undertaken and for large scale studies the value of timely interval cancer registration and review within the screening programme cannot be underestimated and should be considered in future study designs.

The studies on mammography technique (Papers IV and V) took place in a particular screening service and although the supplementary analysis showed considerable between-practitioner differences in technique the overall conventions and culture within the service mean that findings could be different elsewhere, although there was considerable corroboration with similar Dutch and Norwegian studies.

8.4 Conclusions

Each paper included in this thesis draws its own more detailed conclusions but there are a number of more general overarching conclusions that can be drawn from this body of work which are highlighted below:

1) Automated mammogram analysis tools offer benefits beyond their original purpose, which was to provide estimated breast density measurements.

   I show that these tools can be of benefit in a wider context e.g. in the simplest case I show that automated mammographic breast volume estimates offer a valid method for estimating BMI when this information is absent. Although this is of value per se, as BMI is an independent risk factor for breast cancer, the findings also show that automated breast density measurements routinely taken in high-volume screening settings, where BMI data are usually unavailable, can nevertheless be used in risk stratification. This is critical as the

level of breast density, for a woman's age and BMI, is one of the strongest known biomarkers of susceptibility to breast cancer.

2) Novel measures of breast composition, including breast volume and dense volume lateral asymmetry, may offer insight into both breast cancer risks and the possibility that cancers are missed at screening.

   I found evidence to support the hypothesis that fluctuating asymmetry in the breast may be associated with increased susceptibility for breast cancer and also that increased dense volume asymmetry may make a set of mammograms more difficult to interpret in high volume fast paced reading settings. Increased breast fluctuating asymmetry does not appear to be simply a reflection of the presence of a tumour in the breast and although the pathways through which breast fluctuating asymmetry is associated with both inherent breast cancer risk and the risk of masking are still poorly understood, reliable, easily available risk parameters are continually sought for the refinement of risk assessment and screening performance. Even if my findings on fluctuating asymmetry are not in their own right a major breakthrough, the research shows that thorough judicious use of already existing data, we might better understand the reasons why cancers are missed at screening and how subtle aspects of breast composition are associated with breast cancer risk.

3) These novel lateral asymmetry parameters may also provide cues to the most aggressive forms of breast cancer.

   Detecting the most aggressive cancers earlier, when they are more amendable to treatment, can lead to further reductions in mortality and morbidity, without which screening programmes cannot be justified. My tentative findings that increased breast density lateral asymmetry is more strongly associated with triple negative breast cancer than other subtypes suggest that objectively measured mammographic parameters may offer the potential for improving breast screening. For example, these mammographic cues might be used to flag up women for increased scrutiny or who may benefit from an increased sensitivity threshold.

4) There is an association between the way that a mammogram is taken and the effectiveness of screening, but current mammography practice is still inconsistent and some accepted intuitive 'norms' may actually be detrimental to screening performance.

   I found a strong and significant negative association between compression pressure and breast volume suggesting that mammographers do not systematically alter the force they use to account for a woman's breast size during compression. This may be partly a reflection of local culture or guidelines which were expressed in terms of force at the time the study data was collected.

   My thesis challenges the view that using as much force as is tolerated during mammography is the best strategy and suggests there are more subtle associations between breast compression and optimal screening performance since, after adjustment for other risk factors, women who experienced most compression pressure at screening had higher levels of interval cancer than those who experienced medium levels of pressure. Furthermore, flexible paddle use should not be assumed to be a benefit for breast imaging. I found no evidence in the literature that it improves screening experience for women and there is caveated evidence from my research, and from others that rigid paddles or a reduction in paddle tilt may result in better screening performance.

5) It is likely that women of different size and ethnicity have different experiences at screening due to compression technique and this has implications for the equitability of the screening programme which should be a topic for concern particularly in places where screening uptake is poor.

6) Mammography guidelines need to be reviewed when a better empirical understanding of the consequences of excess or inadequate breast compression is available.

8.5 Agenda for future research and improvement

The general aim of this thesis was to provide more understanding about whether the products of automated image assessment tools might contribute to strategies that improve the balance between benefits and harms of breast screening. Three general strategies for improving screening performance are discussed below in relation to my research findings and in the light of my general conclusions:

Stratified screening

It is accepted that a "one-size fits all" approach to screening may be suboptimal and that a risk-stratified approach should produce a better balance between the benefits and harms. However, until now there has been no straightforward way of stratifying the general screening population based on risk. Furthermore, there is, as yet, no strong evidence to suggest that this would save lives, thus the NHSBSP continues largely unchanged since its inception (apart from a tailored programme for a small number of high-risk gene carriers). Large-scale RCTs have commenced in the period since my thesis began, to assess the acceptability and effectiveness of stratified screening approaches. The BRIAID trial (54), offers supplementary imaging to NHSBSP screening women who have higher breast density (in 10 centres). The MyPeBS RCT, is recruiting in 5 European countries and uses a personal risk assessment based on: age, family history, previous history of benign breast biopsy, personal hormone and reproductive history, breast mammographic density and genotyping (polygenic risk score) to offer women found to be at higher risk more frequent screening with different modalities (246). The introduction of stratified screening is thus some way off but there is clearly a need for more accurate and practical risk assessment tools which can be adopted in real-world, high-volume settings. There is an ongoing search for new and practical biomarkers for earlier breast cancer detection especially in women with dense breasts in whom cancers are easily missed and those with more aggressive tumour types. Although dense volume and breast volume asymmetry are not clear-cut risk factors for intrinsic risk or masking, they are easy to derive using automated image analysis tools and could potentially enhance risk assessment tools.

Recently published research by Jiang et al (2023) looked at the relative rate of decline in breast density in each breast between screens and found a slower rate of involution in breasts that developed breast cancer than in the contralateral breast (26). This suggests that the incorporation of longitudinal data into risk assessment models could be of importance. There is scope for further analysis of my data set which could be used to compare with Jiang's findings.

Although studies have shown that breast screening does prevent deaths from aggressive breast tumours (247), there is also evidence to suggest that it may be less effective in detecting and preventing deaths from these more life-threatening tumours (248-250). In a large observational study on 11.3 million screens in the NHSBSP between 2009 and 2016, Blanks et al. estimated that the sensitivity for small grade 3 invasive cancers may be only 26% of that of small grade 1 invasive cancers (251). The authors suggested that this may be associated with the non-specific mammographic features for these small high-grade cancers. A recent audit of interval cancers in the NHSBSP found that high grade, ER negativity and younger age were associated with increased rates

of tumour growth (252) and it follows that breast screening must become more effective in detecting the more aggressive cancers if it is to reduce breast cancer mortality more effectively. My novel finding in Paper III that dense volume asymmetry is more positively associated with triple negative breast cancer than with other subtypes offers a potentially interesting finding in this area. This needs to be verified by further larger studies but if corroborated could offer one cue that women with higher dense volume asymmetry warrant further assessment at screening or may benefit from more regular screening.

Improving image reading, screening sensitivity and specificity

There is debate about the 'true' sensitivity of mammographic screening and against what ground truth it should be measured. Whereas sensitivity of mammograms to cancers present at the time of imaging is estimated at ~87% (253), a more pragmatic assessment of screening sensitivity in the NHSBSP programme as a whole, based on the relative numbers of interval cancers diagnosed within 3 years of a negative screen, puts the overall sensitivity in the NHSBSP around 70% overall (69). Over time with the availability of more sensitive tools such as MRI the overall sensitivity of mammographic screening may be revised down (254) and provide more subtle estimates for women of different ages, breast density and different stages in the screening cycle. Hollingsworth argues that "modest mortality reductions ascribed to screening mammography have been accomplished by identifying only half of the of the detectable cancers, based on 40-year-old technology …." He concludes that "by identifying those cancers missed by screening mammography, a major reduction in mortality could occur well above what is seen today with routine screening". Whilst this is a perhaps contestable view on the gains that can be made by improving sensitivity, it is true that there are great benefits if sensitivity can be improved. When considering screening performance, sensitivity alone is insufficient since many of the harms from screening accrue when specificity is compromised because cancer-free women are called back for unnecessary and invasive tests.

One possible improvement might come through enhanced mammographic reading. Since the screening programme was instigated the most significant improvement in this area was the introduction of double reading. However around 25% of the interval cancers are still classified as ''missed" by readers, despite the fact that screening programme readers are individually regularly objectively monitored, and all interval cancers are reviewed. Computer aided reading tools have been available for some time but there is little evidence that they have improved screening programme performance (255, 256). Artificial Intelligence (AI) tools are being developed and have the potential to be used at different stages of the screening pathway, but a 2021 review concluded that there is still much work to be done before they meet the standards required for use current

practice. In time these tools should be able to classify image patterns with a level of sensitivity and specificity that will improve the screening programme, but the balance of benefits and harms requires, not just improvements in ability to detect abnormal mammograms, but the ability to detect the most life-threatening cancers. At present the consensus is that there is little immediate prospect that AI tools will be available for use in routine screening practice soon. In the meantime, any small improvements may come from using existing technologies more effectively.

One hypothesis generated by my thesis was that the mammographic images that displayed greater left:right asymmetry may be more difficult for a radiologist to read. If this is the case, we would expect to find evidence of more equivocal reading and more recall to assessment where more asymmetry is displayed. Although outside the original scope of my thesis this was investigated as a short supplementary analysis (Appendix F) that found that the presence of high levels of dense volume asymmetry in the mammographic image was associated with higher odds of recall to assessment in cancer-free women even after adjusting for dense volume. This may be because readers are consciously or sub-consciously aware of the additional risks associated with left:right differences across the breast images or that they may simply be more difficult to read. These findings suggests that a future interesting line of study may be to further investigate the association between left:right image asymmetry and reader response. Unnecessary recalls to assessment are detrimental to screening performance, decrease specificity and increase the stress and harms associated with breast screening. Furthermore, as noted previously, an analysis of the reason for the interval cancer (i.e. whether it was missed by film readers, was occult or whether it developed in the interval between screens) was not available at the time of the studies but in future would enable us to clarify associations between L:R breast asymmetry and intervals cancers of different origin, clarifying the extent to which asymmetry is associated with the presence of a cancer in the breast, an inherent BC risk, and potentially the risk of increased tumour obfuscation in images.

Addressing technological weaknesses in mammography

Good quality mammographic images are required for successful screening but the guidance about the technical process of image acquisition is currently vague. Interestingly, in contrast to film reading where PPV, NPV and standardized cancer detection rate (257) is monitored empirically through the NHSBSP QA process, there is little empirical assessment available for mammographers to learn from. A recent review of metrics in the NHSBSP suggests that “After years of monitoring performance at a service level … the ability to analyse individual practitioner level data can provide a greater insight into practice” (258). This thesis has shown that practitioner choices during image acquisition are important for determining screening outcomes and it is possible that a better understanding of the

association between directly measurable image acquisition parameters and tumour conspicuity could inform new guidelines and potentially improve consistency in imaging approach and overall screening performance. Film readers have many tools for self-assessment and learning and therefore there is a precedent for improvement and learning that could possibly be extended to mammography practitioners if the appropriate tools and empirical measures were made available. Introduction of any changes may face some challenges since mammography is not purely a technical exercise and practitioners prefer to use their own intuitive judgement to make finer adjustments to force rather than rely on machine readouts (184). Furthermore, there is, as yet, no clear way of standardising the wide range of factors that contribute to capturing a 'good-enough' mammographic image. Since the start of my studies manufacturers have developed new tools to aid mammographic practice. For example, a mammography quality control tool that provides real-time feedback at the gantry has recently been developed. It calculates the patient positioning score, dose, and compression pressure for each image within 30 seconds of acquisition (22). Although these tools provide information at the time of screening the results are provided *after* the screening event and may lead to a disproportionate number of images being repeated. Since successful compression is related to an accurate estimate of breast size (volume), a possible innovation would be to use the low dose pre-scan, that is currently used for assessing appropriate radiation exposure, to estimate breast volume and check that a viable force is being applied before the image is taken. These tools open up more possibilities for making real-time adjustments during mammography but the long-term benefits of using such tools is yet to be fully evaluated.

During this thesis I did not specifically look at the possibility of reducing the number of false positives in the screening programme through improved mammography technique. By producing better images, fewer women might be subjected to additional tests and the potential harms of the screening programme would thereby be diminished. In additional fewer women would be recalled for a technical recall where the radiologist reader deems the images of insufficient diagnostic quality. There is scope for further study in this area using the main study data set.

My thesis also tentatively suggests that the design and use of mammography equipment may better suited to White or Afro-Caribbean women since women of different ethnicities, but with the same breast measurements and density, experienced different levels of compression pressure. In Chinese women, the breast appears to react differently to compression pressure, resulting in relatively high values of compression thickness and radiation dose in comparison to women of White and Afro-Caribbean women with the same breast measurements. This supports the findings of Lau et al. from a Malaysian study that concluded force-standardized protocols have largely been optimized for

Caucasian women, thus Asian women, who generally have smaller breasts, are subjected to protocols that might not be suitable for them (259). Studies in the Netherlands have also found that women with smaller breast volume experienced severe pain more commonly than other subjects (198) suggesting that protocols are not always appropriate for women of smaller breast volume. Practitioners at SWLBSS were aware of these differences and anecdotal discussions suggested that they do have more difficulty imaging women of Chinese ethnicity because of difficulty moving the breast away from the chest wall especially in the CC view. The implications of this are not clear-cut, although McCarthy et al have also suggested that inferior screening specificity in certain groups could partly be a reflection of ethnic differences in the screening process which may be technology dependent since the differences appear to be exacerbated when using newer (FFDM) technology (260). There is a need for more research in this area.

8.6 Scope for further use of the data sets

Outside the scope of my thesis, it is possible that other researchers that would find the data collected a useful resource, possibly as a source for cohort studies in the longer term. To date a number of other smaller studies led by clinicians working at SWLBSS have taken place utilising the large data sets created for this thesis (238, 261-263).

This research shows that there are potentially rich sources of empirical data on mammographic images and image acquisition that can be used innovatively to provide insights into screening performance and imaging technique. We should be open to using empirical research to develop a better understanding of underlying weaknesses in screening and continually seek to improve our protocols and guidelines and to challenge equipment manufacturers such that the benefits of breast screening can continue to outweigh the harms associated with population-based screening programmes until better forms of breast cancer prevention are available.

A APPENDIX A - Literature review on BMI and mammographic measurements

Search questions and strategy

Search questions were designed to answer the research questions: "How are breast measurements associated with BMI?" and "Can breast measurements be used as a proxy for BMI in estimating the association between BD and BC risk". In order to cover this the search was broken down into various sub-queries as show in Figure A.1

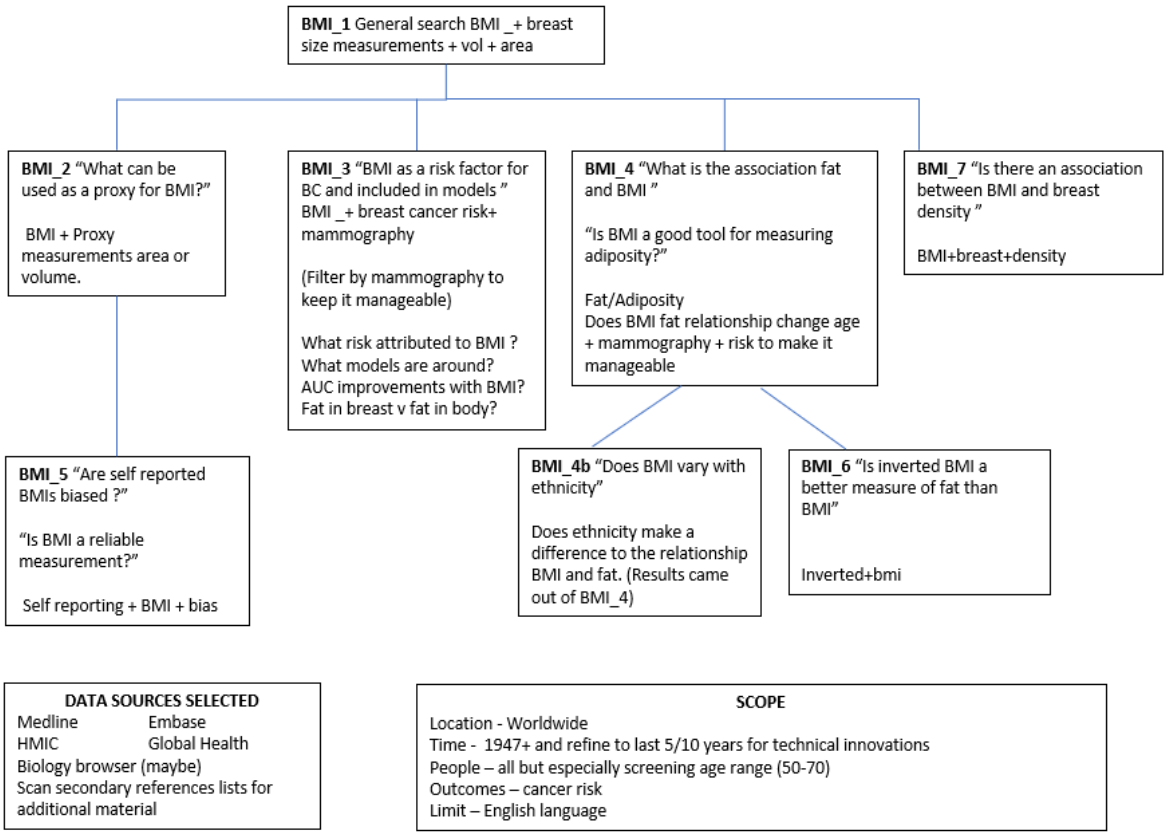


Figure A.1 Search structure and strategy overview – BMI and breast measurements

Search terms

Search combined databases 1947+ to March 2023, limit to English language and deduplicate. A summary of the search terms used is shown below for each search designed. The total number of search results after deduplication but before filtering are shown in brackets.

Search BMI_1 – General search BMI and breast size (328)

(BMI and Breast) and ((Breast adj3 Volum*) or (Breast adj3 Area))

Search BMI_2 – Proxies for BMI (0)

(((Breast adj1 Volum*) or (Breast adj1 Area)) and BMI and Proxy).

Search BMI_3 – BMI as a risk factor for BC in mammography (365)

((breast adj3 cancer adj1 risk) and BMI and Mammogr*)

Search BMI_4 – Association body fat and BMI (38)

(((Breast ADJ1 Volume) or (Breast ADJ1 Area)) AND ((adipos*) OR (fat)) AND (Mammography) AND (Cancer) and (Risk))

Search BMI_5 – Bias in self-reported BMI (32)

(((self adj1 report*) or self-report*) and BMI and Bias* and reliabil*).

Search BMI_6 – Inverted BMI (10)

(Inverted adj1 BMI)

Search BMI_7 – Associations BMI and breast density (407)

(BMI and (breast adj1 density) and (breast adj1 measureme*)) or

(BMI and (breast adj1 density))

Search results and summary table

The findings were filtered by an appraisal of the abstracts and then useful references were downloaded into Endnote for further reading. The table consolidates the main search findings.

Table A.1 Summary of key findings from literature review on BMI and breast measurements

First Author/Year	**Type/Period**	**Location**	**Studies/Cases**	**BD Assessment Methods**	**Key findings – effect size, correlation** (Relative risks (RR), Odds Ratios (OR) and 95% Confidence Intervals)	**Comments**
Body fat measurement and BMI						
Gallagher (1996) (119)	Cross-sectional	USA	202 Black 504 White (Males and females)	BMI Versus Total body fat model	BMI adjusted for age explained ~57% ($p<0.001$) of variation in body fat in the female study group.	Found association between body fat and BMI is age and sex dependent but not related to Black or White ethnic groups.
Deurenberg (2001) (116)	Cross-sectional	5 European centres	234 females 182 males	DXA or densitometry Versus BMI or Impedance	Mean bias females BMI = 0.2 %BF (SD: 0.3) Arm impedance = 0.2 %BF (SD: 0.2) Reference was DXA or densitometry.	Found that overall prediction of %BF from BMI was good after correction for age, impedance methods performed similarly well.
Deurenberg (2002) (117)	Review of literature	Asian studies	Not clear how many studies considered	DXA Skin folds, waist circumference (WC) BMI	All Asian populations studied had a higher BF% at a lower BMI compared to Caucasians. DXA used for reference.	But other methods normally require specialized equipment and techniques that can be prohibitive in terms of cost and time.
Rush (2007) (118)	Observational	South Africa (SA) and New Zealand (NZ)	721 women Five ethnic groups	BMI Waist circumference Whole body Fat (%BF) via DXA	In NZ For %BF of 43% : European women BMI = 30 kg/m^2 Asian women = 26 kg/m^2 SA . Central fat mass lower in black SA than in European SA women ($P<0.001$)	The relationship between %BF and BMI varies with ethnicity. Use of universal BMI or waist cut-points may not be appropriate for comparison of obesity among differing ethnic groups.
Nevill (2011) (122)	Cross sectional	UK	2,993 subjects	BMI Versus %BF	%BF normally distributed BMI right skewed Relationship BMI %BF non-linear Relationship inverted BMI and %BF linear	Direct measurement of %BF impractical outside study settings. NB not clear how %BF was measured. Suggests that inverted BMI is a better measure of %BF
Gosse (2014) (123)	Literature review	N America Europe Australasia	25 studies	BMI Self-reported and Independently measured	For adult females, the mean underestimate ranged from 0.12 kg/m^2 from 0.46 kg/m^2	19/25 studies found self-reported BMI significantly lower than measured BMI.
Chu (2022) (264)	Cross sectional	S.Korea	8,537 subjects	BMI WC WHR Inbody 720 scanner	BMI was a more strongly associated with BIRADS density grade than other fat measurements in pre-menopausal women: OR= 0.265 (0.204–0.344)	Various methods for measuring body composition including an automated scanner. All body fat-driven obesity parameters had significantly -tive association with BIRADS density grade.

				Manual BIRADS	WC was more strongly associated with grade in post-menopausal women: OR= 0.315 (0.239–0.416)	Simpler methods like BMI and WC were most strongly associated with density.
Breast composition and BMI and BC risk models						
Irwin (2007)(229)		USA	552 postmenopausal women BC survivors	Area based	Correlation: %BD and BMI –tive Dense area (DA) and BMI –tive (non-significant)	Focus on physical activity, BMI, Breast density
Boyd (2006)(115)	Matched case control	Canada	1,114 pairs postmenopausal (75%)	Area based	<u>Post-menopausal controls correlations:</u> BMI and DA weak –tive (-0.08) BMI and %BD –tive (-0.38) BMI and NDA +tive (+0.59) but only explains R^2 = 16% variation <u>Assoc BMI with</u> BC (adjusted for demographic and reproductive factors) Quintiles of BMI: OR unadjusted for %BD 1.17 (0.9-1.6) OR after adjustment for %BD 1.67 (1.2-2.3)	Strong neg correlation BMI and %BD means imperative to control for BMI otherwise underestimate effect of %BD on BC risk. BMI and %BD are independent BC risk factors: Concludes that BMI i.e. body size is linked to oestrogen but %BD is not predominantly linked to oestrogen.
Stone J, (2010)(265)	Matched, case-control	Cambridge & Norwich Breast Screening Programme	318 cases and 899 age-matched controls	Area based	DA was best single predictor of BC (χ_1^2 = 53.2 versus 44.4 for %BD). Adjusting for NDA did not improve the fit in the absolute DA model (both P > 0.3). Adjusting for NDA did improve fit of model with %BD (χ_1^2 = 11.6; P <0.001).	Assumed NDA is used as a proxy measure of body size to adjust %BD BC risk model but it did improve the fit. (This contrasts with Pettersson's later findings below)
Lokate (2011)(124)	Nested case-control study	Netherlands	358 cases and 859 Post-menopausal.	Area based Cumulus MLO images	BMI was positively correlated with NDA (Pearson correlation = 0.59) BMI was negatively correlated with DA (Pearson correlation = -0.21) NDA associated with **i**ncreased risk BC OR (Q5 vs Q1) = 2.4 (1.3 to 4.2) Associations changed minimally when included both absolute dense area and NDA area in the same statistical model.	Investigated independent effects of dense and fat tissue on postmenopausal breast cancer risk. Concluded: NDA and DA are independent risk factors

Pettersson (2011)(125)	Nested case-control study	USA	960 cases and 1,662 controls.	Area based (CC images)	NDA associated with decreased risk of BC Premenopausal women OR (tertile 3 vs 1) = 0.51, (0.36 to 0.72) postmenopausal women (OR (quintile 5 vs 1) = 0.46 (0.34 to 0.62). Associations changed minimally when included both absolute DA and NDA in the same statistical model.	Investigated separately absolute dense area and NDA in relation to BC risk. For NDA found opposite to Lokate above Concluded: NDA and DA are independent risk factors NB %BD area was the strongest risk factor for breast cancer in both groups.
Shepherd (2011) (130)	In a case-control study	USA	275 cases 825 controls	Volumetric SXA and area-based Cumulus	%BD using area-based and BMI r= -0.35 %BD and Volume: r= -0.27 Dense Area r= negative assoc Dense Volume r= 0.44	Also concluded that volumetric measures of breast density are more accurate predictors of BC risk than %BD measured by area based methods.
Shepherd (2012) (231)	Discussion in editorial		Discussion of Lokate and Pettersson above	Area	Pettersson (2011) found fatty breasts NDA –tive assoc BC risk. CC views Lokate (2011) found fatty breasts NDA +tive assoc BC risk. MLO views	Attempt to explain conflicting findings. Suggests MLO views have greater NDA and include body subcutaneous adipose tissue -> reflects BMI CC reflects breast fat.
Schetter, S. E. (2014)(128)	Cross sectional	USA	552 post-menopausal women	Volumetric Volpara	Correlation: %BD and BMI (Rho = -0.5, $p < 0.001$) FGV and BMI (Rho = 0.41, $p < 0.001$)	Looked at DB wrt demographics, dietary and physical activity variables Also suggest that absolute BD is a more accurate biomarker of BC risk than %BD.
Eng et al (2014) (132)	Matched, case-control (A study of association BC risk and BD by various estimating methods)	UK	436 BC cases 727 controls	Area based: ImageJ Cumulus BIRADS Volume based: Volpara Quantra SXA	Standardised Regression coefficients Volpara: BMI>30 kg/m² have 1.25 (-1.41 to -1.09) times lower %BD than those with BMI<20 kg/m²	For all methods found a strong negative correlation of %BD with BMI, driven by positive association of BMI with absolute non-density, as well as negative association of BMI with absolute density for the two area-based methods. In contrast, a trend of increasing dense volume with increasing BMI was observed for all volumetric methods

Baglietto L, (2014) (127)	Matched, case-control	Australia	590 BC cases 1695 controls	Area based Controlled for demographic and reproductive factors	BMI was positively correlated with NDA (Spearman's rank correlation = 0.62) BMI was negatively correlated with DA (Spearman's rank correlation = -0.32) DA associated with BC risk (RR (per 1 SD) = 1.50, 95% CI:1.32, 1.70). NDA associated with BC risk RR (per 1 SD) = 0.75, 95% CI: 0.65, 0.86) (under the assumption that fat in the body and fat in the breast cause breast cancer through independent mechanisms) NDA not associated with risk under the assumption that they are both proxies of adiposity.	Aimed to clarify the relationship between mammographic BD and BMI and BC risk. Concluded that risk positive association between DA and BC risk, which does not depend on the choice of the causal model. The role played by NDA is more complex. Meta-analysis includes and agrees with Stone that DA not influenced by NDA.
Pettersson (2014)(26)	Meta analysis	UK, Sweden, NL, Australia, USA Ethnicity White	13 case control studies Post and Pre-menopausal women.	Area Cumulus CC and MLO view	Post--menopausal results shown: NDA –tive associated with BC risk OR (per 1 SD) = OR 0.79 (0.73 to 0.85) (post-menopausal) <u>DA Models:</u> DA (Age adjusted) OR (per 1 SD) = 1.37 (1.33 to 1.40) DA (Age and NDA adjusted) OR (per 1 SD) = 1.37 (1.33 to 1.41) DA (Age parity and BMI adjusted) OR (per 1 SD) = 1.38 (1.31 to 1.44) <u>%BD models</u> %BD (Age adjusted) OR (per 1 SD) = 1.37 (1.32 to 1.42) %BD (Age parity and BMI adjusted) OR (per 1 SD) = 1.53 (1.44 to 1.4)	When building models for association between BD and BC risk 1. Unclear whether NDA is an independent risk factor 2. Adjusting for BMI is important for %BD models 3. Adjusting for NDA has same effect 4. Adjusting for BMI not needed in DA models 5. Adjusting for NDA is not needed for DA models NB found %DA is stronger BC risk factor than or absolute DA

Krishnan (2016) (112)	Nested case-control	Australia	>390 cases >1100 controls	Cumulus	OR = 1.76 (1.39 – 2.22) per 1 OPERA %BD for Interval Cancer: screen-detected cancer	'Inherent' (screen-detected) cancer risk was best explained by %BD or dense area after adjustment for age and BMI, whereas 'masking risk' (interval cancers) was best explained by %BD and the association was stronger.
Soguel (2017) (126)	Systematic review	Comprehensive review of all available literature	Relationship adiposity, breast composition and BC risk	Area and Volume	Studies have found paradoxical association adiposity and NDA are positively correlated but NDA seems to be negatively associated with BC risk. Discusses why fat may be protective (e.g. because it stores vitamin D, or because it has an involution effect on breast cells) but also may promote cancer growth (e.g. through its aromatase activity which is a source of endogenous oestrogens)	Conclude that association breast fat and BC should be further investigated. In particular that more volumetric studies are required. Concluded that although studies are inconsistent, in postmenopausal women the overall pattern is in line with breast fat having a protective effect on risk.
Vik Hjerkind (2018) (266)	Cross sectional	Norway (2007 – 2014)	~46,000 Screening programme	Volpara Looked at associations volumetric density and other risk factors (including age)	BMI<20 kg/m² have ~3 times greater %BD than those with BMI>33 20 kg/m² BMI>33 kg/m² have 1.5 times greater absolute VBD than those with BMI<20 kg/m² Assoc VBD and Age is linear pre-menopause and then plateaus post (-0.08% decline per 5 year age band post and -0.18% pre-menopause)	Found BMI +tive correlation with absolute volumetric density but -tively correlated with % volumetric density. Discusses Volpara limitations (underestimates density in very dense breasts). Shows the plateau in density post menopause

B APPENDIX B - Literature review on breast fluctuating asymmetry and breast cancer risk

Search questions and strategy

Search questions were designed to answer the research questions: “What are the descriptive characteristics of breast asymmetry?”, “Is there an association between Fluctuating Asymmetry (FA) and cancer”, “Do breast volumes/areas/Asymmetry vary within menstrual cycle/age ? ”, “Is there an association BC risk and Breast vol and or size / area?” Searches were designed and specified as outlined below:

Search specification for sub search ASYM_1

1. breast AND ((*symmetr*) or (size ADJ1 diff*) or (vol* adj1 diff*) or (area adj1 diff*))
2. Breast/
3. 1 AND 2
4. Remove duplicates
5. Filter
6. breast and *symmetry ---- in title
7. Remove duplicates
8. 5 OR 7 to get combined set
9. Filter and combine AND with 3
10. Filter and export

Search specification for sub search ASYM_2

Clemmesen – search for details of Clemmesen’s hook – resorted to a text search of google and citations already found through opportunistic search

Search specification for sub search ASYM_3

11. (Fluctuating ADJ Asymmetry) and (Cancer)
12. De-duplicate filter and save
13. Import to endnote (re-run in 2022 showed only 2 relevant additions)

Search specification for sub search ASYM_4

14. ((breast adj vol*) or (breast adj size) or (breast adj area)) and menstrua*

Search specification for sub search ASYM_5

15. ((breast adj vol*) or (breast adj area)) and (breast adj cancer)
16. Breast in title
17. 15 and 16
18. De-duplicate
19. Export

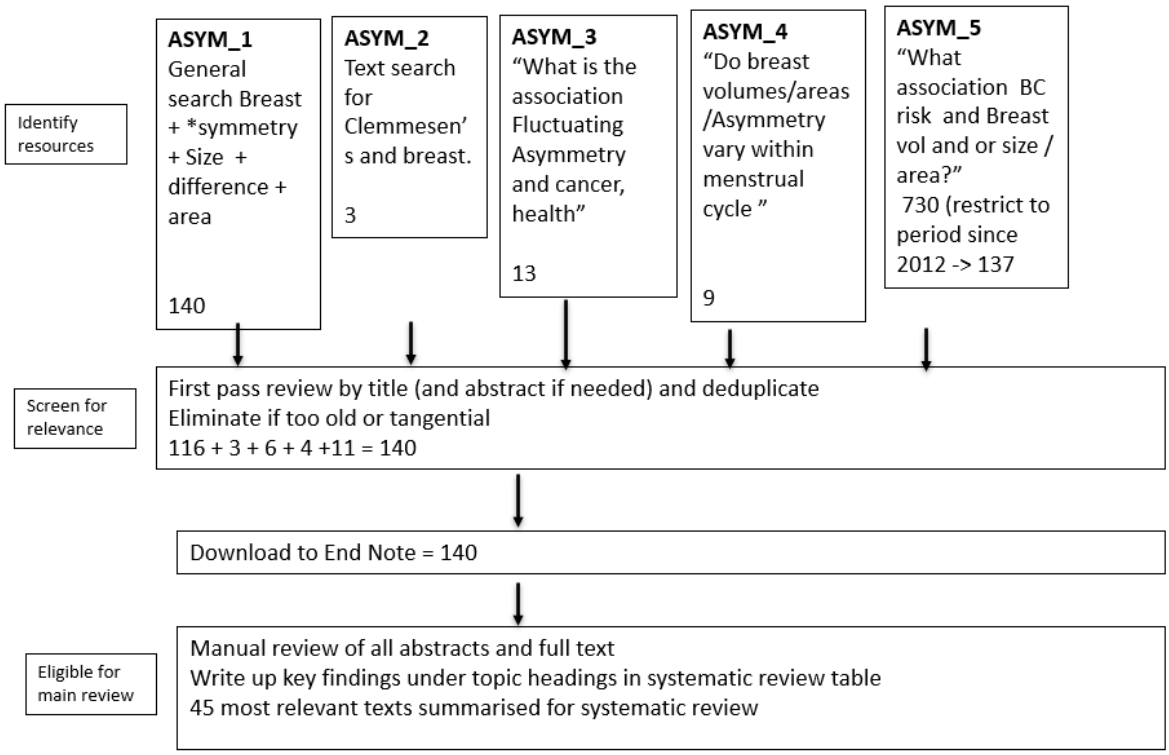


Figure B.1 Literature search strategy for fluctuating asymmetry, health and breast cancer risk

Search results and summary table

The findings were filtered by an appraisal of the abstracts and then useful references were downloaded into Endnote for further reading. The table consolidates the main search findings. Unfortunately, most of the results were not directly comparable therefore there was no attempt to undertake any meta-analyses. Many of the results looked at the association between hand-patterns (dermatoglyphics) and BC risk and many were of relatively poor quality due to small sample size.

Table B.1 Summary of key literature search findings on breast asymmetry, breast cancer and association cancer and fluctuating asymmetry

First Author /Year	**Type/Period**	**Location**	**Studies/ Cases**	**Assessment Methods/ Tools**	**Key findings – effect size, correlation** (Relative risks (RR), Odds Ratios (OR) and 95% Confidence Intervals (CI), Area under curve (AUC) Spearman's correlation Rho.	**Comments**
Breast cancer and breast asymmetry						
Scutt (1997) (97)	Case-control	UK Probably symptomatic clinic patients Mean age 57.6 years	225 cases 225 age matched controls	BV asymmetry Visual assessment	BV Asymmetry (ml) b=0.0045, SE=0.0014, R^2=0.039, p<0.0011 BV Asymmetry is positively correlated with BV in controls: Rho=0.2857, p<0.0001 BV was not correlated to tumour size (Rho=0.028, p=0.74).	First indication that BV asymmetry and BV both associated with increased odds of cancer detection at the time mammogram was taken. Authors found that risk not related to the size of the tumour. No evidence of 'directional asymmetry' i.e. no evidence that FA is related to cancer laterality.
Harvey (2000)(267)	Cases only	USA	30	BV Asymmetry Visually assessment	Infiltrating lobular carcinoma (ILC) is associated with an ipsilateral mammographic decrease in breast size between screens.	Rare tumour type that is difficult to detect using mammography.
Scutt (2006)(156)	Nested case-control Baseline mammogram 1979 – 1986 (Cancer diagnosed before 2002)	UK Age 33 – 70 UK	252 cases and 252 age-matched controls. Women self-referred for screening and normal at baseline screen.	BV BV asymmetry Estimated by visual assessment of CC images.	OR of BC = 1.50 (1.10 - 2.04) for a one 100 ml increase in absolute BV asymmetry OR = 1.09 (1.01 - 1.18) for a 1% increase in relative BV asymmetry after adjusting for the other potential risk factors (including BV).	First evidence that BV asymmetry is higher in healthy women who are free of breast disease but go on to develop BC than in women who remain disease-free in the same period.
Eltonsy (2007)(157)	Observational	USA.	268 cancer cases and 82 normal cases. From Digital Database for Screening Mammography (DDSM), public source.	BV Asymmetry Segmented images	BC patients demonstrate statistically significantly higher fluctuating asymmetry in their screening mammograms than patients with normal screening outcomes. Using an artificial neural network to combine Age and Density resulted in discrimination of cases and controls AUC: 0.80±0.03.	This 'preliminary' study found that bilateral morphological breast asymmetry in screening mammograms correlated with the presence of breast cancer.

Wang (2010)(268) (2011)(269)		USA	100 controls, 100 cases (Screen detected, Interval cancers or recalled for tests)	Bespoke algorithm	Calculated 20 values based on subtracted (absolute difference) values of the 20 different 'features' from each breast image. AUC 0.78. At 90% specificity the method detected 58% of the cases where cancer was detected (at next screen or as IC) within 18 months of the original screen.	Involved complex bespoke methods and some manual corrections to aid breast segmentation (identification of breast tissue) in the images. This method differs from CAD because not targeting specific lesions but performing a general analysis of the whole breast tissue.
Zheng (2012) (160)	Nested case-control Screened 2006-2008	USA	187 cases, 155 normal controls, 109 benign controls.	Bespoke BD asymmetry algorithm. BIRADS visually assessed.	Near term risk of BC (within 18 months of negative baseline screen). AUCs: to classify between two groups Age: 0.633±0.030 BIRADS 0.535±0.036 Bespoke asymmetry rating 0.719±0.027.	Suggests Breast density asymmetry aids diagnostic power. Highly bespoke measurement technique and not clear how repeatable this would be.
Williams (2013) (159)	Observational	USA	800 screens 394 cancers 324 benign 82 normal	Own bespoke software for calculating relative BV asymmetry. MLO views.	Relative asymmetry = (Difference sides/sum sides*2). Compared average asymmetry: Cancers cases 0.047 ± 0.003 Normal 0.028 ± 0.019 P<0.0001 Benign 0.041 ± 0.025 In MLO views average BV asymmetry in cancer cases significantly greater than in benign and normal cases.	Authors concluded that average BV asymmetry in cancer cases significantly greater than in benign and normal cases. But highly bespoke measurements and weak on how population was sampled and population characteristics. Appears to be no controlling for possible confounders.
Tan (2013)(161)	Retrospective Cohort 2012-14	USA	944 women 283 cancers CC images	CAD Type approach Bespoke algorithm. AUC	9 bilateral mammographic feature differences were calculated and added to a model along with age. AUC=0.725±0.018 was obtained for positive and negative/benign prediction for the next screen (at 12-36 months)	Features based on local and global asymmetry in mammographic images are associated with detection of BC. Short term BC risk. The method used a bespoke classification algorithm and acknowledged weaknesses included size of this 'lab-based' study and the fact that the algorithm was trained on same data that it was tested on.
Zheng (2014) (163)	Nested case-control All women had negative baseline screen 2006-2008	USA	230 cases, 230 normal controls, 230 benign controls Cancer cases diagnosed 12-36 months post baseline screen.	Bespoke BD asymmetry estimate (neural network) algorithm. BIRADS visually assessed.	Logistic regression to assess association between near term BC risk and BD asymmetry 'score' made up of 5 aspects of density asymmetry. Range 0 (low) to 1 (high). OR for women in highest 5th of score distribution compared with those in lowest 5th of asymmetry score distribution OR= 9.07 (4.65 – 17.7)	Supports Zheng's previous findings (2012) that BD asymmetry is a biomarker of near-term BC risk. The algorithm is highly bespoke and the experiment difficult to reproduce in practical settings. Surprisingly in the study group there was no association between manually assessed BIRADS density category and BC risk (within 12-36 months) and no association with family history of BC. This suggests that the BD asymmetry is not a proxy for BD but also suggests that the study group may be somewhat unusual.

Kayar (2015) (158)	Observational (cases recorded at clinic attendance controls cancer free at 12 months).	Turkey Symptomatic outpatient clinic	201 cases 466 controls Cases mean age 50 Controls mean age 41	BV assessed using Grossman-Roudner Discs	Positive correlation between a BV breast asymmetry ratio of over 20% and breast cancer detection. OR 2.01(1.24–3.34) in the 40–69 age group No relationship BC and BV asymmetry when asymmetry ratio < 15% No association was found between laterality of largest breast and BC laterality.	Study has serious design limitations e.g. age and BMI differences in cases and controls. Did not control for age or breast density or BMI. The study took place in a symptomatic clinic setting thus the sample was biased. Unusually mean right breasts volume > Mean left breast volume.
Li (2018)(270)	Retrospective case control	USA	566 prior screens, 283 cases 283 controls.	CAD Type approach. Bespoke algorithm. AUC	Used stepwise linear regression to select features that predict BC within 12-18 months. Features including breast density (global and local) asymmetries. Reported AUC 0.6870±0.0220.	A lab-based algorithm trained to discriminate using similar data to above. The method used a bespoke classification algorithm trained on same data as it was tested on. The authors concluded that although results promising the AUC discriminatory power was not sufficient for stratification in screening. Only looked at short term BC risk.
Asymmetry and health and phenotypic quality						
Parsons (1990)(136)	Discussion				Fluctuating asymmetry (FA) is a useful trait for monitoring stress in the laboratory and in natural environments. In our own species, FA of an increasing range of traits has been related to both environmental and genomic stress.	Increased FA is a reflection of poorer developmental homeostasis at the molecular, chromosomal and epigenetic levels
Moller (1995)(137)		USA n=50 Spain n=172	222 plastic surgery patients.	Assessed from physical measures , tape measure (Spain) or photographs (USA)	(1) large breasts have higher levels of fluctuating asymmetry than small breasts, (2) breast fluctuating asymmetry is higher in women without children than in women with at least one child, (3) breast fluctuating asymmetry is a reliable predictor of age-independent fecundity, and (4) breast fluctuating symmetry appears to be associated with sexual selection.	Early study suggesting that breast size FA is associated with a number of reproductive factors, e.g. parity and age-independent fecundity. Women who had lower breast FA scored higher on general reproductive health. However the sample was biased toward plastic surgery candidates and therefore not representative of the general population.
Manning (1997) (143)	Observational	Liverpool Mean age ~40	500 women Non-cancers at self-referred screening.	BV measured manually from mammogram. Wolfe patterns for breast density.	Looked at measured of phenotypic quality (fecundity, age at menarche, height, weight) and measures of breast asymmetry. Log transformed BV asymmetry was negatively associated with number of offspring and positively associated with BV and Age in mutually controlled linear regression models P<0.05 in all cases.	Authors noted that there was evidence of negative allometry (women with large breasts had smaller asymmetry than expected for their breast size) hence not appropriate to correct FA for BV. Study population is younger than screening age population but study was well designed and controlled for most known BC risk factors at the time including weight, height, breast density and age.

					Left breasts were larger than right but difference non- significant at 5% level.	Proposed that exposure to oestrogen leads to greater asymmetry but that higher phenotypic quality makes women better able to produce symmetrical breasts as oestogen levels increase. Hence FA is negatively associated with markers of phenotypic quality.
Thornhill (1997)(271)	Discussion paper				Argues that FA is the most sensitive indicator of the ability to cope with stresses during ontogeny because it covaries negatively with performance in multiple fitness domains in many species, including humans.	Paper argues that there is extensive biological knowledge and proposes that FA is a useful measure of phenotypic and genetic quality.
Milne (2003)(138)	Observational Cross-sectional	New Zealand	965 population sample of 26-year-old men and women	7 health measures assessed	The association between FA and an array of health measures was determined. FA was significantly positively associated with having two or more an identified health conditions in females. Number of reported medical conditions were more likely to report they had two or more medical conditions.	The sample was of young white Caucasians. Found that in young people there was an association between some measures of health (BMI and general number of medical conditions) but not with others such as cardio-vascular health, blood pressure. In conclusive study but in a young age group where there had been limited time for exposure to environmental stressors.
Jasienska (2006)(272)	Observational	Poland	171 women	FA assessed by inequality in the fourth-digit length on right v left hands.	Women who are more symmetrical have 13% higher average levels of estradiol over the menstrual cycle than less symmetrical women (19.4 and 17.1 pmol/l, respectively; *p*=.0001).	Authors suggested that higher hormone levels in women may lead to a substantial rise in the probability of conception, low dermatoglyphic FA may therefore be associated with increased fertility.
Campoy (2016) (273)	Observational	Argentina	119 cases	Cancer traits inferred from methylation and gene expression profiles		Found that in humans that BC arising on different sides of the body present differential cancer traits inferred from gene expression profiles.
L:R Breast Asymmetry general including image reading and asymmetry						
Losken (2005)(152)	Observational Descriptive	USA	87 asymptomatic women. Average age 50 (19-77)	Visually assessed and calculated asymmetry using 3-D surfaces images	Natural breast asymmetry does exist. 10% of women were found to have “severe” breast asymmetry on subjective evaluation, The left breast is on average larger than the right.	Little quantitative analysis and intended for plastic surgery Only looked normal women not BC cases.

Chen (2013)(274)	Observational	Taiwan	24 subjects Pre-menopausal	MRI	Mean absolute BD measured via MRI higher in right breast than left across 24 subjects but findings non-significant.	Contrasts with L>R in most other studies and sample was very small, but highlights the fact that we cannot assume that findings in Caucasians will apply to women of Chinese ethnicity.
Lee (2015)(211)	Observational	S. Korea	860 subjects Mean age 54.7	Volpara and BIRADS Classified difference in bilateral density as a difference of at least one Volpara (4th ed) grade (VDG) between sides. Average MLO CC	Study aimed to assess factors that may contribute to discrepancies between automated and radiologists' density readings. 692 women had breasts of same VDG grade 168 (20%) women had breasts with at least 1 grade difference. Significantly higher probability of agreement radiologist and Volpara in women with breast density FA versus those with none. OR = 1.98 (1.32 – 2.97) p<0.001.	20% women have bilateral density asymmetry of 1 or more VDG. Because radiologists found it more difficult to assess density when breasts were asymmetrical it opens the possibility that breast FA makes mammograms more difficult to read.
Laterality and Incidence						
Senie (1980)(151) Senie (1993)(275)	Case study Case control	Switzerland	980 symptomatic cases, Screening 1973-1980 291 controls 261 cases	Breast size (area) (planimetry applied to images Degree asymmetry measured by 100- smaller breast /larger breast	L breast is larger than R breast with a ratio that is consistent with the incidence in either breast. L:R ratio of 1:1.26. In healthy women 55% had larger left breast (ratio 1: 1.22). BC developed in the larger breast of 57 percent of women with left-sided and 46 percent of women with right-sided disease. Second study found that breast size from 4 years ago is not associated with finding of cancer. Age at cancer diagnosis, at menarche, at first birth, or at menopause did not differ between women with left or right BC.	Found the ratio left to right sided BC 1:1.26, in white women. Suggested that the relative differences are a reflection of differential sensitivity of mammary glands in left and right breasts to hormonal stimulation. The findings were not conclusive, and some hormonal factors were not significant (e.g. age at menarche) whereas parity was with more left sided tumours in women who were parous.
Weiss (1996)(150)	Observational	United States. 1973-92	250,000 cases from the Surveillance, Epidemiology, and End	Laterality of tumour and demographic data recorded in registry.	Confirmed results from other studies of an overall five percent excess of left-sided disease in women. The excess found for all ethnic groups and stages of disease, the excess increases with age.	USA population only but found that ethnicity was not a factor.

			Results (SEER) program			
Perkins (276)(2004)	Observational	USA (1994-1998) Covering 40% of the cancer registries	400,000+ cases.	Laterality of tumour and demographic data recorded in registry.	Significant excess of BC (invasive and in-situ) in L Breast. ~5% more likely to be diagnosed. True across both males and females and all ethnic groups in the USA. Found that age was not a predictor of laterality. Tumors in most common, location the upper outer quadrant, occurred with equal frequency in both breasts.	BC in Left B ~5% more likely to be diagnosed than right side. Age was not a predictor of laterality Authors outline various hypotheses for predominance of L sided cancers including detection bias through predominance of right handedness, breast feeding preference and amount of tissue volume but none of these were tested.
Roychoudhuri (2006)(277)	Observational	UK	250,000 cancer records from Thames Cancer registry	Laterality of tumour and demographic data recorded in registry.	Looked at laterality of different cancer types. Cancer incidence differed significantly by laterality at all sites studied ($p < 0.01$) but substantially in the lung (left-right incidence-rate ratio [IRR] 0.87), breast (IRR 1.07), testis (IRR 0.87) and in ovarian cancer (IRR 0.86).	Many cancers are more common on Right (but BC is more common on left side). Large sample size makes overall IRR reliable but no ethnic breakdown of results.
Amer (278)(2014)	Observational	USA. 2005 and 2012,	687 cases. Clinical data were reviewed		50.9% patients presented with left breast cancer, 46.1% with right breast cancer, and 20 (3.0%) with simultaneous bilateral malignancy. High similarities between patients and their first-degree relatives. i.e. Same breast side cancer (30/66, 45.5%), opposite breast (9/66, 13.6%), and bilateral cancer (27/66, 40.9, P=0.01163). They found no significant differences between the three groups in terms of tumour histology.	Authors point out that laterality may influence treatment options, especially in elderly patients with heart disease who may require radiation therapy. The findings on the concordance of cancer laterality between close relatives may suggest an underlying inherited genetic predisposition to left or right sided BC.
Cheng (2018)(279)	Observational	China (Guangdong) 2013-2017	2,049 cases	Included analysis of hormone status based sub-types by laterality	BC more likely to be diagnosed in left than right breast at a ratio of 1.05 to 1. In contrast to Perkins found that age may be a predictor. Looked at molecular subtypes. Age and laterality were probable predictors of HER-2+ type BC. There was a R sided predominance of HER+ subtypes.	Confirms that BC in left breast ~5% more likely to be diagnosed in Chinese women. First to suggest that HER-2 subtype risk differs by side.

Abbreviations: **BV: breast volume; BC** breast cancer; BD: breast density; BV: breast volume, CC: cranio-caudal view; FA fluctuating asymmetry (FA); IRR: incidence-rate ratio; MLO: mediolateral oblique view; VDG (Volpara density grade).

C APPENDIX C - Literature review on mammographic technique and imaging outcomes

Search questions and strategy

Search questions were designed to answer the research questions: “How consistent is compression in mammography?”, “Is Breast cancer detection related to compression thickness?”, “Is Breast cancer detection/screening performance related to compression pressure or force?” “use of flexible paddles in mammography do they improve pain”, “flexible paddles and detection/performance”

In this case a broad-brush approach was taken in order not to limit the search unduly. There were 2 main searches firstly for MAMM_1 mammography and compression and then a further search MAMM_2 for mention of flexible or rigid paddles. As search MAMM_1 was initially large (1,135 results) it was narrowed down by a series of specific exclusions as shown in the search terms below.

Search terms

Search combined databases 1947+ to March 2023, limit to English language and deduplicate. A summary of the search terms used is shown below. The total number of search results are shown in Figure YY.1

MAMM_1

((Mammog* and compress*) not implant* not deep-learning not AI not calibration not tomography not dosim* not ultras* not tomosyn* not neoadj* not needle not (deep adj1 learning)).ab.

MAMM_2

(((flexib* adj1 paddle*) or rigid*adj1 paddle* or (tilt* adj1 paddle*)) and mammogr*).mp

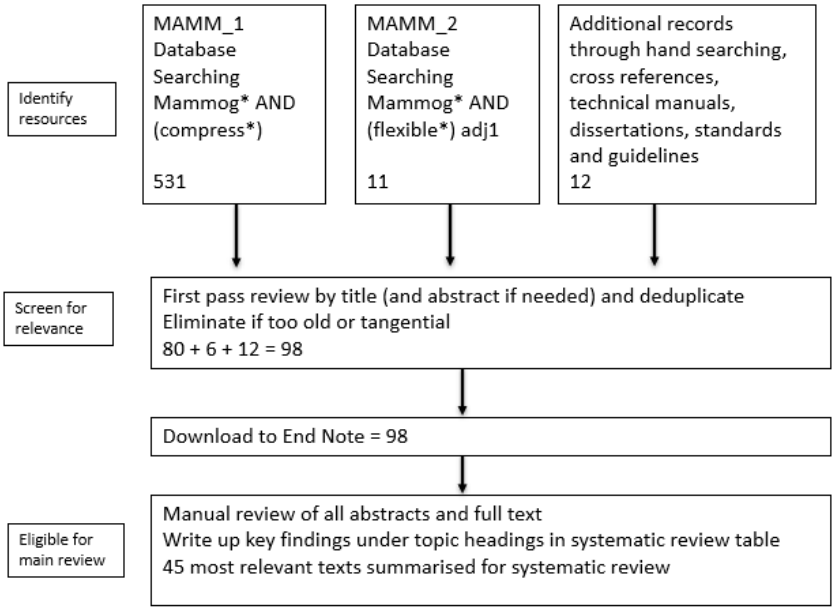


Figure C.1 Literature search strategy Mammography and Compression technique

Search results and summary table

The findings were filtered by an appraisal of the abstracts and then useful references were downloaded into Endnote for further reading. The table consolidates the main search findings. Unfortunately, most of the results were not directly comparable therefore there was no attempt to undertake any meta-analyses.

Table C.1 Summary of key literature search findings on mammography technique and breast compression

First Author/Year	**Type/Period**	**Location**	**Studies/Cases**	**Tools/ Methods/ Compression protocols**	**Key findings – effect size, correlation** (Area under the curve (AUC), Relative risks (RR), Odds Ratios (OR) and 95% Confidence Intervals), Positive predictive values (PPV)	**Comments**
Consistency in Mammography						
Lau (2017)(259)	Retrospective Cohort 2012-14	Asia (Malaysia)	3772 Women screened/ symptomatic Aged 35-80	No target value. "Compress until 'taut'	1) variability of mammographic compression parameters. Median force 12.0daN Median force, pressure, dose, thickness varied greatly across sample. 2) The effects of reducing compression force on image quality and mean glandular dose (MGD) tested on 105 women and a phantom study. Decreasing compression force from 12.0 daN to 9.0 daN increased thickness by 3.3±1.4 mm Increased the MGD by 6.2–11.0% and caused no significant effects on image quality ($p>0.05$).	Based on phantom study, it is feasible to reduce compression force up to 32.5% with minimal effects on image quality and MGD. NB Range of force and pressure values much higher than those observed UK studies.
Dustler (2012) (280)	Observational	Sweden	103 women Aged 41 – 74	Pressure sensors used under the compression plate. "Standard force" Based on mean, 9.5daN (SD 0.69)	Women subjected to two additional breast compressions (at standard force and 50% reduced force). Reducing compression force by 50% increased average breast thickness by 1.8mm ($P<0.0001$). The distribution of pressure differs greatly between breasts and the plate did not provide adequate compression across all areas of the breast (especially on juxtathoracic structures).	Suggests that reported force is a crude estimate of compression received by different areas of the breast. Hypothesizes that for many women the breast tissue is well compressed already at a relatively low applied level of force. Further force resulting in a very small reduction in thickness and minimal impact on glandular dose. Increased compression force will mainly reduce the thickness of the pec- toral muscle and the juxtathoracic area.
Mercer (2013) (174)	Observational 2004 – 2010	UK	14 practitioners 488 screens	Analogue Force DICOM	Differences in mean compression force between mammographers was significant ($p<0.0001$) ANOVA	For the first time, demonstrated that practitioners vary in the amount of compression applied in routine mammography.

				"Force should not exceed 20daN"	Force was weakly correlated with breast volume (this was demonstrated by figures in study based on different pressure groups regression coefficients not reported)	
Mercer (2013) (175)	Cross sectional Longitudinal 2004 – 2010	UK	3 consecutive screens of 344 clients	Analogue Force DICOM "Force should not exceed 20daN"	Significant thickness and compression force differences over the 3 screens for the same client (<0.0001). For the same client, and same mammographer for the 3 screens, variations not significantly different (p > 0.31) Mammographer mean force in CC view ranged from 6daN to 14daN.	Compression force used is highly dependent upon practitioner rather than client.
Mercer (2014)(129)	Cross sectional 2004 – 2010	NW England 3 different screening services	40 Mammographers >2900 screening visits >11,000 MLO images	Analogue Force DICOM "Force should not exceed 20daN" Visual assessment of BIRADS	MLO views ranged as follows: Site one: 6.5daN – 13.6daN , site two: 4.8daN – 13.9daN , site three: 10.3da N – 16.3daN . MLO averages: site one 9.7daN, site two 8.8daN, site three 13.2daN. Analysis of variance (ANOVA) of mean compression force values of practitioners demonstrated a difference between sites (p < 0.0001). Two of the three sites demonstrate variability within themselves.	Mammographers vary in the compression forces they apply to women over sequential screens. NB One site where some guidance on force existed showed better internal consistency.
Boyce (2015) (168, 281)		UK, Norway	112 images each country 5 mammogram scorers each country	PGMI (Perfect, Good, Moderate, Inadequate) scale	Scoring agreement within and between centres analysed using non-weighted kappa statistic. Best agreement Norway raters scoring MLO views from both UK (Right MLO k = 0.57, Left MLO k = 0.490) and Norway (Right MLO k = 0.48, Left MLO k = 0.470).	PGMI QA procedure findings varied between centres in both number and interpretation of criteria. NB Most common faults on MLO views

					Least agreement between UK raters scoring CC views from both UK(RCC k = 0.007, LCC k = 0.01) and Norway(RCC k = −0.04, LCC k = −0.003).	
Branderhorst (2015) (176)	Observational Retrospective Compares across countries	Netherlands USA	~37,500 images NL ~ 7,170 images USA	Volpara for Force (F), Dose (D), Pressure (P), Thickness (T)	Compared mammographic practice across different populations: NL: Mean (SD) F: 13.8(2.7) daN P: 13.7 (5.9 kPa) USA: Mean (SD) F: 7.4 (3.1) daN P: 8.1 (4.1 kPa) Average of average glandular dose per view and SD were higher in USA: 1.83 (0.73) mGy versus NL: 1.54 (0.35) mGy Thickness: USA 59.9 (13.9), NL 60.7 (11.8)mm	On average the forces and pressure applied in NL were higher than in the USA and the average glandular dose lower than in USA
Waade (2017)(282, 283)	Observational Force between centres	Norway Screening 14 centres	17,951 women	Volpara for Force (F), Dose (D), Pressure (P), Thickness (T) Recommended range of compression force (10.8–17.7 daN).	Compared Force (F) by Mammographer, screening centre and machine: F: Mean(SD) = 11.6 daN (range 9.1 to 14.7 between screening centres. Left CC and MLO views averaged. Mean Dose for each image was 1.09 mGy (CC: 1.04mGy, MLO: 1.14mGy), varying from 0.55 mGy to 1.31 mGy between the centres. Compression force alone had a negligible impact on radiation dose (r^2 = 0.8%, p = < 0.001).	~59% of the mammograms in the study complied with the recommended Force range Substantial variations in mean compression forces between the breast centres. Variation between centres greater than within centres or for different machines
Ng (2017)(182)	Observational	Images from 17 countries Worldwide	136,700 images	Volpara for Force (F), Pressure (P), Contact Area (CA)	F: Women in NL received highest median F 13.8daN (but had >BV) Switzerland lowest 6.6daN. P: Highest median P in Malaysia (15.7kPa) and lowest Puerto Rico (7.6kPa). Women in NL, Italy and Malaysia more likely to experience higher P USA, UK and Belgium lower P.	Found large variation in practice due to differences in age composition and compression forces applied. Concluded that beyond breast compression behaviour, age, and breast composition (volume and density) there are other factors influencing compression.
Waade (2018)(284)	Observational	25,143 women Norway Screening		Volpara for Force (F), Dose (D),	Compression F, P and T increased relatively by 18.3%, 14.4% and 8.4% respectively, from 1st to 4th	Observed increasing values of breast compression parameters across consecutive screening examinations

	Between 4 consecutive screens for women	2 centres		Pressure (P), Thickness (T) Recommended range of compression force (10.8–17.7 daN).	screening examination (CC) 12.3%, 9.9% 6.9% (MLO).	
Moshina (2018)(188)	Observational 2014-2015	Norway	14,689 women	Volpara for Force (F), Dose (D), Pressure (P), Thickness (T), Dense volume (DV), Percentage density (%MD) Recommended range of compression force (10.8–17.7 daN).	Described correlations between compression parameters and breast density measurements. P -tive correlated with DV (rho = - 0.37 for CC and rho = - 0.34 for MLO). T -tive correlated with %MD (rho = - 0.56 for CC and rho = - 0.62 for MLO). R-squared values from the regression analyses indicated that breast compression parameters explained a very low proportion of variance in DV or %MD.	High quality, large descriptive study with from one screening service with similar protocols. Better designed than many studies because linear regression was used to corroborate associations between breast compression parameters and density (DV, %MD) adjusting for breast volume, age and BMI. Found that women with smaller breasts received higher pressure. Women with large breast volume have larger absolute dense volume but lower % density. They undergo lower pressures at screening. Raised a query about the way that Volpara calculates DV suggesting that variations in compressed thickness may result in differences in calculated DV. This implies that the further studies on the association between thickness and density and hence BC risk are required before tools like Volpara can be used as a means for stratification in screening.
Voight (2021) (180)	Observational	Sweden	11 women. Underwent compression by 2 different mammographers each at 2 different times. MLO and CC	Bespoke pressure sensor matrix and DICOM	Measured force, compressed breast thickness, force in field of view, contact area, mean pressure. ICC for both intra and interrater reliability were good for thickness (ICC=0.82) but poor for force applied (ICC= 0.42).	Small study but unusually has repeat measures on same individual. Found force was inconsistently applied but thickness was consistent.
Serwan (2021) (179)	Observational retrospective 2019	Australia	1,972 mammograms. Age 36 to 90	Volpara for Force (F), Pressure (P), Contact Area (CA)	Hypothesised that pressure of 10 kPa was optimum. Looked at variation in F and P. F: 98.6% of compressions were >5 daN, 16.6% were >10 daN, 0.0% >15 daN.	Study was descriptive (graphical) only with limited attempts to control for breast size and density. Concluded that pressure was more variable than force and that practical guidelines may increase the reproducibility of image acquisition.

				Compress until 'taut'.	P:94.5% of compressions were >5 kPa 36.0% >10 kPa 6.3% were >15 kPa.	
Dzidzornu (2021) (181)	Observational retrospective 2018-2019.	Ghana, 3 locations.	1,071 women Mean age 54 7 Practitioners	No target value.	Mean compression force (F): CC = 17.2 daN MLO = 18.2 daN F varied significantly (p = 0.0001) among practitioners. F increased significantly (p = 0.0001) with the years of work experience.	Although values were within the maximum target used by European countries they were much higher on average.
BC Detection and compression thickness						
Helvie (1994) (183)	Observational	Routine mammograms	250 paired MLO and CC images	DICOM Analogue film Thickness (T) and Force (F) from DICOM Phantoms used to measure dose	Geometric unsharpness increased by 8% and 19% when a 4.4-cm-thick breast was compared to a 4.8- and 5.4-cm-thick breast. A 5% and 12% loss of contrast was noted when a 4.4-cm-thick breast was compared to a 4.8- and 5.4-cm-thick breast respectively. Mean glandular radiation dose at 4.4, 4.8, and 5.4 cm was 1.40, 1.70, and 2.33 mGy, respectively. MLO views have on average greater thickness (+8%) versus CC view.	Early study on analogue film mammography. Controlled for breast characteristics by paired analysis. Found Increased compression leads to more geometrical unsharpness, and reduced contrast and increased dose.
Saunders (2008) (166)	Medical Physics simulation	USA		Digital Imaging Comparing simulations at different thicknesses (4 and 4.5 and 6 cm and 6.75cm). Under typical dose scenarios, using Monte Carlo modelling	Comparing simulations standard compression (as determined by thickness) and 12.5% reduced compression thickness for two different thicknesses (4 and 4.5 and 6 cm and 6.75cm). Decrease in breast compression (i.e increased thickness) led to linear increase in x-ray scatter. BUT By adjusting machine for ~ 10% increase in total tube output and 10% decrease in detector signal, found that breast compression be reduced by about 12% in breast thickness with little impact on image quality or dose.	This study analyzed how the inherent quality of diagnostic information in digital mammography could be affected by breast compression. As breast compression decreases (thicker breast) scatter increases as expected BUT The results suggested that that breast compression plays a less important role in lesion conspicuity for digital mammography than for screen-film mammography Incidental finding that reduced compression can lead to higher scatter fractions in some circumstances reducing pressure can even reduce dose because the extra distortion results in an increased fraction of the tissue in the higher dose area. Concluded that significant complexities inherent in dose calculations.

Salvagnini (2016) (165)	Medical Physics calibration using simulated lesions	Belgium	520 Images (130 CC images for each of 4 thickness strata) 4 radiologist readers	Volpara BIRADS (For matching images on density) Thickness (T) Categorised (1-4)	AUC for the four readers were 0.80, 0.65, 0.55 and 0.56 going from T1 (low) to T4 (high), ($p<0.05$ except for difference T3 and T4)	Detectability falls as breast thickness increases.
BC Detection and compression pressure or force						
Holland (2016) (189)	Cross sectional Images taken 2003 – 2011	Netherlands	>600 cases >100,000 screens MLO Images only Screening age population (50-75) Excluded interval cancers	Volpara for Force (F), Dose (D), Pressure (P), Thickness (T), Dense volume (DV), Percentage density (%MD) Breast volume (BV) Five pressure quintiles (P1-P5) (Low to high). Most images acquired using flexible paddles	For the five pressure quintiles (Low to high). PPV = 25.4 (21.5–29.2), 31.2 (27.1–35.4), 32.7 (28.6–36.9), 25.8 (21.9–29.7), 22.0 (18.4–25.6), Pearson's $\chi 2$ P=0.001 Cancer detection rate Cancers/1000 5.5 (4.6-6.5), 6.5 (5.5–7.6), 7.1 (6.0–8.2), 5.4 (4.4 – 6.3) 4.9 (3.9-5.8) Pearson's $\chi 2$ P = 0.011 Cancer detection and PPV are better in the moderate pressure group 9.18 kPa < pressure ≤ 10.71 kPa. The recall rate, false positive rate and specificity were not statistically significant from the expectation across the groups. Increased %MD was associated with increased P (12% of the variation in pressure could be explained by higher density R^2 value).	Only looked at screen-detected cancer and not interval cancers but study was large. First to that show cancer detection rate changes according to mammographic pressure applied. They showed that the best cancer detection occurred in the **middle ranges** of breast compression pressure NB median pressure is likely to be lower in the UK where different guidelines apply.
Holland (2017) (190)	Cross sectional	Netherlands Screening	>130,000 exams >57,000 women Included Intervals	As above Adjusted for %MD and BV	Screen detected cancers, Interval cancers false positives, and true negatives (based on interval cancer status) were determined ^%MD ^ pressure	If too much pressure is applied may reduce sensitivity If pressure is too low may reduce specificity.

			GEE to account for women imaged >1 time		Sensitivity based on interval cancers increased up to P3 middle range then declines (82%, 77%, 80%, 71%,71%) Specificity increased with increased pressure.	NB women are invited to screening every 2 years in the Netherlands therefore interval cancers are detected within 24 months of screen (in the UK this is withing 36 months).
Moshina (2017)(191)	Cross sectional 2007-2015	Norway	261,641 exams 93,444 women Average of MLO and CC used (separate rates also given) GEE to account for women imaged >1 time.	Volpara for Force (F), Dose (D), Pressure (P), Thickness (T), Dense volume (DV), Breast volume (BV), Percentage density (%MD) Recommended range of compression force (10.8–17.7 daN)	F and P categorized using tertiles as low, medium, or high. Adjusted for DV, BV and age. Included screen detected cancers and interval cancers. With increased F: Recall rate to assessment decreased PPV increased Specificity increased (P for trend 0.05 for all) With increased P: Recall rate increased PPV decreased Sensitivity, and specificity decreased (P for trend 0.05 for all) High compression pressure was associated with higher odds of interval breast cancer compared with low compression pressure (1.89; 95% CI 1.43-2.48).	Concluded that high compression force and low compression pressure were associated with more favorable early performance measures in the screening program. Linear across 3 groups of pressure/force level. Found an association between pressure and Interval cancers, which suggests that high pressure is associated with higher odds of BC. Interestingly the study adjusted for absolute breast density (rather than %MD as in the case of Holland above). This could explain the slightly different findings with respect to interval cancers in the highest pressure category. NB women are invited to screening every 2 years in the Norway therefore interval cancers are detected within 24 months of screen (in the UK this is withing 36 months).
Hill (2022) (192)	Matched case control	UK	326 Interval cancer (IC) cases 965 screen detected controls (SDC) matched on age and density grade	Volpara for Force (F), Pressure (P), Dense volume (DV), Volpara Density Grade (5th Ed) (DG)	P was categorised into tertiles. Logistic regression analyses used to identify significant predictors of IC versus SDC. Comparing the third and first tertile, P was associated with lower risk of IC versus SDC [0.64 (0.47-0.87)]	Higher pressure was associated with a lower risk of detecting an interval versus a screen detected cancer. Design of study does not look at screening performance as such but the characteristics of screens where cancer was detected.
Flexible Paddles (including pain)						
Hauge (2012) (196)	Medical Physics Calibration	Oslo	2 different sized paddles	Thickness measuring device and	For flexible paddles the largest difference between actual thickness and reported	Found that flexible paddle had higher thickness measurements error. This may mean that the studies that

			2 different compression forces (60 and 100 N).	breast phantoms.	thickness occurred for Hologic Lorad Selenia (18 cm x 24 cm paddle: +26.0%).	use average thickness measures are compromised when flexible paddles are used.
W K Ma (2014)(285)	Experiment	n/a	Using breast phantoms 4 machines 8 flexible, 8 fixed paddles were evaluated.		To determine if movement external to the patient occurring during mammography may be a source of image blur. Compression force readings for both fixed and flexible paddles decreased exponentially with time, while fixed paddles had a larger drop in compression force than did flexible paddles.	Compared with flexible paddles, fixed paddles have a shorter 'settling time'. Provides a possible explanation to mammography image blurring caused by extra patient movement. Therefore, to reduce the risk of blur, use fixed paddles if possible.
Broeders (2015) (286)	Observational	Netherlands	288 women screened 3 practitioners and 3 radiologists reviewed all images	Flexible paddle or Rigid paddle Screen using flexible paddle then took addition image with ridged paddle on same woman. Pain scores Image quality subjective	No difference in pain experience between both paddles (mean difference: 0.08 ± 0.08, p = 0.32). Mean radiation dose was 4.5 % lower with Flexible paddles (0.09 ± 0.01 p = 0.00). Image quality qualitative judgements – Rigid paddles show more breast tissue and have better contrast. Flexible paddles push more dense tissue to the chest wall.	Recommended Rigid breast compression paddles for MLO and CC views based on no difference in the pain scores. Based on subjective assessments concluded that whilst flexible gave better consistency for inclusion of pectoral muscle they tended to push fibroglandular tissue off the image posteriorly which could mean that cancers are missed. Early suggestions that the use of flexible paddles could compromise screening performance for no gain in terms of pain reduction.
Diaz (2017)(219)	Observational	n/a	Using breast phantoms	Hologic Selenia dimension equipment. DBT (tomosynthesis)	Looked at how tilted paddles influence the distribution of undesirable scatter in the receptor. Found different scatter patterns when Flex Pad rather than rigid paddle was used,	Concluded that post processing algorithms should take the paddle type into consideration when dealing with scatter.

Moshina (2019) (287)	Observational	Norway May-November 2017	4675 women 3 different compression paddles: Standardised pressure (10kPa) fixed Fixed paddle Flexible paddle	Pain score: 0:10. Severe pain = 7+ Questionnaire to assess pain scores	No significant difference between pain scores for pressure standardised paddle and the flexible paddle (average score 2.5) Mean pain score lower for the standardised paddle compared to the fixed paddle (2.4 versus 2.6 on NRS, $p < 0.05$) Adjusted for pressure and breast characteristics	No differences in pain scores between use of a pressure standardised fixed paddle and a flexible paddle.
Pressure (and pain and quality)						
Poulos (2003)(185)	Prospective Observational Experiment.	Australia	114 women Aged 50-69 6 Radiologists compared outcomes for diagnostic quality.	Additional CC image taken at a reduced level (-3 daN) Force and thickness measured from mamo unit.	Radiologists found no significant differences in image quality except for contrast resolution within the fatty area of the breast ($P < 0.05$). No linear relationship between applied compression force and compressed breast thickness and no relationship between the amount of compression force applied and reported discomfort.	A relatively small sample but first to show that many women did not experience a difference in thickness when the compression was reduced. Assigned the extra image a random location on viewer to avoid bias.
O'Leary (2011) (288)	Observational	Ireland	4,790 Images	Force and Image Quality assessed PGMI	Suggested that compression forces were too low because of the desire to avoid pain. Proposed mean force of 12.2 daN required for a perfect CC image and 13.0 da N for an MLO image.	Univariate analyses, not controlled for BV but early mention of the issue
de Groot (2013)(198)	Prospective Observational	Netherlands	196 women 291 CC images 299 MLO	Volpara for Breast volume (BV), Pressure (P), Pain score: scale 0-10 Severe pain = 7+ Compression protocol with	Mean pressure CC = 21.3 kPa ± 54% MLO 14.2 kPa ± 32% Negative correlation between BV and pain score ($\rho = -0.19$, $p < 0.01$). Logistic regression modelling estimations showed that mammographic breast compression with 10 kPa may significantly reduce the number of severe pain complaints with an average increase in breast thickness	Suggested that that mammographic breast compression with a standardized pressure of 10 kPa, corresponding with normal arterial blood pressure would reduce pain without compromising quality. Women with smaller breasts had greater pressures and greater pain.

				18.0 daN target force	of 9% for small breasts and 2% for large breasts. For an average 16.5% thickness difference in prior-current mammogram pairs, no differences in image quality and dose.	
de Groot (2015) (187)	Observational	Netherlands	433 women (screening age) 3 radiologists independently identified inadequate images	Bespoke area measurement to calculate pressure. Pain score 0-10	Compared different methods of standardization. With pressure standardization (10kPa) versus force (14daN) standardization : Thickness increased on average 4.2% (MLO) - 6.3% (CC) Pain scores were reduced by 10% (MLO) - 17% (CC) all p-values <0.05.	Suggested that pressure-standardised mammography gave better 'thickness' measures and lower pain scores than force-standardised approaches, concluding that pressure provides a better metric. Sample was too small to look at the effect on cancer conspicuity. but this is the first major study to propose pressure standardisation.
Agasthya (2016)(289)	Observational	USA	Tomosynthesis (DBT) 72 women	Average 'standard' force was 13.92daN (SD 3.52)	Compared pain at different compression levels. Took images using DBT at 'standard' and 50% reduced force and radiologists compared quality Mean force reductions of 48% and 47% for the CC and MLO views: Reduced objectively scored perceived pain level (P<0.05) Increase in compressed breast thickness (mean 0.38 cm (MLO view)) with no change in tissue coverage or increase in motion blurring.	Concluded that pain and discomfort can be reduced substantially with no loss in tissue coverage or increase in motion blurring by decreasing the compression force by 45–50% of the standard. 'Standard' was defined by a subjective value i.e. "force the patient asked compression to be stopped"
de Groot (2017)(290)	Observational 2009, 2014	Netherlands	188 stable lesions imaged using force standardisation in 2009 and Pressure standardised in 2014	Radiologists rated images side by side for lesion visibility, contrast, sharpness and preference	2014 compression forces were lower, (17% (MLO) 29% (CC) but thickness greater by average 2.4% (+1.4mm)	Concluded that 10kPa pressure-standardised compression has non-inferior visibility, contrast and sharpness for stable lesions compared to pain-tolerance limited 18daN target force compression. Well-designed study where observers were blinded for meta data and images presented in random sequence but did rely on subjective opinions of radiologists and stable lesions are not-typical of screening images.
Mammographer and film reading outcome						

Henderson (2015) (291)	Observational Cohort (1994-2009)	USA >1 million screens	300+ Mammographers	Screening Single reading 900k analogue 100k FFDM Outcome measures: RR, PPV, Sensitivity, Specificity	Measured sensitivity using interval cancers within one year of initial screen. Mixed effects logistic regression (clustered data) Variability in RR, sensitivity, specificity (all P<0.005) after controlling for radiologist. Variability in PPV was only evident in screening film not FFDM. Outcome was independent of equipment used.	Investigated the influence of mammographer on the radiologist's ability to interpret film. Concluded that mammographer did influence the interpretative (reading) outcome. Slightly compromised design by including mammographers taking as few as 50 images in period who could be inexperienced. Only study that found to explicitly look at the mammographer as the exposure rather than the technique as objectively measured.
Standards/Guidelines						
Perry (2008) (292)	Guidelines 2008+	Europe	n/a	For quality assurance in breast cancer screening and diagnosis.	Acceptable dose is <2.5 mGy Desirable dose <2.0 mGy "The radiographer must ensure that the breast is properly compressed, but no more than is necessary to achieve good image quality" "Compression reduces the radiation dose, scattered radiation, movement blur and the thickness of the breast which reduces overlapping of tissue shadows"	These guidelines were set out in 2006 and have not been altered substantially in the subsequent updates. Guideline on compression is largely subjective.
Perry (2013) (171)	Guidelines supplement 2013+	Europe	n/a	As above	Supplementary material assumes that there is a 'standard' force of 10daN	
NHSBSP (2006) (2011) (2017)(169, 173, 257)	Guidelines 2006+	UK		As above	Earlier guidance states that the force on the x-ray machine should not exceed 20daN. Later guidance is limited to more subjective: The compression" should be applied slowly and gently to ensure that the breast is held firmly in position"	
Waade et al (2017) (172)	Guidelines	Norway	n/a	As above	Norwegian Breast Cancer Screening Programme recommends a breast compression force range between 11 to 18 daN	
Comparative Values FFDM/Tomosynthesis etc, Pressure-controlled force-controlled paddles						

Waade (2020)(293)	Observational Cohort 2016-2017	Norway	21,729 Screens 11,056 FFDM 10,673 Tomo Aged (50-69)	Volpara for Force (F), Dose (MGD), Pressure (P), Thickness (T), Dense volume (DV), Breast volume (BV), Percentage density (%MD)	For comparison the FFDM mean values were as follows: Mean (left/right combined) P CC = 13.6 kPa P MLO = 9.5 kPa F CC = 107 N F MLO = 121 N T CC = 58.8mm T MLO = 60.4mm MGD (overall) 2.95 mGy Study used linear regression adjusted for Age, Breast Vol, Breast dense vol. to compare pressure (P) force (F) thickness (T), dose (MGD) in tomosynthesis versus standard FFDM Women undergoing tomosynthesis received lower force and pressure and thickness was lower. Dose was greater than women undergoing standard FFDM.	Concluded that need more studies to investigate how differences impact image quality and screening outcome. Also need standards
Jeukens (2019)(294)	Observational	Netherlands	3,188 examinations.	Sensitive Sigma Paddle Force standard 17–18 daN Pressure target of 10 kPa	Compared pressure-controlled paddle with standard force controlled paddles. Differences between groups were tested using the Mann-Whitney U test. No clinically relevant differences in compression thickness, force, pressure, dose, or pain score between the force- and pressure-controlled paddle.	Wide variations in pressures and forces applied were observed but it was not dependent upon which paddle was used
Other related						
Taplin (2002) (193)	Retrospective cross sectional	USA	492 screen-detected and 164 interval-detected	Small study on positioning	Interval cancer detected cases failed the positioning criteria significantly more than screen detected cancer cases. (OR, 2.57; 95% confidence interval, 1.28–5.52%)	Small study suggests that BC detection by mammography may be improved through attention to correct positioning.
Dustler (2016)(295)	Thesis	Sweden	Slant Paddles Pressure		Designed and fitted pressure sensors The flexible compression plate appears to quite effectively redistribute compression force from the juxtathoracic area to the central breast. Still, most of the compression force is not distributed to the breast in MLO.	Designing a new compression plate that is flexible in two degrees of freedom rather than just one might be an improvement.
Metsälä (2017)(296)	Systematic Review	Europe	Summarises challenges for European		The aim of this study was to identify European radiographers' challenges in clinical performance in mam-	Concluded that the introduction of harmonized mammography guidelines across Europe may serve as an

			Mammographers		mography and the main areas that require more and/or improved training. Includes discussion of compression standards, PGMI	evidence-based tool to be implemented in practice and education.
Serwan (2020)(297)	Systematic Review to December 2019	Worldwide 83 articles found 18 included in qualitative synthesis	Force and Pressure Standardisation		Searched for to 'mammograph*' as a database entry. Further keywords added to the search strategy included (AND) 'compress*' OR 'force' OR 'pressure', AND 'standard*'. Compression pressure of approximately 10kPa was found to decrease pain, with a negligible effect on breast thickness, dose and resultant image quality Could ultimately aid in detection of early stage BC (evidence limited to Poulos A, McLean D. Mercer)	Supports an alternate standardised compression protocol based on pressure; this approach accounts for individual breast characteristics in a 'personalised' manner. Majority of high--quality studies included in qualitative analysis were from one country i.e. Netherlands. There is still limited evidence about widespread clinical application.
Murphy (2015) (184)	Qualitative interviews (2014)	UK	41 practitioners split into Focus groups		The findings reveal common themes in mammographer thinking, including client empowerment, white-lies, time for interactions, uncertainty of own practice, culture, power, compression controls, digital technology, dose audit-safety nets, numerical scales. The culture and the practice of the units themselves influenced beliefs and attitudes.	A structured qualitative study. Offers insights that may explain the wide variation in compression practice. Compression force was applied in many different ways due to individual practitioner experiences and behaviour.

Abbreviations: ANOVA: Analysis of Variance; **BV: breast volume; BC** breast cancer; BD: breast density; BV: breast volume, CC: cranio-caudal view; FFDM: full field digital mammography; DICOM: Digital Imaging and Communications in Medicine; IRR: incidence-rate ratio; MGD: Mean glandular dose; MLO: mediolateral oblique view; PGMI: Perfect, Good, Moderate, Inadequate; VDG (Volpara density grade).

D APPENDIX D Literature review of Volpara tool with respect to similar assessment tools

Search questions and strategy

A broad search was carried out in 2015 to find published literature where the Volpara tool had been used and to locate comparative reviews where the performance, validity, and reliability of Volpara was assessed in comparison to similar tools. A simple search using the term “Volpara” was conducted. Additional relevant research was identified from the references in the Volpara papers. 58 candidate papers were identified, and they were filtered for relevance against the criteria above and summarised in table D.1.

Search results and summary table

Table D.1 Summary of key papers which compare and evaluate density assessment methods pertinent to Volpara

First Author /Year	**Type/Period**	**Location**	**Studies/ Cases**	**Assessment Methods/ Tools**	**Key findings – effect size, correlation** Hazard ratio (HR),Relative risks (RR), Odds Ratios (OR) and 95% Confidence Intervals (CI), Area under curve (AUC), Spearman's correlation Rho, Pearson's R correlation	**Comments including strengths/limitations**
Shepherd (2011)(130)	Case-control (2004-2006) Comparing area and volumetric methods	USA Screening programme	275 interval cancer cases 825 age matched controls Models to discriminate cancers versus non-cancers	SXA (Volumetric) CC views only %BD by area %BD volume and DV	After adjustment for age, BMI, reproductive and familial risk factors found. ORs for breast cancer risk in the highest versus lowest measurement quintiles were 2.5 (95% CI: 1.5–4.3) for percent dense area, 2.9 (95% CI: 1.7–4.9) for FGV, and 4.1 (95% CI:2.3–7.2) for percent dense volume.	Concluded that volumetric measures of breast density are more accurate predictors of interval BC risk than percent dense area.
Kallenberg (2013)(298)	Compared temporal pairs of digital screening exams	Netherlands Screening exams	Compared 42,414 density estimates. 1.5 to 2.5 years apart	Volpara %BD DV	Pearson's correlation coefficient: 0.90 (0.90-0.91 95 %CI). On average, density decreased slightly over time.	Found high temporal stability between Volpara estimates in longitudinal study.
Wang. (2013) (214)	Compare the agreement of automated breast composition measurement with MRI measurement	USA Screening Setting	99 women with no cancer who had both MRI and Mammograms	SXA Quantra Volpara MRI %BD BI-RADS	Volpara showed highest correlation log FGV and MRI Pearson r^2 0.63 but SXA showed higher correlation in %BD Quartile groupings for %BD measures were compared using weighted kappa statistics Volpara performed best with substantial agreement (K=0.74) between MRI and Volpara categories.	Authors concluded that classification by volumetric density using any of the three techniques is comparable to classifications by MRI density.
Seo (2013) (218)	Compare the agreement of automated BI-RADS and radiologists BI_RADS classification	S. Korea screening setting	193 normal images	Volpara versus Visual BIRADS	ICC between Volpara BIRADS classification and 3 radiologists was good ICC = 0.757 DV showed a highly significant positive correlation with visual assessment (Spearman's Rho 0.754, $p < 0.001$).	This early study was small and all radiologists came from the same screening location but showed that Volpara BI-RADS classification showed strong correlation with assessment by experienced clinicians.

					Found reduced agreement when there were areas of scattered density.	
Engelken (2014)(299)	Assessed reproducibility using temporal pairs of digital screening examinations	Germany screening setting	174 pairs of normal images Up to 2 years apart	Quantra %BD DV	Pearson correlation coefficients of for DV = 0.947 (P<0.05) for %BD = 0.920 (P<0.05)	Found high temporal stability between Quantra volumetric estimates in longitudinal study to a similar degree to Kallenberg above.
Eng (2014) (132)	Case control Compared 6 different methods of breast density estimation	UK breast symptomatic clinic for cases and screening programme for controls	414 cases 685 controls	Quantra, SXA Volpara BI-RADS, Cumulus ImageJ %BD DV	After adjustment for age, BMI and reproductive variables, increase in risk per standard deviation increment in %BD was highest for Volpara Volpara: 1.83 (95% CI: 1.51-2.21) Cumulus: 1.58 (95% CI 1.33-1.88) ImageJ-based method had a slightly higher ability to discriminate between cases and controls for %BD: Image-j AUC = 0.68, (95% CI: 0.63-0.73) Volpara AUC = 0.67, (95% CI: 0.62-0.72) Cumulus AUC 0.65, (95% CI 0.60-0.70) Image-J failed to process 11% of the images Volpara and Cumulus had 0% failure rate.	Concluded that fully-automated methods are valid alternatives to the labour-intensive "gold standard" at that time i.e. Cumulus for quantifying density in FFDM. All methods were also positively associated with BC risk but the association with Volpara was strongest. Volpara was the most robust and practical in clinic settings.
Brand (2014)(209)	Prospective cohort 2011-2014 Compared Volpara density distributions with Cumulus. Examined the agreement by side and view.	Sweden screening setting	>40,000 women 206 cancer cases. Mean age at entry = 55	Volpara Cumulus %BD DV	Correlation between total breast area (Cumulus) (Pearson's R) and total breast volume (Volpara) was high: Absolute density R= 0.91, (95% CI: 0.87–0.94), %BD R= 0.93 (95% CI: 0.89–0.96) HRs for BC in the highest versus lowest measurement quartiles were 2.93 (95% CI: 1.73–4.96) for %BD 1.63 (95% CI: 1.10–2.42) for DV After adjustment for reproductive, demographic, age and BMI risk factors. .	Volpara is practical a high-throughput setting. Comparison of density distributions across different automated analysis systems found good agreement between Volpara and more established methods of BD assessment. Volpara breast density was a strong predictor of BC risk which concurs with previous studies.
Gubern-Merida (2014) (215)	Observational 2000-2011 Compared Volpara	Netherlands	250 studies (132 women) Non-cancers	Volpara MRI (median time 6 days	Correlation between MRI and Volpara BV R= 0.97 %BD R= 0.93 DV R= 0.85	Found that Volpara is an accurate and practical method for assessing breast volumetric measurements using MRI as a gold standard but tends to underestimate breast density in high density breasts compared to MRI.

	estimates with MRI estimates.		Mean age 46.5 years	between images) %BD BI-RADS	Pearson's correlation coefficients (%BD) for Low Medium and High density tertiles: 0.93, 0.97 and 0.85 respectively.	The women in this study have a lower average age to UK screening populations and included more premenopausal women.
Alonzo-Proulx (2015)(133)	Test-retest Reliability tests. March-May 2013	USA Screening setting.	31 CC images taken and repeated on same day Non-cancers	Cumulus ABD, CumulusV Quantra Volpara	Within-breast BD standard deviations were Volpara: 0.99%, (95% CI: 0.79, 1.33), Quantra: 1.64%, (95% CI: 1.31, 1.39) Cumulus ABD: 3.32% (95% CI: 2.65, 4.44) CumulusV: 3.59% (95% CI: 2.86, 4.48),	Compared test retest estimates for and found that Volpara was most reliable tool. The study has the advantage that it uses the same women on the same day but the sample size is quite small.
Lee (2015) (211)	Observational 2011-2012 Compared Radiologist findings to Volpara BIRADS	S Korea Screening and symptomatic setting.	860 women Mean age 54.7 Non-cancers only.	Volpara BIRADS	Agreement between experienced radiologist and Volpara BI_RADS classification: Weighted Kappa K=0.80 Where there was a difference in bilateral density the agreement was lower.	Substantial agreement between radiologists and Volpara. Findings suggest that bilateral density asymmetry may make it harder to visually assess density.
Winkel (2015)(94)	Case control 2007 2 radiological readers 3 methods of assessing BD.	Denmark Screening setting	122 cases 262 controls	BIRADS Tabar Bespoke computerised Interactive Thresholding (%BD) to assist in density assessment.	Radiologists had substantial inter-observer agreement for BI-RADS (K=0.68) Tabár (K=0.64) High/low-risk BC agreement (determined by defining the following categories as high-risk: BI-RADS's D3 and D4, Tabár's PIV and PV and upper quartiles of %BD) RR of BC was estimated using logistic regression to calculate OR adjusted for age, which were compared between the two readers. The two readers judged 5% (Automated %BD tool), 10% (Tabár) and 13% (BI-RADS) of the women to different high/low-risk Categories	Study looked at the need for automation by investigating inter-observer agreement of three different methods one that was computer assisted. ICC agreement was substantial but the summary BC risk categories showed less agreement and if these tools are to be used for stratification this could be problematic. The Authors concluded that "Automated computerized techniques are needed to fully overcome the impact of subjectivity"
Holland (2016)(210)	Observational Inter examination correlation	Netherlands screening setting	500 women each with 2 screening examinations	BIRADS Volpara versus	Percentage agreement between exams for (BIRADS high or low): Volpara = 90.4% (CI 87.9–92.9%) Human readers ranged from 86.2% to 89.2%	Found that inter-examination agreement for Volpara was higher and hence concluded that it is more reliable than visual assessment.

	between 2 sequential screens		2 years apart.	Radiologist		However some change in category is expected over time as women get older and %BD declines or if women gain/lose weight.
Brandt (2016)(217)	Case control AUC comparisons 2006-2012	USA Screening setting	1,911 cases 4,170 controls	BIRADS Quantra Volpara	Clinical BI-RADS AUC=0.60 (95% CI: 0.58, 0.61) Volpara AUC=0.58 (95% CI: 0.56, 0.59) Quantra AUC=0.56 (95% CI: 0.54, 0.58) Volpara classified 51% of women as having dense breasts, Quantra 37%, and clinical BI-RADS 43%.	Clinical BI-RADS assessment showed best discrimination of case status but AUC was similar for all methods. Authors concluded that use of BI-RADS classifications may not be best approach when using volumetric methods. The number of women who fall into each category of BD depends upon the tool used. This could have implications for clinical practice and for stratification. When assessing longitudinal changes in BD then it is important to use the same technology.
Jeffers (2017)(300)	Case control Compared three metrics of DB as predictors of future BC risk 2004-2013	USA Screening setting	125 cases 274 age and race matched controls	Volpara Cumulus 6 for area-based Assessment (i.e. visual analogue)	Model were adjusted for age, race, BMI, parity, and menopausal status. ORs for women with extremely dense breasts compared with those with scattered areas of fbroglandular density: Cumulus BIRADS = 2.06 (95% CI: 0.85, 4.97) Volpara BIRADS = 2.05 (95% CI: 0.90, 4.64) Cumulus BIRADS AUC = 0.68, (95% CI: 0.63, 0.74) Volpara BIRADS AUC = 0.64 (95% CI: 0.58, 0.70) Volpara %BD AUC = 0.66 (95% CI: 0.60, 0.72), Volpara DV AUC = 0.65 (95% CI: 0.59, 0.71)	Volpara (volumetric) automated classification was as accurate as area-based computer-assisted methods for discrimination of patients from control subjects. For AUC Cumulus performed better than Volpara estimates but differences were not statistically significant.
Astley (2017)(301) Astley (2018) (302)	Observational from cohort of women were participants in "Predicting Risk Of Cancer At Screening" (PROCAS) study	UK Screening setting	366 screening cases 338 interval (114) or next screen detected cancers (224) 3 controls per case were matched on age, BMI, hormone	Cumulus (i.e. visual analogue) Volpara Quantra Densitas	(OR) between the highest and lowest quintile, based on the density distribution in controls adjusting for classic risk factors:: For screen detected cancers: Cumulus OR = 4.37 (95% CI 2.72–7.03) Volpara OR = 2.42 (95% CI 1.56–3.78), Densitas OR = 2.12 (95% CI 1.30–3.45) Quantra OR = 1.02 (95% CI 0.67–1.54) For interval cancers: Cumulus OR = 4.48 (95% CI 2.79–7.18) Volpara OR = 2.87 (95% CI 1.77–4.64),	This study shows like Jeffers above that semi-automated visual methods have a strong association with BC detected at screen or in the medium term after screen. The study chose to combine interval cancers and next round screen detected cancers which is somewhat unusual because they are detected in a different way but the results are consistent with other studies. The authors concluded that visual methods were not practical for largescale use and that Volpara provides a practical method for risk stratification and performed better than other automated tools tested.

			replacement therapy use and menopausal status.		Densitas OR = 2.34 (95% CI 1.50–3.68) Quantra OR = 1.32 (95%CI0.85–2.05).	

Abbreviations: AUC: Area under the curve; **BC:** breast cancer; BIRADS: Breast Imaging and Reporting Data System; BMI: Body mass index; BV: Breast volume ; DV Dense volume; FGV Fibro-glandular tissue volume %BD Percentage breast density either area of volume based as specified; ICC: Inter class correlation ; NDV: Non-dense volume**;** CC: cranio-caudal view; MLO: mediolateral oblique view; VDG: Volpara density grade.

E APPENDIX E – Validity and reliability of Volpara breast fluctuating asymmetry estimates

Introduction

Previous studies have shown a high correlation between Volpara volumetic measurements and volumetric measurements derived from sophisticated imaging methods such as CT and MRI, which are considered ground truth values (Chapter 2). The literature search however, yielded no research into the validity or reliability of breast *fluctuating asymmetry* estimates derived from Volpara or other automated quantitative tools. Therefore, a small study was undertaken to investigate this as a precursor to the use of asymmetry measurements in my main studies.

Asymmetry estimates are more susceptible to error than overall averaged density estimate since they are based on single readings rather than aggregated breast volume and breast density measurements. The aims of this small study were to assess:

a) how sensitive Volpara asymmetry estimates are to outliers and to investigate whether extreme values represent genuine cases of true asymmetry or whether they are a result of imaging issues such as poor positioning or inadequate pressure caused by either operator error or by constraints relating to the woman (such as a frozen shoulder or wheelchair use). A secondary objective was to propose a practical method for differentiating between genuine cases of extreme asymmetry and cases of asymmetry that are likely to be due to imaging/positioning problems,

b) whether there is concordance between estimates of asymmetry calculated using the CC view and the MLO view,

c) the reliability of asymmetry measurements using test-retest examinations from women screened twice in the early study period.

Methods

The study participants were females aged 50–70 years as described fully in Paper II (Chapter 6), who underwent the NHSBSP standard 2- view (craniocaudal (CC) and mediolateral- oblique views (MLO)) mammography of each breast at the South-West London Breast Screening Service between March 2013 and August 2016. Raw digital mammographic images were processed via the automated algorithm Volpara® DensityTM v. 1.5.11 (Volpara), (Matakina Technology Limited, Wellington, New Zealand) (208); this algorithm provided fully automated estimates (in cm3) of the volume of the breast (BV) and the volume of the radio-dense tissue (DV) separately for each of the four [left (L) and right (R) breasts/CC and MLO views] images and average values were derived using all available measures.

Both absolute and relative measures of left-right asymmetry were calculated: absolute asymmetry (in cm3), i.e. the unsigned difference between left BV (or DV) and right BV (or DV). This was calculated for CC and MLO views separately.

The distributions of the asymmetry values were plotted. Natural-log transformations were applied to normalize the distributions.

Potential outliers were identified using the conventions described by Tukey (303) who used quartiles (Q1-Q4) and interquartile ranges (IQR) to propose that the 'fences' for major outliers lie at Q1 – (3*IQR) and Q3 + (3*IQR). In a similar way moderate outliers adopt the 1.5*IQR convention for identifying fences.

A subset of 50 women with the greatest DV asymmetry was selected for further investigation. These extreme values had not been rejected on the basis of internal checks by Volpara and did not have a screen detected cancer at this screen. First the electronic screening records were examined for each case by an experienced radiologist and any cases where technical issues had arisen were excluded. For the remaining cases the images were retrieved and assessed subjectively by a consultant breast radiologist who examined both the CC and MLO views and checked visually for concurrence between views. The clinician classified the observed asymmetry according to the perceived cause: imaging issues, clinical conditions or natural asymmetry.

To test for concordance between MLO and CC estimates, and to look for any systematic biases in the asymmetry measurements, intra-class correlation coefficients (ICC) were calculated, and Bland-Altman plots were constructed for BV and DV asymmetry to plot the differences between CC and MLO asymmetry estimates against mean asymmetry. Images from women with cancer detected at this screen or previous history of BC were excluded as were women with breast implants and those where the standard 2-view mammography was not possible (e.g. women with multiple mosaic images). Images rejected by Volpara were also excluded as were images taken during non-routine appointments (e.g. second stage screening "assessment" examinations). This left 54,591 examinations for this preliminary study as described in Paper II Table 1 (Chapter 6).

To assess the reliability of the asymmetry measurements a subset of 464 women were selected who had undergone two routine mammography examinations in the period. Between test differences were calculated for BV and DV asymmetry. ICC was calculated and Bland-Atman plots constructed to look for systematic bias.

Results

The characteristics of the study population of over 50,000 women are shown in Paper II table 1 (Chapter 6). The distributions of BV and DV absolute asymmetry estimates were right skewed as (Paper II Figure 1) with a small number of large outliers.

Outliers and radiologist's review

The subset of 50 extreme outliers had DV asymmetry estimates ranging from 237 cm^3 to 86.80 cm^3 whereas the whole study group had a median absolute DV asymmetry of 5.6 cm^3 (Paper II Table 1).

Table E.1 Summary of Radiologists' review of cases of extreme asymmetry

Radiologist's Finding	**Examinations**
Type 1 – Technical/Imaging Issues	**n=25**
Mosaic images – multiple images required and large breasts	4
Positioning issue	4
Technical recall imaging error	1
Equivocal - Non-concurrence between views MLO/CC but no obvious reason	16
Type 2 – Clinical issues	**n=5**
Previous cancer surgery	2
Special appointment client had mobility issues	2
Special case - woman declined but had cancer	1
Type 3 – Natural asymmetry	**n=20**
True clinical reason (e.g. large cysts, hamartoma)	2
Concurrent MLO/CC but no clear reason for asymmetry	18

Table E.1 shows that 25 cases of extreme asymmetry were explained by technical reasons of these 9 cases (the mosaic images and the technical recalls) can be excluded as non-standard image sets. In 16 cases the MLO and CC records did not concur i.e., the asymmetry was evident from one view only. There was no obvious reason for the difference but in general imaging issues such as positioning, pressure, thickness deemed to be the most likely cause by the radiologist. These are the equivocal cases which are probably, but not certainly, invalid asymmetry estimates.

We found 20 cases of natural asymmetry of which 2 were explained by specific clinical findings. The extreme asymmetry was evident in both MLO and CC imaging. Therefore, the asymmetry finding is valid in these cases given the available evidence and we should not exclude these cases from any

analysis. They appear to represent genuine natural extreme asymmetry. A further 5 cases were explained by clinical issues that were uncovered during the record review. These could be excluded from future studies.

The proportion of examinations that would be classified as 'major' outliers under Tukey's convention would be relatively small (~0.6%, data not shown) and therefore represent a relatively small number of cases.

MLO Asymmetry versus CC Asymmetry concordance

MLO/CC concordance across the whole dataset was quite low (ICC: 0.34 (95% CI: 0.33, 0.35)) but there is no evidence of differential bias as shown in Fig E.1. The average MLO asymmetry values for DV are slightly higher than the corresponding CC asymmetry values.

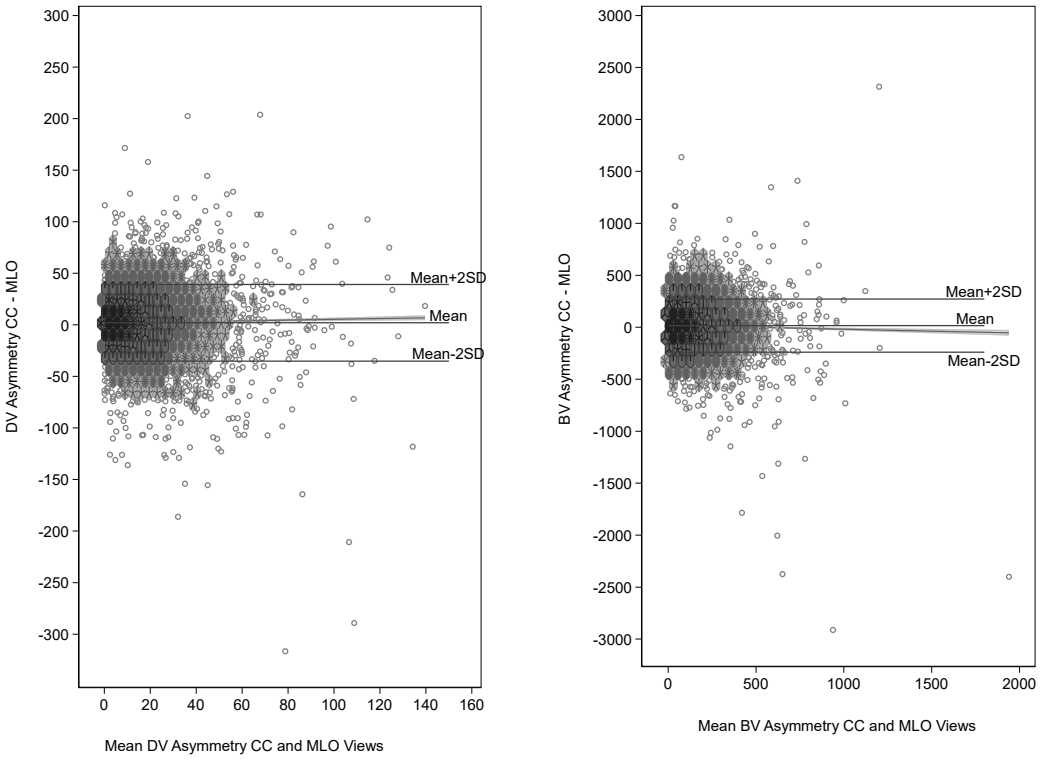


Figure E.1 Bland-Altman plots of dense volume (DV) and breast volume (BV) asymmetry for both CC and MLO views taken in same imaging session

Footnotes: All Dense volume (DV) and Breast Volume (BV) estimates are in cm^3. The lower and upper limits of agreement are shown (i.e. mean difference CC-MLO ± 2 standard deviations).

Temporal asymmetry concordance

The 928 mammograms (464 pairs) in the temporal study were taken on different occasions at least 6 months apart and with a maximum of 28 months difference in timing. In 14 cases a cancer was detected at the second screen. The mean age of the women was 53.97 which is younger than the average screening age of the full data set mean age 58.64 (see Paper II Chapter 6 Table 1). This is due to the fact that the temporal-pairs include a higher proportion of younger women on the high or medium risk screening protcols (i.e. screened more frequently). This however does not negate the general applicability of the reliability test. The Bland Altman Plots (Figure E.2) show that across all tests there is no systematic bias with the outliers scattered non-differentially around the mean. In each case the mean difference between the two tests is approximately zero

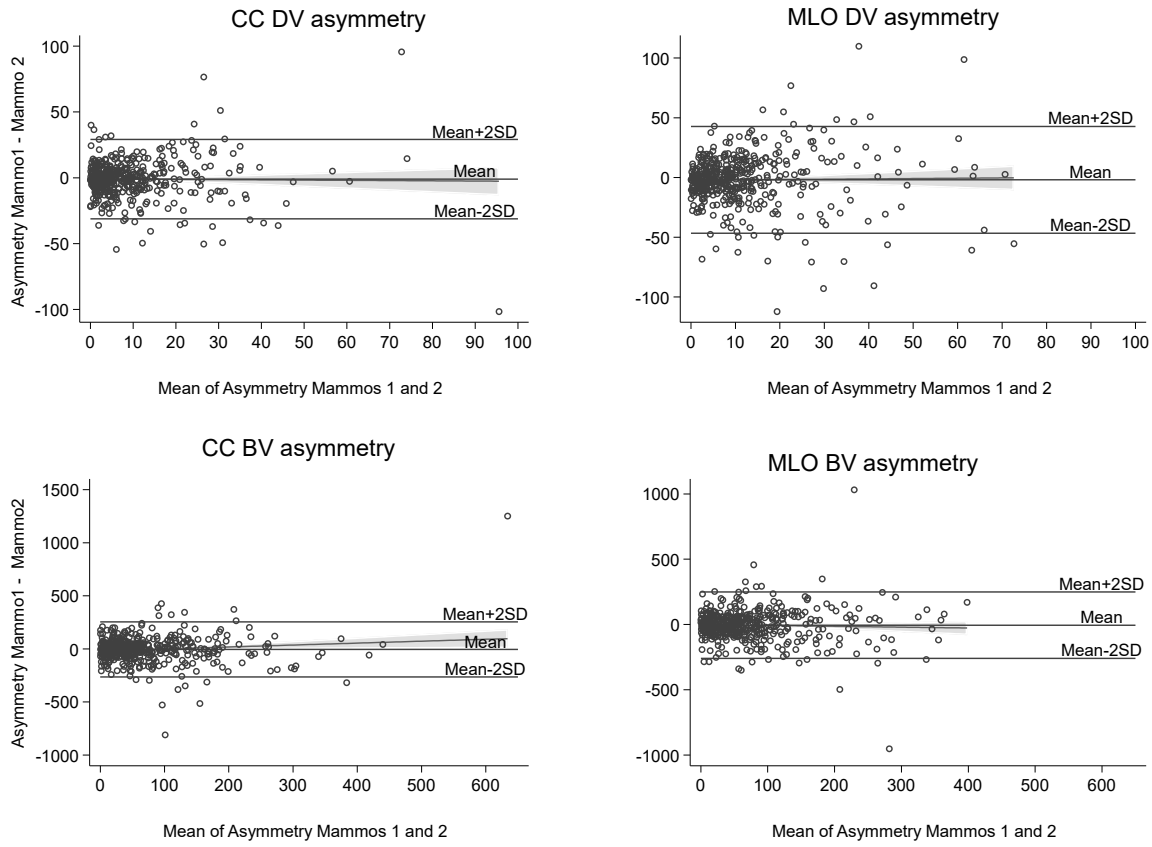


Figure E.2 Bland-Altman plots of dense volume (DV) and breast volume asymmetry (BV) asymmetry for both CC and MLO views taken at sequential screens (Mammo 1 and Mammo 2)

Footnotes: All Dense volume (DV) and Breast Volume (BV) estimates are in cm^3 .The lower and upper limits of agreement are shown i.e. mean difference Mammo 1-Mammo 2 ± 2 standard deviations).

A comparison of the DV Bland Altman plots and the ICC tests for CC and MLO views show that the upper and lower limits of agreement are narrower for the CC view asymmetry.

Table E.2 Intra class correlation for asymmetry test-retest examinations

Asymmetry measure[a]	**Number of pairs (n=464 examinations)**[b]	**Intra-class Correlation Coefficient (ICC)**	**Bland-Altman limits of agreement 95%**
		(95% CI)	**Mean (95% CI)**
DV asymmetry	464	0.566 (0.504, 0.628)	(-30.6, 28.6)
BV asymmetry	464	0.425 (0.351, 0.500)	(-259.1, 249.4)
%DV asymmetry	464	0.341 (0.260, 0.421)	(-5.76, 5.36)
Sensitivity analysis[c]			
DV asymmetry	440	0.771 (0.644, 0.757)	
BV asymmetry	446	0.643 (0.589, 0.697)	
%DV asymmetry	445	0.601 (0.542, 0.660)	

Footnotes:

[a] Asymmetry calculated from Volpara estimates as absolute value of (CC view left – CC right side) for DV BV and %DV for each examination (i.e. Test 1 and Test 2) separately. Between test differences calculated for DV BV and %DV separately as Value for Test 1 – Value for Test 2..

[b] Each pair represents images taken from the same woman but at two separate screening events. For screening examination to be included it must have with exactly 4 images taken, only screening appointments are included but high risk and moderate risk women were included.

[c] Sensitivity analysis carried out by excluding any tests pairs that were outside the Bland Altman limits of agreement

Discussion

Removing major asymmetry outliers would reduce the number of examinations by around 0.6% but some of these asymmetry cases represent natural asymmetry. It is not possible to automatically classify all exception cases from our available data, for example where there is a true clinical reason for asymmetry such as large cysts, Poland syndrome (rare), hamartoma etc. They occur naturally in the screening population and it is appropriate to retain these cases. From the MLO versus CC concurrence tests, we conclude that although CC asymmetry and MLO asymmetry may not always strongly concur, there is no evidence for differential bias.

Reliability was assessed using the intraclass correlation coefficient (ICC) which quantitatively summarizes how strongly units in the same group resemble each other. In this study we assessed case how strongly the two asymmetry measurements from the same woman, taken at two different times, are correlated. ICC values close to 1 indicate a high degree of measurement consistency. The most reliable measurements in our study were from CC views. The ICCs from this study are quite sensitive to extreme values for example, for CC view DV asymmetry ICC is increased from 0.566 to 0.711 by the exclusion of just 16 cases which have test differences which fall outside the Bland

Altman 95% limits of reliability. It is interesting to note that 2 of the 16 cases excluded using these limits are cancer cases, leading to a suggestion that temporal changes in asymmetry could be due to tumour development in one breast which has increased breast volumetric asymmetry. There was no evidence for systematic bias in the asymmetry estimates but individual estimates of asymmetry assessed using the Volpara algorithm are not as reliable or repeatable as measures of average overall breast density and breast volume reported in similar studies (133) although it is worth noting that the difference in time between asymmetry test measurements in our study, was much greater than those in similar test-retest studies using breast density only, where tests were repeated on the same day.

Conclusions

Asymmetry studies should exclude cases where asymmetry measurements are likely to be biased by known technical and clinical issues only i.e.

- Technical recalls – imaging deemed not adequate by film reader,
- Examinations where just one side was imaged or the CC images were missing,
- Examinations where the Volpara algorithm rejected either the CC image or the MLO image (on the basis of its own internal checks (208))
- Examinations taken at non-routine events (e.g., assessment clinics),
- Examinations on women who had previously been diagnosed with BC,
- Examinations where multiple >4 (mosaic images) were required.

There was no evidence for systematic bias in the asymmetry estimates calculated using the outputs from the Volpara algorithm, but individual estimates of asymmetry are not as reliable as measures of breast density and breast volume reported in similar studies.

F APPENDIX F – Supplementary Analysis - Assessment and arbitration rates stratified by dense volume asymmetry

In Paper III we found an association between DV asymmetry and interval cancers i.e. FN at screening. It has been reported that radiologists may find it more difficult to interpret bilateral mammograms that display greater asymmetry between the breasts (211, 304), due to the 'obfuscation' effect of increased L:R asymmetry. Therefore, one plausible, contributory factor, for our finding is that the degree of asymmetry hindered the film readers in some way. Although this could not be tested directly it was possible to calculate the assessment and arbitration rates and assess whether there was any variation in these rates between women with different degrees of DV asymmetry. When a radiologist or film reader is equivocal about a set of images they will recall for assessment. Because all screening films are double read the other reader may or may not agree with the recall. Where readers do not concur an arbitration process takes place with a third reader or consensus read deciding whether the woman is recalled for assessment. I therefore hypothesised that a film reader will approach perceived L:R asymmetry in the breasts with more uncertainty about the findings and more likelihood of recall for further assessment tests even in women whose outcome is normal i.e. no cancer found at screening and not an interval cancer case. A short supplementary analysis was conducted to look at preliminary findings outside the scope of the original objectives of my thesis.

The raw odds and rates for recall to assessment and arbitration for normal outcomes at different levels of DV asymmetry show an increase in recall and assessment rate across the DV asymmetry gradient:

Table F.1 Raw Arbitration and Recall to assessment rates in non-cancer screens stratified by DV asymmetry quintile

	DV Asymmetry Category [a]					
	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	*P* trend
Screened [bc]	16,229	16,133	15,991	15,939	14,596	
Assessed [d]	708	740	724	769	807	
Assessment Rate	4.36%	4.59%	4.53%	4.82%	5.53%	<0.001
Arbitrated [e]	3,125	3,204	3,213	3,220	3,130	
Arbitration Rate	19.87%	20.41%	20.70%	20.80%	22.12%	<0.001

Footnotes:

[a] Quintiles of the distribution in non-cancer cases used to categorise absolute DV asymmetry where volume was derived from CC images

[b] Screens included must have at exactly 4 images taken, only screening images are included, excluded are images that were rejected as outliers by the Volpara algorithm.

[c] Screens for women known to have previous breast cancer were also excluded.

[d] Screens where women were recalled for additional tests on the basis of the imaging but were found to be normal after assessment.

[e] Screens where the first and second reader did not concur or where both readers recalled for assessment but were found to be normal after assessment.

DV asymmetry is however, positively associated with DV and women with dense breasts are more likely to be recalled for further assessment. After adjustment for absolute DV the association between assessment rate and DV asymmetry was attenuated (P-trend = 0.06), women in the highest fifth of the asymmetry distribution having 1.35 times the odds of women in the lowest category (p=0.028).

After adjustment for absolute DV asymmetry the association between DV asymmetry and arbitration rate disappeared suggesting that the differences were explained by the variation in DV (Ptrend=0.606).

This preliminary analysis shows that the presence of high levels of DV asymmetry in the mammographic image may lead the film reader to recall even after the effect of DV is taken into consideration. This may be because readers are consciously or sub-consciously aware of the additional risks associated with L:R differences across the breasts.

REFERENCES

1. Gastounioti A, Conant EF, Kontos D. Beyond breast density: a review on the advancing role of parenchymal texture analysis in breast cancer risk assessment. Breast Cancer Research. 2016;18(1):1-12.
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-49.
3. International Agency for Research on Cancer. Breast Cancer Estimated Incidence, Mortality and Prevalence Worldwide in 2020 [Internet]. Lyon, France2020 [International Agency for Research on Cancer Global Cancer Observatory]. Available from: http://globocan.iarc.fr.
4. Li N, Deng Y, Zhou L, Tian T, Yang S, Wu Y, et al. Global burden of breast cancer and attributable risk factors in 195 countries and territories, from 1990 to 2017: results from the Global Burden of Disease Study 2017. Journal of hematology & oncology. 2019;12(1):140.
5. Office for National Statistics (ONS). National Population Projections: 2020-based interim UK: ONS; 2022 [updated Jan 2022. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/bulletins/nationalpopulationprojections/2020basedinterim#changing-age-structure.
6. Midlands Cancer Intelligence Unit. Second All Breast Cancer report. 2011 2011.
7. Cancer Research UK. Breast Cancer Incidence by Age 2016-2018 UK2023 [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive#heading-One.
8. Hoover RN, Hyer M, Pfeiffer RM, Adam E, Bond B, Cheville AL, et al. Adverse Health Outcomes in Women Exposed In Utero to Diethylstilbestrol. New England Journal of Medicine. 2011;365(14):1304-14.
9. Silva IdS, De Stavola B, McCormack V, Collaborative Group on Pre-Natal Risk F, Subsequent Risk of Breast C. Birth size and breast cancer risk: re-analysis of individual participant data from 32 studies. PLoS medicine. 2008;5(9):e193-e.
10. Collaborative Group on Hormonal Factors in Breast C. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. The Lancet Oncology. 2012;13(11):1141-51.
11. van den Brandt PA, Spiegelman D, Yaun SS, Adami HO, Beeson L, Folsom AR, et al. Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. Am J Epidemiol. 2000;152(6):514-27.
12. Collaborative Group on Hormonal Factors in Breast C. Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. Lancet. 2002;360(9328):187-95.
13. Colditz GA, Rosner B. Cumulative risk of breast cancer to age 70 years according to risk factor status: data from the Nurses' Health Study. Am J Epidemiol. 2000;152(10):950-64.
14. Ma H, Bernstein L, Pike MC, Ursin G. Reproductive factors and breast cancer risk according to joint estrogen and progesterone receptor status: a meta-analysis of epidemiological studies. Breast cancer research : BCR. 2006;8(4):R43.
15. Lambe M, Hsieh CC, Chan HW, Ekbom A, Trichopoulos D, Adami HO. Parity, age at first and last birth, and risk of breast cancer: a population-based study in Sweden. Breast cancer research and treatment. 1996;38(3):305-11.
16. Collaborative Group on Hormonal Factors in Breast C. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. Lancet. 1996;347(9017):1713-27.
17. Beral V. Breast cancer and hormone-replacement therapy in the Million Women Study. Lancet. 2003;362(9382):419-27.

18. Collaborative Group on Hormonal Factors in Breast Cancer. Alcohol, tobacco and breast cancer--collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. Br J Cancer. 2002;87(11):1234-45.
19. Allen NE, Beral V, Casabonne D, Kan SW, Reeves GK, Brown A, et al. Moderate alcohol intake and cancer incidence in women. J Natl Cancer Inst. 2009;101(5):296-305.
20. Chan DSM, Abar L, Cariolou M, Nanu N, Greenwood DC, Bandera EV, et al. World Cancer Research Fund International: Continuous Update Project-systematic literature review and meta-analysis of observational cohort studies on physical activity, sedentary behavior, adiposity, and weight change and breast cancer risk. Cancer causes & control : CCC. 2019;30(11):1183-200.
21. Bhaskaran K, Douglas I, Forbes H, dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults. Lancet. 2014;384(9945):755-65.
22. Chen Y, Liu L, Zhou Q, Imam MU, Cai J, Wang Y, et al. Body mass index had different effects on premenopausal and postmenopausal breast cancer risks: a dose-response meta-analysis with 3,318,796 subjects from 31 cohort studies. BMC Public Health. 2017;17(1):936.
23. Papadimitriou N, Markozannes G, Kanellopoulou A, Critselis E, Alhardan S, Karafousia V, et al. An umbrella review of the evidence associating diet and cancer risk at 11 anatomical sites. Nature Communications. 2021;12(1):4579.
24. Wolfe J. Risk for breast cancer development determined by mammographic parenchymal pattern. Cancer. 1976;37:2486 - 92.
25. McCormack V, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev. 2006;15:1159 - 69.
26. Pettersson A, Graff RE, Ursin G, Santos Silva ID, McCormack V, Baglietto L, et al. Mammographic density phenotypes and risk of breast cancer: a meta-analysis. J Natl Cancer Inst. 2014;106(5).
27. McCormack VA, Perry NM, Vinnicombe SJ, dos Santos Silva I. Changes and tracking of mammographic density in relation to Pike's model of breast tissue aging: a UK longitudinal study. International Journal of Cancer. 2010;127(2):452-61.
28. Krishnan K, Baglietto L, Stone J, Simpson JA, Severi G, Evans CF, et al. Longitudinal Study of Mammographic Density Measures That Predict Breast Cancer Risk. Cancer Epidemiol Biomarkers Prev. 2017;26(4):651-60.
29. National Collaborating Centre for Cancer (UK). Classification and Care of People at Risk of Familial Breast Cancer and Management of Breast Cancer and Related Risks in People with a Family History of Breast Cancer. NICE Clinical Guidelines, No. 164. Cardiff UK2013.
30. National Institute of Health and Care Excellence (NICE). Familial breast cancer: classification, care and managing breast cancer and related risks in people with a family history of breast cancer. UK: National Institute of Health and Care Excellence (NICE); 2013 [updated 20/11/2019. Clinical Guideline CG164]. Available from: https://www.nice.org.uk/guidance/cg164.
31. Evans DG, Shenton A, Woodward E, Lalloo F, Howell A, Maher ER. Penetrance estimates for BRCA1 and BRCA2 based on genetic testing in a Clinical Cancer Genetics service setting: risks of breast/ovarian cancer quoted should reflect the cancer burden in the family. BMC Cancer. 2008;8:155.
32. Walsh T, King MC. Ten genes for inherited breast cancer. Cancer Cell. 2007;11(2):103-5.
33. Ellingjord-Dale M, Vos L, Tretli S, Hofvind S, Dos-Santos-Silva I, Ursin G. Parity, hormones and breast cancer subtypes - results from a large nested case-control study in a national screening program. Breast cancer research : BCR. 2017;19(1):10-.
34. Antoni S, Sasco AJ, dos Santos Silva I, McCormack V. Is mammographic density differentially associated with breast cancer according to receptor status? A meta-analysis. Breast cancer research and treatment. 2013;137(2):337-47.

35. Anderson WF, Rosenberg PS, Menashe I, Mitani A, Pfeiffer RM. Age-Related Crossover in Breast Cancer Incidence Rates Between Black and White Ethnic Groups. Journal of the National Cancer Institute. 2008;100(24):1804-14.
36. Gleason MX, Mdzinarishvili T, Sherman S. Breast Cancer Incidence in Black and White Women Stratified by Estrogen and Progesterone Receptor Statuses. PLoS ONE. 2012;7(11):e49359.
37. Copson E, Maishman T, Gerty S, Eccles B, Stanton L, Cutress RI, et al. Ethnicity and outcome of young breast cancer patients in the United Kingdom: the POSH study. Br J Cancer. 2014;110(1):230-41.
38. Gathani T, Ali R, Balkwill A, Green J, Reeves G, Beral V, et al. Ethnic differences in breast cancer incidence in England are due to differences in known risk factors for the disease: prospective study. Br J Cancer. 2014;110(1):224-9.
39. Januszewski A, Tanna N, Stebbing J. Ethnic variation in breast cancer incidence and outcomes—the debate continues. British Journal of Cancer. 2014;110(1):4-6.
40. Song Q, Li J, Huang R, Fan J-H, Zheng R-S, Zhang B-N, et al. Age of Diagnosis of Breast Cancer in China: Almost 10 Years Earlier than in the United States and the European Union. Asian Journal of Cancer Prevention. 2014;15(22):10021-5.
41. Chen C, Sun S, Yuan JP, Wang YH, Cao TZ, Zheng HM, et al. Characteristics of breast cancer in Central China, literature review and comparison with USA. Breast (Edinburgh, Scotland). 2016;30:208-13.
42. Ziegler RG, Hoover RN, Pike MC, Hildesheim A, Nomura AMY, West DW, et al. Migration Patterns and Breast Cancer Risk in Asian-American Women. Journal of the National Cancer Institute. 1993;85(22):1819-27.
43. Antoniou AC, Pharoah PPD, Smith P, Easton DF. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. British Journal of Cancer. 2004;91(8):1580-90.
44. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting Individualized Probabilities of Developing Breast Cancer for White Females Who Are Being Examined Annually. JNCI: Journal of the National Cancer Institute. 1989;81(24):1879-86.
45. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. Statistics in Medicine. 2004;23(7):1111-30.
46. IARC Working Group on the Evaluation of Cancer-Preventive Strategies. Breast Cancer Screening. IARC Handbooks of Cancer Prevention, Volume 15. 2016.
47. Pal Choudhury P, Brook MN, Hurson AN, Lee A, Mulder CV, Coulson P, et al. Comparative validation of the BOADICEA and Tyrer-Cuzick breast cancer risk models incorporating classical risk factors and polygenic risk in a population-based prospective cohort of women of European ancestry. Breast Cancer Research. 2021;23(1):22.
48. Colditz GA, Bohlke K. Priorities for the primary prevention of breast cancer. CA: A Cancer Journal for Clinicians. 2014;64(3):186-94.
49. World Cancer Research Fund / American Institute for Cancer Research. Continuous Update Project Report. Food, Nutrition, Physical Activity, and the Prevention of Breast Cancer. London; 2010. Contract No.: 11/11/2016.
50. Offit K. BRCA mutation frequency and penetrance: new data, old debate. J Natl Cancer Inst. 2006;98(23):1675-7.
51. Anglian Breast Cancer Study Group. Prevalence and penetrance of BRCA1 and BRCA2 mutations in a population-based series of breast cancer cases. Br J Cancer. 2000;83(10):1301-8.
52. Silva I, McCormack V, Jedy-Agba E, Adebamowo C. Downstaging breast cancer in sub-saharan Africa: A realistic target? Cancer Control. 2017.
53. Richards MA, Westcombe AM, Love SB, Littlejohns P, Ramirez AJ. Influence of delay on survival in patients with breast cancer: a systematic review. Lancet. 1999;353(9159):1119-26.
54. US Library of Medicine. Breast Screening - Risk Adaptive Imaging for Density (BRAID): US Library of Medicine; 2019 [Description of BRAID trial]. Available from: https://clinicaltrials.gov/ct2/show/NCT04097366.

55. Richards M, Corner J, Maher J. The National Cancer Survivorship Initiative: new and emerging evidence on the ongoing needs of cancer survivors. Br J Cancer. 2011;105 Suppl 1(Suppl 1):S1-4.
56. Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med. 2005;353(17):1784-92.
57. Demicheli R, Ambrogi F, Weedon-Fekjaer H, Romundstad PR, Vatten LJ, Stout NK, et al. Comparative benefit from small tumour size and adjuvant chemotherapy: clues for explaining breast cancer mortality decline. BMC Cancer. 2014;14(6):702.
58. Evans DGR, Howell A. Breast cancer risk-assessment models. Breast Cancer Research. 2007;9(5):213.
59. Altobelli E, Rapacchietta L, Angeletti PM, Barbante L, Profeta FV, Fagnano R. Breast Cancer Screening Programmes across the WHO European Region: Differences among Countries Based on National Income Level. International journal of environmental research and public health. 2017;14(4).
60. Cancer Research UK. Types of breast cancer and related conditions UK: Cancer Research UK; 2022 [Available from: https://www.cancerresearchuk.org/about-cancer/breast-cancer/stages-types-grades/types.
61. National Cancer Registration and Analysis Service (NCRAS). Routes to Diagnosis: NHS Digital; 2022 [Routes to Diagnosis 2017]. Available from: http://www.ncin.org.uk/publications/routes_to_diagnosis.
62. Screening and Immunisations team Health and Social Care Information Centre. Breast Screening Programme, England - 2013-14 [NS] Health & Social Care Information Centre; 2015 [updated 18/02/2105. v 1.0:[Breast Screening Programme, England, Statistics for 2013-14]. Available from: http://www.hscic.gov.uk/catalogue/PUB16803.
63. NHS Digital Screening & Immunisations Team. Breast Screening Programme England 2018-2019 UK: NHS Digital; 2020 [updated 30/01/2020. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/breast-screening-programme/england---2018-19.
64. Cancer Research UK. Survival statistics UK2020 [Available from: https://www.cancerresearchuk.org/about-cancer/breast-cancer/survival.
65. National Cancer Registration and Analysis Service (NCRAS). Screen-Detected Breast Cancer UK: NHS Digital; 2011 [Available from: http://www.ncin.org.uk/publications/data_briefings/screen_detected_breast_cancer.
66. Moshina N, Falk RS, Botteri E, Larsen M, Akslen LA, Cairns JA, et al. Quality of life among women with symptomatic, screen-detected, and interval breast cancer, and for women without breast cancer: a retrospective cross-sectional study from Norway. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2022;31(4):1057-68.
67. NHS Digital. Routes to Diagnosis, 2018: NHS Digital; 2022 [updated 01/12/2022. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/routes-to-diagnosis/2018.
68. Dibden A, Offman J, Parmar D, Jenkins J, Slater J, Binysh K, et al. Reduction in interval cancer rates following the introduction of two-view mammography in the UK breast screening programme. Br J Cancer. 2014;110(3):560-4.
69. Bennett RL, Sellars SJ, Moss SM. Interval cancers in the NHS breast cancer screening programme in England, Wales and Northern Ireland. Br J Cancer. 2011;104(4):571-7.
70. Warren R, Duffy S. Interval cancers as an indicator of performance in breast screening. Breast Cancer. 2000;7(1):9-18.
71. Holland K, van Gils CH, Wanders JOP, Mann RM, Karssemeijer N, editors. Quantification of mammographic masking risk with volumetric breast density maps: how to select women for supplemental screening2016.

72. Mandelson MT, Oestreicher N, Porter PL, White D, Finder CA, Taplin SH, et al. Breast density as a predictor of mammographic detection: Comparison of interval- and screen-detected cancers. Journal of the National Cancer Institute. 2000;92(13).
73. Cox D, Powell A. PB.12. Audit and root-cause analysis of classification 2 and 3 interval cancers. Breast Cancer Research. 2014;16(1):P3.
74. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. The Lancet. 2012;380(9855):1778-86.
75. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. Br J Cancer. 2013;108(11):2205-40.
76. Lauby-Secretan B, Scoccianti C, Loomis D, Benbrahim-Tallaa L, Bouvard V, Bianchini F, et al. Breast-Cancer Screening — Viewpoint of the IARC Working Group. New England Journal of Medicine. 2015;372(24):2353-8.
77. Maroni R, Massat NJ, Parmar D, Dibden A, Cuzick J, Sasieni PD, et al. A case-control study to evaluate the impact of the breast screening programme on mortality in England. British Journal of Cancer. 2021;124(4):736-43.
78. Morton R, Sayma M, Sura MS. Economic analysis of the breast cancer screening program used by the UK NHS: should the program be maintained? Breast Cancer (Dove Med Press). 2017;9:217-25.
79. Hauge IHR, Pedersen K, Olerud HM, Hole EO, Hofvind S. The risk of radiation-induced breast cancers due to biennial mammographic screening in women aged 50–69 years is minimal. Acta Radiologica. 2014;55(10):1174-9.
80. Warren LM, Dance DR, Young KC. Radiation risk of breast screening in England with digital mammography. Br J Radiol. 2016;89(1067):20150897.
81. Taylor P. Breast Cancer Screening London Review of Books. 2014;36(11):30-2.
82. Duffy SW, Tabar L, Olsen AH, Vitak B, Allgood PC, Chen TH, et al. Absolute numbers of lives saved and overdiagnosis in breast cancer screening, from a randomized trial and from the Breast Screening Programme in England. J Med Screen. 2010;17(1):25-30.
83. Heinavaara S, Sarkeala T, Anttila A. Overdiagnosis due to breast cancer screening: updated estimates of the Helsinki service study in Finland. Br J Cancer. 2014;111(7):1463-8.
84. Gotzsche PC, Jorgensen KJ, Zahl PH, Maehlen J. Why mammography screening has not lived up to expectations from the randomised trials. Cancer causes & control : CCC. 2012;23(1):15-21.
85. Gotzsche PC, Jorgensen KJ, Maehlen J, Zahl PH. Estimation of lead time and overdiagnosis in breast cancer screening. Br J Cancer. 2009;100(1):219-.
86. Public Health England. Breast Screening Programme, England Statistics for 2014-15 2016 [Available from: http://www.hscic.gov.uk/pubs/brstscreen1415.
87. de Gelder R, van As E, Tilanus-Linthorst MMA, Bartels CCM, Boer R, Draisma G, et al. Breast cancer screening: Evidence for false reassurance? International Journal of Cancer. 2008;123(3):680-6.
88. Cooper GC, Harvie MN, French DP. Do negative screening test results cause false reassurance? A systematic review. British Journal of Health Psychology. 2017;22(4):958-77.
89. Programme NBS. Breast Screening Programme England, 2018-19. In: Team NSaI, editor. UK: NHS Digital; 2020.
90. Bond M, Pavey T, Welch K, Cooper C, Garside R, Dean S, et al. Systematic review of the psychological consequences of false-positive screening mammograms. Health Technol Assess. 2013;17(13):1-170, v-vi.
91. Gram IT, Funkhouser E, Tabar L. The Tabar classification of mammographic parenchymal patterns. European journal of radiology. 1997;24(2):131-6.
92. American College of Radiology. ACR BI-RADS Atlas 5th Edition 2013 [Available from: http://www.acr.org/~/media/ACR/Documents/PDF/QualitySafety/Resources/BIRADS/01%20Mammography/02%20%20BIRADS%20Mammography%20Reporting.pdf.

93. Wolfe J. Breast patterns as an index of risk for developing breast cancer. AJR Am J Roentgenol. 1976;126:1130 - 9.
94. Winkel RR, von Euler-Chelpin M, Nielsen M, Diao P, Nielsen MB, Uldall WY, et al. Inter-observer agreement according to three methods of evaluating mammographic density and parenchymal pattern in a case control study: impact on relative risk of breast cancer. BMC Cancer. 2015;15:274.
95. Damases CN, Mello-Thoms C, McEntee MF. Inter-observer variability in mammographic density assessment using Royal Australian and New Zealand College of Radiologists (RANZCR) synoptic scales: Journal of Medical Imaging and Radiation Oncology. 60 (3) (pp 329-336), 2016. Date of Publication: 01 Jun 2016.; 2016.
96. Boyd NF, Lockwood GA, Martin LJ, Knight JA, Byng JW, Yaffe MJ, et al. Mammographic densities and breast cancer risk. Breast Dis. 1998;10.
97. Boyd N, Guo H, Martin L, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. N Engl J Med. 2007;356:227 - 36.
98. Boyd NF, Martin LJ, Bronskill M, Yaffe MJ, Duric N, Minkin S. Breast tissue composition and susceptibility to breast cancer. J Natl Cancer Inst. 2010;102(16):1224-37.
99. Cummings SR, Tice JA, Bauer S, Browner WS, Cuzick J, Ziv E, et al. Prevention of breast cancer in postmenopausal women: approaches to estimating and reducing risk. J Natl Cancer Inst. 2009;101(6):384-98.
100. Huo CW, Chew GL, Britt KL, Ingman WV, Henderson MA, Hopper JL, et al. Mammographic density-a review on the current understanding of its association with breast cancer. Breast cancer research and treatment. 2014;144(3):479-502.
101. Ursin G, Hovanessian-Larsen L, Parisky YR, Pike MC, Wu AH. Greatly increased occurrence of breast cancers in areas of mammographically dense tissue. Breast cancer research : BCR. 2005;7(5):R605-8.
102. Sherratt MJ, McConnell JC, Streuli CH. Raised mammographic density: causative mechanisms and biological consequences. Breast Cancer Research. 2016;18(1):45.
103. Guo Y, Martin L, Hanna W, Banerjee D, Miller N, Fishell E, et al. Growth factors and stromal matrix proteins associated with mammographic densities. Cancer Epidemiol Biomarkers Prev. 2001;10:243 - 8.
104. Barcellos-Hoff MH, Medina D. New highlights on stroma–epithelial interactions in breast cancer. Breast Cancer Research. 2005;7(1):33-6.
105. Pinto Pereira SM, McCormack VA, Hipwell JH, Record C, Wilkinson LS, Moss SM, et al. Localized fibroglandular tissue as a predictor of future tumor location within the breast. Cancer Epidemiol Biomarkers Prev. 2011;20(8):1718-25.
106. Boyd N, Martin L, Yaffe M, Minkin S. Mammographic density and breast cancer risk: current understanding and future prospects. Breast Cancer Research. 2011;13(6):223.
107. Yaghjyan L, Colditz GA, Rosner B, Tamimi RM. Mammographic breast density and subsequent risk of breast cancer in postmenopausal women according to the time since the mammogram. Cancer Epidemiol Biomarkers Prev. 2013;22(6):1110-7.
108. Whitehead J, Carlile T, Kopecky KJ, Thompson DJ, Gilbert FI, Jr., Present AJ, et al. Wolfe mammographic parenchymal patterns. A study of the masking hypothesis of Egan and Mosteller. Cancer. 1985;56(6):1280-6.
109. Pisano ED, Hendrick RE, Yaffe MJ, Baum JK, Acharyya S, Cormack JB, et al. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. Radiology. 2008;246(2):376-83.
110. Wanders JOP, Holland K, Veldhuis WB, Mann RM, Pijnappel RM, Peeters PHM, et al. Volumetric breast density affects performance of digital screening mammography. Breast cancer research and treatment. 2017;162(1):95-103.

111. Carney PA, Miglioretti DL, Yankaskas BC, Kerlikowske K, Rosenberg R, Rutter CM, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Annals of internal medicine. 2003;138(3):168-75.
112. Krishnan K, Baglietto L, Apicella C, Stone J, Southey MC, English DR, et al. Mammographic density and risk of breast cancer by mode of detection and tumor size: a case-control study. Breast Cancer Research. 2016;18(1):1-13.
113. Melnikow J, Fenton JJ, Whitlock EP, Miglioretti DL, Weyrich MS, Thompson JH, et al. U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. Supplemental Screening for Breast Cancer in Women With Dense Breasts: A Systematic Review for the US Preventive Service Task Force. Rockville (MD): Agency for Healthcare Research and Quality (US); 2016.
114. Bakker MF, de Lange SV, Pijnappel RM, Mann RM, Peeters PHM, Monninkhof EM, et al. Supplemental MRI Screening for Women with Extremely Dense Breast Tissue. New England Journal of Medicine. 2019;381(22):2091-102.
115. Boyd NF, Martin LJ, Sun L, Guo H, Chiarelli A, Hislop G, et al. Body size, mammographic density, and breast cancer risk. Cancer Epidemiology Biomarkers and Prevention. 2006;15(11):2086-92.
116. Deurenberg P, Andreoli A, Borg P, Kukkonen-Harjula K, de Lorenzo A, van Marken Lichtenbelt WD, et al. The validity of predicted body fat percentage from body mass index and from impedance in samples of five European populations. European journal of clinical nutrition. 2001;55(11):973-9.
117. Deurenberg P, Deurenberg-Yap M, Guricci S. Asians are different from Caucasians and from each other in their body mass index/body fat per cent relationship. Obesity reviews : an official journal of the International Association for the Study of Obesity. 2002;3(3):141-6.
118. Rush EC, Goedecke JH, Jennings C, Micklesfield L, Dugas L, Lambert EV, et al. BMI, fat and muscle differences in urban women of five ethnicities from two countries. International journal of obesity (2005). 2007;31(8):1232-9.
119. Gallagher D, Visser M, Sepulveda D, Pierson RN, Harris T, Heymsfield SB. How useful is body mass index for comparison of body fatness across age, sex, and ethnic groups? Am J Epidemiol. 1996;143(3):228-39.
120. Federico B, D'Aliesio F, Pane F, Capelli G, Rodio A. Body mass index has a curvilinear relationship with the percentage of body fat among children. BMC research notes. 2011;4:301.
121. Duncan MJ, Martins C, Silva G, Marques E, Mota J, Aires L. Inverted BMI rather than BMI is a better predictor of DEXA determined body fatness in children. European journal of clinical nutrition. 2014;68(5):638-40.
122. Nevill AM, Stavropoulos-Kalinoglou A, Metsios GS, Koutedakis Y, Holder RL, Kitas GD, et al. Inverted BMI rather than BMI is a better proxy for percentage of body fat. Annals of human biology. 2011;38(6):681-4.
123. Gosse MA. How accurate is self-reported BMI? Nutrition Bulletin. 2014;39(1):105-14.
124. Lokate M, Peeters PH, Peelen LM, Haars G, Veldhuis WB, van Gils CH. Mammographic density and breast cancer risk: the role of the fat surrounding the fibroglandular tissue. Breast cancer research : BCR. 2011;13(5):R103.
125. Pettersson A, Hankinson SE, Willett WC, Lagiou P, Trichopoulos D, Tamimi RM. Nondense mammographic area and risk of breast cancer. Breast cancer research : BCR. 2011;13(5):R100.
126. Soguel L, Durocher F, Tchernof A, Diorio C. Adiposity, breast density, and breast cancer risk: epidemiological and biological considerations. Eur J Cancer Prev. 2017;26(6):511-20.
127. Baglietto L, Krishnan K, Stone J, Apicella C, Southey MC, English DR, et al. Associations of mammographic dense and nondense areas and body mass index with risk of breast cancer. Am J Epidemiol. 2014;179(4):475-83.

128. Schetter SE, Hartman TJ, Liao J, Richie JP, Prokopczyk B, DuBrock C, et al. Differential impact of body mass index on absolute and percent breast density: implications regarding their use as breast cancer risk biomarkers. Breast Cancer Research & Treatment. 2014;146(2):355-63.
129. Mercer CE, Szczepura K, Kelly J, Millington SR, Denton ERE, Borgen R, et al. A 6-year study of mammographic compression force: Practitioner variability within and between screening sites. Radiography. 2014;21(1):68–73.
130. Shepherd J, Kerlikowske K, Ma L, Duewer F, Fan B, Wang J, et al. Volume of mammographic density and risk of breast cancer. Cancer Epidemiol Biomarkers Prev. 2011;20:1473 - 82.
131. Aitken Z, McCormack VA, Highnam RP, Martin L, Gunasekara A, Melnichouk O, et al. Screen-film mammographic density and breast cancer risk: a comparison of the volumetric standard mammogram form and the interactive threshold measurement methods. Cancer Epidemiol Biomarkers Prev. 2010;19(2):418-28.
132. Eng A, Gallant Z, Shepherd J, McCormack V, Li J, Dowsett M, et al. Digital mammographic density and breast cancer risk: a case-control study of six alternative density assessment methods. Breast cancer research : BCR. 2014;16(5):439.
133. Alonzo-Proulx O, Mawdsley GE, Patrie JT, Yaffe MJ, Harvey JA. Reliability of Automated Breast Density Measurements. Radiology. 2015;275(1):366-76.
134. Ellison-Loschmann L, McKenzie F, Highnam R, Cave A, Walker J, Jeffreys M. Age and ethnic differences in volumetric breast density in new zealand women: a cross-sectional study. PLoS One. 2013;8(7):e70217.
135. van der Waal D, Emaus MJ, Bakker MF, den Heeten GJ, Karssemeijer N, Pijnappel RM, et al. Geographic variation in volumetric breast density between screening regions in the Netherlands. Eur Radiol. 2015;25(11):3328-37.
136. Parsons PA. Fluctuating asymmetry: an epigenetic measure of stress. Biol Rev Camb Philos Soc. 1990;65(2):131-45.
137. Møller AP, Soler M, Thornhill R. Breast asymmetry, sexual selection, and human reproductive success. Evolution and Human Behavior. 1995;16(3):207-19.
138. Milne BJ, Belsky J, Poulton R, Thomson WM, Caspi A, Kieser J. Fluctuating asymmetry and physical health among young adults. Evolution and Human Behavior. 2003;24(1):53-63.
139. Natekar PE, DeSouza FM. Fluctuating asymmetry in dermatoglyphics of carcinoma of breast. Indian Journal of Human Genetics. 2006;12(2):76-81.
140. Manning JT, Leinster SJ. re: The ratio of 2nd to 4th digit length and age at presentation of breast cancer: a link with prenatal oestrogen? The Breast. 2001;10(4):355-7.
141. Muller DC, Baglietto L, Manning JT, McLean C, Hopper JL, English DR, et al. Second to fourth digit ratio (2D:4D), breast cancer risk factors, and breast cancer risk: a prospective cohort study. Br J Cancer. 2012;107(9):1631-6.
142. Bunevicius A. The Association of Digit Ratio (2D : 4D) with Cancer: A Systematic Review and Meta-Analysis. Disease markers. 2018;2018:7698193.
143. Manning JT, Scutt D, Whitehouse GH, Leinster SJ. Breast asymmetry and phenotypic quality in women. Evolution and Human Behavior. 1997;18(4):223-36.
144. Key T, Appleby P, Barnes I, Reeves G. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. J Natl Cancer Inst. 2002;94(8):606-16.
145. Key TJ, Appleby PN, Reeves GK, Travis RC, Alberg AJ, Barricarte A, et al. Sex hormones and risk of breast cancer in premenopausal women: a collaborative reanalysis of individual participant data from seven prospective studies. Lancet Oncol. 2013;14(10):1009-19.
146. Key TJ, Appleby PN, Reeves GK, Roddam AW, Helzlsouer KJ, Alberg AJ, et al. Circulating sex hormones and breast cancer risk factors in postmenopausal women: reanalysis of 13 studies. Br J Cancer. 2011;105(5):709-22.

147. Lambertini M, Santoro L, Del Mastro L, Nguyen B, Livraghi L, Ugolini D, et al. Reproductive behaviors and risk of developing breast cancer according to tumor subtype: A systematic review and meta-analysis of epidemiological studies. Cancer Treat Rev. 2016;49:65-76.
148. Bao J, Yu K-D, Jiang Y, Zhao N, Di G-H. The Effect of Laterality and Primary Tumor Site on Cancer-Specific Mortality in Breast Cancer: A SEER Population-Based Study2014. e94815 p.
149. Ekbom A, Adami HO, Trichopoulos D, Lambe M, Hsieh CC, Ponten J. Epidemiologic correlates of breast cancer laterality (Sweden). Cancer causes & control : CCC. 1994;5(6):510-6.
150. Weiss HA, Devesa Ss Fau - Brinton LA, Brinton LA. Laterality of breast cancer in the United States. Cancer causes & control : CCC. 1996;Sep;7(5)(0957-5243 (Print)):539-43.
151. Senie RT, Rosen PP, Lesser ML, Snyder RE, Schottenfeld D, Duthie K. Epidemiology of breast carcinoma II: factors related to the predominance of left-sided disease. Cancer. 1980;46(7):1705-13.
152. Losken A, Fishman I, Denson DD, Moyer HR, Carlson GW. An objective evaluation of breast symmetry and shape differences using 3-dimensional images. Annals of plastic surgery. 2005;55(6):571-5.
153. Chen S. Aromatase and breast cancer. Frontiers in bioscience : a journal and virtual library. 1998;3:d922-33.
154. Ghosh S, Kang T, Wang H, Hu Y, Li R. Mechanical phenotype is important for stromal aromatase expression. Steroids. 2011;76(8):797-801.
155. Scutt D, Manning JT, Whitehouse GH, Leinster SJ, Massey CP. The relationship between breast asymmetry, breast size and the occurrence of breast cancer. British Journal of Radiology. 1997;70(OCT):1017-21.
156. Scutt D, Lancaster GA, Manning JT. Breast asymmetry and predisposition to breast cancer. Breast cancer research : BCR. 2006;8(2):R14.
157. Eltonsy HN, Elmaghraby A, Tourassi G. Bilateral Breast Volume Asymmetry in Screening Mammograms as a Potential Marker of Breast Cancer: Preliminary Experience2007. V-5 p.
158. Kayar R, Cilengiroglu OV. Breast volume asymmetry value, ratio, and cancer risk. Breast Cancer: Basic and Clinical Research. 2015;2015(9):87-92.
159. Williams AC, Hitt A, Voisin S, Tourassi G, editors. Automated assessment of bilateral breast volume asymmetry as a breast cancer biomarker during mammographic screening. SPIE Medical Imaging; 2013 2013: International Society for Optics and Photonics.
160. Zheng B, Sumkin JH, Zuley ML, Wang X, Klym AH, Gur D. Bilateral mammographic density asymmetry and breast cancer risk: a preliminary assessment. European journal of radiology. 2012;81(11):3222-8.
161. Tan M, Zheng B, Ramalingam P, Gur D. Prediction of Near-term Breast Cancer Risk Based on Bilateral Mammographic Feature Asymmetry. Academic Radiology. 2013;20(12):1542-50.
162. Sun W, Zheng B, Lure F, Wu T, Zhang J, Wang BY, et al. Prediction of near-term risk of developing breast cancer using computerized features from bilateral mammograms. Comput Med Imaging Graph. 2014;38(5):348-57.
163. Zheng B, Tan M, Ramalingam P, Gur D. Association between computed tissue density asymmetry in bilateral mammograms and near-term breast cancer risk. The breast journal. 2014;20(3):249-57.
164. The Royal College of Radiologists. Guidance on screening and symptomatic breast imaging. London; 2013 June 2013.
165. Salvagnini E, Bosmans H, Van Ongeval C, Van Steen A, Michielsen K, Cockmartin L, et al. Impact of compressed breast thickness and dose on lesion detectability in digital mammography: FROC study with simulated lesions in real mammograms: Medical Physics. 43 (9) (pp 5104-5116), 2016. Date of Publication: 01 Sep 2016.; 2016.
166. Saunders RS, Jr., Samei E. The effect of breast compression on mass conspicuity in digital mammography. Med Phys. 2008;35(10):4464-73.
167. Yaffe MJ, Mainprize JG. Risk of radiation-induced breast cancer from mammographic screening. Radiology. 2011;258(1):98-105.

168. Boyce M, Gullen R, Parashar D, Taylor K, editors. Comparing the use of PGMI scoring systems used in the UK and Norway to assess the technical quality of screening mammograms: a pilot study. British Society of Breast Radiology Annual Scientific Meeting 2012; 2012 12/11/2012; Leeds, UK.: BioMed Central; 2012.
169. NHS Breast Screening Programme. Quality Assurance Guidelines for Breast Cancer Screening Radiology 2011 [2nd:[Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/470579/nhsbsp59_QA_radiology_uploaded_231015.pdf.
170. Perry N, Broeders M, Wolf C, Törnberg S, Holland R, Karsa L. European guidelines for quality assurance in breast cancer screening and diagnosis. Luxembourg: Office for Official Publications of the European Communities; 2006 2006//.
171. Perry N, Broeders M, de Wolf C, Tornberg S, Holland R, von Karsa L. European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition--supplement Luxembourg: Office for Official Publications of the European Communities; 2013 [updated 2020. 4th:[Available from: http://www.euref.org/european-guidelines.
172. Waade GG, Moshina N, Sebuodegard S, Hogg P, Hofvind S. Compression forces used in the Norwegian Breast Cancer Screening Program. Br J Radiol. 2017;90(1071):20160770.
173. NHS Breast Screening Programme. NHS Breast Screening Programme Guidance for breast screening mammographers. London, UK: Public Health England; 2017.
174. Mercer CE, Hogg P, Lawson R, Diffey J, Denton ER. Practitioner compression force variability in mammography: a preliminary study. The British journal of radiology. 2013;86(1022):20110596.
175. Mercer CE, Hogg P, Szczepura K, Denton ERE. Practitioner compression force variation in mammography: A 6-year study. Radiography. 2013;19(3):200-6.
176. Branderhorst W, de Groot JE, Highnam R, Chan A, Bohm-Velez M, Broeders MJ, et al. Mammographic compression--a need for mechanical standardization. European journal of radiology. 2015;84(4):596-602.
177. Waade GG, Sanderud A, Hofvind S. Compression force and radiation dose in the Norwegian Breast Cancer Screening Program. European journal of radiology. 2017;88.
178. Waade GG, Sebuodegard S, Hogg P, Hofvind S. Breast compression across consecutive examinations among females participating in BreastScreen Norway. Br J Radiol. 2018;91.
179. Serwan E, Matthews D, Davies J, Chau M. Mechanical standardisation of mammographic compression using Volpara software. Radiography. 2021;27(3):789-94.
180. Voigt M, Bolejko A, Dustler M. Intra- and inter-rater reliability of compressed breast thickness, applied force, and pressure distribution in screening mammography. Acta radiologica open. 2021;10(12):20584601211062078.
181. Dzidzornu E, Angmorterh SK, Ofori-Manteaw BB, Aboagye S, Ofori EK, Owusu-Agyei S, et al. Compression force variability in mammography in Ghana - A baseline study. Radiography (London, England : 1995). 2021;27(1):150-5.
182. Ng K, Hill M, Johnston L, Highnam R, Tomal A. Large variation in mammography compression internationally. European Congress of Radiology; Vienna: European Society of Radiology; 2017.
183. Helvie MA, Chan HP, Adler DD, Boyd PG. Breast thickness in routine mammograms: effect on image quality and radiation dose. AJR Am J Roentgenol. 1994;163(6):1371-4.
184. Murphy F, Nightingale J, Hogg P, Robinson L, Seddon D, Mackay S. Compression force behaviours: An exploration of the beliefs and values influencing the application of breast compression during screening mammography. Radiography. 2015;21(1):30-5.
185. Poulos A, McLean D, Rickard M, Heard R. Breast compression in mammography: how much is enough? Australas Radiol. 2003;47(2):121-6.
186. De Groot JE. Pressure-standardized breast compression in mammography. Amsterdam: Amsterdam; 2015.

187. De Groot JE, Branderhorst W, Grimbergen CA, Den Heeten GJ, Broeders MJM. Towards personalized compression in mammography: A comparison study between pressure- and force-standardization. European journal of radiology. 2015;84(3):384-91.
188. Moshina N, Roman M, Waade GG, Sebuodegard S, Ursin G, Hofvind S. Breast compression parameters and mammographic density in the Norwegian Breast Cancer Screening Programme. Eur Radiol. 2018;28(4):1662-72.
189. Holland K, Sechopoulos I, Den Heeten GJ, Mann RM, Karssemeijer N, editors. Performance of Breast Cancer Screening Depends on Mammographic Compression. Breast Imaging: 13th International Workshop, IWDM 2016, Malmö, Sweden, June 19-22, 2016, Proceedings; 2016 July 2016; Malmö: Springer International Publishing.
190. Holland K, Sechopoulos I, Mann RM, den Heeten GJ, van Gils CH, Karssemeijer N. Influence of breast compression pressure on the performance of population-based mammography screening. Breast Cancer Research. 2017;19(1):126.
191. Moshina N, Sæbuødegård S, Hofvind S. Is breast compression associated with breast cancer detection and other early performance measures in a population-based breast cancer screening program? Breast Cancer Res Treat. 2017;163.
192. Hill M, Martis L, Halling-Brown M, Highnam R, Chan A. Mammographic compression pressure as a predictor of interval cancer: SPIE; 2022.
193. Taplin SH, Rutter CM, Finder C, Mandelson MT, Houn F, White E. Screening mammography: clinical image quality and the risk of interval breast cancer. AJR Am J Roentgenol. 2002;178(4):797-803.
194. Broeders MJ, Voorde M, Veldkamp WJ, Engen RE, Landsveld-Verhoeven C, Nl't Jong-Gunneman M, et al. Comparison of a flexible versus a rigid breast compression paddle: pain experience, projected breast area, radiation dose and technical image quality. Eur Radiol. 2015;25.
195. Moshina N, Sebuodegard S, Evensen KT, Hantho C, Iden KA, Hofvind S. Breast compression and experienced pain during mammography by use of three different compression paddles. European journal of radiology. 2019;115.
196. Hauge IH, Hogg P, Szczepura K, Connolly P, McGill G, Mercer C. The readout thickness versus the measured thickness for a range of screen film mammography and full-field digital mammography units. Med Phys. 2012;39(1):263-71.
197. Ma WK. Mammography machine compression paddle movement and observer performance analysis: University of Salford (United Kingdom); 2021.
198. de Groot JE, Broeders MJ, Branderhorst W, den Heeten GJ, Grimbergen CA. A novel approach to mammographic breast compression: Improved standardization and reduced discomfort by controlling pressure instead of force. Medical physics. 2013;40(8):081901.
199. Whelehan P, Evans A, Wells M, Macgillivray S. The effect of mammography pain on repeat participation in breast cancer screening: a systematic review. Breast (Edinburgh, Scotland). 2013;22(4):389-94.
200. Meyer J, Maxwell A, Harkness E, Astley S, Mercer C, Wilson M, et al. PB.22. Does mammographic compression force at breast screening influence the likelihood of subsequent screening attendance? Breast Cancer Research. 2014;16.
201. Hogg P, Mercer C, Maxwell A, Robinson L, Kelly J, Murphy F. Controversies in compression, Imaging and oncology. 2013 2013 [cited 13/11/2015]. In: Imaging & Oncology [Internet]. UK: The Society and College of Radiographers. 2013. Imaging & Oncology, [cited 13/11/2015]; [28-36]. Available from: http://www.sor.org/system/files/article/201305/I%26O%202013.pdf.
202. Brentnall AR, Harkness EF, Astley SM, Donnelly LS, Stavrinos P, Sampson S, et al. Mammographic density adds accuracy to both the Tyrer-Cuzick and Gail breast cancer risk models in a prospective UK screening cohort. Breast Cancer Research. 2015;17(1):1-10.
203. Inc. AYD. Are You Dense? Exposing the best-kept secret.™ 2013 [Available from: https://www.areyoudense.org/.

204. Malkov S, Wang J, Kerlikowske K, Cummings SR, Shepherd JA. Single x-ray absorptiometry method for the quantitative mammographic measure of fibroglandular tissue volume. Med Phys. 2009;36(12):5525-36.
205. Alonzo-Proulx O, Packard N, Boone JM, Al-Mayah A, Brock KK, Shen SZ, et al. Validation of a method for measuring the volumetric breast density from digital mammograms. Phys Med Biol. 2010;55(11):3027-44.
206. Hologic Inc. Understanding Quantra™ 2.0 User Manual. USA: Hologic Inc.: Bedford, MA, USA,; 2012.
207. Volpara. Automated measurement of volumetric mammographic density: a tool for widespread breast cancer risk assessment. 2014 Sep.
208. Matakina Technology Ltd. VolparaDensity™ User Manual Version 1.5.11. [User Manual Volpara Software]. In press 2014.
209. Brand JS, Czene K, Shepherd JA, Leifland K, Heddson B, Sundbom A, et al. Automated Measurement of Volumetric Mammographic Density: A Tool for Widespread Breast Cancer Risk Assessment. Cancer Epidemiology Biomarkers & Prevention. 2014;23(9):1764-72.
210. Holland K, van Zelst J, den Heeten GJ, Imhof-Tas M, Mann RM, van Gils CH, et al. Consistency of breast density categories in serial screening mammograms: A comparison between automated and human assessment. The Breast. 2016;29:49-54.
211. Lee HN, Sohn YM, Han KH. Comparison of mammographic density estimation by Volpara software with radiologists' visual assessment: Analysis of clinical-radiologic factors affecting discrepancy between them: Acta Radiologica. 56 (9) (pp 1061-1068), 2015. Date of Publication: 01 Jan 2015.; 2015.
212. Gweon HM, Youk JH, Kim JA, Kim H, Park YJ, Son EJ. Mammographic density assessment in digital mammography: Comparison of BI-RADS categories with a fully automated volumetric method: Journal of Medical Imaging and Radiation Oncology. Conference: Royal Australian and New Zealand College of Radiologists, RANZCR 2012, 63rd Annual Scientific Meeting, and Asian Oceanian Congress of Radiology, AOCR 2012. Sydney, NSW Australia. Conference Start: 20120830. Conference End: 20120902. Conference Publication: (var.pagings). 56 (pp 99), 2012. Date of Publication: August 2012.; 2012.
213. Singh T, Sharma M, Singla V, Khandelwal N. Breast Density Estimation with Fully Automated Volumetric Method: Comparison to Radiologists' Assessment by BI-RADS Categories: Academic Radiology. 23 (1) (pp 78-83), 2016. Date of Publication: 01 Jan 2016.; 2016.
214. Wang J, Azziz A, Fan B, Malkov S, Klifa C, Newitt D, et al. Agreement of mammographic measures of volumetric breast density to MRI. PLoS One. 2013;8(12):e81653.
215. Gubern-Mérida A, Kallenberg M, Platel B, Mann RM, Martí R, Karssemeijer N. Volumetric Breast Density Estimation from Full-Field Digital Mammograms: A Validation Study. PLOS ONE. 2014;9(1):e85952.
216. Astley SM, Harkness EF, Sergeant JC, Warwick J, Stavrinos P, Warren R, et al. A comparison of five methods of measuring mammographic density: a case-control study. Breast cancer research : BCR. 2018;20:10.
217. Brandt KR, Scott CG, Ma L, Mahmoudzadeh AP, Jensen MR, Whaley DH, et al. Comparison of clinical and automated breast density measurements: implications for risk prediction and supplemental screening. Radiology. 2016;279(3):710-9.
218. Seo JM, Ko ES, Han BK, Ko EY, Shin JH, Hahn SY. Automated volumetric breast density estimation: A comparison with visual assessment: Clinical Radiology. 68 (7) (pp 690-695), 2013. Date of Publication: July 2013.; 2013.
219. Flohr TG, Lo JY, Gilat Schmidt T, Díaz O, García E, Oliver A, et al. Scattered radiation in DBT geometries with flexible breast compression paddles: a Monte Carlo simulation study. 2017;10132:101324G.

220. Ekpo EU, McEntee MF. Measurement of breast density with digital breast tomosynthesis--a systematic review: The British journal of radiology. 87 (1043) (pp 20140460), 2014. Date of Publication: 01 Nov 2014.; 2014.
221. Ciatto S, Bernardi D, Calabrese M, Durando M, Gentilini MA, Mariscotti G, et al. A first evaluation of breast radiological density assessment by QUANTRA software as compared to visual classification. The Breast. 2012;21(4):503-6.
222. Sickles EA, D'Orsi CJ, Bassett LW, D'Orsi CJ, Sickles EA, Mendelson EB, et al. ACR BI-RADS Atlas, Breast Imaging Reporting and Data System2013.
223. Health and Social Care Centre. Breast Screening programme England 2016-2017. UK: NHS Digital; 2018 4/4/2018.
224. Office for National Statistics (ONS). 2011 Census Guidance and Methodology 2015 [Overview of methods and codes used for 2011 census]. Available from: https://www.ons.gov.uk/census/2011census/2011censusdata/2011censususerguide/variablesandclassifications.
225. Government CaL. The English Indices of Deprivation 2010 [Excel]. 2010 [Table of IMD codes for given LSOAs]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6871/1871208.pdf.
226. Hologic Inc. Selenia Quality Control Manual. Bedford, MA, USA; 2014.
227. West Midlands Cancer Screening QA Reference Centre. NHS Cancer Screening Programme Confidentiality and Disclosure Policy update 20112011.
228. Hudson S, Hjerkind KV, Vinnicombe S, Allen S, Trewin C, Ursin G, et al. Adjusting for BMI in analyses of volumetric mammographic density and breast cancer risk. Breast Cancer Research. 2018;20.
229. Irwin ML, Aiello EJ, McTiernan A, Bernstein L, Gilliland FD, Baumgartner RN, et al. Physical Activity, Body Mass Index, and Mammographic Density in Postmenopausal Breast Cancer Survivors. Journal of Clinical Oncology. 2007;25(9):1061-6.
230. Sun X, Gierach GL, Sandhu R, Williams T, Midkiff BR, Lissowska J, et al. Relationship of Mammographic Density and Gene Expression: Analysis of Normal Breast Tissue Surrounding Breast Cancer. Clinical Cancer Research. 2013;19(18):4972-82.
231. Shepherd JA, Kerlikowske K. Do fatty breasts increase or decrease breast cancer risk? Breast Cancer Research. 2012;14(1):1-3.
232. Hopper JL. Odds per Adjusted Standard Deviation: Comparing Strengths of Associations for Risk Factors Measured on Different Scales and Across Diseases and Populations. American Journal of Epidemiology. 2015;182(10):863-7.
233. Hopper JL, Nguyen TL, Schmidt DF, Makalic E, Song YM, Sung J, et al. Going Beyond Conventional Mammographic Density to Discover Novel Mammogram-Based Predictors of Breast Cancer Risk. Journal of clinical medicine. 2020;9(3).
234. Hudson SM, Wilkinson LS, Denholm R, Stavola BLD, dos-Santos-Silva I. Ethnic and age differences in right-left breast asymmetry in a large population-based screening population. The British Journal of Radiology. 2019;0(0):20190328.
235. Li T, Li J, Heard R, Gandomkar Z, Ren J, Dai M, et al. Understanding mammographic breast density profile in China: A Sino-Australian comparative study of breast density using real-world data from cancer screening programs. Asia-Pacific Journal of Clinical Oncology. 2022;18(6):696-705.
236. Hudson SM, Wilkinson LS, De Stavola BL, Dos-Santos-Silva I. Left-right breast asymmetry and risk of screen-detected and interval cancers in a large population-based screening population. Br J Radiol. 2020:20200154.
237. Hudson SM, Wilkinson LS, Stavola BLD, dos-Santos-Silva I. To what extent are objectively measured mammographic imaging techniques associated with compression outcomes. The British Journal of Radiology.0(0):20230089.

238. Heller SL, Hudson S, Wilkinson LS. Breast density across a regional screening population: effects of age, ethnicity and deprivation. Br J Radiol. 2015;88(1055):20150242.
239. Hudson S, Wilkinson L, de Stavola B, dos Santos Silva I. Are mammography image acquisition factors, compression pressure and paddle tilt, associated with breast cancer detection in a large population-based screening population? British Journal of Radiology - submitted for peer review. 2023.
240. Cancer Research UK. Common cancers Incidence Projections 2023 [updated February 2023. Calculated by the Cancer Intelligence Team at Cancer Research UK]. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/common-cancers-compared#heading-Four.
241. Saunders CL, Abel GA, Turabi AE, Ahmed F, Lyratzopoulos G. Accuracy of routinely recorded ethnic group information compared with self-reported ethnicity: evidence from the English Cancer Patient Experience survey. BMJ Open. 2013;3(6):e002882.
242. Maheswaran R, Pearson T, Jordan H, Black D. Socioeconomic deprivation, travel distance, location of service, and uptake of breast cancer screening in North Derbyshire, UK. Journal of epidemiology and community health. 2006;60(3):208-12.
243. Bhola J, Jain A, Foden P. Impact of index of multiple deprivation and ethnicity on breast screening uptake in the North West of England. Breast Cancer Research. 2015;17(1):P24.
244. Douglas E, Waller J, Duffy SW, Wardle J. Socioeconomic inequalities in breast and cervical screening coverage in England: are we closing the gap? Journal of Medical Screening. 2016;23(2):98-103.
245. Mottram R, Knerr WL, Gallacher D, Fraser H, Al-Khudairy L, Ayorinde A, et al. Factors associated with attendance at screening for breast cancer: a systematic review and meta-analysis. BMJ Open. 2021;11(11):e046660.
246. US National Library of Medicine. My Personalized Breast Screening (MyPeBS) USA: US National Library of Medicine,; 2018 [Available from: https://clinicaltrials.gov/ct2/show/NCT03672331?term=myPebs&draw=2&rank=1.
247. Tabár L, Fagerberg CJ, Gad A. Reduction in mortality from breast cancer after mass screening with mammography: randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. Lancet. 1985;1:829.
248. Autier P, Boniol M, Koechlin A, Pizot C, Boniol M. Effectiveness of and overdiagnosis from mammography screening in the Netherlands: population based study. Bmj. 2017;359:j5224.
249. Welch HG, Prorok PC, O'Malley AJ, Kramer BS. Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness. N Engl J Med. 2016;375(15):1438-47.
250. Harding C, Pompei F, Burmistrov D, Wilson R. Long-term relationships between screening rates, breast cancer characteristics, and overdiagnosis in US counties, 1975-2009. Int J Cancer. 2019;144(3):476-88.
251. Blanks RG, Wallis MG, Alison RJ, Given-Wilson RM. An analysis of screen-detected invasive cancers by grade in the English breast cancer screening programme: are we failing to detect sufficient small grade 3 cancers? European Radiology. 2021;31(4):2548-58.
252. MacInnes EG, Duffy SW, Simpson JA, Wallis MG, Turnbull AE, Wilkinson LS, et al. Radiological audit of interval breast cancers: Estimation of tumour growth rates. Breast (Edinburgh, Scotland). 2020;51:114-9.
253. Gilbert F, Tucker L, Gillan M, Willsher P, Cooke J, Duncan K, et al. The TOMMY trial: a comparison of TOMosynthesis with digital MammographY in the UK NHS Breast Screening Programme - a multicentre retrospective reading study comparing the diagnostic performance of digital breast tomosynthesis and digital mammography with digital mammography alone. Health Technol Assess. 2015;19(4).
254. Hollingsworth AB. Redefining the sensitivity of screening mammography: A review. American journal of surgery. 2019;218(2):411-8.

255. Azavedo E, Zackrisson S, Mejàre I, Heibert Arnlind M. Is single reading with computer-aided detection (CAD) as good as double reading in mammography screening? A systematic review. BMC Medical Imaging. 2012;12(1):22.
256. Kohli A, Jha S. Why CAD Failed in Mammography. Journal of the American College of Radiology : JACR. 2018;15(3 Pt B):535-7.
257. NHS Cancer Screening Programmes. QUALITY ASSURANCE GUIDELINES FOR MAMMOGRAPHY Including Radiographic Quality Control. Sheffield: NHS Cancer Screening Programmes; 2006. Contract No.: NHSBSP Publication No 63.
258. Cohen SL, Blanks RG, Jenkins J, Kearins O. Role of performance metrics in breast screening imaging - where are we and where should we be? Clin Radiol. 2018;73(4):381-8.
259. Lau S, Abdul Aziz YF, Ng KH. Mammographic compression in Asian women. PLoS ONE. 2017;12(4):e0175781.
260. McCarthy AM, Yamartino P, Yang J, Bristol M, Conant EF, Armstrong K. Racial differences in false-positive mammogram rates: results from the ACRIN Digital Mammographic Imaging Screening Trial (DMIST). Medical care. 2015;53(8):673-8.
261. Heller SL, Hudson SM, Wilkinson LS. Indicators of Future Breast Cancer Risk at Prevalent Round Screen. ECR European Society of Radiology; Vienna2015.
262. Dumonteil S, Wilkinson LS, Burnside ES, Heller SL, Hudson S, Mohammadi S. The relationship between quantitative breast density, age and cancer detection rate in a large UK breast screening population. RSNA 20172017.
263. Ward L, Heller S, Hudson S, Wilkinson L. Parenchymal pattern in women with dense breasts. Variation with age and impact on screening outcomes: observations from a UK screening programme. Eur Radiol. 2018;28(11):4717-24.
264. Chu A, Sung P, Lee J, Cheun J-H, Hwang K-T, Lee K, et al. Association of body composition fat parameters and breast density in mammography by menopausal status. Scientific reports. 2022;12(1):22224.
265. Stone J, Ding J, Warren RM, Duffy SW, Hopper JL. Using mammographic density to predict breast cancer risk: dense area or percentage dense area. Breast cancer research : BCR. 2010;12(6):R97.
266. Kirsti Vik Hjerkind, Ellingjord-Dale M, Johansson ALV, Hildegunn Siv Aase, Hoff SR, Solveig Hofvind, et al. Volumetric mammographic density, age-related decline, and breast cancer risk factors in a national breast cancer screening program. Cancer Epidemiology Biomarkers & Prevention. 2018;June 2018.
267. Harvey JA, Fechner RE, Moore MM. Apparent Ipsilateral Decrease in Breast Size at Mammography: A Sign of Infiltrating Lobular Carcinoma. Radiology. 2000;214(3):883-9.
268. Wang X, Lederman D, Tan J, Wang XH, Zheng B. Computerized detection of breast tissue asymmetry depicted on bilateral mammograms. A preliminary study of breast risk stratification. Academic Radiology. 2010;17(10):1234-41.
269. Wang X, Lederman D, Tan J, Wang XH, Zheng B. Computerized prediction of risk for developing breast cancer based on bilateral mammographic breast tissue asymmetry. Medical Engineering & Physics. 2011;33(8):934-42.
270. Li Y, Fan M, Cheng H, Zhang P, Zheng B, Li L. Assessment of global and local region-based bilateral mammographic feature asymmetry to predict short-term breast cancer risk. Phys Med Biol. 2018;63(2):025004.
271. Thornhill R, Moller AP. Developmental stability, disease and medicine. Biological Reviews of the Cambridge Philosophical Society. 1997;72(4):497-548.
272. Jasienska G, Lipson SF, Ellison PT, Thune I, Ziomkiewicz A. Symmetrical women have higher potential fertility. Evolution and Human Behavior. 2006;27(5):390-400.
273. Campoy EM, Laurito SR, Branham MT, Urrutia G, Mathison A, Gago F, et al. Asymmetric Cancer Hallmarks in Breast Tumors on Different Sides of the Body. PLOS ONE. 2016;11(7):e0157416.

274. Chen JH, Chan S, Yeh DC, Fwu PT, Lin M, Su MY. Response of bilateral breasts to the endogenous hormonal fluctuation in a menstrual cycle evaluated using 3D MRI. Magn Reson Imaging. 2013;31(4):538-44.
275. Senie RT, Saftlas AF, Brinton LA, Hoover RN. Is breast size a predictor of breast cancer risk or the laterality of the tumor? Cancer Causes & Control. 1993;4(3):203-8.
276. Perkins CI, Hotes J, Kohler BA, Howe HL. Association between Breast Cancer Laterality and Tumor Location, United States, 1994–1998. Cancer Causes & Control. 2004;15(7):637-45.
277. Roychoudhuri R, Putcha V Fau - Moller H, Moller H. Cancer and laterality: a study of the five major paired organs (UK). Cancer causes & control : CCC. 2006;Volume 17, Issue 5, (0957-5243 (Print)):655–62.
278. Amer MH. Genetic factors and breast cancer laterality. Cancer Management and Research. 2014;6:191-203.
279. Cheng SA, Liang LZ, Liang QL, Huang ZY, Peng XX, Hong XC, et al. Breast cancer laterality and molecular subtype likely share a common risk factor. Cancer Manag Res. 2018;10:6549-54.
280. Dustler M, Andersson I, Brorson H, Frojd P, Mattsson S, Tingberg A, et al. Breast compression in mammography: pressure distribution patterns. Acta Radiol. 2012;53(9):973-80.
281. Boyce M, Gullien R, Parashar D, Taylor K. Comparing the use and interpretation of PGMI scoring to assess the technical quality of screening mammograms in the UK and Norway. Radiography. 2015;21(4):342-7.
282. Waade GG, Moshina N, Sæbuødegård S, Hogg P, Hofvind S. Compression forces used in the Norwegian breast cancer screening program. Br J Radiol. 2017;90.
283. Waade GG, Sanderud A, Hofvind S. Compression force and radiation dose in the Norwegian Breast Cancer Screening Program. European journal of radiology. 2017;88:41-6.
284. Waade GG, Sebuodegard S, Hogg P, Hofvind S. Breast compression across consecutive examinations among females participating in BreastScreen Norway. Br J Radiol. 2018;91(1090):20180209.
285. Ma WK, Brettle D, Howard D, Kelly J, Millington S, Hogg P. Extra patient movement during mammographic imaging: an experimental study. The British Journal of Radiology. 2014;87(1044):20140241.
286. Broeders MJM, ten Voorde M, Veldkamp WJH, van Engen RE, van Landsveld – Verhoeven C, ’t Jong – Gunneman MNL, et al. Comparison of a flexible versus a rigid breast compression paddle: pain experience, projected breast area, radiation dose and technical image quality. European Radiology. 2015;25(3):821-9.
287. Moshina N, Sebuodegard S, Evensen KT, Hantho C, Iden KA, Hofvind S. Breast compression and experienced pain during mammography by use of three different compression paddles. European journal of radiology. 2019;115:59-65.
288. O'Leary D, Grant T, Rainford L. Image quality and compression force: the forgotten link in optimisation of digital mammography? Breast cancer research : BCR. 2011;13(Suppl 1):P10-P.
289. Agasthya GA, D'Orsi E, Kim YJ, Handa P, Ho CP, D'Orsi CJ, et al. Can Breast Compression Be Reduced in Digital Mammography and Breast Tomosynthesis? AJR Am J Roentgenol. 2017;209(5):W322-w32.
290. de Groot JE, Hopman IGM, van Lier M, Branderhorst W, Grimbergen CA, den Heeten GJ. Pressure-standardised mammography does not affect visibility, contrast and sharpness of stable lesions. European journal of radiology. 2017;86:289-95.
291. Henderson LM, Benefield T, Marsh MW, Schroeder BF, Durham DD, Yankaskas BC, et al. The Influence of Mammographic Technologists on Radiologists' Ability to Interpret Screening Mammograms in Community Practice. Academic Radiology. 2015;22(3):278-89.
292. Perry N, Broeders M, de Wolf C, Tornberg S, Holland R, von Karsa L. European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition--summary document. Ann Oncol. 2008;19(4):614-22.

293. Waade GG, Holen Å, Sebuødegård S, Aase H, Pedersen K, Hanestad B, et al. Breast compression parameters among women screened with standard digital mammography and digital breast tomosynthesis in a randomized controlled trial. Acta Radiologica. 2020;61(3):321-30.
294. Jeukens C, van Dijk T, Berben C, Wildberger JE, Lobbes MBI. Evaluation of pressure-controlled mammography compression paddles with respect to force-controlled compression paddles in clinical practice. Eur Radiol. 2019;29(5):2545-52.
295. Dustler M. Pressure distribution in mammography. Mechanical imaging and implications for breast compression [PhD]:

Lund University, Faculty of Medicine; 2016.
296. Metsala E, Richli Meystre N, Pires Jorge J, Henner A, Kukkes T, Sa Dos Reis C. European radiographers' challenges from mammography education and clinical practice - an integrative review. Insights into imaging. 2017;8(3):329-43.
297. Serwan E, Matthews D, Davies J, Chau M. Mammographic compression practices of force- and pressure-standardisation protocol: A scoping review. Journal of medical radiation sciences. 2020;n/a(n/a).
298. Kallenberg M, Karssemeijer N. Temporal stability of fully automatic volumetric breast density estimation in a large screening population. European Society of Radiology; 20132013.
299. Engelken F, Singh JM, Fallenberg EM, Bick U, Bottcher J, Renz DM. Volumetric breast composition analysis: reproducibility of breast percent density and fibroglandular tissue volume measurements in serial mammograms. Acta Radiol. 2014;55(1):32-8.
300. Jeffers AM, Sieh W, Lipson JA, Rothstein JH, McGuire V, Whittemore AS, et al. Breast cancer risk and Mammographic Density assessed with semiautomated and Fully automated Methods and Bi-raDs. Radiology. 2017;282(2):348-55.
301. Astley S, Harkness E, Sergeant J, Stavrinos P, Warren R, Wilson M, et al. Proffered Paper: A comparison of four methods of mammographic density measurement in the UK Predicting Risk Of Cancer At Screening (PROCAS) study-on behalf of the PROCAS Study team. European Journal of Cancer. 2016;61 (1):S6.
302. Astley SM, Harkness EF, Sergeant JC, Warwick J, Stavrinos P, Warren R, et al. A comparison of five methods of measuring mammographic density: a case-control study. Breast Cancer Research. 2018;20(1):10.
303. Tukey J. Exploratory Data Analysis: Addison-Wesley; 1977.
304. Onega T, Smith M, Miglioretti DL, Carney PA, Geller BA, Kerlikowske K, et al. Radiologist agreement for mammographic recall by case difficulty and finding type. Journal of the American College of Radiology : JACR. 2012;9(11):788-94.