Mammographic (Breast) Density as a Potential Biomarker for Endocrine Trial Treatment Efficacy in Breast Cancer Prevention

Judith Marie Jobling
B. Sci. (Chem.), Univ. Pittsburgh, PA, USA
B. Comp. Sci., Flinders Univ., SA, Australia

This thesis is submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy (Community Medicine and Clinical Epidemiology),
School of Medicine and Public Health
University of Newcastle, NSW, Australia
February 2017

Statement of originality

The thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to the final version of my thesis being made available worldwide when deposited in the University's Digital Repository, subject to the provisions of the Copyright Act 1968.

Judith M. Jobling

Statement of Collaboration

I hereby certify that the work embodied in this thesis has been done in collaboration with other researchers. I have included, below, a statement clearly outlining the extent of collaboration, with whom and under what auspices:

Professor John F. Forbes, Professor Jack Cuzick, Professor Tony Howell and other members of the IBIS-II international steering committee are responsible for the design and conduct of the IBIS-II trial. Calvary Mater Newcastle IBIS-II mammograms were collected by trial coordinators from the Calvary Mater Newcastle hospital, the Australia New Zealand Breast Cancer Trials Group and myself. I de-identified and measured the mammograms utilised in this project. I designed and conducted the analyses in this thesis with assistance from Professor Forbes and Professor Catherine D'Este. The Primary analysis of this thesis was also performed with the assistance of Dr Adam Brentnall, Ms Emma Christine Atakpa and Dr Ivana Sestak from the Queen Mary University of London.

Judith M. Jobling

Acknowledgements

I am very appreciative of the support and encouragement received from numerous people and organisations whilst undertaking this project.

I would like to thank my principal supervisor Professor John F. Forbes for the opportunity to undertake a PhD and to better understand the challenges involved in the treatment and prevention of cancer. I am grateful for his facilitation of this project and other activities, and for introducing me to the fascinating subject of mammographic (breast) density. His enthusiasm, insight and forward thinking have helped make great inroads into treatment improvements for breast cancer and enhanced the lives of many women and their families globally. I am grateful to be able to make a small contribution towards eradication of this disease.

Expert and judicious guidance from my co-supervisor Professor Catherine D'Este was instrumental in providing a statistical framework for this project, as well as organisation of the thesis. She knew when to give me the freedom to resolve an issue independently, but also when more explicit direction was required. Thanks very much especially Cate for undertaking the tedious job of organising and repetitively reviewing the thesis chapters prior to submission.

I am grateful to the IBIS-II breast cancer prevention trial participants at the Calvary Mater Newcastle hospital and the Southern Highlands Cancer Centre for agreeing to collection of their mammograms for use in approved studies, and for volunteering to take part in the trial. Thanks very much also to the ever helpful Yvonne Harrower and Vicki Sproule who coordinated the IBIS-II trial at the Calvary Mater and facilitated mammogram collection for this study, and to Helen Garner for her cheerful administrative support. I am also very grateful to the Department of Radiology at the Mater for providing the numerous digital mammograms for this project.

My sincere thanks to the Australia New Zealand Breast Cancer Trials Group (ANZ BCTG) for their support my projects, and to the lovely ANZ BCTG IBIS-II trial central coordinating staff Rochelle Thornton and Nicole Francis who procured mammograms and data for this project. Thanks also to Prof Jack Cuzick, Dr Adam Brentnall, Dr Iva Sestak, Ms Emma Atakpa and Ms Michelle Sleeth at the Queen Mary University of London, and the IBIS-II trial international steering committee, for their assistance with the primary analysis for this project.

Many thanks to the lower Hunter New England (Newcastle) BreastScreen, for the opportunity to enjoy the camaraderie of a busy but lively and humour-filled clinical workplace. Thanks to Michael Symonds, Director of HNE Imaging, and BreastScreen Managers Karen Clark and Rebecca Delaporte for their support of my projects. I appreciated the clinical insights into radiography and screening, as well as the generosity of Chief Radiographer Rhonda Brown and the other BreastScreen radiographers, nurses, radiologists and staff. Thanks to Kathy Fulcher for organising the regular birthday lunch feasts, these were most enjoyable. Additional thanks also to NSW BreastScreen, Cancer Institute NSW, my supervisors, and in particular Dr Arthur Hung, for their support of the BreastScreen and 45 and Up Study linkage mammographic density project and application process.

This project would also not have been possible without a student support grant from the Cancer Institute NSW, as well as a NHMRC/University of Newcastle equipment grant and post-

graduate student training funding from the Australian Government. A special thank you is directed to the Masons at Morisset NSW for organising a fund-raising dinner towards purchase of the Array x-ray scanner. I regret the tragic circumstances which precipitated the dinner, but am grateful that you helped the ANZ BCTG obtain a critical piece of equipment for this project.

Additional thanks to the School of Medicine and Public Health at the University of Newcastle for providing additional research funding, access to post-graduate courses in epidemiology and biostatistics, as well as the repeated opportunity to tutor first year medical students during their problem based learning and other tutorials. This experience helped broaden my understanding of the health system, medical practice, and the physiology of health and disease. Thanks also to the Wollotuka Institute for the opportunity to assist many of the First Australian medical students with the public health course during their first year of study. A particular thank you to A/Prof Mark McEvoy for the opportunity to tutor the first year medical public health course; this reinforced many of the concepts and ideas which underpin my PhD studies. It was fantastic getting to know the other regular public health tutors Meg, Khanrin, Neta, and Pari, as well as the PBL tutors and medical students. Best wishes to all the students and tutors for success in their future endeavours.

Thanks also to Dr Anita Bourke, the Sir Charles Gairdner Hospital and WA BreastScreen for their help with an IBIS-I mammographic density project, and the Hunter Medical Research Institute for providing funding towards this project. I would also like to thank my former co-supervisors Professor Afaf Girgis and Professor David Sibbritt and their colleagues for their support early in my PhD career. I am grateful to Dr Jennifer Stone for her instruction in the use of the Cumulus program.

Many thanks to my friends and colleagues Marg, Coralie and Annette and their families for the enjoyable times as well as support throughout my candidature. I appreciate the encouragement received from my family during all my tertiary studies. In particular, my husband Phil has been very supportive throughout my earlier computer science degree in South Australia and during my PhD candidature in Newcastle. Our family would not be complete without our lovely sons Evan and Hugh. Evan, you've grown into a gorgeous, dedicated and talented adult; your father and I are so proud of you. Hugh, you've always made everyone laugh as you run circles around them. Thanks to both of you for bringing such joy to our lives.

Table of Contents

Table of Contents

List of Tables

Table of Tables

Table of Tables

List of Figures

Lexicon of terms, abbreviations, symbols and acronyms

±	plus or minus
+	positive, e.g. a substance is present in tissue during pathological testing; BC+ = history of breast cancer; hormone+ = positive test for hormonal receptors
-	negative, e.g. a substance is not present in tissue during pathological testing; BC- breast cancer free; ER- estrogen receptor negative
~	approximately
Δ	delta: change in, or difference between. Δ2 is a change of 2
>, ≥	greater than, greater than or equal to
< , ≤	less than, less than or equal to
2D	two dimensional
3D	three dimensional
95%CI	95% Confidence Interval, approximately two times the standard deviation (SD) above and below the mean of a normally distributed variable. See also SD, normally distributed, CI.
ACR	the American College of Radiology
adjuvant	in addition to, additional.
AI	aromatase inhibitor, e.g. anastrozole, letrozole, exemestane. Used to reduce the amount of estrogen available to promote tumour growth by preventing the aromatase enzyme (a protein) to convert androgens (hormones) into estrogen. See also SERM
AIHW	Australian Institute of Health and Welfare
anastrozole ANAS	an aromatase inhibitor (AI). Anastrozole inhibits the production of estrogen in the body by blocking the enzyme aromatase. In post-menopausal women, the primary source of estrogen is from the conversion of androgens to estrogen in the peripheral tissue via the enzyme aromatase. See also estrogen
ANZ BCTG	Australia New Zealand Breast Cancer Trials Group; original name for the (Australian) Breast Cancer Trials (group)
ATAC	breast cancer trial ‘Arimidex, Tamoxifen, Alone or in Combination’

AUC	area under the curve, an estimate of the goodness of fit of 'receiver operator characteristic' plots. An AUC of 0.5 is equivalent to chance (i.e. the risk model is no better than chance at estimating the likelihood an outcome will occur). An AUC of 1.00 is a perfect fit, i.e. the model is able to predict the outcome with 100% accuracy. Most current BC risk models (e.g. the Gail or Tyrer-Cuzick) yield AUC of about 0.6 in most populations.
BC	breast cancer. Typically refers to invasive breast cancer, as opposed to non-invasive (in-situ) cancers. See also in-situ, DCIS, LCIS
BD	breast density
BI-RADS	the Breast Imaging – Reporting and Data System, American College of Radiology (ACR)
Bland Altman plot	a statistical method used to assess agreement between two measurement techniques
BMI	body mass index: weight (kg)/(height (m))^2. Also known as the Quetelet index. A BMI < 17 is typically considered to be underweight, 18-25 normal weight, 26-29 overweight, 30-34 obese, 35+ very obese. BMI is also a significant modifier of percent density: as weight increases the amount of the fat in the breast also tends to increase. Thus BMI is inversely related to percent density.
BMP	BitMaP, a common image format
BRCA	BReast CAncer gene, e.g. BRCA1 or BRCA2 genetic mutations
BreastScreen	the Australian breast screening program, which is federally funded and available to all female Australian citizens and permanent residents aged 40 and over. Women aged 50-74 are specifically targeted to attend screening every two years
BScreen	BreastScreen (above)
CI	Confidence Interval, e.g. a 95% CI. A range of values within which we believe, with a specific probability, that the population parameter will lie. A 95% CI, will include the population parameter 95% of the time (i.e. 19 in 20 cases). See also IQR, standard deviation, 95%CI
CC	cranio-caudal (one of two standard views taken during mammography); see also MLO
CCD	charge coupled device. a type of photosensor used in digital imaging
CINSW	the Cancer Institute New South Wales, Australia
CMN	Calvary Mater Newcastle hospital (Newcastle, Australia)
CMN MD and AI substudy	The name for the project which is the subject of this thesis

confidence interval	see CI
CR	computed radiography- x-ray imaging technique, which produces a digital (electronic) image via use of removable phosphor plates in a cassette. Use of CR cassettes in lieu of film cassettes enables a film mammography machine to produce digital images. See also DR (digital radiography) and film-screen mammography
CRF	case record forms: used to record (participant) data in clinical trials
CRUK	Cancer Research, United Kingdom
CSIRO	the Commonwealth Scientific and Industrial Research Organisation, Australia
DA	dense area (of the breast), often measured in mm^2 or cm^2
DCIS	Ductal Carcinoma In Situ, a non-invasive type of breast cancer which originates in the ducts of the breast. The ducts provide a conduit from the lobular (milk-producing) tissue to the nipple. See also LCIS
dense tissue	(of the breast): the connective (stromal) and glandular (epithelial) tissues of the breast which are opaque during mammography. The dense tissues appear white on mammograms.
density	breast density, in particular mammographic or percent (mammographic) density
DEXA	dual energy x-ray absorptiometry: bone density scan (also known as DXA)
DICOM	digital imaging and communication in medicine: a standard for distributing and viewing medical images (e.g. x-rays, ultrasounds, MRI scans)
DNA	di-nucleic acid, one of the chemicals which comprises the genetic material in living things
DoF	digital on film (an acronym created for use in this thesis): digital mammograms which have been printed to film instead of being retained in digital (fully electronic) format; DoF therefore differ from "original digital" mammograms. See also original digital (mammogram)
DR	direct (digital) radiography: a type of x-ray imaging which is fully digital (fully electronic); a sensor in the mammography machine converts the photons generated by the x-ray beam directly into an (electronic) image; an intermediary cassette is not required to produce the electronic image. See also CR (computed radiography) and film-screen mammography.
DXA	dual energy x-ray absorptiometry: bone density scan (also known as DEXA)
DV	dense volume, the estimate of the total volume (cm^3) of the breast occupied by dense tissue

DVD	digital video disc
ECM	extracellular matrix, the connective tissues between cells in the body
episode	for this thesis, an episode is a visit to an imaging facility to undertake mammography (x-ray images of the breast)
ER	estrogen receptor
ER+	estrogen receptor positive: signifies tissues that have tested positive (+) for the presence of estrogen receptors (ER)
ER-	estrogen receptor negative: signifies tissues that do not have, or are below the pathological testing thresholds for, estrogen receptors
estrogen	a hormone produced in humans which is responsible for the female sex characteristics.
EXE	exemestane, an aromatase inhibitor (AI)
F test	statistical test named after statistician R. A. Fisher. The F-test determines if two variances are equal. The F-test uses the F-distribution (a probability distribution), which is the ratio of two chi-square distributions.
film-screen mammography	mammographic (breast imaging technique utilising x-ray imaging) technique in which x-rays are passed through a compressed breast. The x-ray image is captured on a cassette under the breast which holds traditional photographic (silver coated) film. See also CR (computed radiography) and DR (digital radiography).
first degree relative	a person's mother, father, brother, sister, son or daughter. See also second degree relatives
first mammograms	a term used in this thesis to signify the earliest mammograms for a participant collected for this project, i.e. baseline mammograms, or the earliest available mammograms if baseline images were not obtained
follow up	a clinic visit by IBIS-II participants to receive trial medication and provide trial follow up information such as treatment compliance
genotype	the genetic makeup of an organism. Due to genetic interaction (e.g. 'dominant' vs 'recessive' genes, imprinting and/or DNA methylation) not all genes are expressed and others vary in how much they are expressed; additionally, the environment of an organism affects how its genotype is expressed. This results in each expression of a unique phenotype (observed characteristics) for each organism; phenotype can differ for two organisms even if genotype is identical. See also phenotype
HER	human epidermal growth factor receptor

HER2+	human epidermal growth factor receptor 2 positive (+): tissue that tests positive (meets pathological minimum standards) for HER2 protein overexpression and/or HER2 gene amplification
high PD	percent density ≥ 50%
HNE	Hunter New England, a geographical region in New South Wales (NSW), Australia
HNEH	Hunter New England Health (one of 15 health districts in the state of New South Wales)
HR	hazard ratio. A name for the relative risk ratio utilised in survival analysis.
HREC	human research ethics committee. All research involving human participants must be approved by a human research ethics committee before the project commences.
HRT	hormone replacement therapy. Used to reduce the side-effects of menopause such as hot flushes. HRT is associated with an increase in breast density. Combination HRT (estrogen and progestin) is associated with an increase in breast cancer risk.
IBIS	breast cancer prevention trial 'International Breast cancer Intervention Study'
IBIS-I	breast cancer prevention trial 'International Breast cancer Intervention Study –I' placebo vs tamoxifen
IBIS-II	breast cancer prevention trial 'International Breast cancer Intervention Study –II'; consists of the IBIS-II Prevention trial (IBIS-IIP) of anastrozole vs placebo for women at high risk of breast cancer, an IBIS-IIP bone-density substudy, and the IBIS-II Ductal Carcinoma In-Situ (IBIS-II DCIS) trial of anastrozole vs placebo in women with diagnosed DCIS. IBIS-II (unless otherwise specified) in this thesis refers only to the IBIS-II Prevention trial
ICC	intraclass correlation. This assesses the similarity/relatedness of observations within the same group (e.g. measurement time or mammographic technique).
ID	identification, e.g. an identification (ID) number replaces participant name in clinical trials
incidence	the number of new events during a particular span of time. Cancer incidence typically is described as new cases of cancer diagnosed during a year per 100,000 members of population. See also prevalence
in-situ	a cancer which does not typically invade the surrounding tissue. See also BC, DCIS, LCIS
IQR	intra-quartile range: the range of values that fall between the lowest quartile (lowest 25% of values) and highest quartile (upper 25% of values); i.e. the middle 50% of an ordered range of values.

L	left
LCC	cranio-caudal (CC) view of the left breast taken during mammography
LCIS	Lobular Carcinoma In Situ, a non-invasive form of breast cancer which originates in the lobular (lobes, milk-producing) tissue of the breasts. See also DCIS
LET	letrozole, an aromatase inhibitor (AI)
LMLO	medio-lateral oblique view of the left breast taken during mammography
LOA	limits of agreement, used to describe the highest and lowest values of the middle 95% of a set of values in a Bland Altman plot
MAP	breast cancer 'Mammary Prevention Trial' 1 (.1), 2 (.2) or 3 (.3)
mammography	a breast imaging technique in which x-rays are passed through a compressed breast. The x-ray image is captured on a cassette (film-screen and computed radiography (CR)) or directly via a digital sensor (digital radiography (DR)).
mammographic density	the dense (white appearing) tissues of the breast as they appear on an x-ray of the breast (mammogram). See also dense tissue, density
MD	mammographic (breast) density. See also dense tissue.
mean	a measure of central tendency which is the average of a variable. Used for parametric (normally distributed data). See also median, CI, IQR.
median	a measure of central tendency which is the equivalent of the mean for non-parametric (non-normally distributed, skewed) numeric data. The median is the value at the 50% percentile (quartile 2) of an ordered set of numeric data. See also mean, CI, IQR.
MI, mi	multiple imputation (a type of statistical estimation to account for missing data)
MIRC	Medical Imaging Resource Centre of the Radiological Society of North America
MLO	medio-lateral oblique (one of two standard views taken during mammography); see also CC
MRI	medical resonance imaging
n, N	number (of), e.g. number of participants or mammograms
NSABP	National Surgical Adjuvant Breast and Bowel Project, USA
NHMRC	the National Health and Medical Research Council, Australia
NHS	the National Health Service, United Kingdom

NIH	the National Institutes of Health, USA
non-parametric	numeric data which are non-parametric do not have a probability completely specified distribution which is not symmetrically bell-shaped. The types of statistical tests utilised on non-parametric data differ from those utilised on parametric (normal) data. See also parametric
normally distributed	numeric data whose probability distribution is in the shape of a symmetric bell. The middle of the bell is the mean of the data. See also parametric, non-parametric
NSW	New South Wales, Australia
OD	optical density. An optical density of 0 is white; higher numbers (e.g. 4.3) denote deeper blacks
original digital (mammogram)	a digital mammogram which was collected in its (original) fully electronic format; See also DoF (digital printed on film)
PACS	picture archiving and communication system: computer system used to view and store digital (electronic) medical images
parametric	numeric data which have a specified distribution, often a'normal' distribution. Normally distributed data have a probability distribution which is bell-shaped. Parametric distributions have special properties which enable use of specific statistical tests, such as the well known 't-test'. See also non-parametric.
PD	percent (mammographic) density. The proportion of the total breast area which is covered by dense (connective and glandular) tissue
phenotype	the expressed characteristics of an organism, as determined by the genetic makeup of that organism and its interaction with the environment. See also genotype.
PLAC	placebo
PNG	Portable Network Graphic, a common image format
PR	progestin receptor
PR+	tissue that tests positive for progestin receptor
prevalence	the number of events which exist per head of population at a particular time. Cancer prevalence is a count of all the people who are alive (at a particular point in time) after a cancer diagnosis, divided by the total population, expressed as a percentage or #/100,000. For instance, BC prevalence is increasing, not only because incidence (new BC cases) is increasing but because more and more women are surviving for many years post-diagnosis due to improvements in screening and treatment. See also incidence
Prob, prob	probability

Project team	Professor Forbes, Professor D'Este and Mrs Jobling
Q1	quartile 1, the value which is located at the 25% percentile in an ordered set of numeric values; see also IQR
Q3	quartile 3, the value which is located at the 75% percentile in an ordered set of numeric values; see also IQR
QMUL	the Queen Mary University of London, UK
RCC	cranio-caudal view taken of the right breast during mammography
relative risk	see RR
research team	the project team: Professor Forbes, Professor D'Este and Mrs Jobling
RLX	raloxifene, a serum estrogen receptor modulator (SERM)
RMLO	medio-lateral oblique view taken of the right breast during mammography
RR	relative risk, a ratio (comparison) of two probabilities of risk
RSNA	Radiological Society of North America
SCC	six category classification for percent density: 0%, 1 to 10%, 10 to 25%, 25 to 50%, 50 to 75%, >75% [Boyd 1995]
SD, sd	standard deviation: the square root of the variance of a normally distributed (parametric) variable. The sd is an estimate of the variability of a numeric quantity. See also mean, median, confidence interval (CI)
second degree relatives	a person's aunts, uncles, grandparents, grandchildren, first cousins (children of aunts and uncles). See also first degree relative
SERM	serum estrogen receptor modulator, e.g. tamoxifen, raloxifene. Used to prevent breast cancer by reducing the amount of estrogen available to promote tumour growth. SERMs work by competing with estrogen for estrogen receptors on cells, and in many cell types this reduces the growth promoting effects of estrogen on the tissues. See also AI
SNP	single nucleotide polymorphism ('snip'). SNPs are differences at single points in DNA; a single nucleotide is changed at a particular point in a DNA strand (set of genes). SNPs are markers of genetic differences between organisms. Some SNPs are associated with increased risk of disease, others may indicate a reduced chance of developing an illness.
standard deviation	see SD

TAM	tamoxifen, a serum estrogen receptor modulator (SERM)
TIFF	Tag Image File Format, a common image format
Tx	treatment, treated
Type	a term utilised in this thesis to denote film vs digital mammograms (i.e. film and digital mammograms are different Types of mammograms); see also Version
UK	the United Kingdom
US	ultrasound or United States
U.S., USA	the United States (of America)
Version	a term utilised in this thesis to denote mammograms which were produced via different acquisition strategies. Digital mammograms produced from different equipment (e.g. different mammography machines) as well as different versions of software (e.g. different post-processing algorithms) are considered to have different mammogram 'Versions' in this thesis. Film mammograms are also a different Version from digital mammograms. see also Type
very high PD	percent density $\geq 75\%$
VPD	Volumetric percent density (cm^3), the percentage of the breast volume which is occupied by dense tissues; VPD is typically estimated using 'raw' (pre-processed) digital mammographic data
VPD, AI and TAM project	The name given in this thesis to a project which compared volumetric longitudinal MD measurements of AI and tamoxifen treated women diagnosed with breast cancer with longitudinal MD measurements of healthy controls: Engmann, N. J., et al. (2017). "Longitudinal changes in volumetric breast density with tamoxifen and aromatase inhibitors." Cancer Epidemiology Biomarkers & Prevention.

Synopsis

Mammographic (breast) density is comprised of the glandular and connective tissues of the breast which appear dense (white) on mammograms, x-rays of the breast. Reductions in both mammographic density and breast cancer risk are associated with estrogen lowering therapies (endocrine treatments) such as tamoxifen and the aromatase inhibitors. The Primary Aim of this thesis was to compare the mammographic density response in women at high (2-fold) risk of breast cancer treated with the aromatase inhibitor anastrozole for breast cancer prevention in the International Breast (cancer) Intervention Study II (IBIS-II) trial with the mammographic density response in similar women randomised to placebo treatment, to ascertain if mammographic density may be a biomarker during endocrine therapy for breast cancer (prevention).

The literature review undertaken for this thesis confirmed breast cancer is a heterogenous disease with many risk factors; many breast cancer risk factors are also factors associated with mammographic density. A measurement technique reliability analysis revealed mammographic density measurements made using visual assessment were less consistent than measurements made with a semi-automated thresholding technique. A longitudinal statistical model of mammographic density change for120 Calvary Mater Newcastle hospital IBIS-II participants was developed using a mixed regression model. The results for the Primary Aim show longitudinal change in mammographic density for anastrozole and placebo treated IBIS-II participants does not differ; this result however is constrained by the small number of participants sampled and confounded by frequent changes of the film and digital mammography equipment used to take the trial mammograms. A sensitivity analysis undertaken with participants' film mammograms only hints at a possible reduction in the rate of annual change in mammographic density for anastrozole treated participants relative to controls. This latter result supports the recent findings of a volumetric mammographic density study which found significantly greater annual decreases in volumetric mammographic density for breast cancer cases treated with aromatase inhibitors compared to the rates of decrease in healthy controls.

Thesis Overview

The first four chapters of this thesis comprise an introduction, a literature review, a project methods chapter, and a review of the techniques used to measure mammographic (breast) density. Four analysis chapters follow. The first analysis chapter investigates reliability of three different mammographic density measurement techniques (Chapter Five). In Chapter Six, Calvary Mater Newcastle hospital (CMN) International Breast Intervention Study II (IBIS-II) breast cancer prevention trial participant baseline characteristics are summarised, and assessed for associations with baseline measurements of mammographic density. A longitudinal model of mammographic density change is developed for the aggregate of both treated and control CMN IBIS-II participants in Chapter Seven. The final analysis (Chapter Eight) examines differences in longitudinal mammographic density for CMN anastrozole treated participants compared to placebo treated IBIS-II participants, the Primary Aim of this thesis. The thesis concludes with a Discussion chapter which is comprised of the overall conclusions from this body of work, future directions for research, as well as the strengths and limitations of this project.

1. Introduction

Breast cancer is the most common cause of mortality from cancer for Australian women [1], and is the second most frequent cause of death from cancer for women globally [2]. The introduction of new treatments for cancer since the 1940s, such as chemotherapy and radiotherapy, to surgical therapy has greatly enhanced the life expectancy for people diagnosed with cancer, including women diagnosed with breast cancer (BC). Despite a 'war on cancer' being declared in the 1970s in the United States [3, 4], cancer remains a problem for all societies across the globe and a major cause of morbidity and mortality. Great inroads have been made into the screening, diagnosis and treatment of breast cancer with early detection a key aspect in successful treatment.

There are many well established risk factors for breast cancer, such as a family history of breast cancer (BC), being female, and increasing age [5]. Mammographic density is one of the strongest breast cancer risk factors [6] . Mammographic density (MD) is composed of the ductal (glandular) and connective (stromal) tissues of the breast which appear white on mammograms (x-rays) of the breast. The adipose (fatty) tissues of the breast are transparent to x-rays and appear dark on a mammogram. The proportion of the total breast area covered by the dense white tissues is known as percent density (PD). Higher amounts of dense tissue in the breast are associated with a higher risk of breast cancer, but it is not clear why this is so.
For women of similar age and menopausal status, women with ≥75% PD are at a 3 to 5-fold risk of BC compared to women who have very low levels of PD (≤10%). This makes breast density (BD) one of the strongest breast cancer risk factors, and one of the few strong (≥ 3-fold risk) risk factors that is modifiable. However, most women have at least some dense breast tissue; highly adipose breasts are uncommon. Hence when women with very high (≥ 75%) PD are compared at the population level to women with average density (25% to 49%), BC risk is doubled [7].

MD had been a controversial subject clinically since it was first described as a BC risk factor by John Wolfe in 1976 [8]. At that time, many radiologists thought MD solely masked BC, and did not consider it an independent BC risk factor [9]. Masking of BC by MD remains an important clinical issue, exemplified by the mandatory BD reporting legislation in many USA states [10]. The question of how to best image and care for women identified as having high MD remains open as no broad, population based clinical trials have reported on these issues [11]. A number of very large trials (>30,000 women), e.g. in Italy [TBST (Tailored Screening for BC in Premenopausal Women) [12]], the Netherlands [Breast Cancer Screening With MRI in Women Aged 50-75 Years With Extremely Dense Breast Tissue: the DENSE Trial [13]], the US [BCSC-ADVANCE (Assessing BD's Value in Imaging) [14], and WISDOM (Women Informed to Screen Depending on Measures of Risk (Wisdom Study)[15]] are working towards redressing the issue of how to best screen women with dense breasts. .

Use of MD and other measures of BD in BC prediction models such as the Gail model have not substantially increased the accuracy of BC prediction at the level of the individual [16]. However, even amongst BReast CAncer 1 (BRCA1) mutation carriers who have an up to 80% lifetime risk of BC, it is not possible to accurately predict which individuals will develop BC, and which women will remain BC free. Improvement in the ability to predict BC for all women remains an ongoing challenge. Longitudinal MD changes during hormonal treatment to prevent BC may be a useful biomarker to improve BC risk prediction [17, 18].

Recent research has helped elucidate some of the cellular mechanisms that cause high MD. Low levels of the transmembrane receptor CD36 are associated with high mammographic density [19]. Low levels of CD36 are also associated with the dense tissues which often surround invasive breast cancer [19]. Additionally, overexpression of the inflammatory cytokine CCL2 is associated with increased density of murine breast stroma as well as an increased susceptibility to BC in mice; higher levels of CCL2 are also associated with high vs

low areas of mammographic density in human breasts [20]. These associations of CCL2 with both high MD and increased BC risk provide a possible mechanism for the observed protective effect of aspirin on BC incidence [21] and a potential target to reduce BC risk and MD.

Tamoxifen, a serum estrogen receptor modulator (SERM), reduces risk of breast cancer recurrence and also prevents breast cancer by blocking the stimulation of breast tissue by estrogen [22]. Tamoxifen reduces mammographic density, and strong (≥10%) decreases in PD in women treated with tamoxifen for BC prevention are associated with a 63% reduction in BC risk [17]. Tamoxifen has been shown to increase CD36 expression in vivo.

The three aromatase inhibitors (AI) in common clinical use—anastrozole, letrozole and exemestane—prevent the aromatase enzyme from converting androgens to estrogens in peripheral tissues such as the breast [23]. Contralateral breast cancer is reduced by ~75% in women with hormone-sensitive early BC treated with five years of AI therapy [24]. AI treatment is associated with modest reductions in MD, however it is not yet known whether measureable changes in MD during AI therapy will be useful as an early biomarker for treatment efficacy. This thesis aims to explore whether changes in MD during preventive treatment for BC with anastrozole in healthy, high-risk women differ from those on placebo treatment. This is important because early (1 year) reductions in MD associated with AI treatment may also be associated with reductions in future BC occurrence; alternatively if MD does not change during AI treatment, this may indicate other therapies should be pursued.

Because up to 80 % of BC is hormone sensitive [25], AI have been trialled as BC preventive agents for post-menopausal women. The IBIS-II trial is an international BC prevention trial which compares five years of treatment with anastrozole to five years of placebo treatment for BC prevention in women at high (approximately 2-fold) risk of BC. At a median of five years of follow up, participants in IBIS-II randomised to active treatment with anastrozole were half as

likely to develop BC as participants randomised to placebo treatment [26]. A 60% relative reduction in the development of hormone-sensitive BC was seen for IBIS-II anastrozole treated participants relative to controls. Similar improvements in disease free survival were seen during the Mammary Prevention Trial 3 (MAP.3) at three years of median follow up: participants randomised to exemestane treatment had a 65% relative reduction in BC incidence compared to those on placebo treatment [27].

Longitudinal MD changes in response to AI treatment lasting more than one year are not well characterised. When this project commenced in 2010, limited information was available on the effect of AI treatment on MD for women with hormone sensitive BC [28] as well as women undertaking AI treatment for BC prevention (MAP.1 trial) [29]. Since 2010, nine additional studies investigated the effect of AI upon longitudinal MD. Notably, only two of these studies evaluated MD at 12 and 24 months of AI treatment [30, 31]; one other evaluated MD at 12, 24 and 26 months [32]. The median time between measurements of volumetric density in another study was 32 months [33]. All others evaluated change in MD solely at approximately 12 months of AI treatment.

The Primary Aim for this thesis was to determine if participants in the IBIS-II breast cancer prevention trial randomised to anastrozole experienced measureable decreases in MD relative to trial controls. This may indicate that MD is a potential biomarker for the prevention of BC in high risk women. Because MD changes for AI treatment longer than 12 months duration have not been well studied, the Primary Aim included characterisation of longitudinal changes in MD for treated vs control participants during all five years of trial treatment, as well as two years post-treatment.

Mammographic density has been described as “the best–kept secret” in the medical community [34, 35]. One of the reasons for this is MD is difficult to measure repeatably and reliably.

Hence a secondary Aim for this project was to determine which of the available measurement techniques provided the most repeatable and reliable MD measurement of CMN IBIS-II participant mammograms. The expected average decrease in MD due to AI treatment was small (1 to 2%) [29], hence the technique also needed to be capable of measuring potentially small longitudinal differences in MD.

Mammographic density is associated with many other BC risk factors, such as age and body mass index (BMI). These other BC risk factors, particularly age and BMI, confound the association between MD and BC risk. They may also confound estimation of longitudinal change in MD, and need to be controlled for during statistical analysis.

An additional secondary Aim of this project was to characterise the associations between baseline MD and other baseline BC risk factors in the population of CMN IBIS-II participants. This ascertained if the expected relationships between MD and BC risk factors were present for this population of high-risk women, in keeping with the MD literature.

Prior to incorporating important covariates such as BC risk factors into longitudinal statistical models of change, a representative model for mean change over time is often developed [36]. Subsequent to developing an 'unconditional' (without covariates) mean model of change, important covariates such as confounders are added to the model before the addition of covariates of interest such as treatment group.

The IBIS-II trial is ongoing, and trial treatment status remains double-blind to everyone except the IBIS-II trial statistician. Because treatment status could not be divulged, the final secondary Aim of this project was to develop a suitable model for MD longitudinal change, adjusted for important covariates, prior to incorporating treatment allocation into the model. This was

necessary in order to provide a suitable set of statistical models to the IBIS-II trial statistician, who then performed the unblinded analysis of treated vs control groups (thesis Primary Aim).

This study (comparison of MD response to AI vs placebo in Calvary Mater Newcastle IBIS-II participants) is important because few MD studies have examined longitudinal MD and AI in (high risk) women in contrast with a high quality comparison group (i.e. similar women randomly allocated to a control group receiving no treatment). Single arm trials of MD in women treated with AI are helpful, as are case-control and cohort studies, but these do not provide the superior data possible in an RCT. If early (1 year) decreases in MD are proven to be a biomarker which shows treatment with five years of AI is effective for these women (i.e. decreases in MD due to AI treatment are associated with decreases in BC incidence), this may save lives by encouraging high-risk women to take up preventive hormonal treatments. Similarly, it may encourage women to continue adjuvant endocrine therapy despite side-effects because the short-term discomfort will be worth the long-term benefit.

1.1 Thesis AIM summary, CMN IBIS-II MD and AI substudy

The Primary Aim of this thesis is to compare longitudinal MD change in IBIS-II participants randomised to treatment with anastrozole and placebo. Before the Primary Aim could be addressed, four Secondary Aims were completed:

Secondary Aims:

AIM 1: To review and select MD measurement techniques for this thesis

AIM 2: To investigate intra- and inter-technique measurement reliability for MD

AIM 3: To describe baseline characteristics of the participants included in the CMN MD and AI substudy and investigate associations between MD and baseline characteristics

AIM 4: To develop an adequate model for mean change in MD over time for the aggregate (treated + control) groups— a "blinded" longitudinal analysis

Primary Aim (AIM 5):

To assess differences in longitudinal MD for treated vs control participants during the five years of treatment during the IBIS-II trial, and for two years post-treatment— an "unblinded" longitudinal analysis

1.2 Thesis Chapter Overview

Chapter 1: Thesis introduction including Aims and Chapter Overview

Chapter 2: Breast cancer background and MD literature review

Chapter 3: Project methods (IBIS-II trial background, approvals, statistical method overview)

Chapter 4: Review of techniques used to measure mammographic density

Chapter 5: Intra- and inter-technique reliability analysis

Chapter 6: Baseline characteristics and MD association analysis

Chapter 7: Aggregate IBIS-II group (treated + control) longitudinal MD change analysis

Chapter 8: Treated vs control (unblinded) MD longitudinal analysis

Chapter 9: Discussion

The next chapter examines MD within the broader context of breast cancer. Breast development is briefly discussed, as well as BC treatments, staging, prognostic markers, and risk factors. The relationship of MD with BC risk factors and therapies which affect MD are reviewed.

2. Breast Cancer, Risk Factors and Mammographic Density

This chapter has six main sections. The first three sections discuss the breast and breast cancer, to provide a context for the mammographic density review in latter sections. The first section provides a brief description of breast development and breast density. The second section includes a summary of breast cancer incidence and mortality (worldwide and within Australia), as well as a brief review of BC treatments, staging and prognosis. The third section discusses the relationship of BC and established risk factors for BC. The fourth section reviews the relationship of MD with BC risk factors. The fifth section describes the changes which occur in MD during treatment with hormonal medications; this section contains a review of AI and BD studies (as of April 2018). The sixth section contains a discussion of MD in the context of modelling BC risk. The chapter concludes with a summary of the factors which are associated with MD.

2.1 Introduction

The breast is a complex organ. Development in females begins in utero and ceases soon after birth until it resumes during puberty [37, 38]. The breast completes maturation during the last stages of pregnancy when type 4 (milk-producing) lobules become dominant [37]. During menopause the glandular (lobular) and inter-lobular fibrous tissues are replaced by fat (menopausal involution) [39]. The lower breast density observed for post-menopausal women compared to pre-menopausal women is due to breast tissue involution during menopause.

2.1.1 Breast Density

Structures such as bone or teeth appear white on x-rays and are called dense because they attenuate (absorb) x-rays more strongly than surrounding tissues. Similarly, the dense breast tissues appear white on mammograms (x-rays of the breast) because they attenuate x-rays more than fat, Figure 2-1. Fat is translucent to x-rays and appears dark on mammograms. Both

lobular epithelial (glandular) and connective breast tissues absorb more x-ray energy than adipose tissue and appear dense; however, it is primarily the inter-lobular fibrous connective (stromal) tissue which gives rise to the dense appearance of the non-lactating breast.

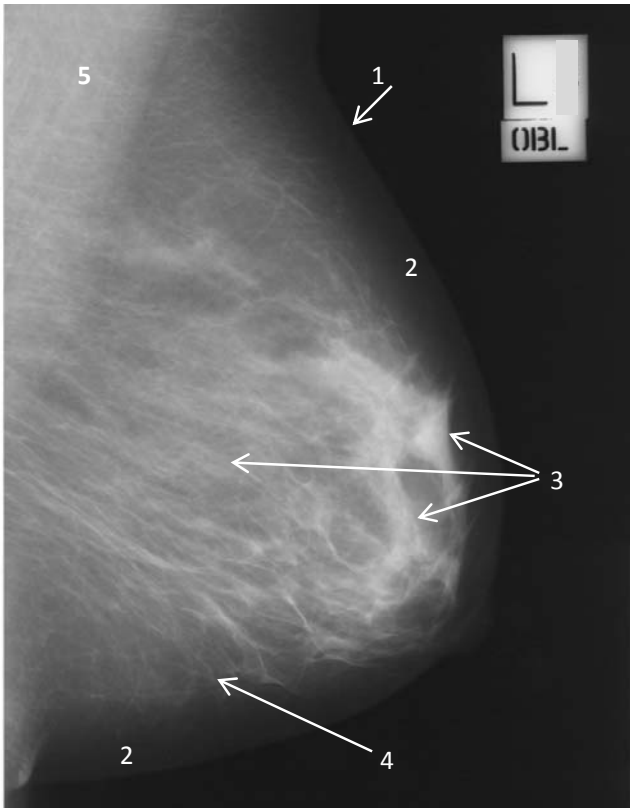


Figure 2-1 IBIS–II[1] [26] digitised film-screen mammogram.

About 50% of the breast area is covered by dense white 'fluffy' tissues = ~50% percent density (PD). Fat (adipose tissue) is translucent to x-rays and appears dark on mammograms. The pectoral (chest) muscle can also be seen crossing the upper left corner of the mammogram (left medio-lateral oblique (LMLO) view).

1.) Outer edge of breast (skin)
2.) Subcutaneous fat
3.) Dense breast tissues
4.) Cooper's ligaments
5.) Pectoral muscle

2.2 Breast Cancer – background

Breast cancer is the most common cause of mortality from cancer in Australian women, causing over 2,500 deaths per year [40]. Australian women have a one in nine chance of developing breast cancer within their lifetime [5]. The World Health Organisation (WHO) estimated over 536,000 women in 2012 worldwide died from breast cancer, the second leading cause of death from cancer in women (1% of all global deaths) [2]. By 2030, the WHO predicts BC will be the leading cause of cancer death for women (over 800,000 deaths; 1.1% of all deaths). Breast

[1] IBIS-II is an international prospective randomised controlled trial of an aromatase inhibitor (anastrozole) for the prevention of breast cancer in high-risk women. This trial is further described in Chapter 3.

cancer can have wide ranging effects on most aspects of a woman's life [41]; the emotional, familial, and economic impacts of breast cancer are profound [42, 43].

Almost all (99%) of breast cancer occurs in females; ~1% occur in males [44]. Due to the much greater incidence of BC in females, this thesis only reviews breast cancer and breast density in females. Breast density in males is not well described.

Most BC arise from the ductal and lobular epithelial tissues of the breast. Invasive ductal carcinoma is the most common BC (60-90%), followed by lobular tumours (~10%), unspecified (invasive) carcinomas (~5%) and other "special" subtypes (5 to 10%) [44]. BC types tend to vary with by age, reproductive history, family BC history, ethnicity and country of residence [44-49].

Breast cancers undergo additional pathological classification for prognostic (survival) and therapeutic (treatment) purposes. Staging, which measures cancer size and whether the cancer has spread to the lymph nodes and more distant sites (metastases), is important for prognosis and surgical treatment options. Smaller cancers which have not metastasised at the time of diagnosis have better prognosis than larger non-metastatic tumours. BC survival rates are also affected by ethnicity, socio-economic status and other factors. The presence or absence of BC histo-pathological markers (e.g. estrogen-receptor status) are also associated with survival rates, and guide treatment options. Newer BC classification schemes such as molecular and genomic expression are being used to inform prognosis and treatment.

2.2.1 Impediments to BC screening and treatment in low resource settings

Breast cancer is the most frequent type of cancer for women, and the leading cause of cancer death for women worldwide [50]. Incidence is higher in developed vs developing regions of the world, but mortality rates are comparatively higher in developing areas. BC mortality is higher

in less well–resourced countries because communicable diseases typically present more of a burden in these areas [49, 51]; treatment and prevention of non-communicable diseases including cancer is a lower priority. Overall there are fewer resources to assist with healthcare needs in lower income countries, adding to the difficulties of treating cancer [52]. Breast cancer, along with gynaecological cancers, present particular problems to women in the male-dominated societies which prevail in the developing world [53, 54]. Geographical issues (e.g. great distances) and lack of education add to the difficulties posed both by personal (at the level of the individual) and regional financial constraints [52, 54]. Hence many women in developing countries present with late stage (larger size, poor prognosis) BC compared to their counterparts in better resourced areas [49, 55].

The combination of screening with the ever growing array of treatments in developed countries has led to BC mortality reductions for women living in these regions. Adequate cancer screening and treatment requires funding for and coordination of a broad range of clinical disciplines [56]. Among the non-communicable diseases, cancer presents a particular challenge in developing countries due to these requirements [57]. Screening programs are only effective in areas which have the appropriate infrastructure and medical facilities to adequately care for persons diagnosed with cancer. Whilst the Breast Health Global Initiative recommend opportunistic screening with clinical breast examination in low resource settings [56], allocating scarce resources to widespread mammographic screening to detect small (early stage) BCs is a low priority in these settings if the women are subsequently unable to receive satisfactory treatment.

2.2.2 BC incidence and mortality in low vs high resource settings, and Australia

The World Health Organization's (WHO) International Agency for Research on Cancer

GLOBal CANcer (GLOBOCAN) project[2] estimates that BC is the second leading cause of cancer death in women in developed countries (15.4%, second to lung cancer), and is the leading cause of cancer death of women in less developed countries (14.3%) [50]. The annual global absolute incidence (number new cases of BC) in 2012 was similar in developed (794,000 women) vs less developed regions (883,000 women). Although the absolute number of new BC cases in 2012 was similar between developed and less developed nations, 2012 BC incidence *rates* were much higher in developed (74.1/100,000) vs less developed regions (31.3/100,000). This is because life expectancy is higher in developed vs less developed regions and cancer incidence rates increases with age.

Despite the approximately equal numbers of women diagnosed with BC in developed vs developing regions in 2012 (~850,000), the estimated absolute number of women who died in 2012 from BC was much higher in developing regions. The GLOBOCAN project estimated 324,000 women in less developed regions succumbed to BC in 2012 versus 198,000 in developed regions [50]. When the number of BC deaths to the number of new BC cases diagnosed in 2012 are compared, the proportion of BC deaths to diagnoses in developing nations is approximately twice as high (324,000/883,000 = 41%) relative to developed nations (198,000/794,000 = 22%).

The disparity between absolute BC incidence and absolute number of deaths is echoed when the 2012 BC incidence *rates* and BC mortality *rates* are compared. Although the 2012 BC incidence rate was more than twice as high in developed regions vs less well developed regions (74.1 vs 31.3/100,000), mortality rates are similar (14.9/100,000 (developed) vs 11.5/100,000 (developing)). The similarity in death rates despite the more than double incidence rate in developed regions is likely due to disparities in adequate health care access and availability of effective treatment.

[2] http://globocan.iarc.fr, with interactive data visualisation tools at http://gco.iarc.fr/today/home

The GLOBOCAN estimates reveal that Australia and New Zealand women have one of the highest incidence rates in the world (85.8/100,000), second only to Northern Europe (89.4/100,000) [50]. Mortality rates in Australia are comparatively low (14.5/100,000) given the high incidence rate, which is probably a reflection of access to early detection and treatment. The Australian Institute of Health and Welfare (AIHW), using Australian-based age-standardisation, calculated the incidence rate of BC as 115/100,000 in 2008 (13,567 women).

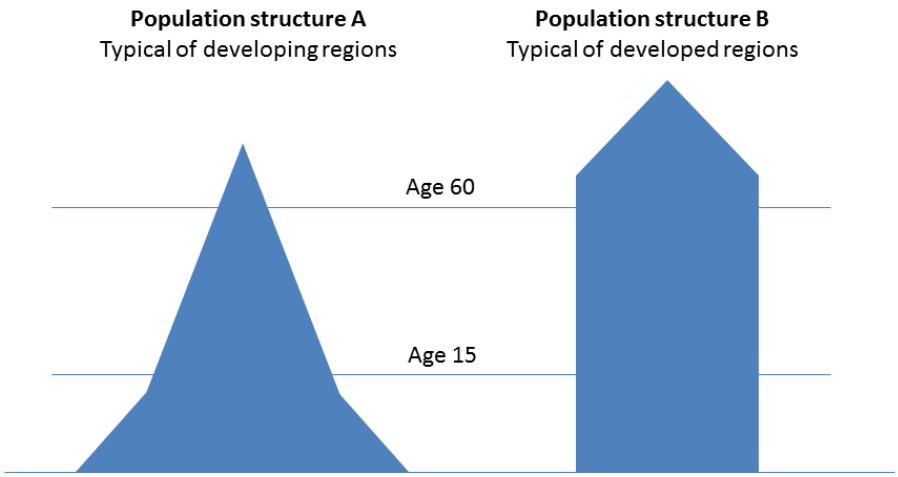


Figure 2-2 Low income country population structure (structure A) vs high income country (structure B) Cancer particularly is a disease which most commonly occurs in older people; most people in developing nations are unable to reach an age to succumb it or other conditions commonly associated with ageing. Differences in population structure must be accounted for during comparisons of disease and death rates across geographical regions and time, else comparisons will likely be misleading.

Differences in population structure between the average global population (more younger people and fewer older people (pyramid shape, Figure 2-2, left) and the distribution of ages in Australia (more balanced among all age groups: cylindrical shape, Figure 2-2, right) accounts for the differences in the incidence between the GLOBOCAN and AIHW rates. Because the Australian population structure differs from the global population structure, direct comparisons of rates (for any disease) for any age group are likely to be misleading [58]. Therefore different population structures are converted (standardised), typically to a common (global) reference population structure, to facilitate comparisons across different geographic regions as well as different points in time.

Relative survival compares the survival rates of people with a condition to similar people without the condition. It is important that the comparator group of unaffected individuals is similar, else the number calculated for relative survival will be inaccurate. Five-year relative survival for BC in Australia is estimated to be 89.4% [44]. Hence when all the women diagnosed with BC in Australia from 2006 to 2010 were compared to all the other women in Australia (matched by age and sex in the same calendar year, adjusted for remoteness area and SES), 89.4% of the women diagnosed with BC lived for 5 years compared to their unaffected counterparts. This compares favourably with five-year relative survival in the US and Europe, and is higher than five-year relative survival in lower resource settings, e.g. ~71% five year survival rates in Thailand and Yunnan province in China [59].

In developed countries, both the absolute number of women diagnosed per year with BC has increased, as well as the yearly rate (e.g. incidence per 100,000 women) [44, 60]. In Australia, 5,310 women diagnosed with BC in 1982, 8,059 in 1992, 10,765 in 1998, and 13,567 in 2008. Age-standardised rates of BC in Australia per 100,000 women were 81.1, 100.8, 114.9 and 115.4 for the same years.

The remainder of this chapter primarily discusses breast cancer and breast density in developed regions. Breast density is only detectable via medical imaging techniques such as mammography, ultrasound and magnetic resonance imaging (MRI). Imaging facilities with these techniques are not wide-spread in lower resource countries; hence information on mammographic/breast density is not widely available. However, research investigating the relationship between breast density and BC in developing regions has recently commenced, as well as an international consortium to pool MD research results from many countries [61]; the results from these projects are awaited with interest.

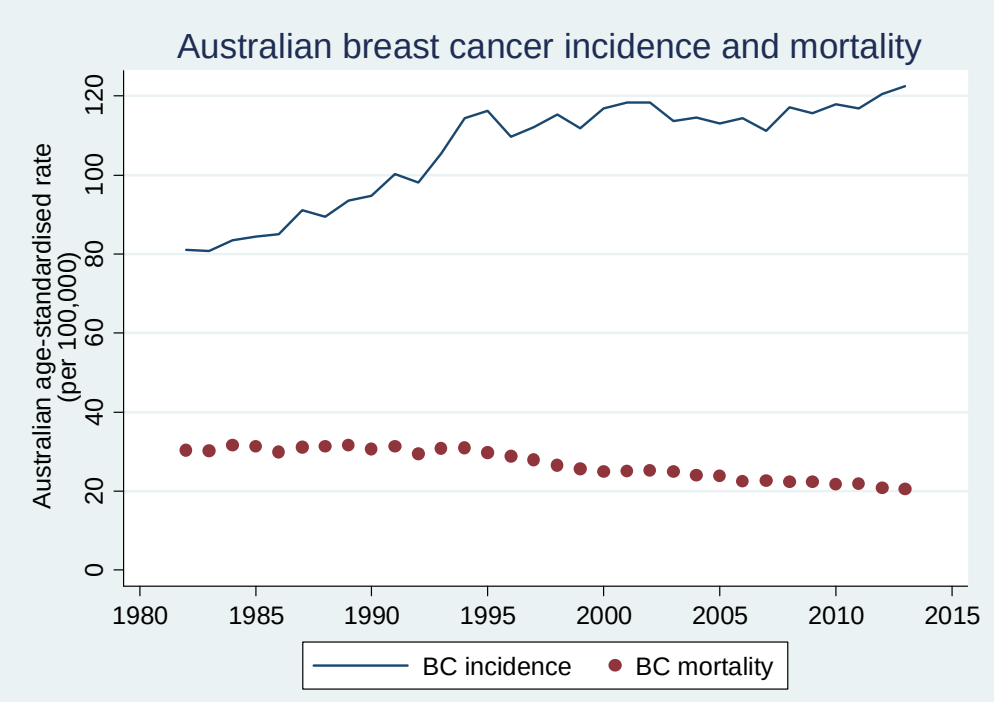


Figure 2-3 Australian standardised incidence & mortality rates per 100,000 women, 1982 to 2013
The pattern of increasing incidence rate, but decreasing mortality rate is typical for BC in most Western countries over the last 70 years. Data sourced from the Australian Institute of Health and Welfare [62]

2.2.3 BC Treatment

Cancers amenable to surgical removal— e.g. small BC— are typically treated with surgery which can include lymph node removal if cancer has been detected in the nodes (nodal involvement). Because BC is a solid tumour (as opposed to haematological (blood) cancers, which are dispersed throughout the fluid in the body), surgical removal of the tumour is considered the primary treatment. Therapies which occur in addition to primary therapy (surgery) are called adjuvant therapies [63]. Adjuvant therapies include chemotherapy, radiation therapy (radiotherapy), hormonal therapy and targeted therapy.

The term chemotherapy (the use of chemicals to treat disease) was first associated with cancer treatment from 1940 through the 1980s [3, 4]. Cancer chemotherapy has broad (multi-system) effects upon the body by disrupting the ability of both cancer and normal cells to grow and

divide; chemotherapy doses are based on the ability of the toxins to kill the cancer cells (which tend to grow more rapidly than normal tissues) before killing normal cells.

Cancer "chemotherapy" is considered to be an entity distinct from "hormonal" therapies (e.g. estrogen, tamoxifen, aromatase inhibitors such as anastrozole) and "targeted" therapies such as herceptin. In the context of breast cancer, hormonal (endocrine) therapy refers to treatments which affect the production or action of estrogen, progesterone or other sex hormones (naturally occurring chemicals in the body). Unlike the broad effects of chemotherapy, hormonal therapy tends to only affect cells with receptors to the sex hormones. Hormonal treatment for advanced (metastatic) BC in the 1960s included treatment with estrogen, androgens and progestin [64]. Current treatments for hormone-sensitive BC include the serum estrogen modulators (SERMS) tamoxifen and raloxifene, as well as aromatase enzyme inhibitors such as anastrozole, exemestane and letrozole. These modern hormonal therapies are further described in section 2.3.3.

2.2.4 BC Staging, Histopathological/Molecular/Genetic Markers and Prognosis

Tumour size is an important staging, prognostic and therapeutic factor; smaller cancers are less likely to cause mortality [44], and can be treated with lumpectomy (breast conserving surgery) more readily than large tumours. Invasive cancers which have grown out of their tissue of origin are much more problematic than carcinoma in-situ which remains within the originating tissue. Lymph nodes draining the breast tissues are pathologically examined for the presence of and size of BC during staging. Other organs are also often imaged and/or sampled for the presence of BC (metastases).

Cancers are also classified according to how much the cancer has changed the histological appearance of the cell (differentiation from normal cells), the presence (+) or absence(-) of

estrogen and progesterone receptors (ER/PR), and presence of human epidermal growth factor receptor 2 (HER2) receptors on the cancer cells [65]. Newer methods of classifying tumours use molecular and genetic markers [66], such as use of gene-expression profiling to classify BC into Luminal A, Luminal B, basal, and normal-type tumours [67]. The different classification systems provide groupings which often overlap, but help distinguish cancers based on these different (marker) characteristics. The markers provide insight into prognosis as well as which treatment/s best suit the cancer. In general, BC more closely resembling normal tissue (less differentiated) has a better prognosis than cancer which has increased mitotic activity, mitochondria, nuclei and microtubules (highly differentiated).

Most (60 to 80%) invasive BC is ER+ and/or PR+ [68, 69]; the prevalence of ER+/PR+ BC depends upon the characteristics of the population studied including ethnicity, reproductive status and age [68-70]. ER+ and/or PR+ BC (i.e. 'hormonal sensitive' BC) is typically treated with anti-endocrine (hormonal) therapies to decrease the amount of estrogen available to promote the growth of the cancer; prognosis is favourable (90% 5-year survival for early stage (I-II) BC) [71].

HER2+ cancer is typically very aggressive, and was highly fatal prior to discovery of targeted monoclonal antibody treatments. Breast cancers without receptors for ER/PR/HER2 are known as triple negative cancers. These poor prognosis cancers generally can only be treated with chemotherapy in addition to any surgical and/or radiotherapy options available.

At diagnosis, a primary (first, original) cancer may be composed of identical cells, however most primary tumours are composed of a variety of cells with different mutations. One type of mutation may dominate. Due to the physiologic environment in different parts of the body, secondary (metastatic) cancers can differ from those of the primary tumour (clonal evolution). Treatment based upon markers within the primary tumour influences which cancer cells perish

during treatment, and which cells survive (within the primary tumour, and secondary tumours if present).

The tumour markers expressed by primary and secondary tumours often change during cancer treatment, presumably due to evolutionary pressures exerted by the treatment and location of the tumour. Change in markers indicates altered susceptibility to treatment. The ongoing ability of cancer to mutate makes this heterogenous disease extremely difficult to eradicate, especially if it has metastasised. Our understanding of DNA transcription, how this is affected by environmental & developmental influences (epigenetics), and the complex interplay of molecules within each cell is growing, but very limited compared to the level of complexity within the body.

Some women fare very well on the treatment selected for their BC stage and tumour markers; other women have similar cancers which do not respond to the same treatment. Why this is so remains a complex and ongoing area of research [72].

In-depth 'personalised' (precision) medicine is still in its infancy, because of the complexity of each individual, as well as the complexity within the cancers themselves. More evidence is coming to light that the tumour microenvironment (e.g. stromal tissue surrounding the cancer) plays an important role in tumour initiation and progression [73]. The biological, genetic and environmental factors which affect a woman's response to BC treatment may be reflected in the amount of dense tissue in the breast before, during and after treatment. Hence mammographic density may make an effective biomarker for BC prognosis, treatment selection, as well as the efficacy of particular treatment regimens.

2.3 Breast Cancer Risk Factors

The average risk woman has a lifetime BC risk (to age 80) of 12.5%. This equates to a 1 in 8 lifetime risk of breast cancer. BC risk increases with age; older women have a higher baseline risk than younger women [74]. Australian BC incidence was 8 per 100,000 women aged 30 to 34; 230 per 100,000 women aged 50 to 54; and 298 per 100,000 women aged 70 to 74 [75].

Breast cancer risk factors may be subdivided many ways, e.g. into strong vs weak vs protective factors, or those that can be changed (are modifiable) and those that are not [5]. An inherited BRCA mutation is an example of a strong BC risk factor that cannot be changed. Mammographic density is one of the few strong (≥3 fold) risk factors which can be modified.

2.3.1 Non-modifiable risk factors for breast cancer

The strongest risk factor for breast cancer is being female. 99% of BC occurs in women; only about 1% of BC develops in males [44]. Many of the risk factors for male BC are similar to that for female BC, but due to the low prevalence of male BC this disease is not widely studied.

The BRCA1 and BRCA2 mutations are two of the strongest known risk factors for BC [5]. The lifetime risk of BC for a woman with a BRCA1 mutation is around 65 to 80% (20-fold risk increase compared to an average risk woman); lifetime BC risk for BRCA2 mutation carriers is about 50 to 60% [76]. The BRCA genes are involved in DNA repair, and also confer an increased lifetime risk for ovarian cancer (15 to 40%) [76] and other cancers. Preventive strategies developed specifically BRCA-mutation carriers are being explored, e.g. RANKL (receptor activator of nuclear factor kappa-B ligand) blockade to inhibit mammary tumoriogenesis in atypical luminal progenitor cells present in BRCA1 mutation carriers [77].

Increasing age is also a strong BC risk factor. The majority of BC (75%) occurs in women over

age 50 [44]. Treatment with radiation at young age to treat Hodgkin's disease is also strongly associated with BC, as are previous history of the breast conditions DCIS (ductal carcinoma in-situ), LCIS (lobular carcinoma in-situ), atypical hyperplasia and BC [5].

Weaker, non-modifiable BC risk factors include family history of breast cancer (e.g. BC history in first and/or second degree relatives), early age at menarche and thelarche, late age at menopause, high levels of circulating androgens, and (for post-menopausal women) high circulating levels of estrogens [5]. Exposure to ionising radiation at a young age, especially < 20 years, a history of any other cancer, and tall stature (>1.8m) also increase BC risk [5] .

2.3.2 Potentially modifiable risk factors for BC

Affluent country of residence is strongly associated with BC risk. A substantial proportion of this risk is associated with dietary and physical activity factors, but is also intimately intertwined with parity and breast feeding [78, 79].

A review performed in 2002 estimated the rate of BC to age 70 in affluent countries would be halved from 6 to less than 3 per 100 women if women in developed countries gave birth to 6 children which they each breastfed for 24 months like their counterparts in less developed countries [80]. Currently western women tend to have 2 to 3 children which they breastfeed for ~8months. The number of children per woman in developed countries is effectively not modifiable for a variety of social, economic and environmental reasons. However, increasing breast feeding duration can potentially be addressed through improved education about the benefits of breastfeeding, and societal changes such as acceptance of breast feeding in public and facilities for breast feeding and/or expressing milk upon return to the workforce.

The lowering of risk with increasing parity is closely associated with age at first birth. Lower age (< 24 years) at first birth is associated with decreased risk of post-menopausal BC (relative

risk 0.5), compared to women who give birth for the first time at age 30 or older [81]. Parity generally confers a decrease in BC risk [5], however, women who first give birth at age 35 or older have a greater BC lifetime risk than nulliparous women [82]. BC risk is transiently elevated for up to 10 to 15 years after the first full-term pregnancy at any age [83]. Hence first giving birth at an older age coincides with increased BC risk due to age; these two effects appear to be additive.

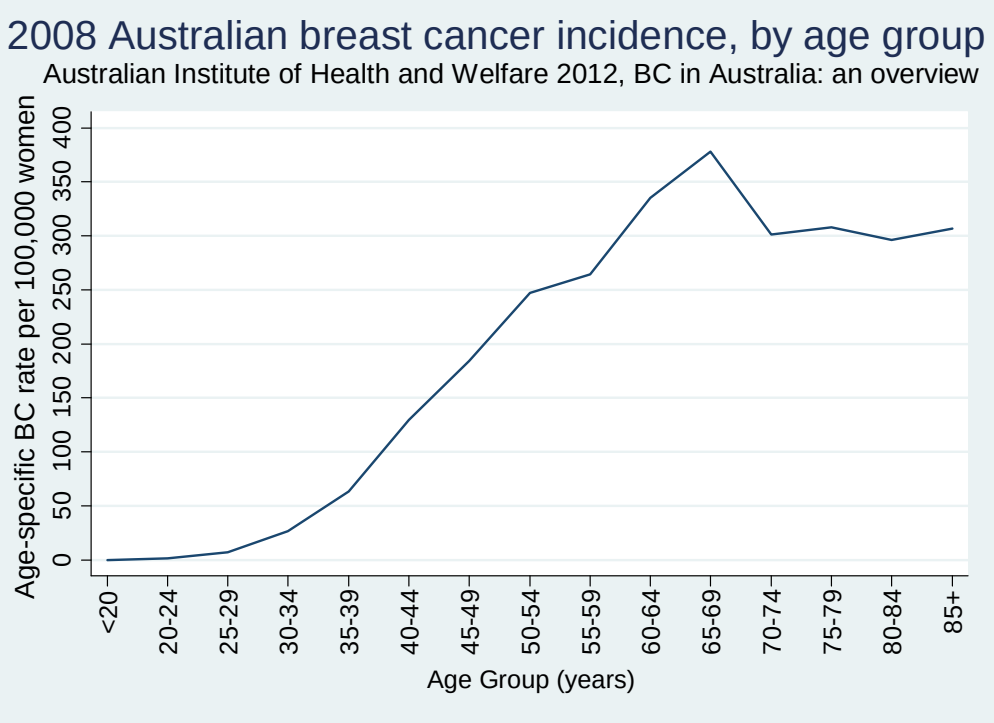


Figure 2-4 Australian age-specific rates of BC in 2008
BC incidence increases with age. The decrease incidence rate after ages 65 to 69 is due to a previous cease of invitations to screen after age 70. The age at which invitations to screen now cease has recently increased to age 74. Data sourced from the Australian Institute of Health and Welfare [62]

The breast tissues proliferate during pregnancy in preparation for breast feeding, and the immune system is also suppressed [83, 84]. Involution of the breast tissues post-lactation may also be a pro-tumorigenic environment [83]. As per Figure 2-4, the absolute risk of BC per 100,000 women is very low during the traditional childbearing years: less than 10 under age 30, fewer than 50 for ages 30-39 versus more than 250 for women over 50 [44]. The transient increase in BC risk incurred during younger age at pregnancy is balanced by the very low BC

risk for younger women; the transient increase in BC is at younger age is also greatly offset by its later protective effects when BC risk is much higher (i.e. age 40+). Older mothers (age 35+) experience the transient risk increase due to pregnancy; however, their risk does not drop again because their age-related risk is increasing.

Pike theorised that overall BC risk is related to lifetime cumulative exposure to estrogen [85]. This theory helps explain the elevated risk of BC seen in older age, when estrogen exposure is low. Pike noted that whilst the log incidence rates for most cancers remain constant with increasing log age, the same is not true for breast cancer (Figure 2-5 part B, below). The linear relationship between the log of age and the log incidence rate for breast cancer becomes markedly less strong at about age 50, the approximate age of menopause.

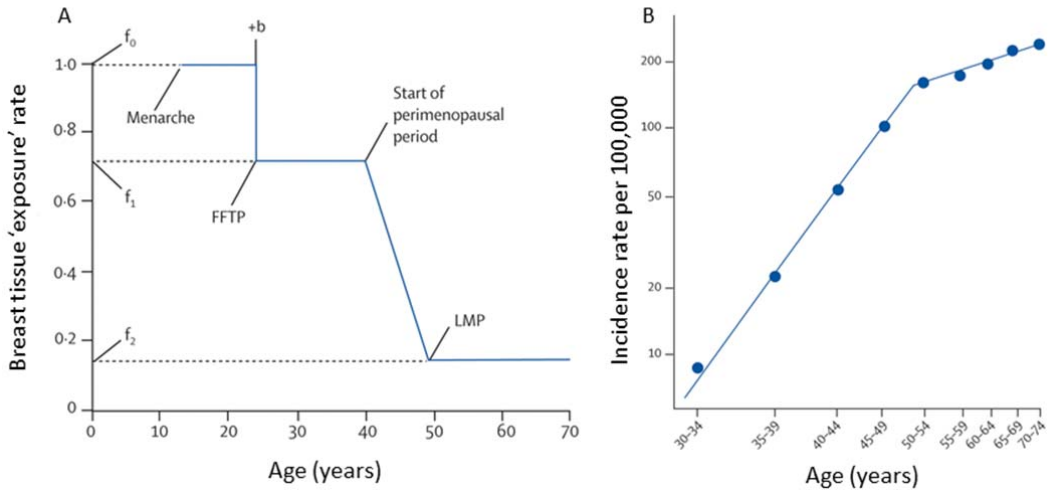


Figure 2-5 (A) Pike model of mammary carcinogenesis. (B) Age-specific incidence of breast cancer – observed and predicted by the Pike model
The breast tissue exposure rate in the model (a) prior to the start of the peri-menopausal period is very high compared to the rate of exposure in the peri-menopausal and menopausal period. When fitted to age-specific log incidence rates in (b), the log incidence rate predicted from the model (curve) closely approximates the actual log age-specific incidence rates observed in the female population (dots).
FFTP First full term pregnancy; LMP Last menstrual period.
Sourced from Pike et al 1983 [85] via Boyd et al 2009 [86]

Pike's 'breast tissue ageing' model theorises that the amount of estrogen exposure at different times of life determines a woman's cumulative risk of developing breast cancer (Figure 2-5, part

A). The estrogen exposure rate in the peri-menopausal and menopausal periods is much less than rates of exposure in the pre-menopausal period; this could explain the marked decrease in the log incidence rate for breast cancer observed at menopause. The curve from the predicted values of this model is well fitted to the observed age-specific breast cancer log incidence rates (Figure 2-5 part B). Therefore, it is quite possible that exposure to estrogen throughout life is strongly related to breast cancer risk.

The above model also helps to explain the population level associations of increasing breast cancer risk seen with earlier menarche, later age at first pregnancy/nulliparity, and late age of menopause. For all of these conditions, exposure to higher levels of estrogens occurs for a longer part of the lifespan compared to women with later menarche, young age at first pregnancy and early menopause. The model appears to be at odds with the transient increase in BC seen post-partum, but this is accounted for by adding an increase in estrogen exposure depicted by +b at the first full-time pregnancy (FFTP). This factor is included in the model (line) that closely tracks the log age-specific BC rates (dots) seen in part B of the figure.
The biological underpinnings of these effects may be due to susceptibility of the incompletely differentiated terminal ducts of the breast prior to first full term pregnancy. The ducts contain a higher number of stem cells which are more susceptible to the carcinogenic effects of estrogen exposure [81, 83, 84, 87, 88]. Pregnancy and lactation are associated with fewer menstrual cycles and therefore fewer cycling hormones, which may also reduce BC risk. Older first-time mothers would have more exposure to these carcinogenic factors, and hence would be more at risk post-pregnancy as well as from the additional pro-tumourigenic effects of breast involution [83].

There are a number of dietary, physical activity and leisure activities associated with increased or decreased risk of BC. It is estimated that up to 30% of BC risk could be eliminated with weight control and increased physical activity [89]. Higher body mass index (BMI: weight /height^2) in pre-menopausal women (>30) is associated with a decreased risk of BC compared to

thin women (BMI < 21), whilst higher BMI is associated with increased BC risk in post-menopausal women. A number of mechanisms may be involved for these pre- vs post-menopausal differences. High BMI in pre-menopausal women may be associated with fewer ovulatory menstrual cycles and thus reduced estrogen levels/exposure to hormone cycles. Higher BMI is associated with more adipose (fat) tissue in the breast. For pre-menopausal women this may equate to more fat interspersed amongst the stromal (connective) tissues of the breast, changing the microenvironment of the breast. The tissue may be less stiff, and/or the fat may change the signals produced by the surrounding connective tissues, which may reduce tumourigenesis. In post-menopausal women the increase in breast adipose tissue is likely to cause an increase in local estrogen production by the aromatase enzyme. Estrogen is a known carcinogen, and may be the source of increased BC risk for post-menopausal women with high BMI.

Alcohol consumption is associated with a 7% increase in BC risk per standard drink per day vs. no alcohol consumption. Cigarettes appear to have a mixed effect upon BC risk; some studies show increased risk whilst others show a decreased risk. Given that all of these studies are observational (i.e. are not RCTs), potential methodological confounding may be present. One mechanism proposed for the supposed protective effect of cigarette smoking is suppression of estrogen production. However, most studies find a dose-response effect with increasing exposure to cigarette smoke. Cigarettes are strongly associated with a wide variety of other cancers and serious medical problems, therefore women wishing to lower their breast cancer risk are best to refrain from smoking.

Exercise is protective against BC independent of its effect upon BMI. Diet does not appear to have a strong effect upon BC risk, however a diet high in fruit and vegetables and low in fats is protective against many other cancers. Phytoestrogens, found in foods such as soy products, and green tea have been explored as one of the reasons for lower rates of BC in Asian countries; conclusive evidence does not exist to support consumption of these foods for BC prevention.

2.3.3 Hormonal therapies associated with prevention and treatment of BC

Estrogen is a known tumour promoter. Estrogen not only stimulates growth of ductal and other breast tissues during puberty and pregnancy, but also enhances the growth of many breast cancers. As described earlier, up to 80% of BC has receptors for estrogen and/or progesterone [90]. The survival of these 'hormone sensitive' cancers are often dependent upon a cascade of intracellular signals generated by activation of estrogen and/or progesterone receptors on the tumour cells. A number of therapies which reduce the availability of estrogen to body tissues are associated with reductions in breast cancer risk. The selective estrogen receptor modulators (SERMs) tamoxifen (TAM) and raloxifene (RLX) disrupt the ability of cellular estrogen receptors to bind estrogen [91]. The aromatase inhibitors (AI) exemestane, anastrozole and letrozole reduce estrogen by inactivating the aromatase enzyme which converts androgens to estrogens in the peripheral tissues of the body [92]. The ovaries are the primary source of estrogen in premenopausal women, hence SERMs are effective in both pre-menopausal and post-menopausal women due to the blockade of estrogen at the cellular level. AI are only therapeutic in post-menopausal women whose primary source of estrogen is via the aromatase enzyme [23].

SERMs and AI are used to treat BC [93], and are also effective for the prevention of BC [72, 94, 95]. These therapies are typically given for five years; but recent data indicate longer therapy post BC (10 years or more) prevent recurrence of BC effectively without a great increase in other adverse events [96]. TAM in particular is associated with a very small but significant increase in endometrial cancers when given for BC prevention. RLX is slightly less effective than TAM in preventing BC but has a better side-effect profile: fewer endometrial cancers, thromboses (clots), and cataracts [97]. AI treatment has fewer serious (life-threatening) adverse effects but bone weakening (osteoporosis), fractures and arthralgia (joint pain) are common (>5% of women). Generally, AI are preferred to TAM for treatment post-menopausal endocrine sensitive (ER+) BC because of higher efficacy and a favourable side-effect profile [98].

TAM does not cause an excess of endometrial cancer in pre-menopausal women treated with five years of TAM for early, hormone sensitive BC; the excess of endometrial cancer occurs primarily in post-menopausal women who appear to be the most greatly affected by the proliferative effects of TAM upon the lining of the uterus [22]. It is probable that use of TAM for BC prevention in young (<45 years) pre-menopausal women would not be associated with an excess of risk of endometrial cancer. TAM has a carryover effect on BC prevention, which means that BC is suppressed for an additional five years after therapy ceases compared to controls in randomised controlled trials (RCTs) who received no treatment. However, the excess risk of adverse events associated with TAM cease at the end of treatment. This means the BC prevention effects from TAM persist for 5 years after treatment, but the possible side-effects stop. A possible BC prevention treatment strategy could therefore be TAM until 45 years of age (five years before the average menopause at age 50) followed by preventive treatment with AI at menopause.

Whilst used very widely for BC treatment of hormone-sensitive (ER+/PR+) BC, both AI and SERMs are rarely utilised for BC prevention in part due to concerns about possible side effects [99, 100]. Low levels of awareness in both general practitioners (GPs) and the general public are the major impediment however [101-103]; availability and cost, and difficulties in assessing BC risk have also limited the use of AIs and SERMs for BC prevention [104]. One substantial barrier to the use of TAM for BC prevention was lowered for Australian women in 2016: the Australian Pharmaceutical Benefits Scheme (PBS) indications for TAM now include BC prevention for high risk women. Australian women at high risk of BC can now be prescribed TAM to prevent BC by their doctors at a greatly reduced cost via the PBS.

Evidence suggests that stronger side effects whilst on hormonal therapy are associated with higher efficacy of treatment (i.e. lower risk of BC) in many [105, 106] but not all clinical trials

[107, 108]. Low vitamin D levels (< 30ng/ml) were found in 88% of post-menopausal women with arthralgia and myalgia treated with third generation AIs [109, 110].

Many other BC preventive therapies have been proposed [95], including short-term use of human chorionic gonadotropin hormone released by the uterus during pregnancy in nulliparous women to induce the genomic signature in breast tissues acquired during pregnancy [111]. However, no effective method exists to prevent the typically aggressive triple-negative subtype of BC; this remains an area of need to be addressed [94, 112-114]. A recent review indicated that longer length of breast feeding is associated with lower risk of developing triple negative BC, however, which may provide clues as to how to best prevent this particularly difficult group of cancers [78].

2.4 Mammographic Density, BI-RADS and BC masking

As described earlier in this chapter, mammographic density (MD) is comprised of the ductal (glandular) and stromal (epithelial) tissues of the breast and appear white on mammograms (x-rays of the breast), whilst adipose tissue is transparent to x-rays and appears dark on mammograms, Figure 2-6. Higher levels of MD are associated with higher levels of BC risk.

Breast density has been recognised as a masking factor during mammography for at least 40 years. The American College of Radiology (ACR) Breast Imaging - Reporting and Data System (BI-RADS) was created to standardise radiologists' reports when mammographic imaging became widespread in the 1970s and 1980s in the US. A parameter for MD was included from the first edition of BI-RADS because it was recognised that dense breast tissue could hide cancers. The BI-RADS density parameter is comprised of four categories which have undergone several evolutions since they were first developed in the late 1980s [115]. John Wolfe's initial articles in 1988 and 1989 disclosing BC as an independent BC risk factor were

controversial because many radiologists considered density solely as a factor that impeded cancer detection.

The original BI-RADS (1992) density categories were purely descriptive (qualitative): 1) fat containing/almost entirely fat, 2) low density/scattered fibroglandular densities, 3) equal density/heterogeneously dense, and 4) high density/extremely dense [116]. These categories were intended to give the reader of the mammography report (who did not have mammograms available to view) the likelihood a cancer was masked by the dense tissues. Most radiologists tended to assign 90% of mammograms into categories 2 and 3 using these descriptors [117] . This was in part because it is difficult to assign the categories systematically, repeatably and reliably using qualitative descriptions.

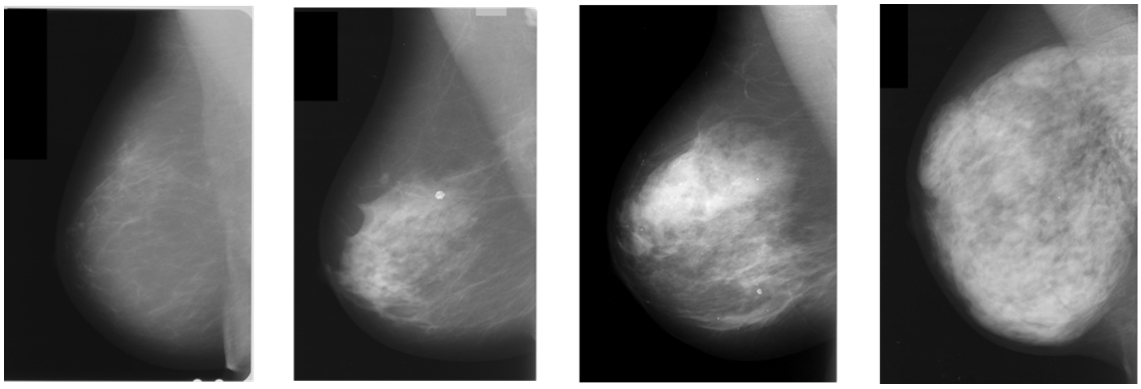


Figure 2-6 Digitised film-screen mammograms from participants in the ANZ Breast Cancer Trials Group's IBIS-II Breast Cancer Prevention Trial, categorised into the BI-RADS (4th Edition, 2003) density categories
The mammograms show varying levels of percent density (PD), defined as the percent area of the breast covered by radio-opaque (white) dense epithelial and connective tissues.

Numeric ranges for PD were first published in the 4th edition of BI-RADS (2003) in an attempt to standardise the BI-RADS density categories. These density categories use quartiles of PD to provide guidance for the assignment of each density category: (0 to 24%, 25 to 49%, 50 to 74% and 75%+) [118]. This was helpful epidemiologically because it provided a numeric (quantitative) estimation of the dense tissues on a mammogram. However, the 4th edition

density categories did not fully address the primary BI-RADS clinical problem of BC masking by MD.

The most recent version of BI-RADS (5th Edition, 2013) has returned to a qualitative description of the four density categories [119]. To lessen confusion between the BI-RADS density score (4th edition: 1 2 3 4) and the overall BI-RADS assessment categories (a numeric score of 1 to 6 which is assigned by the radiologist to advise the likelihood of BC), the 5th edition version of BI-RADS assigns a letter (A B C D) to the four density categories. Mammograms with little or no density are classified as category A (almost entirely fat). Categories B, C and D reflect the increasing likelihood cancer is masked by dense tissue: B. scattered fibroglandular tissue, C. heterogenous fibroglandular tissue, D. extreme fibroglandular tissue. Hence the quantitative (PD) BI-RADS density categories have been replaced by new qualitative categories that reflect the likelihood BC is masked on the mammogram.

Although information on breast density was routinely included in the US radiology reports generated for clinicians, it was not routinely divulged to the women in their mammography outcome letter. This has changed because of mandatory reporting legislation in a growing number of US states [120]. The reasons women were not informed of their MD include: concern about alarming women that mammography may not be able to fully detect BC; concerns about variability in the application of the BI-RADS density categories; and uncertainty how to best image breasts affected by the masking effects of MD. Breast density is not routinely assessed or recorded at BreastScreen in Australia, nor is information on MD post-mammography typically provided to Australian women or their general practitioners.

The recent legislation has shifted the focus of MD in the US from a closely scrutinised risk factor within the research community to one of widespread concern to women and their clinicians. Whilst the US MD legislation mandates that women are informed, guidelines as how

to best care for women with dense or highly dense breasts were few when legislation first commenced. Both doctors and patients required succinct and accurate guidance, which the clinical and research community have endeavoured to provide [7, 121].

Most MD research prior to enactment of the US legislation used standard epidemiological modelling techniques where the lowest level of risk is used as the comparison category. Women undertaking their own enquiries into how MD affects their BC risk will encounter these comparisons in the breast density literature. However, providing clinical advice to women with high (50%+) or very high (75%+) levels of density that they are at 3 to 5 times the risk of BC compared to women with very low levels of MD is not compatible with other models of disease risk.

Both women with highly dense breasts and those with very low density are unusual, and together comprise only about 20% of all women. Most women have more than 10% dense tissue in their breasts [122], and use of women with very low PD as the reference category for BC risk due to MD is clinically problematic because these women are not at average risk due to their MD. Use of moderate density (~25 to 49% dense) as the reference category to approximate the average lifetime risk of BC due to MD is clinically appropriate. In keeping with this, the online guidance provided to clinicians by the California Breast Density Information Group advises that the ~10% of women with extremely dense breasts are at approximately double the risk of developing BC (~x2 relative risk (RR)) [123]. Very low levels of density (<10%) become protective: 0.8 RR compared to women of average density.

MD is also difficult to quantify and describe. The lack of a well validated, repeatable, reliable and easy to administer means to measure BD has been a core issue both clinically and epidemiologically. Reliable, easy measurement of BD is needed clinically to help women and their physicians make informed decisions. Epidemiological studies have also been hampered by

the lack of precise tools to measure BD. Studies tend to use different categories of density, the methods used to assess density often differ, and the techniques are typically subjective. For example, the freedom to assign a BI-RADS density category lies with the radiologist, and application of the categories has been shown to vary amongst practitioners. This adds complexity to the question of the additional risk posed by high MD, as well as which distinct characteristics of MD (i.e. patterns) are of greatest concern.

Use of a simple 4-category system such as the BI-RADS density categories, or a single numeric estimation of density such as PD, may overlook additional risk information present in the pattern of density on the breast [124-127]. Not all patterns of density are the same. Some density is sheet like, other density is wispy (scattered) and some resembles fluffy cotton balls. MD may be thickly clustered in a single quadrant of the breast, or evenly distributed over the whole breast area. The well-established semi-automated research tool to assess PD, the Cumulus program, was named to reflect the cloud-like nature of MD with its varied patterns that resemble clouds in the sky. The original density categories described by radiologist John Wolf attempted to describe and categorise these patterns. However the Wolfe density patterns were very difficult to implement repeatably and reliably.

Visual means of assessing mammographic density such as BI-RADS density [117] and semi-quantitative methods such as the Cumulus program [86, 128] are being supplanted by faster, fully automated clinical software integrated with the mammography equipment such as Matakina's Volpara [129] and Hologic's Qantra [130, 131]. Continued improvements to the techniques to assess breast density should lead to ongoing improvements in the identification of women at high risk of BC and better clinical outcomes.

Masking by MD is an important clinical issue. High levels of MD are associated with risk of interval BC (cancer diagnosed between screening mammograms). Consumer-driven legislation

to inform women of their mammographic density has been passed in at least 30 USA states. Women who are informed that they have high MD may subsequently pursue alternative methods of imaging if desired. However, it is still too early for the large screening trials currently underway to provide guidance regarding the best alternative imaging method/s (e.g. ultrasound, MRI) which best detect cancers in women with mammographically dense breasts [12-15]. Hence it is difficult to know which imaging modality or combination of modalities provides the most effective screening and diagnostic capabilities in women with dense breasts.

Whilst the issue of how to best screen and diagnose breast cancer in women with mammographically dense breasts is a worldwide problem, as yet, no mandate exists in Australia or other countries outside the USA to advise women of their breast density after imaging. This is in part due to the lack of a "gold standard" for density assessment. As yet there is no consensus on the best way to assess the amount of dense tissue in the breast.

2.4.1 Mammographic Density, a proven 'Independent' BC Risk Factor

Mammographic density was first described as an independent risk factor for breast cancer by Wolfe in 1976 [8]. Wolfe described four "parenchymal patterns"— N1, P1, P2, and DY— which correspond to increasing amounts of ductal patterns and 'dysplastic' (dense) tissues on a mammogram. Wolfe's research was controversial, because although about half the researchers who subsequently applied Wolfe's parenchymal patterns to their data did support his findings that mammographic density was an independent risk factor for breast cancer, the rest did not find an association between increasing breast density and breast cancer risk [132]. This latter group claimed that high levels of breast cancer found in women with extensive density resulted merely from the masking of breast cancer by the dense tissue [9, 133].

In the mid-1980s several researchers, including Norman Boyd, Audrey Saftlas and Pam Goodwin, applied epidemiological standards such as the Bradford Hill criteria to all twenty-two available studies of mammographic density [132, 134, 135]. Studies that were robust (i.e. met most epidemiological criteria) supported the hypothesis that mammographic density was an independent risk factor for breast cancer. Studies which were not supportive tended to lack methodological quality [132, 135]. Robust studies correlated intra- and inter-observer mammographic density assessments, avoided referral bias by selecting asymptomatic subjects as controls, and accounted for associations of mammographic density with other breast cancer risk factors such as age. It also became apparent that one of the difficulties with mammographic density research was the highly subjective nature of Wolfe's parenchymal patterns.

A more recent meta-analysis of 42 mammographic density studies in 2006 found an association of around two to five times the risk of breast cancer for women with increasing levels of mammographic density compared to women with very little breast density (5 to 24%, 25 to 49%, 50 to 74% and 75%+ PD vs <5% PD) [136]. This analysis provided additional evidence of the ability of dense tissue to mask BC detection: prevalence studies had lower gradients of risk for breast cancer compared to incidence studies. The higher rates of risk seen in incidence studies was due to the increased number of cancers found in the first year compared to later years due to masking by dense tissue. Adjustment for this masking effect by excluding cancer detected in the first year of incidence studies yielded lower gradients of risk which approximated the risk gradients seen in prevalence studies. As described in the previous section, models utilising average density for the comparison category yield a doubling of risk for women with very high levels of density vs average density (25 to 49% PD).

MD is related to numerous BC risk factors, including age, body mass index and reproductive history. These relationships are described in the next section (2.5 Mammographic Density and breast cancer risk factors). A paradox exists whereas MD decreases with age, whilst BC risk

increases with age. Not all women experience the same decrease in MD as they age, and it is not yet known whether women who retain high MD, for instance, after childbearing and menopause are at increased risk compared to their counterparts whose high MD decreased during these reproductive events.

MD is also one of the few strong BC risk factors that can be modified. Therapies which increase estrogen levels in the body tend to increase MD; therapies which decrease estrogen levels in the body tend to decrease MD. The mutability of MD may make it a useful biomarker of BC risk during hormonal therapy for the prevention and/or treatment of BC. These relationships are discussed in Section 2.6 Drug therapies that affect mammographic density.

2.5 Mammographic Density and breast cancer risk factors

Many of the factors which affect BC risk also are associated with breast density. In general, increased parity, longer duration of breast feeding and lower age at first birth are associated with lower PD. Whilst BC risk increases with age, paradoxically PD tends to decrease with increasing age. Another BC risk factor, body mass index (BMI, kg/m^2) is also strongly and inversely associated with lower PD.

2.5.1 Increasing age is associated with decreasing MD

Breast density is not normally distributed in women aged over forty and is age dependent [137, 138]. Percent density in women of mammographic age (age 40+) tends to have a right skew (long right tail), as illustrated by the distribution of PD in baseline mammograms from a group of postmenopausal women who are at high risk of breast cancer (Figure 2-7, below).

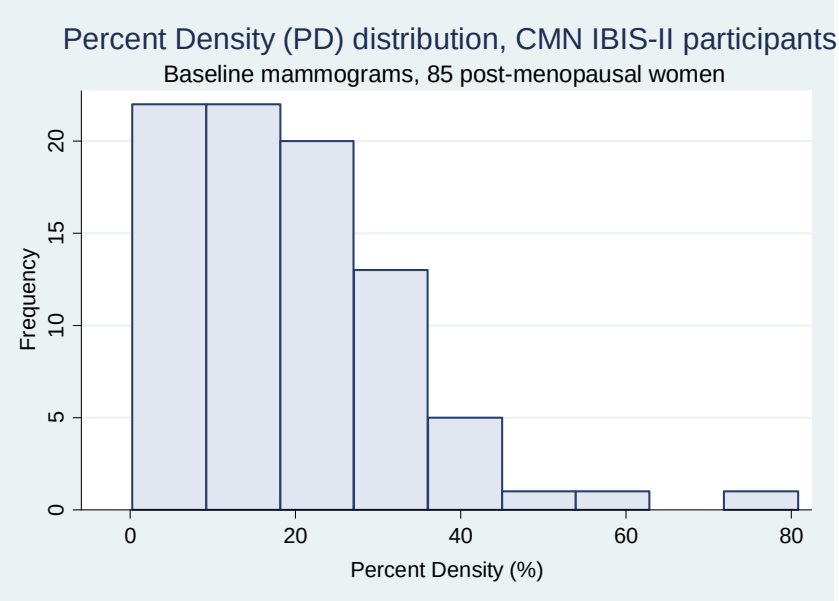


Figure 2-7 PD distribution, baseline mammograms from 85 Calvary Mater Newcastle IBIS-II participants

Density was assessed using the Cumulus program. The distribution has a right skew. All participants in the IBIS-II breast cancer prevention trial are post-menopausal and at high risk of BC (≥2 relative risk).

The right skew for PD in older (e.g. post-menopausal) women is further illustrated in density distributions of women from a normal-risk population (Figure 2-8, below). These post-menopausal women (aged ~50+) tend to have low levels of density [139].

Mammographic density decreases with age [136, 137, 140-146]. PD typically reduces by about 1% per year [147]; higher rates of decrease are generally observed in younger women compared to older women. For women free of breast cancer aged 40 to 44 in the Canadian National Breast Screening Study, median percent mammographic density was ~38% (Figure 2-9, below) [122]; this declined to an 18% for postmenopausal women aged 55 to 59. The younger women not only had higher average percent density, but the distribution is broader and more symmetric (i.e. more normal) compared to stronger skew observed in their older counterparts. Additionally, relatively few women of mammographic age (age 40+) have dense breasts (>50% density). Most women, even those in the age 40 to 44 group, have breast density less than 40% as shown in Figure 2-9: median density for each age group is indicated by the white band at the waist of each vertical bar; 40% PD is greater than the median for all age groups. Extremely dense breasts (≥75% density) are even less common. Like average percent density, the proportion of

women with highly (≥50%) and very highly (≥75%) dense breasts also decreases with age [137, 138, 144, 148, 149].

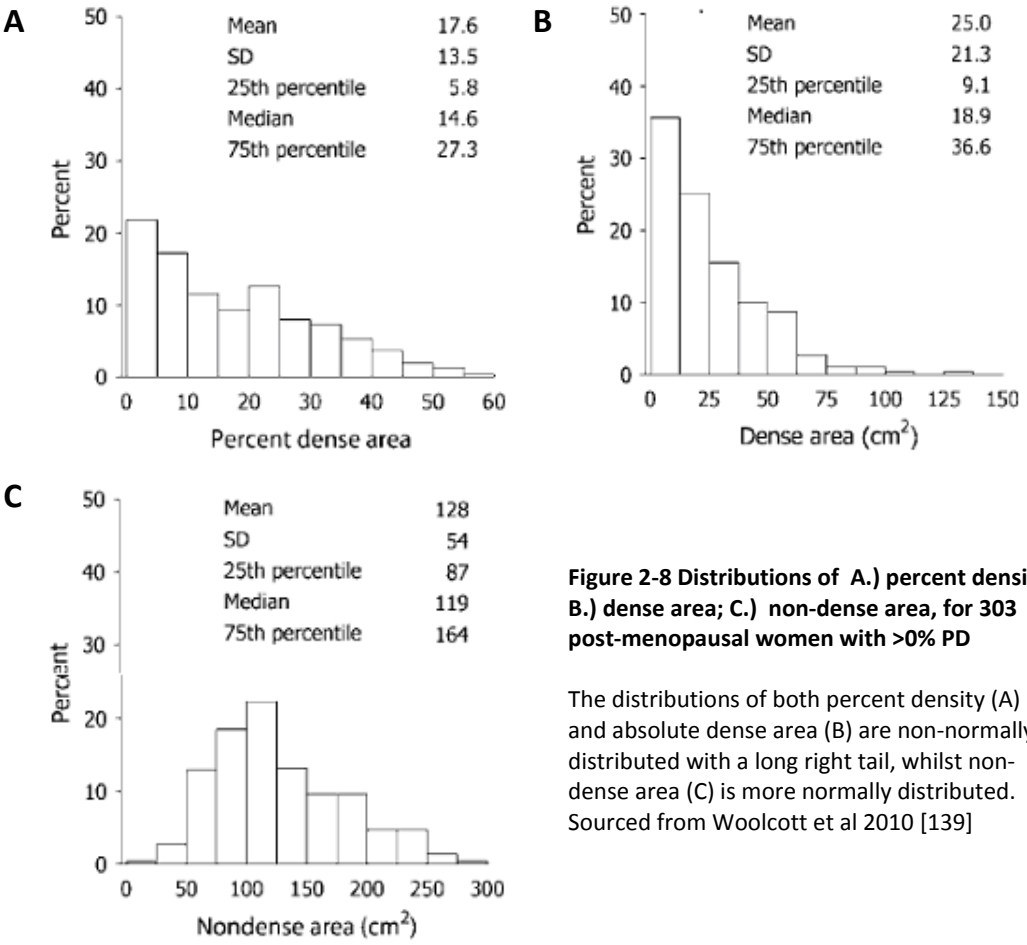


Figure 2-8 Distributions of A.) percent density; B.) dense area; C.) non-dense area, for 303 post-menopausal women with >0% PD

The distributions of both percent density (A) and absolute dense area (B) are non-normally distributed with a long right tail, whilst non-dense area (C) is more normally distributed. Sourced from Woolcott et al 2010 [139]

Mammography is generally not performed in women under age 40. Diagnostic accuracy is reduced by the dense breast tissues, and the breast cannot be suitably compressed to obtain good image quality at a low radiation dose. The restrictions on performing mammography in younger women have prevented widespread epidemiological studies of MD in young women. A few BD studies in young women have been performed using magnetic resonance imaging (MRI) [150-159] and dual x-ray absorptiometry (DXA) [160-164].

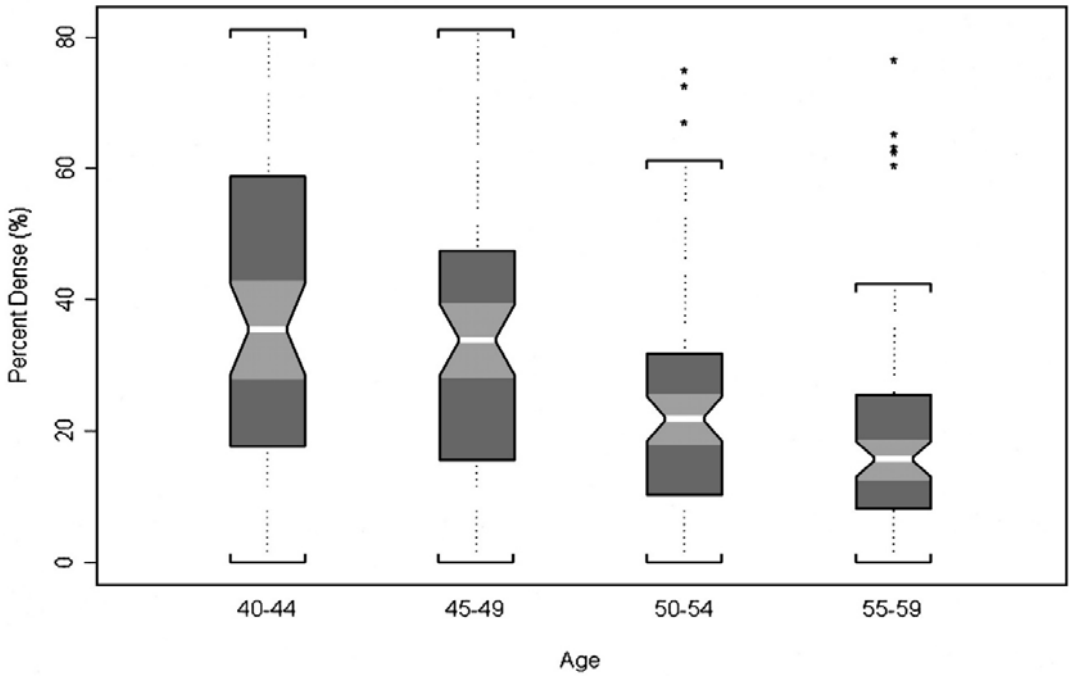


Figure 2-9 Distributions of percent mammographic density by age in 354 breast cancer free control subjects
The distribution of percent density decreases with age. A marked decrease is seen for women in the menopausal transition (age 50 to 54). A small proportion of women have mammographic density over 50%; this proportion decreases with age. Very few women have very high percent density (over 75%), especially those who are post-menopausal (age 50+). Sourced from Boyd et all 2002 [137]

Percent breast water in MRI is likely equivalent to percent mammographic density. The fatty tissues of the breast contain little water and appear dark on MRI; these same tissues are also transparent to x-rays and do not contribute to density during mammography. MRI breast water and mammographic density are strongly correlated (ICC 0.8) [152, 155]. Both MRI and DXA assessments of mammographic density are reproducible [152, 154].

MRI, unlike mammography and DXA, does not utilise radiation and hence does not expose young women to additional radiation at a time when they may be most susceptible to it carcinogenic effects [165-168]. Use of MRI as a routine imaging technique has been limited due to high costs (thousands of dollars) per episode, but this hurdle grows smaller each year as costs decrease to hundreds instead of thousands of dollars per episode. Standard breast MRI requires use of a potentially harmful contrast agent to differentiate breast tissues; this renders

MRI unsuitable for general screening. Additionally, MRI is problematic for some individuals due to the need to lie very still as well as claustrophobic fears. Effective methods of using non-contrast MRI to assess breast density are being developed. The DXA assessment technique is well tolerated.

Due to the difficulty with assessing breast density in young women, little is known about the natural history of breast density in women under 40 years and how this relates to breast cancer risk. A study in older women [147] has shown that women tend to remain in similar groups of percent density over time— i.e. women in the upper quintile (upper 20%) of percent mammographic density tended to remain in the upper quintile over time. Greater rates of decline in percent mammographic density have been found in older women with higher baseline mammographic density [144]; it is theorised that this may be true for younger women as well [169].

A (non-contrast) MRI study of young women and their mothers also supports the theory that greater declines in breast density are seen for women with higher initial levels of density [155]. Figure 2-10 (below) shows the distributions of percent water for daughters aged 15 to18, for daughters aged 19 to 30 and in a subset of 100 of their mothers. Median percent breast density as represented by MRI percent water decreases with increasing age in these groups. Similarly, the inter quartile range (IQR), a measure of the variability in a population, declines with age. Several models of how breast density might change over time are shown in the left column of the figure. The MRI data best fits model C, where women with the highest percent density have the steepest decline in percent density as they age. In this model, both median percent density and the IQR of the distribution decline with age, as observed in the Boyd MRI study. Therefore, it is likely that women who are in the upper quartile of percent density at a young age will remain in the upper quartile for density when they are older.

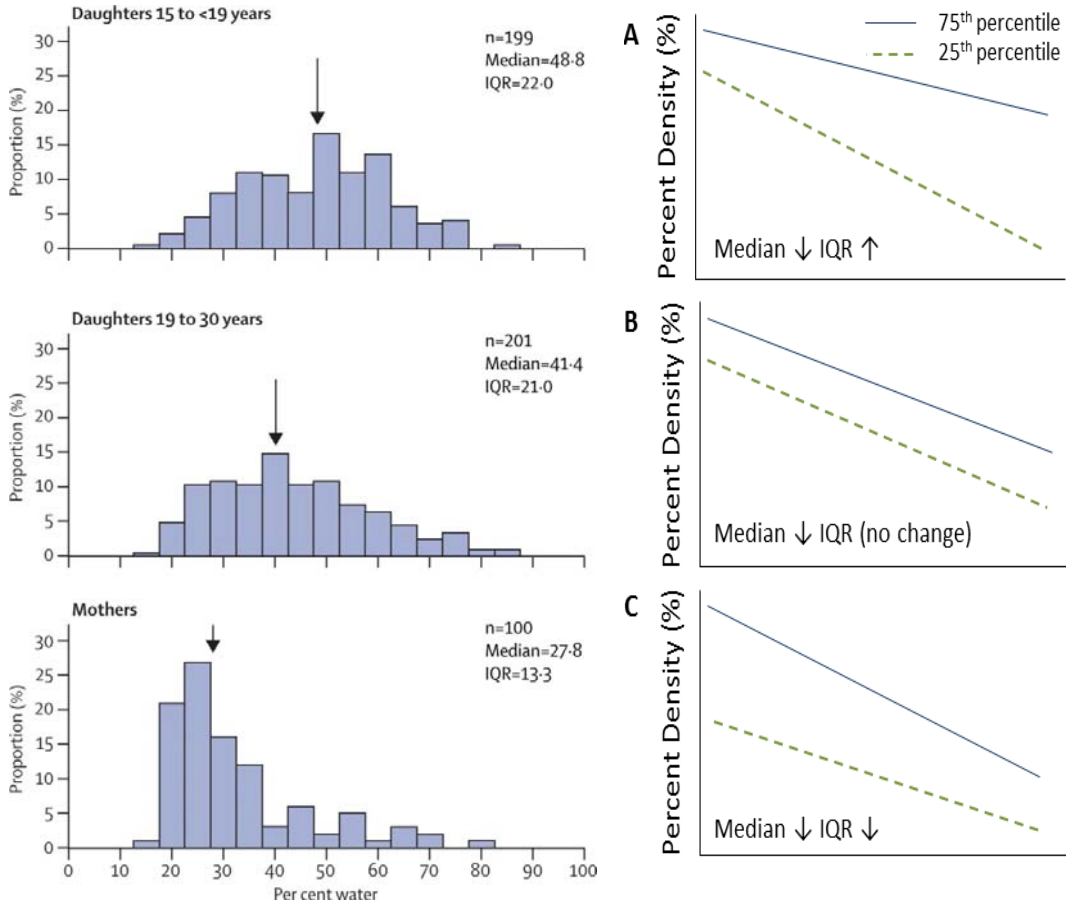


Figure 2-10 Left column: Distributions of percent water in Daughters aged 15 to 18 years, Daughters 19 to 30 years, and their Mothers. Right column: Theoretical distributions of percent density with age Left: Median percent density (arrow) declines with age, as does the IQR (inter quartile range). The youngest group of women have a nearly normal distribution of percent water; the distribution becomes more and more skewed to the right for women in older age groups. The data from the left column most closely matches the theoretical distribution shown in (C) right column, where both median density and IQR decline with age. Sourced from Boyd et al 2009 **[155].**

Interestingly, the distributions of MRI percent water in the figure change markedly with age. The distribution of percent water in daughters aged 15 to18 years of age is roughly normal, but the distribution of daughters aged 19 to 30 years becomes slightly skewed to the right. The 100 mothers' MRI percent breast water distribution is much less normal, and has a pronounced skew to the right. This latter distribution is similar to the strongly right skewed percent density distributions seen for mammograms of post-menopausal women (Figure 2-7, Figure 2-8, & Figure 2-9 ages 50+). The strong right skew is likely due to the cumulative effects of ageing, parity and breast feeding in the population of mothers. After menopause, the distribution of percent water in mothers may become even more similar to that observed for percent

mammographic density in postmenopausal women, due to the rapid decline in mammographic density expected for most women during the menopausal transition. The differences in MD distributions for younger and older women underscores the need to take age into account when assessing MD and breast cancer risk.

As for MRI, DXA (dual-energy x-ray absorptiometry) has been used to show that daughters have higher % fibroglandular tissue in their breasts (median 69.4 % fibroglandular volume (FGV)) and the distribution is more symmetric than their mothers; the mothers have lower % density (median 35.8% FGV) with a right skewed distribution [161]. Higher BMI in adolescent girls is associated with larger breast size (area and volume), higher FGV and lower breast density [162, 170], whilst total body fat distribution is inversely associated with absolute FGV [164]. Increasing Tanner stage is positively associated with breast area, breast volume and FGV [162]. Breast density increases until Tanner Stage IV, but is lower at Tanner Stage V [162]. Consumption of more sweetened milk-based drinks is associated with higher %FGV in adolescent girls, whilst higher yoghurt intake is associated with lower %FGV [171]. Adolescent saturated fat intake is positively associated with breast density (measured by MRI) in women aged in their 20's, whilst lower mono- and polyunsaturated fat intake is inversely associated with breast density [172].

Greater use of non-contrast MRI and other techniques that do not involve ionising radiation such as ultrasound tomography may eventually lead to widespread characterisation of breast density in young women [173]. This may help the discovery of additional factors that cause breast density to decrease over time for most women, and what factors may cause some women to retain high levels of breast density despite the cumulative effects of age, menopause and child bearing. Since it is also theorised that exposure to estrogen and other potential oncogens at a young age are responsible for increased levels of breast cancer in later life, early detection of high mammographic density may make it possible to subsequently intervene more effectively in women at a younger age to better implement breast cancer risk reduction [169].

2.5.2 Weight and BMI are inversely associated with MD

Mammographic density is often quantified using PD, which is the proportion of the total breast area covered by dense tissue. As a consequence, varying weight can greatly influence PD because breast size also varies [174, 175]. Body mass index (BMI)— weight in kilograms divided by the square of height in metres— has a strong, but inverse, association with PD in most [175-182] , but not all [183] studies. The amount of adipose tissue in the breast tends to increase as BMI increases (i.e. additional fat accumulation in the breast is seen with increasing BMI), but the area of the breast occupied by dense tissue typically does not show a concomitant increase [177, 184, 185]. Hence PD decreases because the absolute area of dense tissue does not increase with BMI. As a result, BC risk due to the percentage of tissue classified as dense in the breast may be underestimated if BMI is not taken into account [178].

The observation that BC risk is decreased for obese women (BMI >30) before menopause, but increased for obese women after menopause may not be mediated through MD. However, the amount of adipose tissue in the breast may complement MD as a biomarker for BC risk [186].

2.5.3 Parity and breast feeding tend to reduce MD

Parous women tend to have lower breast density than non-parous women [138, 145, 149, 187-194], although as discussed in a later section, age at first birth also has an effect upon both BC risk and MD. Hence the lower life-time risk of BC associated with parity may be mediated through MD [195].

Grove [196] first described the inverse relationship between the number of children (full-term pregnancies) and the proportion of P2 + DY (highly dense) Wolfe patterns. Other researchers have found similar negative associations between Wolfe's parenchymal patterns or PD with increasing parity [138, 146, 179, 180, 194, 197, 198]. PD assessed before and after first birth at

an average age of 33 in (primarily symptomatic) young women showed their average percent mammographic density decreased by 12% compared to 3% in non-parous age-matched controls over the three year average period between the first and second mammograms [199]. Approximately half the new mothers experienced a 10% or greater decrease in PD after birth (Figure 2-11, below). The effect of each full term pregnancy also appears to be additive— each full term birth confers an additional decrease in mammographic density as well as lifetime reduction in BC risk.

Similarly, length of time breast feeding provides not only a reduced risk of breast cancer, but is associated with a decrease in PD [179]. Thus it is possible that the long term decrease in breast cancer risk associated with each full term birth and subsequent breast feeding noted at the population level may be moderated through mammographic density. BD may also be a biomarker at the level of the individual for reduction in BC risk: marked reductions (>10% PD) in BD after the first pregnancy and subsequent breast feeding may indicate a reduced risk of BC compared to women whose BD does not show marked reductions.

As noted previously, the gradual decrease in mammographic density observed at the population level between puberty and menopause is likely influenced greatly by the effects of parity and breastfeeding most women experience during this time period.

2.5.4 Menopause reduces average MD

On average, mammographic density decreases by about 5 to 8% over the menopausal period; this period of life is associated with the greatest annual decrease in change of PD [137, 144, 200]. As noted previously, the involution of the breast during menopause has similarities to the involution of the breast after cease of breast feeding. The proportion of ductal and connective tissues (extracellular matrix, ECM) decreases, and the percentage of fat in the breasts increases.

Both involutions involve interactions of the stroma (connective tissues) with the ductal tissues, and produce inflammatory components linked to carcinogenesis.

It is not yet known why some women do not experience breast involution during the menopausal transition as strongly as other women. The breast cancer risk for women who retain more than 50% mammographic density after menopause is quite high (~3-fold RR); the risk is even more elevated (~6-fold) for post-menopausal women with more than 75% mammographic density compared to those with less than 10% density [136, 201]. The retention of high or very high MD post menopause may indicate the need for more frequent mammographic screening (e.g. annual attendance at BreastScreen) and/or use of additional imaging techniques (MRI, ultrasound) during the screening process. These women may also be candidates to undertake treatment for BC prevention.

2.5.5 Age at first birth, menarche and menopause

Women who give birth at younger ages tend to have lower PD [138, 149, 190, 192, 193, 202]. Late age at first birth is associated with increased breast density [149, 177, 194, 196, 203, 204], and the rate of age-related decrease in MD may be reduced for first-time mothers older than age 30 [145].

Later menarche is associated with higher breast density [194, 202], but this relationship is often attenuated when adjustment for other breast cancer risk factors is made [176, 203, 205, 206]. In contrast, late age at menarche is associated with a decrease in breast cancer risk, so the increased BC risk due associated with early menarche may not be modulated through BD.

Later age at menopause is associated in some studies with increased mammographic density [147, 177]. Therefore, age at menopause may be related to breast cancer risk via mammographic density. For a group of women over age 70, surgical menopause was positively related to PD (33% PD vs 25%PD for surgical vs natural menopause).

These relationships suggest, as per first live birth, that certain events significantly change MD, permanently, and the age at which they occur is very important for modifying MD. It is not known 1.) if the change in MD induced by these reproductive events occurs only in select women; 2.) which breast cancer risk factors and/or genetic markers predict which women will respond to menopause, childbearing, etc. with measurable decreases in MD; and 3.) whether decreases in MD always correlate with decreases in BC risk at the level of the individual (as opposed to population-level BC risk). It is also not yet known whether the amount of change induced by these reproductive events is proportional to reduction in BC risk, and how much change in MD is possible given the amount of MD which develops in the breast at puberty and is subsequently retained through the reproductive years.

The mutability of MD and some of the possible effects of pregnancy on MD are illustrated in Figure 2-11 . One (symptomatic) young woman showed a marked PD decrease after her first full-term pregnancy, from 65% to 10% (Figure 2-11 A, below). Presumably this woman's BC risk is now much lower than a similar woman whose MD remained high after her first pregnancy. One other parous woman in this study had an MD increase of >15% after her first pregnancy. Density was assessed by two independent readers using an interactive thresholding method (Madena), so this increase is unlikely to be a reading artefact. Assuming that the woman had completed involution post lactation prior to her second mammogram, presumably this increase in density is due to an incomplete involution of the connective and glandular tissues which developed during pregnancy. It is possible that increases in MD (post-lactation) may be a biomarker for developing pregnancy-associated BC and/or later occurring BC. Perhaps women whose density increases after first birth also experience very little decrease in MD over the peri-menopausal and menopausal timeframes, which is associated with higher risk of BC. Whether this plasticity in MD during reproductive events predicts the likelihood of short and long term changes in BC risk requires further investigation.

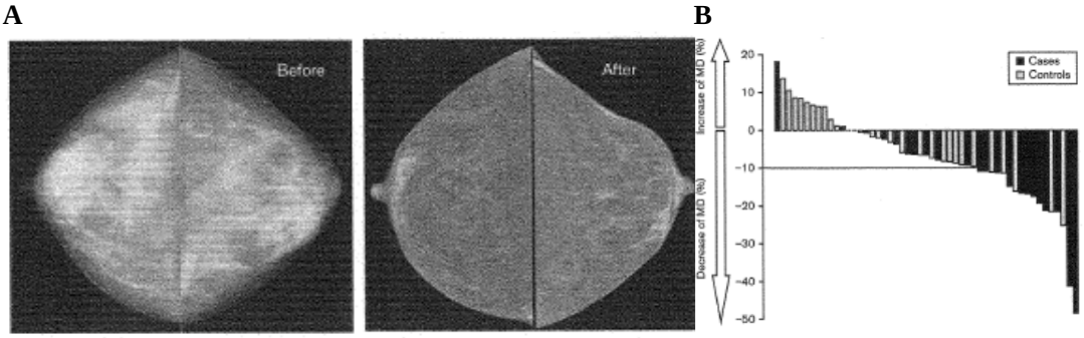


Figure 2-11 A) Before pregnancy and After pregnancy mammograms for a first time mother B) Percent density change in cases (1st time mothers) and non-parous age-matched controls
Sourced from Loehberg et al 2010 [199]

As per the Pike model of breast tissue ageing [85], delayed age of first birth and late age of menopause increases a woman's exposure to hormonal and other factors which may subsequently keep the rate at which her breast tissue ages elevated compared to her similarly aged peers who gave birth earlier and have experienced menopause. Pike's model may help explain how increased age at first birth and menopause may increase mammographic density. Interestingly, though the Pike model predicts the protective effect of late menarche with breast cancer risk, the possible increase in mammographic density with late menarche opposes the Pike model.

2.5.6 Taller women tend to have higher PD

Tall women are at higher risk for breast cancer, and generally have increased percent mammographic density [174, 190, 192, 197]. It is possible that a shared mechanism (e.g. growth factors [207]) is responsible for this positive association between height and percent mammographic density with breast cancer risk. Given that it is unlikely that widespread use of estrogen or other therapy to reduce adult height in young, healthy (adolescent) females will again be considered a viable treatment for 'constitutional tall stature' [208, 209], this BC risk factor is an unlikely target for reduction in BC risk. However, utilisation of a woman's height in risk assessment may help improve the tailoring of screening and BC risk prevention strategies for individual women.

2.5.7 Endogenous hormone levels

An increase in BC risk for post-menopausal women is associated with higher levels of endogenous estradiol and testosterone; BC risk is lower in post-menopausal women who have higher levels of endogenous sex-hormone binding globulin (SHBG). This latter molecule binds strongly to estradiol and testosterone, limiting their bioavailability [210]. The relationship between endogenous hormone levels and pre-menopausal breast cancer is not as well characterised [211, 212]. It is likely that the heritability of breast density (comprising dense area, adipose area, and PD) is a separate trait from the heritability of endogenous sex-hormone levels [210, 213].

Increased mammographic density has been observed for women in the top percentage levels for estrogen and other endogenous sex hormones such as testosterone in the blood [214-218]; however this relationship has not been observed in all studies [219-221]. For post-menopausal women in particular, the lack of association is not surprising given that serum levels of hormones may not reflect the hormone levels in the breast tissue [222, 223]. This is because most of estrogens and androgens produced by the cells of post-menopausal women are utilised by the same cells (intracrine action) [223, 224]. Very little, perhaps only 10%, of the product of this local synthesis is released into the circulatory system. Therefore, plasma levels of endogenous estrogens and androgens are unlikely to reflect the levels of sex hormones present in the breast tissue of post-menopausal women. The serum metabolites of androgens have been found to accurately reflect intracrine production of these hormones in women [223]. Similarly, the blood and urinary metabolites of estrogen also appear to reflect the levels of estrogen in post-menopausal breast tissue; these metabolites also have an association with mammographic density and breast cancer risk [225, 226].

The serum levels of sex hormones in pre-menopausal women differ markedly from those of post-menopausal women because the ovaries in pre-menopausal women produce high levels of

estrogens which enter the circulatory system [224]. A slight association may exist for mammographic density and progesterone, sex hormone binding globulin (SHBG) and estrone in pre-menopausal women [216, 227].

The association with serum and/or urinary sex hormones or their metabolites and mammographic density in pre- or post-menopausal is not yet fully characterised nor understood. If proven to be true, some of these associations may provide some insight into the biology underpinning mammographic density. Better methods by which to measure sex hormones levels within the breast tissues may lead to better estimates of BC risk for women with high vs low levels of these hormones, as well as the interaction of these hormone levels and MD.

2.5.8 Heredity

Twin studies have shown that there is a highly heritable component to mammographic density; up to 60 or 70% of the variation in mammographic density can be explained by heredity [228-232]. Genetic factors which influence mammographic density may also influence other breast cancer risk factors [198, 233]. Similarly, the relationship between mammographic density and breast cancer has been shown to vary by ethnicity in many [234-237] but not all [238, 239] studies.

Since family history appears to play a large role in breast cancer risk [5], the heritable component of breast density is not unexpected, nor are the possible links with breast density and ethnicity. Genome wide association studies (GWAS) have produced a number of new candidate genes for breast cancer [240, 241]. Some of the BC genes are associated with mammographic density [242, 243]. MD also has novel (different) genes associated with it which are not associated with BC. Due to the large numbers of people required to obtain genetic associations using GWAS, it is likely that more BC candidate genes are associated with MD than is currently

apparent; the same could be true for the genes associated with MD which currently lack an association with BC.

Interestingly, MD does not appear to differ between women with and without BRCA1 and BRCA2 mutations, i.e. MD is not higher in women with BRCA1/2 mutations compared to those without [244, 245]. However, as for women without BRCA1/2 mutations, women with BRCA1 and BRCA2 mutations are also at increasing risk of BC as MD increases [193].

2.6 Drug therapies that affect mammographic density

Treatment with therapies that increase hormone levels such as combination hormone therapy with estrogen and progesterone tend to increase mammographic density [215, 246]; in contrast, endocrine therapies that decrease the effects of estrogen such as tamoxifen tend to decrease mammographic density [247-251]. The exact mechanisms by which mammographic density is affected by these treatments is unknown; further investigation into these processes is needed.

Because an association with BC and MD exists for these medications, they provide clues as to how breast density may be modified by hormonal changes. This association also provides an opportunity to utilise changes in mammographic density as a biomarker for the breast cancer risk associated with these treatments.

2.6.1 Hormonal contraceptive use

The impact on MD from use of oral and other hormonal contraceptives is not well defined. PD and dense breast area have been shown to both increase and decrease [190] with use of hormonal contraceptives, or remain unaffected [144, 186]. The effects may vary by ethnicity/race and country [252]. Premenopausal women unaffected by BC had a trend toward higher PD with OC use, but lower PD if diagnosed with BC [253].

2.6.2 Hormone replacement therapy and mammographic density

Hormone replacement therapy (HRT) historically has been used to reduce the menopausal side-effects such as hot flushes, and was thought to potentially reduce the risk of cardiovascular disease in post-menopausal populations [254, 255]. HRT is typically formulated with a combination of estrogen and progesterone, or with estrogen only for women without a uterus. Progesterone is added to estrogen in HRT to reduce the proliferative effects on the uterus caused by estrogen-only (unopposed) HRT.

By 1980 investigations into the effects of HRT on MD had commenced [256]. Additional studies were undertaken in the 1990s as HRT became more commonplace [257, 258]. Most studies found a significant increase in the proportion of women with high MD for HRT users compared to non-users [259-264]. A visibly detectable change in mammographic density may take place as soon as 4 months after commencing HRT [258]. Kaufman [265] theorised that HRT appeared to interfere with the normal tissue involution of the breast post-menopause.

Combination hormone therapy with estrogen and progesterone is significantly associated with increased mammographic density [215, 257, 266]. The WHI (Women's Health Initiative) was a randomised controlled trial in which over 16,000 women were randomised to either combination hormone therapy or placebo. Both a statistically significant increase in breast cancer risk and mammographic density were found in women treated with combination HRT [246]. After this discovery was published, hormone therapy use was reduced in the general population [267, 268]. This reduction in HRT use may have helped lower the worldwide incidence of breast cancer in post-menopausal women [269-271]. HRT has also been associated with an increase in PD as well as an increase in BC in other studies [253].

Interestingly, in the WHI estrogen-only HRT randomised controlled trial of over 10,000 women without uteri, breast cancer risk was reduced in women randomised to the estrogen-only arm

even though mammographic density modestly increased during treatment [272]. The magnitude of the MD increase for estrogen-only HRT was smaller than the magnitude of the density increase for combination HRT.

The WHI studies were not powered to further investigate whether the women who experienced strong (≥10%) mammographic density increases due to estrogen-only or combination hormone replacement therapy also experienced changes in breast cancer incidence. It is not known whether increased MD is a personal-level biomarker of increased breast cancer risk for the women who took combination HRT.

Increases in MD may be a biomarker for decreased BC risk for estrogen-only HRT. This paradox could result from estrogen's antagonistic effects upon BC growth, such as observed historically in women with advanced hormone-sensitive BC treated with estrogen. Changes in estrogen availability (both lower and higher) appear to negatively affect hormone sensitive BC growth.

Another randomised controlled trial, the PEPI (Post-menopausal Estrogens/Progestins Interventions) trial, showed similar MD changes in response to HRT [273, 274]. Mean PD for controls decreased non-significantly by ~0.1% from baseline. Adjusted percent density increased non-significantly by 1.2% for women randomised to estrogen-only HRT during the same period. In contrast, adjusted percent density for women randomised to one of three types of combination HRT increased significantly by 3.1%, 4.6% and 4.8%.

The response to treatment with HRT in the PEPI trial was variable. Some women did not experience changes in density. Density decreased up to 25% for some women, whilst density increased up to 38% for others.

It is probable like the WHI trial, the majority of the average increase in density was due to a subset of women who experienced strong changes in density whilst on HRT. The PEPI trial also indicates (like an earlier case-control study of density and HRT, confounded by selection bias of cases [275]) that women with lower baseline percent density experienced larger increases in PD from HRT compared to women with higher levels of baseline density.

As discussed earlier, endogenous levels of serum sex hormones are not strongly associated with mammographic density in post-menopausal women. However, the increased levels of serum sex hormones from HRT are associated increases in mammographic density [266, 276-278]. Additionally, increases in prolactin— a hormone which has many functions and is essential for breast development and lactation [279]— was associated with increases in MD for women participating in the PEPI trial.

MD tends to decrease after cessation of HRT. The typical risk factors for BC such as age, race, BMI and family history of BC are not associated difference in change in MD after cease of HRT [280]. Differences in genetic makeup, perhaps via hormone metabolism and growth factor pathways, are likely responsible for the variable response of MD to HRT [281].

2.6.3 GnRH agonists and mammographic density

Gonandotrophin releasing hormone agonists (GnRHAs) are potent suppressors of ovarian function [282]. GnRHAs are useful for contraceptive purposes, and are also widely used in assisted reproduction [251, 282]. GnRHAs are also efficacious in treating hormone-sensitive cancers in pre-menopausal women because they suppress ovarian function [283, 284]. GnRHA therapy induces a menopause-like state that, in conjunction with other endocrine treatments like tamoxifen or aromatase inhibitors, may improve survival in this group of breast cancer patients compared to tamoxifen only treatment [285-288].

A number of small mammographic density studies have been performed in premenopausal female subjects examining the potential of GnRHAs as a contraceptive and cancer prevention agent [251, 289-291]. Average mammographic density decreased whilst the patients received the GnRHA treatment. The density reductions as well as the contraceptive effects of the treatment completely reversed 12 months after therapy ceased. Like the reversible changes in mammographic density experienced by women on HRT (and tamoxifen treatment, as described in the next section), these experiments show that GnRHA therapy also induces reversible changes in mammographic density, and that these changes might be modulated through the influence of estrogen.

2.6.4 Tamoxifen and mammographic density

As described earlier in this Chapter (section 2.3.3), tamoxifen is a SERM (serum estrogen receptor modulator), and has both estrogenic and non-estrogenic effects in different areas of the body. Tamoxifen is an effective adjuvant endocrine therapy for hormone sensitive breast cancer because of its anti-estrogenic properties in breast tissue. Tamoxifen has also been proven to reduce the incidence of breast cancer in healthy women at high risk of breast cancer [292, 293], but its side effect profile is such that the risks outweigh the benefits for breast cancer prevention in women at normal risk from breast cancer [294].

A large of number of studies have investigated the effects of tamoxifen on mammographic density [295] . In general, use of tamoxifen causes a reduction in mammographic density during the first 12 to 18 months of use [296]; these effects are reversible after therapy ceases. For some women, the change in mammographic density whilst on tamoxifen treatment is striking; up to 50% or more of the dense tissue in the breast disappears upon treatment with tamoxifen [250].

Closer examination of the preventive effects of tamoxifen in over 1000 high-risk IBIS-I participants revealed that women treated with tamoxifen who had a strong (≥10%) reduction in MD had a 63% reduced risk of breast cancer compared to controls [17] . In contrast, women treated with tamoxifen who did not have a strong reduction in mammographic density (<10% change) had a breast cancer risk similar to that of IBIS-I controls who received placebo (no change in risk). This study and others [30, 297, 298] provide evidence that mammographic density may be a useful biomarker for treatment efficacy of breast cancer adjuvant endocrine therapies.

2.6.5 Other SERMS and mammographic density

Raloxifene is the most other studied SERM in relation to mammographic density, however a small amount of density research has also been published for the SERMs tibolone and bazodoxifene. A review published in 2010 on the effects of raloxifene found that its effects on mammographic density are inconclusive [299]; a more recent [2014] review found raloxifene decreased BD in two studies whilst a further seven found no effect [295]. A phase II trial is currently underway to discover if the fourth generation SERM alcolbifene is a suitable agent for the prevention of breast cancer. Measurement of changes in mammographic density is one of the secondary endpoints for this trial [300].

2.6.6 Aromatase inhibitors and mammographic density

As described earlier in this Chapter (Section 2.3.3), aromatase inhibitors (AI) inhibit the synthesis of estrogen by the enzyme aromatase [92]. AI are effective only in post-menopausal women whose main source of estrogen is from the activity of aromatase in peripheral adipose tissues. Aromatase inhibitors are not effective in pre-menopausal women whose main source of endogenous estrogen is the ovaries.

Two types of third-generation AI are commonly used for prevention and treatment of BC. Two

of the AI, letrozole and anastrozole, are non-steroidal, reversible inhibitors of aromatase; the other, exemestane, is a steroidal irreversible aromatase inhibitor. These three AI are also known by their trade names: Femara (letrozole), Arimidex (anastrozole) and Aromasin (exemestane) [301]. The clinical trial MA.27 revealed that treatment with five years of exemestane or anastrozole for early breast cancer provided similar efficacy [302], i.e. the steroidal and non-steroidal AI are similar in effectiveness. There is good general consensus, as a result of many randomised controlled trials, that aromatase inhibitors are superior to tamoxifen for BC treatment in post-menopausal women [301].

Though often unpleasant, menopausal-like side effects such as hot flashes, arthralgia and sweating from treatment with AI may indicate that the therapy is effective because these symptoms are associated with a reduction in breast cancer events [106]. Similarly, it is possible that strong changes in mammographic density upon treatment with AI is also a biomarker for treatment efficacy. Together, MD change and treatment side effects may provide an even stronger biomarker for treatment effectiveness.

At the time this project began in 2010, few (four) studies had examined the effect of aromatase inhibitors upon mammographic density. Two of these studies examined the effect of AI in conjunction with HRT [303, 304], whilst the others examined letrozole in populations of HRT-free women with and/or without BC [28, 29]. These latter studies indicated the 1 year change in PD associated with AI treatment would be small (–1% to –2%). As MD and AI are the content focus of this thesis, a more formal and up-to-date literature review was undertaken to investigate the effects of AI on MD (Section 2.7 Literature Review of AI and Breast Density).

2.6.7 Summary of endocrine therapies and mammographic density

In general, endocrine therapies which increase levels of estrogen in the body such as HRT with estrogen or estrogen and progesterone also increase breast density. An increase in breast cancer

risk was observed for women on combination HRT in the WHI trial.

Concomitantly, endocrine therapies which reduce the levels of estrogen in the body such as tamoxifen and GnRHAs tend to decrease mammographic density. In the IBIS-I randomised prevention trial, statistically significant, strong decreases in mammographic density were accompanied by a significant 63% decrease in cancer risk for women treated with tamoxifen. Some studies have shown small, yet significant decreases in MD during treatment with AI. Whether these decreases during AI therapy will be useful clinically to determine treatment efficacy is not yet known.

2.7 Literature Review of AI and Breast Density (BD)

2.7.1 Introduction

The longitudinal effects of AI on MD/BD in post-menopausal women are not well characterised. A formal literature search was undertaken in April 2018 to investigate the effect of AI on BD in women with and without BC. The broader term 'breast density' (BD) was used to include potential AI studies utilising imaging methods besides mammography. Conduct of the review was guided by the PRISMA statement (Preferred Reporting Items for Systematic reviews and Meta-Analyses) [305], as well as the Cochrane Handbook (http://handbook-5-1.cochrane.org/) and Cochrane Systematic Review training received in 2000 by the candidate whilst co-developing a protocol for a Cochrane Review [306] .

2.7.2 Methods

Any study which examined the effects of AI on MD and/or BD in post-menopausal women was eligible for this review, including reviews of these studies. Searches were performed for PubMed, ClinicalTrials.gov, and the Cochrane Central Register of Controlled Trials through 23 April 2018. PubMed was searched in English using the following search strategy: ((breast OR

mammographic OR mammography) AND (density OR dense) AND (AI OR aromatase OR exemestane OR tamoxifen OR letrozole OR anastrozole)). The results were reviewed and extracted. The titles of all results were reviewed by a single reviewer (the candidate) and abstracts of articles read if deemed potentially relevant. ClinicalTrials.gov was searched for trials involving breast cancer and density; resulting titles and interventions were screened for (aromatase, letrozole, exemestane, anastrozole). The Cochrane Central Register of Controlled Trials was searched using the terms (breast density or mammographic density and letrozole or exemestane or anastrozole); resulting articles published online from late 2014 onwards were reviewed (because these complemented the most recent (2016) full systematic review in this area [307]). The grey literature and reference lists of extracted articles published (from all three databases) prior to late 2014 were not searched, as the prior full, multi-author formal systematic review [307] was expected to have uncovered any unpublished works. Articles containing reviews of AI and MD/BD were examined for additional studies. Studies were summarised in tables including the type of treatment given, numbers of participants in each treatment arm, and 'absolute' PD or BD change over time. A meta-analysis of RCTs of AI and MD [308], including an assessment of study bias and quality, would be undertaken if sufficient studies existed (i.e. 10 or more, Section 9.6.5.1 Cochrane Handbook (v5.1.0)).

2.7.3 Results

The search of PubMed yielded n=1479 articles titles for review; abstracts from 51 articles were appraised for suitability, yielding twelve AI and MD studies, one AI and (volumetric) BD study, and two systematic reviews. The search of ClinicalTrials.gov yielded 631 studies whose titles and interventions were screened; 86 study descriptions were subsequently searched for breast or mammographic density as an outcome. Three further studies were found. The search of the Cochrane Central Register of Controlled Trials generated n=2277 articles. Article titles of the subset of 850 studies published online from late 2014 onwards were reviewed. Twenty-five abstracts were read for relevance, but no studies not included in the PubMed or ClinicalTrials

database were found. One article [32] containing a review cited a further study published only as a conference abstract [309].

A summary of these studies are presented in Table 2-1, Table 2-2 and Table 2-3. Table 2-1 includes four PD single arm studies [30, 303, 310, 311], two PD case-control studies [304, 312], two PD cohort studies [18, 309] and five PD RCTs [28, 29, 31, 32, 313]. Table 2-2 lists a volumetric BD study [33], and Table 2-3 lists four publications which contain AI and BD reviews [32, 295, 307, 312] including two formal systematic reviews [295, 307]. Insufficient RCTs with evaluable results were found to warrant a meta-analysis (n=4).

Three unpublished trials (not tabulated) were found on the ClinicalTrials.gov database: a single-arm, multi-centre MD study of 140 women with BC treated with anastrozole or exemestane [314]; a single-arm, single centre study of ~400 women with BC treated with AI due for completion in 2018 [315]; and a multi-centre, case-control study due for completion in 2019 of ~3000 comparing health controls to women with BC treated with AI [316].

Eight studies in Table 2-1 were undertaken in women with hormone sensitive (hormone+) early BC (1 single-arm study, 1 case-control study, 2 cohort studies, 4 RCTs) [18, 28, 29, 31, 32, 309, 310, 312], and four were undertaken in high risk BC-free, non-HRT populations (2 single-arm studies, 1 case-control study, 1 RCT) [29, 30, 311, 313]. Most of these studies showed modest average reductions in MD (–1 to –3%) after one or two years of treatment with AI; it is notable some women (both AI treated and those on placebo) showed an increase in PD over time. One study published only as a conference abstract reported much higher visually estimated declines of ~6% with AI treatment, and ~8% with the combination therapy (AI + TAM) [309]. Two AI-only studies reported an association for greater declines in PD for women with higher baseline PD [31, 312]. One study (comprised of 62% TAM only, 16% AI only and 22% TAM then AI treated women) reported a significant association between higher PD and greater baseline PD,

as well as a significant difference in recurrence-free survival for women with a MD reduction of –10% or more compared to women whose reduction was <–5% [18].

Of the five randomised trials of AI and MD, two had active comparator groups (Henry et al [31] (2013), van Nes et al [32] (2014)) whilst three compared treatment with AI to placebo (Vachon et al [28] (2007), Cigler et al [313] (2010) and Cigler et al [29] (2011)). The van Nes study which compared exemestane vs tamoxifen followed by exemestane did not report evaluable PD outcomes; the Boyd SCC (six category classification [128]) utilised to assess MD is not precise enough to evaluate the small (–1 to –2% changes) in MD resulting from AI treatment. The other active comparator trial (Henry) of women treated with exemestane vs letrozole for hormone sensitive BC reported a ~2% decrease in PD in both groups at 2 years. This does not differ from the expected 1% annual decline in PD in the general population.

Of the three placebo controlled RCTs found, Vachon (2007) [28] reported effectively nil difference in (raw) average PD between control and treated groups at 1 year. PD adjusted by age, BMI, time on TAM, nodal status and number of tumours showed larger, but non-significant differences between groups:–1.0% (95% CI –2.7 to 0.7%) for letrozole and –0.3% (95%CI –2.1 to 1.4%) for placebo. Adjusted annual longitudinal change was –0.68% (95%CI –1.3 to –0.02%) for letrozole and –0.12% (95%CI –0.8 to 0.6%) for placebo (a non-significant comparison). The two Cigler RCTs [29, 313] also report non-significant average PD differences at 12 months of treatment with letrozole (n=~30) or exemestane (n=~34) vs control. Although non-significant, differences in average PD of –1% to –2% for AI vs placebo were reported at different time points, with large SD and 95%CI compared to the observed mean change (e.g –1.7% PD (SD 5.7, 95%CI –3.9 to 0.4%) for n=30 women treated with 12 months of letrozole). Because these studies each report similar but non-significant effects, this potentially indicates AI may reduce PD compared to no AI in a single trial with sufficient participants (e.g. 1000s of women). The small difference of –1% to –2%, however, would be

difficult to measure clinically given the variability in most MD measurement techniques and the expected –1% annual decline in PD due to aging.

Table 2-1 Raw (unadjusted) PD (%) change with AI treatment, by type of study

Author	Population	Treatment and sample size	PD (%) change (Δ) summary
	Single Arm studies		
Fabian et al 2007 [303]	Increased risk BC- women receiving HRT and 6 months letrozole (LET)	LET + HRT (n=42)	PD decreased by 2% at 6 months
Prowell et al 2011 [310]	Prospective single-arm cohort, hormone+ BC, 12 months anastrozole	Anastrozole (n=54)	Time point / Median PD / Mean(μ) ΔPD baseline (ref) Baseline 13.4 ref. 6 mo. 13.0 2 12 mo. 10.3 -16
Smith et al 2012 [311]	Prospective single-arm trial of high-risk BC- women, 12 months letrozole	Letrozole (n=20)	Time point / Median PD / Median ΔPD baseline (ref) Baseline 19.9 ref. 6 mo. 19.2 -1.4 12 mo. 16.2 -1.75
Gatti-Mays et al 2016 [30]	High risk BC- women treated with exemestane (EXE) for 2 years	Exemestane (n=35)	Time point / Mean PD / Median ΔPD baseline (ref Baseline 32.5 ref 12 mo * -2.4 -3.4 24 mo * -4.1 -2.8 *PD difference from baseline
	Case-control studies		
Mousa et al 2008 [304]	BC- women receiving low dose HRT with/without letrozole	LET + HRT (n=28) HRT alone (n=28)	Significant lowering of volumetric PD in LET treated women. Radiologist able to visually detect changes with median (volumetric) change of 27.9%, but unable to for median change <11.3%.
Vachon et al 2013 [312]	Women with early hormone+ BC from trials MA.27, N0631 or MC0532; matched to healthy screening controls	AI (n=369) Exemestane 52% Anastrozole 48% Control (n=369)	Time point / AI Mean PD / No Tx Mean PD Baseline 17.8 16.2 1.1 years * (median time) -1.3 -1.1 *PD difference from baseline
	Cohort studies		
Kim et al 2012 [18]	Retrospective cohort of women with TAM and/or AI treated hormone+ BC	TAM → AI (n=233) 5 years AI (n=175)	AI subgroup: -3.1% PD reduction after an average of 13 months. Women (TAM and AI treated) with reductions in MD ≥10% experienced significantly lower rates of recurrence compared to women whose % PD change was <–5%.
Cuzick et al 2006 [309] reported in [32]	128 of 1000 planned BC+ women from the UK enrolled in the ATAC (Anastrozole, TAM or combination) trial (nested cohort)	Anastrozole (53) TAM (46) Combination (46) Film MLO mammograms	Overall (all groups) 5.8% reduction Anas: 6% reduction TAM: 3.5% reduction Combination: 8% reduction No significant differences between groups Visual assessment by radiologist

Author	Population	Treatment and sample size	PD (%) change (Δ) summary
	Randomised Trials		
Vachon et al 2007 [28]	MA.17 RCT (5 years Letrozole vs placebo post 5 years of TAM for hormone+ BC)	Letrozole (n=56) Placebo (n=48) from 2 MA.17 centres	Time point: Letrozole N 35, PD 18.5; Placebo N 33, PD 20.0 12 mo.: Letrozole N 35, PD 16.9; Placebo N 33, PD 19.0 ΔPD 12mo.: Letrozole N 35, PD 0.8; Placebo N 33, PD -0.6
Cigler et al 2010 [29]	High risk BC- women (50%) and BC+ women (50%) with >25% PD randomised 2:1 to 12 months of letrozole or placebo (MAP.1 trial)	Letrozole (n=44) Placebo (n=23)	Time point: Letrozole N / PD; Placebo N / PD Baseline: 31 / 39.6; 19 / 40.0 12 mo.*: 30 / -1.74; 19 / -0.24 24 mo.*: 27 / -0.01; 16 / -1.32 *PD difference from baseline
Cigler et al 2011 [313]	High risk BC- women with PD ≥ Boyd SCC 2-6, randomised to 12 months of exemestane or placebo (MAP.2 trial)	Exemestane (n=49) Placebo (n=49)	Time point: Exemestane N / PD; Placebo N / PD Baseline: 38 / 33.9; 34 / 36.5 6 mo. *: 36 / -1.33; 33 / 0.22 12 mo.*: 34 / 0.56; 31 / 0.58 24 mo.*: 24 / -0.17; 19 / -2.93 *PD difference from baseline
Henry et al 2013 [31]	ELPh trial of Exemestane vs Letrozole Pharmacogenomics for women with hormone+ BC; 2 years AI treatment	Exemestane (n=120) Letrozole (n=139)	Time point: Exemestane PD / ΔPD; Letrozole PD / ΔPD Baseline: 16.1 / ref; 18.0 / ref 24 mo.*: 14.3 / -1.7; 15.9 / -2.1 Time point: Exemestane DA / ΔDA; Letrozole DA / ΔDA Baseline: 2577 / ref; 2809 / ref 24 mo.*: 2405 / -172; 2185 / -187 Dense Area (DA) in mm^2
van Nes et al 2014 [32]	Women from the TEAM trial randomised to 5 years EXE only or TAM→EXE for early hormone+ BC	EXE only (n=197) TAM then EXE (n=181); per protocol analysis	Three assessors utilised the Boyd SCC (PD categories: 0%, 1–10%, 11–24%, 25–49%, 50–74% and ≥75%). No difference was found between baseline, 1 year, 2 years and 3 years of follow up. Change in PD was not significantly associated with BC recurrence.

A summary of the AI treatment results from a recently published volumetric MD and endocrine treatment study [33] is presented in Table 2-2. The study compared the rates of change in volumetric density between BC cases treated with tamoxifen or AI therapy and healthy controls from the general screening population. Adjusted annualised change in volumetric percent mammographic density (VPD) for AI treated cases relative to controls was –0.30%/year for Volpara measured VPD and –0.58%/year for Quantra measured VPD (both significant, $p<0.05$, F-test). Adjusted annualised change in VPD for cases and controls with baseline VPD <10% were not significantly different; however AI treated women with baseline VPD ≥10%

experienced significantly greater declines in adjusted annualised change in VPD compared to controls with baseline VPD ≥10% ($p<0.05$, F-test).

Table 2-2 Volumetric mammographic density change with AI treatment

Author	Population	Treatment	Volumetric MD change summary
	Case-control study		
Engmann et al 2017 [33]	Women with hormone+ BC receiving AI treatment; matched to healthy screening controls	Anastrozole (n=403) Controls (n=1618)	Measurement technique: Adjusted annualised rate of change* for AI treated relative to controls — Volpara: VPD (%) -0.30, DV (mm^3) <-1; Quantra: VPD (%) -0.58, DV (mm^3) ~-3. * Median time between mammograms was 3 years for AI treated cases and 2 years for controls; DV dense volume (mm^3)

Four studies containing reviews of BD and AI treatment, including two formal systematic reviews [295, 307], were found during the search for studies of AI and MD/BD (Table 2-3). The reviews indicate heterogeneity amongst the AI and MD/BD studies, with inconclusive results. However, no further PD (MD) and AI studies have been published since the most recent formal systematic review [307] was undertaken, hence the candidate did not perform an additional, formal systematic review as little could be contributed to the literature.

Table 2-3 Systematic and other reviews of MD/BD and AI treatment

Author	Summary description	Results
	Articles containing non-systematic reviews	
Vachon et al 2013 [312]	9 studies through 2013 are listed in Table 1 of a case-control study matching women with early hormone-sensitive BC from three trials	Authors note the results of their study (no effect of AI on MD) are consistent with 5 of the 9 other studies, however 4 studies did find significant effects. The 9 studies are heterogenous.
van Nes et al 2014 [32]	8 studies of AI and/or TAM through 2014 are listed in Table II of a randomised controlled trial (TEAM trial) article which measured MD using the Boyd SCC	Authors report the TEAM trial results of la small or no effect of AI on MD consistent with other AI studies.
	Systematic reviews	
Lienart et al 2014 [295]	Qualitative systematic review of AI, TAM and raloxifene's (RLX) effects on BD. Seven AI studies were assessed (n=416 participants).	TAM reduces BD, however the effects of RLX and AI are unclear due to conflicting results among studies
Ekpo et al 2016 [307]	Systematic review of AI and BD (n=10 studies), as well as studies reporting the association between BD and SERMs (n=26 studies) and physical activity or diet (n=55 studies).	TAM reduces BD, however the effects of RLX, tibolone and AI on BD are unclear.

2.7.4 Discussion

Most studies showed small average (<–2%) and non-significant PD changes per annum in response to treatment with AI. This small change is difficult to detect in individual women due to the variability in the commonly used (semi-automated) PD measurement techniques, but may be possible to discern with the more modern fully automatic (volumetric BD) methods available [317].

The limitations of this study include lack of a complete (thorough) systematic review of the literature, including the grey literature, conference proceedings and abstracts, and contact with the authors of existing publications to enquire about other possible unpublished studies. The review was undertaken in English only, potentially limiting the studies available for review, given that MD research is rising in popularity outside of the USA, Europe and Australia. The studies were not assessed for quality, beyond separation into study type (e.g. single arm vs case-control vs cohort vs RCTs), partly because too few studies existed to perform a meta-analysis.

Evidence of a potential publication bias exists, as two planned MD studies found during the literature search have not been completed and published [309, 314]: one is presented only in abstract form (from 2006 [309]), whilst the other is listed as completed in 2009 on ClinicalTrials.gov [314], but no publications are listed on the trial listing and a web search using the trial unique identifier did not find any publications. Further planned and/or partially completed AI and MD and/or BD studies may be found if searches of relevant sources are undertaken, e.g. conference proceedings such as: SABCS (San Antonio Breast Cancer Symposium), Why Study MD (Australia), the San Francisco Breast Densiometry Workshops), or searches of other trial registries.

More research is needed to clarify the relationship between AI and BD. It is still unclear if change in PD and/or volumetric BD are significantly associated with AI treatment compared to no treatment, or if the three commonly prescribed AI treatments differ from each other in their

effects upon BD (although the latter seems unlikely). The relationship between AI and BD is important to know at the population level to determine if the mammograms acquired during routine care can be used to acquire information on treatment efficacy. It is also important for risk prediction and treatment efficacy assessment for individual women undertaking AI therapy for BC treatment and/or prevention.

Clearly what is also needed, however, is a precise method with which to measure longitudinal MD and/or volumetric BD for this to have clinical relevance given the treatment effect vs measurement variability is high; this is especially true for measuring BD during AI treatment because these effects tend to be $|\leq 10\%|$ for individual women [28, 311]. Although three studies involving TAM treatment have shown significant improvements in BC-free survival are associated with strong reductions in PD (<–10% +) [17, 18, 297], no AI-only study has successfully measured change in MD/BD in association with BC survival/recurrence to assess: 1) if an association exists; and 2) if this association differs for example by baseline PD, age, and/or the strength of the measured change from baseline. Furthermore, it is unclear why some women on AI (or other endocrine therapy & HRT) experience increases in PD on treatment, whilst others remain the same or decline. The difference in MD response to hormonal treatment is likely attributable to genetic (and possibly environmental) difference between women; additional research is also required to help clarify why women's responses differ to hormonal therapies. Similarly, the physiologic and metabolic pathways by which AI (& other hormone therapies) affect MD are not known; elucidation of these may explain some of the reasons why certain women respond to AI more strongly than others, as well as shed light on treatments which may help improve AI/endocrine treatment response in non-responders.

2.8 Other associations with mammographic density

The associations between MD and other factors is broad and varied. These factors include breast cancer tumour pathology, BC mortality, compressibility and firmness of the breast,

physical activity and diet.

Although some studies have found positive associations between ER+ tumours and highly dense breasts [318-320], a recent meta-analysis showed little association between ER tumour status and MD [321]. Luminal A and/or luminal B type cancers may develop more often in highly dense breasts [322, 323]. Racial differences for breast cancer subtypes vs MD may also exist [324], suggesting the presence of further links between race, heredity and mammographic density. Further research is needed to ascertain if breast density is related to particular subtypes of breast cancer.

Although high MD is associated with larger tumour size at diagnosis (i.e. worse prognosis), high MD is not associated with overall increased risk of death from BC [325]. After adjustment for tumour size and the typical MD modifiers age, BMI and HRT use, interval (non-screen detected) cancers in women with low density breasts (<25%) had significantly worse 5 year survival than women with higher density (≥25%). Low or very low (<10%) MD may be a prognostic factor for poorer survival compared to higher MD [326].

However, women with high MD who do not undergo radiation therapy may be at higher risk of both BC recurrence and death [327, 328]. This association is mitigated by radiotherapy: women with high MD who receive radiotherapy treatment for BC have improved survival rates [328]. Mammographic density does not appear to be associated with breast size, compressibility upon mammography, nor thickness of the breast after compression [329]. Physical activity does not appear to change absolute breast dense area greatly [181]. Because BMI is strongly and inversely associated with PD— and BMI tends to be inversely associated with physical activity— relationships exist between physical activity and mammographic density [330, 331]. However, the likely protective effects of exercise against breast cancer do not appear to be modulated through mammographic density [139, 331].

Diet has an effect upon cancer, and may also have associations with breast density [179, 307, 332, 333], though levels of serum and dietary cholesterol do not appear to have an association with mammographic density [334]. It is not known why women residing in Asian countries such as Japan and China have highly dense breasts but low rates of BC compared to their Western counterparts.

2.9 Mammographic density and BC risk modelling

MD is a very common BC risk factor— breast density is classified as high to very high in a significant proportion of women. Approximately 30% of premenopausal women over age 40 and ~13% of post-menopausal women in a meta-analysis of MD studies had highly dense breasts [136]. About 43% of the US screening population has dense to very dense breasts; this equates to over 25 million women [335]. Women aged 40 to 49 years account for approximately 45% of the US women with dense breasts.

Other moderate (1.5 to 2.5 fold) risk factors are also common (late age at first birth, nulliparity, fewer months of breast feeding). The interactions between MD and other BC risk factors are not fully known [243, 336-342]. It is not clear whether certain women with high MD and specific combinations of common BC risk factors are at higher risk due their MD compared to other women with high MD who have a different combination of BC risk factors.

Area under the curve (AUC, the 'c-statistic') is a method used to assess how well a predictor, or set of predictors, is able to forecast an event (outcome) will occur. An AUC of 50% (0.5) indicates a predictor is no better than chance at forecasting the outcome. An AUC of 100% (1.0) demonstrates the predictor/s are able to accurately predict an outcome will occur 100% of the time. Some routinely used clinical prediction models have modest (AUC 0.6 to 0.8) ability to discriminate between those who will develop an outcome and those who will not [343].

Modelled on its own as a risk predictor, MD has an AUC of ~0.55 to 0.6 to predict the risk of developing BC [16, 344, 345].

Popular clinical models used to predict BC risk for the general (average lifetime risk) population include the Gail, Tyrer-Cuzick and Claus models [346-349], though many others exist [350, 351]. These models incorporate many of the risk factors associated with BC such as age, age at first birth, age menstruation commenced, personal history of breast biopsy and benign breast disease, and family history of breast and ovarian cancer. One of the limitations of the Gail and Claus models is that they do not take into account family history from both sides of the family (mother's and father's sides of the family). The Tyrer-Cuzick model does include both parents' family histories, but is more complicated therefore to implement. For women at high risk— i.e. women with familial BRCA mutations or strong family history of breast and ovarian cancers— other more specific models tailored to estimate risk in these populations are utilised clinically such as the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) [352, 353].

The addition of MD to existing risk models yields only modest to moderate (AUC +0.01 to +0.1) gains to BC prediction at the population level [16, 89, 344, 351, 354, 355]. Area of dense tissue has also been investigated for its utility to improve risk modelling. A 'density residual' (percent dense area, adjusted by age and BMI) was added to the Tyrer-Cuzick model to yield a substantial improvement in risk estimation for a population of high risk women (AUC improved from 0.51 to 0.62) [356]. However, when a similar density residual was added to the Tyrer-Cuzick and Gail models to estimate risk in a general screening population, more modest improvements in AUC were found for both models (+0.04 AUC improvement, yielding an AUC of 0.61 (T-C) and 0.59 (Gail)) [345].

The more modest improvement in risk estimation between the population may be a case of

'overfitting' the density residual parameter to the modelled population, or it could represent a difference in risk phenotype between a general screening population and a high-risk cohort of women. Women at high risk due to family history tend to have higher MD than the general population. MD variability may translate to a higher difference in risk for high-risk women than for those in the general population. Further testing of the density residual's effectiveness in predicting BC in high-risk populations, e.g. the National Surgical Adjuvant Breast and Bowel Project (NSABP) Prevention-1 (P-1) cohort of high risk women, as well as other screening populations is warranted.

The BCSC (Breast Cancer Surveillance Consortium) risk model was developed using common BC risk factors including BI-RADS density [343]. The BCSC model yielded an age-adjusted AUC of ~0.62 in the general screening population, similar to the AUCs for the Tyrer-Cuzick model with the density residual. The addition of genetic markers (76 single nucleotide polymorphisms (SNPs), increased the AUC for the BCSC model from 0.66 to 0.69. The BCSC model was developed for the geographically and ethnically diverse USA population; further testing of the model in populations within and outside the US is warranted.

Because MD is related to many other risk factors such as BC family history and age, the additional contribution to risk beyond the factors used in the models is small even though MD is a strong BC risk factor. As described by Howell et al [89], due to correlations amongst most BC risk factors, additional markers are likely to only modestly improve models which already contain other risk factors.

For a single risk factor to be useful for screening, very large relative risks (>100x)— preferably with different population variances (variability) between those with the risk factor and those without— are needed to distinguish between low and high risk [343, 357]. Even a strong risk factor with a RR of 3 (with similar variances for the groups with and without the risk factor) provides limited separation of the populations (Figure 2-12, below).

This limited separation of the curves for MD (RR~3) is one of the reasons that AUC generally show limited improvement with the addition of MD to risk models. Theoretically however, a model containing enough risk factors will provide sufficient information to distinguish women who will develop BC during their lifetime, even if many of these risk factors might individually contribute only a small amount to the ability of the model to differentiate risk.

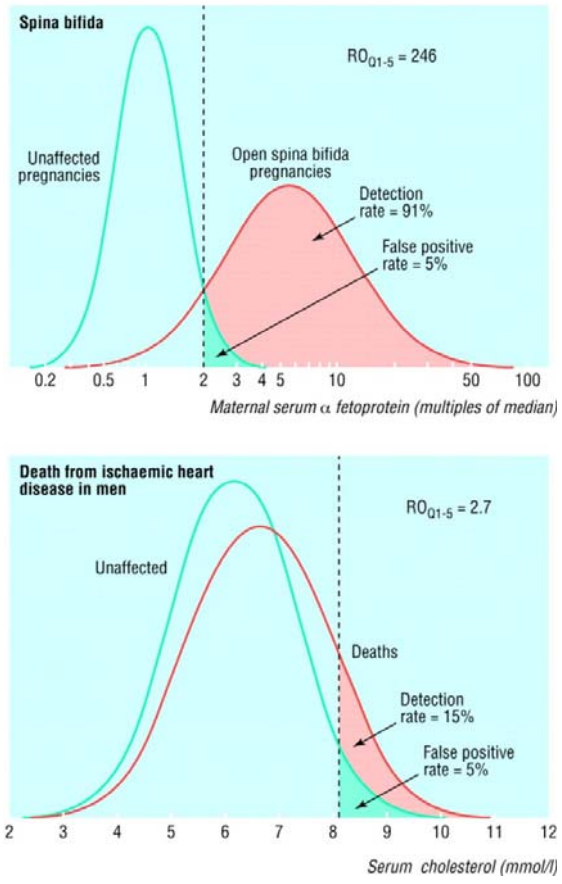


Figure 2-12 Odds Ratio distributions for extreme and typical strong risk factors
Sourced from Wald et al 1999 [357]

Upper graph: Maternal serum α fetoprotein (multiples of the median) has a large relative odds (RO_{Q1-5}) of 246 and different population variances between babies which have spina bifida and those that do not. The value 2 correctly identifies 91% of babies with spina bifida, whilst allowing for a 5% false-positive rate (incorrectly identifying 5% of babies who do not have spina bifida).

Lower graph: Typical probability curves for a normally distributed risk factor with equal variances which has a relative odds (RO_{Q1-5}) of ~3 between the groups with and without the risk factor. These curves illustrate how only 15% of the group with ischaemic heart disease are correctly identified by a serum cholesterol level of 8 mmol/l or greater, which also incorrectly identifies 5% of men without ischaemic heart disease (5% false positive rate).

Abbreviations: RO_{Q1-5} Relative odds (odds ratio) of risk between the lowest 20% of the population (Q1) and the highest 20% of the population (Q5)

PD may be an over-simplification of the risk information within the mammographic image. The different patterns of density and other features on breast images may provide additional features to discriminate between higher and lower risk women [124, 158, 358-360]. Whether MD presents the same biomarker for BC risk in all women is not known [336], partly because all BC risk models have a limited capacity to predict cancer at the level of the individual [89, 361]. This may be in part due to the inherent limitations of modelling.

Epidemiological studies can only examine a certain number of risk factors (e.g. smoking, BMI, alcohol intake) at any one time due to sample size and power considerations, as well as time and financial constraints. BC is a common cancer, but is still a rare event at the population level: nearly 90% of women will not develop BC within the 'lifetime' risk of 80 years. Very large numbers of women (thousands) are required to study BC incidence within a population, as exemplified by the need to recruit thousands of high-risk women for the IBIS-I and IBIS-II breast cancer prevention trials. These large sample sizes are required to include multitudes of risk factors in these studies and still achieve reasonable statistical power, the ability to distinguish differences between groups with different combinations of risk factors, treatments, etc.

However, many of the current limitations in BC risk prediction likely result from yet undiscovered risk factors and biomarkers for these risks. Newer research techniques are continually being created to further investigate biological markers and other factors which may affect cancer risk. The incredible complexity of the body and the unobserved interaction of the proteins, genetic material and other chemicals within an organism makes modelling these inter-relationships extremely difficult (even if all were known). As described above, the more elements which are included in a model the larger the sample size (e.g. number of women) needs to be to be able to distinguish among groups, making epidemiological studies all the more difficult. The large research consortia required to search for genetic identifiers (e.g. SNPs) using genome-wide association study (GWAS) techniques are another example of this [362, 363].

Additionally, use of alternative methods of deriving the strengths of risk factors during modelling — such as OPERA (odds per adjusted standard deviation), a way of describing the "absolute change in risk per standard deviation of its residuals after adjusting for" other parameters in a model [364]— may help to better quantitate the contribution of each risk factor and increase the ability of models to predict both individual and population risk

Generally speaking, the 'low hanging' (easily picked) fruit associated with BC risk have already been discovered, such as age and the BRCA genes. However, finding the BRCA genetic markers was very challenging using the techniques and information available at the time they were discovered. Ongoing discovery of additional markers and techniques with which to model risk and improve cancer prediction will assist with improved risk prediction [365]. Prediction models may therefore become extremely complex in order to reliably foretell which women will develop BC. It may then be possible to undertake even more closely tailored interventions to prevent BC. Early detection of BC via as yet unknown biomarkers may be a viable alternative pathway to improved health outcomes if highly effective treatments with minimal side-effects are developed.

2.10 Summary of factors that are associated with mammographic density

To better understand the etiology of mammographic density, the exploration of the relationship between breast cancer risk factors and mammographic density has been intensely studied. Many— but not all— risk factors which are associated with breast cancer are also associated with mammographic density. In particular, the breast cancer risk factors age, menopausal status, height and weight (BMI), parity, and heredity are all strongly associated with percent mammographic density. Mammographic density usually decreases with age, after menopause and with each birth; the protective effects of these events against breast cancer may be modulated via breast density. Other breast cancer risk factors such as duration of breast feeding, age at first birth and age at menopause have less marked effects upon breast density and breast cancer risk, but these BC risk factors may also be mediated via breast density

It is necessary to collect information on the various factors that affect mammographic density because they can have a marked effect upon the interpretation of the BC risk posed by breast density [192, 198, 203]. In particular, age, menopausal status and BMI (body mass index) can

confound research results. For instance, a breast which is half covered dense tissue (50% dense) may not be uncommon at 40 years of age, but as seen in Figure 2-9, it is rare to find women aged 55 and over with highly dense breasts. 50% density at age 40 is far less likely to be strongly associated with breast cancer risk than it is at age 60.

The effects of age on breast density observed at the population level may be mainly mediated through decreases in breast density experienced by most women due to parity, breast feeding, declining hormone levels during the peri-menopausal period as well as the rapid decline in endogenous hormone levels during the menopausal transition. The effects of body mass index on percent mammographic density are easier to explain. BMI is inversely associated with percent mammographic density because as BMI increases, additional fat accumulates in the breast without a concomitant increase in the amount of dense tissue. Both age and BMI are strongly and negatively associated with mammographic density; models utilising percent mammographic density as the dependent variable should control or account for both these factors at a minimum [178]. Whilst breast density in most women decreases with age, certain factors appear to keep density levels elevated in some women. Despite the combined effects of age, child birth, breast feeding and menopause, some women appear to retain most of the breast density they first developed at puberty. The biology underpinning the loss of breast density in most women as they age, as well as why density appears to be retained in others, is not yet known [366].

Breast density is also modifiable in many women via hormonal manipulation. Therapies such as HRT which increase estrogen tend to increase MD, as well as BC risk. Therapies such as tamoxifen, anastrozole and GnRH inhibitors which decrease the availability of estrogen tend to reduce MD, as well as BC risk. Whether the women who show strong responses to hormonal treatments also exhibit marked changes in MD due to age, child bearing, breast feeding and

menopause is not yet known. Further research into the biology underlying radiologically dense breasts is needed.

The next chapter describes the methods for this project, which include a description of the IBIS-II breast cancer prevention trial, ethics approvals, methods for mammogram collection and deidentification, as well as an introduction to the statistical methods used in this thesis.

3. Methods used in this Thesis

The majority of this thesis involves mammographic data from the IBIS-II trial. This chapter describes the IBIS-II trial and trial outcomes. The statistical methods utilised in the thesis are introduced. The ethics approval and mammogram collection processes are described. The digitisation and anonymisation process utilised on the IBIS-II mammograms is explained.

Additional, potential projects involving other Australia and New Zealand Breast Cancer Trials Group (ANZ BCTG, now known as 'Breast Cancer Trials') studies were discussed with the candidate's principal supervisor, however most of the projects were not deemed practical for various reasons including time elapsed since the trial was undertaken, and too few local (Newcastle) participants in the trial. Two additional projects were undertaken by the candidate in addition to the IBIS-II study reported in this thesis: 1.) an MD linkage study collecting data from healthy screening women attending NSW BreastScreen, the 45 and Up Study and the NSW Cancer Registry; and 2.) a longitudinal MD project utilising trial data and mammograms from Western Australian IBIS-I participants. At this time, data from NSW BreastScreen was not linked with the NSW Centre for Health Record Linkage (CHeReL); the candidate worked closely with a Senior Research Officer at Cancer Institute NSW to ascertain which fields in the BreastScreen database would provide useful information during linkage. The first project (BreastScreen and 45 and Up Study linkage) was taken to the ethics approval stage but was abandoned due to lack of funding, whilst the second received ethics approvals but was not able to collect sufficient mammograms to provide analysable data.

3.1 IBIS-II Prevention trial methods

IBIS-II study rationale

The International Breast cancer Intervention Study II (IBIS-II) trial of anastrozole vs placebo was created to continue the work of the IBIS-I trial (BC prevention with tamoxifen vs placebo).

In IBIS-I, 7,154 breast cancer-free women from Australia, New Zealand, the UK and Europe aged 35 and up at elevated (~2-fold) risk of breast cancer were randomised to five years treatment of tamoxifen (TAM) or placebo. At 96 months median of follow up, IBIS-I participants treated with tamoxifen had an overall 27% reduction in BC incidence: (relative risk (RR) 0.73, 95%CI 0.58 – 0.91)[367]. Most undesirable side-effects and medically important adverse events from TAM— such as hot flushes, deep vein thrombosis and pulmonary embolism— ceased at the end of the 5 year treatment period, however the reduction in BC continued for at least another 5 years after the end of treatment.

The third generation aromatase inhibitors (AI)— anastrozole, exemestane and letrozole— have favourable side effect profiles compared to TAM and other serum estrogen receptor modulators (SERMs). These AI were first shown to be effective in the treatment of hormone sensitive (HS+) advanced (metastatic) BC [95]. Head to head trials of TAM vs AI for adjuvant treatment of HS+ early BC show improved disease free survival for the AI treated groups [98].

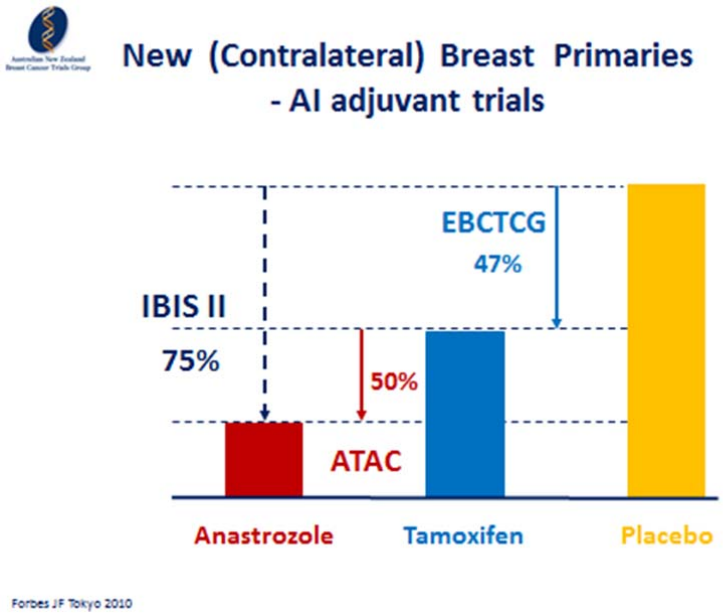


Figure 3-1 Expected reduction in ER+ BC incidence by anastrozole, IBIS-II trial [JF Forbes]
Contralateral estrogen receptor positive (ER+) BC incidence during treatment trials was reduced by half with TAM vs placebo, and halved again by treatment with anastrozole. EBCTCG—Early BC Triallists' Collaborative Group. ATAC—Arimidex, Tamoxifen, Alone or in Combination. Sourced from JF Forbes, 2010 [368]

In particular, incidence of contralateral ER+ BC was approximately halved in the AI treated participants compared to the TAM treated participants [369]. Five years of TAM reduces the risk of contralateral ER+ BC by 50% compared to placebo treatment. A further 50% reduction in ER+ BC with five years of preventive AI treatment was predicted, compared to women who did not undergo therapy [Figure 3-1]. This 75% estimated reduction in HS+ BC, along with the favourable side effect profile, supported the use of AI for BC prevention.

IBIS-II Aims

The IBIS-II (Prevention) trial is an international multi-centre trial of anastrozole vs placebo in post-menopausal women at elevated risk of developing breast cancer. Given the high desirability of preventing BC and improved the efficacy and side effect profile of AI as an adjuvant treatment for hormone-sensitive (HS+) BC, IBIS-II was developed to further investigate the use of anti-estrogen treatment in the prevention of HS+ BC.

The overall aim of the IBIS-II trial was to determine if anastrozole was effective at prevention of breast cancer in women at elevated risk of the disease. Secondary aims were: to examine the role of anastrozole in preventing ER+ BC and BC mortality; to examine the effect of anastrozole on other cancers, cardiovascular disease, fracture rates, and non-breast cancer death; and determine the tolerability and acceptability of side effects from trial treatment.

Eligibility Criteria

Because AI are only effective in post-menopausal women, IBIS-II participation was restricted to post-menopausal women (e.g. over age 60, or 12+ months of amenorrhea with an intact uterus). To enhance the risk to benefit ratio, only women at elevated (~2-fold) risk of BC were eligible for the trial.

The elevated BC risk required to enter the trial differed by age. Women aged 45 to 70 were

eligible if they were at double the risk of BC compared other women their age (e.g. a first degree relative with BC diagnosed at age <50). Women aged 40 to 44 needed to have approximately 4-fold the average risk of developing BC, such as two or more first or second degree relatives who developed BC or ovarian cancer at age ≤ 50. Women aged 60 to 70 are already at elevated risk due to age, hence women in this age group could also enter the trial with slightly lower risk (~1.5RR, e.g. a first degree relative with BC diagnosed at any age) than their younger counterparts. Women aged 40 to 70 at higher risk due to certain breast conditions, such as lobular carcinoma in-situ (LCIS) or benign atypical or lobular hyperplasia, were also eligible to enter the trial.

Interestingly, mammographic density covering at least 50% of the breast area (in the absence of HRT use) for women aged 45 to 70 was included as an entry criterion. However, this eligibility criterion was rarely utilised. This was most likely due to the difficulty in obtaining information about PD by potential participants, as well as low general awareness (both in the general and medical communities) about the status of MD as an independent BC risk factor.

Trial exclusion criteria included pre-menopausal status, most cancers, an intention to continue HRT, osteoporosis, and other conditions which made the candidate unsuitable for trial participation including gluten sensitivity. Previous SERM use for >6 months was also a trial exclusion criteria, except for IBIS-I participants (whose treatment status remains blinded).

In addition to the trial inclusion criteria mentioned above (post-menopausal status, elevated BC risk) participants were required to have a normal (BC-) mammogram within 1 year of trial entry as well as spinal and bone mineral (DXA) scans to prove they did not have osteoporosis. Signed, fully informed consent was obtained. From June 2007, a protocol amendment allowed participation for IBIS-I participants if they met all trial entry criteria and had ceased IBIS-I trial treatment at least 5 years previously.

Trial Activation and Recruitment

International activation of the trial commenced in January 2003. Originally, a trial accrual of 6000 participants was estimated to detect a difference in BC incidence of 50% between the arms of the trial, at the 5% significance level with 95% power. However, the 2011 interim observed incidence of 6.6 BC events/1,000 women in IBIS-II was higher than the 6 BC events/1,000 women projected as per IBIS-I. A protocol amendment in September 2011 subsequently included a recalculated target sample size of 4,000, with a minimum sample size of 3,500 women, to have >90% power to detect the required difference between the arms at the 5% significance level.

Twenty-six IBIS-II (Prevention) centres were activated in Australia, as well as four in New Zealand. The majority of the Australian centres were in the states of New South Wales (13) and Victoria (8), however two centres were activated in Queensland, and one each in South Australia, Tasmania and Western Australia. Recruitment of Australian participants occurred primarily in response to media releases (via print (e.g. newspapers), radio and internet) coordinated by the ANZ BCTG and participating hospitals. Interested women phoned an Australia-wide 1-800 number to contact the ANZ BCTG in Newcastle or their local centre. Trial enquiries received by the ANZ BCTG were forwarded to the nearest centre activated for IBIS-II.

As of the trial accrual closure on 15 February 2012, a total of 818 women were randomised at ANZ BCTG participating centres. New South Wales (NSW) accrued the largest number of participants in Australia (319), approximately half of which (142) were randomised at the Calvary Mater Newcastle hospital (CMN). The next largest accruing centre in NSW was The Breast Centre, Gateshead with 37 participants. As for IBIS-I, Sir Charles Gairdner Hospital (SCGH) in Western Australia accrued the highest number of IBIS-II participants (153). One hundred and eight IBIS-II participants were randomised in Victoria, which had the third highest

accruing Australian centre (44 participants at St Vincent's Hospital, Melbourne). One hundred and ninety-one participants were randomised in New Zealand; almost half of the NZ participants (90) were randomised at Waikato Hospital.

Informed Consent, Data Collection, and Randomisation

The fully informed consent included an explanation about the trial entry and exclusion criteria, pre-trial entry procedures (mammography, DXA bone mineral density scans), medication (1 mg/day anastrozole tablet or placebo tablet for 5 years) and follow up clinic visits (at 6 months, 12 months, then yearly in years 2 to 5). Annual clinic visits or posted questionnaires were undertaken in years 6 to 10.

Subjects were randomised to receive either five years of anastrozole 1mg daily or an anastrozole placebo. Both the participant and the clinical teams at participating centres are blinded to the participants' treatment allocation (double blind trial). Randomisation was stratified by centre, with retrospective stratification by risk factor taking place at the time of analysis. Blocks were used to maintain balance between the groups, and randomisation occurred centrally via a randomisation centre in the UK.

Mammographic data: Trial mammograms were taken annually. The four standard mammographic Views (right craniocaudal (CC), left CC, right medio-lateral oblique (RMLO) and left MLO) were collected at each mammographic episode. An additional, optional consent for permission to use trial mammograms for breast density studies by the ANZ BCTG and Cancer Research UK was requested during the reconsent process undertaken by all participants after a September 2011 protocol amendment.

Australian imaging centres underwent the transition from analogue (film) to digital mammography during the recruitment and follow up period of the IBIS-II trial.

Study endpoints: The primary endpoint for the trial was development of histologically-proven BC (either invasive or non-invasive (DCIS)). Secondary endpoints were mortality from BC, total mortality, and cause-specific mortality.

Adverse events: Anastrozole has been in clinical use for the treatment of BC since 1995 and is generally well tolerated. The following very common (≥10% prevalence) and common (1% to 10% prevalence) 'side effect' adverse events (AE) were collected at the 6 month and 12 month visits, and yearly thereafter to 10 years: arthralgia, hot flushes/night sweats, vaginal changes, eye disease/cataracts, osteoporosis/fracture, and 'other' side effects. Grade of severity (none, mild, moderate, severe) was also recorded. Potentially life-threatening, primarily uncommon (<1% prevalence), medically serious AE (SAE) were also recorded for each visit: myocardial infarction, cardiovascular events, thrombo-embolic disease, stroke/transient ischaemic events/cerebrovascular events, gynaecologic events, cancer other than breast, and 'other' SAE.

Follow up is still underway, with an anticipated completion date of February 2022.

Ethics and Governance

Ethics approval was first received from North West Multicentre Research Ethics Committee in the UK. Each participating Australian and New Zealand centre received approval for the study from their local human research ethics committee from June 2005 onwards. Activation of the IBIS-II trial in 2005 pre-dates the use of the 'harmonised' NHMRC single-ethics review and approval process in Australia. Hence each participating centre needed to receive approval for original IBIS-II protocol as well as all six IBIS-II protocol amendments from their respective ethics committees. Each centre also underwent local review of trial governance— a site specific assessment— to ensure that the resources needed to undertake the trial (e.g. principal investigator, trial coordinators, pharmacy staff, space to house trial data and meet with participants) were available and that the trial complied with local regulatory requirements (e.g. trial indemnity/insurance, trial agreement contracts, responsible research conduct). Participants whose mammograms were utilised in this thesis were recruited from the Calvary Mater

Newcastle (CMN) hospital. CMN received ethics approval for the IBIS-II trial from the Hunter New England Human Research Ethics Committee (HNE HREC) in June 2005.

The September 2011 protocol Amendment mentioned above also included a reconsent request to all IBIS-II participants. Reconsent to continue participation in the IBIS-II trial was requested due to the positive results of the MAP.3 BC prevention trial of exemestane vs placebo. MAP.3 trial participation was unblinded at a median of 3 years of follow up; MAP.3 participants randomised to placebo were allowed to crossover to five years of exemestane treatment. The IBIS-II trial steering committee considered the MAP.3 results too early to be definitive. IBIS-II participants were advised of the MAP.3 results and asked to reconsent to continued participation in the IBIS-II trial.

This IBIS-II trial reconsent process allowed participants agree to optional consent for mammographic density studies and optional consent for blood and tissue biomarker studies. ANZ BCTG participants were asked for their permission to utilise their trial mammograms for (ethically approved) studies at both the ANZ BCTG and CRUK.

All trials undertaken by the ANZ BCTG are investigator-led; whilst the trial funding may be provided by pharmaceutical companies, the trial design, implementation, data collection, analysis and publication are the responsibility of the ANZ BCTG study chairs and other investigators. Responsibility for trial conduct is undertaken via three levels of management. Each centre's principal investigator is responsible for the conduct of the trial at their centre. An ANZ BCTG trial study chair is responsible for trial conduct at all ANZ BCTG centres. Because IBIS-II is an international trial, three international study chairs are also responsible for coordination and implementation of the trial worldwide, via one or more of each local country's (investigator-led) oncology research organisation/s such as the ANZ BCTG and the German Breast Group (GBG)).

3.1.1 IBIS-II trial results

The study closed to recruitment on 15 February 2012. Total international accrual to the study was 3864 participants: 1920 in the Anastrozole group and 1944 in the placebo group. Final accrual at all ANZ BCTG centres was 818 participants, including 142 at the Calvary Mater Newcastle hospital.

Selected baseline characteristics for all 3864 participants are shown in Table 3-1. Baseline characteristics not shown in Table 3-1 were also balanced between the treatment groups. For example, the number of first or second degree relatives who developed BC by a certain age— e.g. a first degree relative with BC at age ≤ 50 years— were similar for both treatment groups.

Table 3-1 Selected IBIS-II participant baseline characteristics by Treatment Group.
Adapted from Cuzick, Sestak, Forbes et al 2014 [26]

Baseline characteristic	Anastrozole group (n=1920)	Placebo group (n=1944)
Age (years)	59.5 (55.0 to 63.5)	59.4 (55.1 to 63.3)
Age at menarche (years)	13.0 (1.2 to 14.0)	13.0 (12.0 to 14.0)
Parous (yes)	1601 (83%)	1637 (84%)
Age at first birth (years)	24.0 (21.0 to 27.0)	24.0 (21.0 to 27.0)
Age at menopause (years)	50.0 (45.0 to 52.0)	49.0 (45.0 to 52.0)
Height (cm)	162 (158 to 166)	162 (158 to 167)
Weight (kg)	71.8 (64.0 to 82.2)	72.1 (64.0 to 83.5)
Body mass index (kg/m2)		
<25	581 (30%)	568 (29%)
25 to 30	699 (36%)	732 (38%)
>30	640 (33%)	644 (33%)
Previous HRT	893 (47%)	910 (47%)
HRT within 12 months prior to randomisation	128 (7%)	152 (8%)
Hysterectomy	631 (33%)	656 (34%)
10–year Tyrer–Cuzick risk (%)	7.6% (5.8 to 9.9)	7.8 (5.1 to 10.2)

Data are median (Q1 to Q3) or n (%)

At 5 years median follow-up, 85 IBIS-II participants assigned to placebo developed breast cancer (4%) versus 40 women in the anastrozole group (2%) [26]. The number of invasive BC was significantly halved on anastrozole: 32 vs 64 for placebo, a hazard ratio (HR) of 0.50 (95% CI 0.32–0.76); p-value from Cox proportional hazards regression model = 0.001. Invasive ER+ tumours were also significantly reduced by anastrozole: 20 vs 47 for placebo, a HR of 0.42 (95% CI 0.25–0.71); p-value 0.001 from Cox proportional hazards regression model. However the proportion of ER- tumour numbers was not significantly affected: 11 for anastrozole vs 14 on placebo (HR 0.78, 95%CI (0.35–1.72), p=0.53, Cox proportional hazards regression model).

Two deaths from BC were reported in the anastrozole group versus none in the placebo treated group. Ten deaths from other cancers were reported for the placebo group, whilst seven were reported for the anastrozole treated group. However, no specific causes of mortality were more common in one group versus the other (p=0.836, relative risk model). Overall mortality was similar between the groups: 18 for anastrozole (1%) and 17 for placebo (1%).

The frequency of adverse events between treatment groups was compared using relative risks (RR), with Fisher's exact tests where appropriate. Interestingly, this double-blind study showed no overall difference in adverse event frequency between the groups, Table 3-2. However, small but statistically significant increases of ~10-15% in certain types of very common (~50% prevalence) adverse events— arthralgia, hot flushes/night sweats— were present in the anastrozole treated group. A dose-response effect of grade (mild, moderate, severe) for arthralgia appeared to be present, with those in the anastrozole treated group at significantly higher risk of moderate and non-significantly at higher risk of severe symptoms. Other musculoskeletal symptoms (joint stiffness and carpal tunnel syndrome/nerve compression) were also significantly elevated in the anastrozole treated group.

The frequency of fracture, including subgroups of fracture by location (arm, leg, pelvic/hip, rib/spine/collarbone, skull) was not statistically different between the groups. Overall gynaecological symptoms, vascular effects (including myocardial infarction, cardiac failure and thrombosis), eye symptoms and infections were not statistically different between the treatment groups. However, the relative risks for certain subgroups of these symptoms were significantly higher in the anastrozole treated group (RR (95% CI)): vaginal dryness 1.19 (1.03 to 1.37); hypertension 1.64 (1.18 to 2.28); dry eyes 1.45 (1.04 to 2.01), influenza 2.11 (1.06 to 4.19) and otitis media 3.04 (1.21 to 7.64).

Table 3-2 Selected IBIS-II AE frequencies by Treatment Group

Adapted from Cuzick, Sestak, Forbes et al 2014 [26]

Adverse Event	Anastrozole (n=1920)	Placebo (n=1944)	Risk Ratio (95% CI)
Any AE (medically non-serious and serious)	1709 (89%)	1723 (89%)	1.00 (0.98 to 1.03)
Fractures, all	164 (9%)	149 (8%)	1.11 (0.90 to 1.38)
Musculoskeletal, all	1226 (64%)	1123 (58%)	1.10 (1.05 to 1.16)
Arthralgia	972 (51%)	894 (46%)	1.10 (1.03 to 1.18)
Mild	385 (20%)	386 (20%)	1.01 (0.89 to 1.15)
Moderate	422 (22%)	363 (19%)	1.18 (1.04 to 1.33)
Severe	151 (8%)	123 (6%)	1.24 (0.99 to 1.56)
Joint stiffness	143 (7%)	96 (5%)	1.51 (1.17 to 1.94)
Pain in hand or foot	178 (9%)	147 (8%)	1.23 (0.99 to 1.51)
Carpal tunnel syndrome/nerve compression	67 (3%)	43 (2%)	1.58 (1.08 to 2.30)
Vasomotor (hot flushes, night sweats), all	1090 (57%)	961 (49%)	1.15 (1.08 to 1.22)
Mild	550 (29%)	504 (26%)	1.10 (1.00 to 1.22)
Moderate	390 (20%)	330 (17%)	1.20 (1.05 to 1.37)
Severe	150 (8%)	127 (7%)	1.20 (0.95 to 1.50)

Data are n (% of group)

The estimated five-year adherence to protocol treatment was 68% in the anastrozole group and 72% in the placebo group. Adverse events (20% of anastrozole participants vs 15% placebo) and patient refusal (5% in both groups) were the main reasons for discontinuation of treatment.

3.1.2 The International IBIS-II MD and AI study

Mammograms from IBIS-II international centres are being collected by Cancer Research UK (CRUK) and Queen Mary University of London (QMUL) for an international mammographic density project. An additional, modest reimbursement is paid to participating centres upon receipt of a set of entry, midpoint and final mammograms for each participant. Mammograms are considered entry mammograms if they are within 18 months of randomisation. Entry mammograms at baseline are ideal (i.e. taken on the mammographic date recorded on participants' randomisation forms). If entry mammograms are not available then midpoint (18 to 47 months post-randomisation) and final mammograms (18 to 66 months post-randomisation) are not required. Medio-lateral oblique View mammograms are preferred, however medio-lateral View mammograms are also accepted.

Mammogram collection for the IBIS-II International MD and AI project commenced after the IBIS-II September 2011 protocol Amendment and trial reconsent mentioned previously, which enabled participants to provide additional optional consent for mammographic density and blood and biomarker studies. The final IBIS-II participants will complete five years of randomised treatment in February 2017; hence it is anticipated that the final mammograms for the international IBIS-II mammographic density study will soon be collected. It is anticipated that the international IBIS-II MD and AI study will compare longitudinal MD changes for the anastrozole treated group vs the control group, as well as whether any differences in MD change over time between the treatment groups are associated with differences in BC incidence. Results from the International IBIS-II MD and AI study have not yet been disseminated.

3.2 The CMN IBIS-II MD and AI substudy

The remainder of this thesis focuses on the CMN IBIS-II Mammographic Density and AI substudy. The overall goal for this thesis was to characterise the relationship between MD and

an AI in the BC prevention setting. This had not previously been undertaken for the third generation AI anastrozole. Hence it was necessary to collect and measure MD on IBIS-II trial mammograms, and fit statistical models of the associations of MD for this nested cohort of IBIS-II participants. As discussed further below, the project for this thesis differed from the international IBIS-II MD and AI study in that it utilised annual IBIS-II trial mammograms from a single centre only, and examined MD growth (but not, presumably, differences in BC outcomes) between the treated and control groups.

Two Aims were relevant to the investigation of the relationship of MD with other factors. One Aim was to quantify the relationship between baseline MD and established, associated BC risk factors such as body mass index and age. The other Aim, the Primary Aim (AIM 5) of this thesis, was to characterise longitudinal changes in MD for IBIS-II trial participants randomised to anastrozole and placebo, and ascertain if changes in longitudinal MD differed between treatment groups.

Both of these Aims were dependent upon selection of a consistent, reliable and repeatable method to measure MD. This led naturally to AIM 1 of this thesis— selection of a method or methods suitable to measure longitudinal MD on IBIS-II participant mammograms. AIM 1 was subdivided into three sequential goals: 1.) undertake a review of the methods to measure MD; 2.) determine which methods could be applied to the breast images available for the CMN IBIS-II participants; and 3.) decide which method/s were likely to provide an easy to implement, reliable and accurate means of characterising MD for the sampled group.

AIM 2 of this thesis was to quantify the reliability of the methods selected in AIM 1. Until very recently, most methods for measuring MD had a subjective component (i.e. user input) which introduces additional variability into the measurements. Because the estimated effect of anastrozole was small (1 to 2% decrease in MD) compared to the likely variability of the

measurements [29], selection of the technique with the lowest variability was important to the success of the project. AIM 2 of this thesis was subdivided into separate sub-aims, according to the type of reliability to be tested: intra-observer reliability, inter-method reliability, acquisition technique reliability and repeated measurements inter-method reliability.

AIM 3 of the thesis—descriptive statistics of the CMN IBIS-II participant characteristics and a cross sectional analysis of baseline MD with characteristics likely to be associated with MD—was accomplished by performing statistical measures of central tendency such as t-tests and Wilcoxon rank-sum tests, and regression analyses.

AIM 4 of the thesis— development of a longitudinal model of change in mean MD for the aggregate of the treated and control groups— was completed by preparing a 'mixed' linear statistical model (one with both fixed and random effects) of mean change ('growth') of MD over time, adjusted for significant covariates and age. Models for repeated measures data account for the violation of the assumption that all observations in the sample are independent from each other. This is achieved by incorporating groups (also known as levels or hierarchies) into the models. This type of model has a number of names: mixed model, multi-level model, nested model, random-coefficient model [370]. For the longitudinal data in this project, the mixed model description which best suits is 'hierarchical linear' model. This is the most commonly used type of multilevel model, and is an extension of multiple linear regression [371]. After this mixed model was formulated, data and scripts to run a set of statistical models were forwarded to the IBIS-II trial statistician to test for differences in MD change over time between the treated and control groups. This accomplished the Primary Aim (AIM 5) for this thesis.

3.2.1 Approvals for the CMN IBIS-II MD and AI substudy

This project commenced in March 2010. A subsequent internal approval process for the project was undertaken with the ANZ BCTG was completed in May 2011. The Confirmation process at the University of Newcastle was successfully completed in February 2011. An approval for the thesis Primary Aim (Chapter 8— MD average annual change in IBIS-II treated vs control groups) was sought and obtained in October 2016 from the IBIS-II international steering committee.

After completing the internal approvals process with the ANZ BCTG in May 2011, the ANZ BCTG advised that an IBIS-II protocol amendment would be submitted shortly to the Hunter New England Health Human Research Ethics Committee (HNE HREC). As previously described, IBIS-II participants were asked to reconsent to continue trial participation in their randomised groups because the trial steering committee believed it was too soon to consider the results of the MAP.3 trial definitive. Optional consent for tissue, biomarker and mammographic density studies were included as part of the reconsent process.

The ANZ BCTG requested that the ethics application for this project was deferred until after the IBIS-II protocol amendment was approved by the HNE HREC for CMN. This request was made to lesson confusion between the International IBIS-II and local CMN IBIS-II MD and AI projects. Although the protocol amendment was not released until September 2011, the upside of this request was that a waiver of consent for IBIS-II mammogram collection was not required from the HREC, as only participants who consented to the optional mammographic density studies would be included in both studies. Consequently, only mammograms from Calvary Mater Newcastle participants who explicitly consented to use of their mammographic data were used in this project.

3.2.1.1 *Ethics approval – Hunter New England Health HREC (HNE HREC)*

The IBIS-II protocol amendment was released by CRUK in September 2011, and forwarded to participating ANZ BCTG centres by the ANZ BCTG in October. The amendment was approved for CMN by the HNE HREC in November 2011. The approval for this CMN IBIS-II MD and AI project was granted by the HNE HREC in December 2011.

Three variations to approval for this project were also approved by the HNE HREC:

1. Expansion of mammogram collection to NSW centres other than the Calvary Mater Newcastle (2012)
2. Data sharing with the CSIRO in relation to the AutoDensity program for measurement of MD (2013)
3. Collection of IBIS-II data on covariates known to be potential confounders from the ANZ BCTG to improve statistical models (2014)

3.2.1.2 *Ethics approval – University of Newcastle HREC*

The approvals for this project from the HNE HREC were registered with the University of Newcastle's HREC as per standard policy for University staff and students. The ethics approval and variations for IBIS-II were registered and approved the University of Newcastle's HREC in January 2012.

3.2.2 Constraints on mammogram collection outside of CMN

Initially, it was anticipated that mammograms from most IBIS-II ANZ BCTG participating centres would be available for use in this project. However, various constraints precluded this. One of the principal constraints was imposed by the ANZ BCTG; others reflect the difficult nature of retrospective mammography collection.

In 2011, the ANZ BCTG requested that all IBIS-II mammogram collection for centres outside of CMN be performed by the ANZ BCTG. This was to lessen confusion for ANZ BCTG IBIS-II participating centres about which project/s would be utilising IBIS-II mammograms. Direct contact by the candidate and IBIS-II participating centres was not possible; all mammogram collection and requests for ethics and other approvals were managed by the ANZ BCTG and trial coordinators at ANZ BCTG IBIS-II participating centres.

Although the HREC approvals for the September 2011 protocol amendment included an approval for the International MD and AI project, approvals for this project were not included. This additional overhead further decreased the possibility that many IBIS-II mammograms outside of CMN would be available for this local MD and AI substudy, because each IBIS-II centre would need to gain local HREC approval to contribute mammograms to this project.

Whilst an Australian mutual HREC acceptance process for multi-centre interventional clinical trials research commenced in early 2012, interstate multi-centre interventional clinical trials approved prior to 2012 such as IBIS-II were not eligible for this process. Projects approved by the earlier process as described above for IBIS-II (i.e. each site submitted separate approvals to its own local HREC) continued to operate in accordance with the earlier process. However, mutual acceptance for HREC approvals within the NSW public health system has been operational since late 2007. Ethics committees in NSW are permitted to accept the ethics approval from another NSW HREC.

HREC approval was received in 2012 for a letter sent to NSW IBIS-II Principal Investigators to request their assistance with collection of mammograms for this local IBIS-II MD project. The ANZ BCTG trial coordinators forwarded this request to NSW centres in 2012 and 2013. One centre in NSW with seven IBIS-II participants, Southern Highlands Cancer Centre, kindly completed the NSW mutual acceptance process.

Given that all clinical trial activities undertaken at participating centres need to be fully funded, funding to cover the costs of the additional ethics approvals needed for this project may have made participation more attractive for IBIS-II centres. Similarly, trial coordination activities undertaken by the ANZ BCTG need to be supported. Although funds were forwarded to the ANZ BCTG for mammogram collection for this project, funding for ANZ BCTG personnel dedicated to mammogram collection for this project would likely have increased access to IBIS-II mammograms.

3.2.3 Sample size calculations

Sample size estimates for this project were initially informed by the MAP.1 study, a randomised control trial of one year of treatment with letrozole vs placebo for BC prevention in high-risk women [29]. Percent (mammographic) density is the proportion (percentage) of the total breast area covered by the dense tissues of the breast. Table 3-3 lists percent density (PD) measured from mammograms of the MAP.1 participants at baseline, and 12 and 24 months post-randomisation. Differences in PD for both between- and within-treatment group comparisons for MAP.1 were made using a Wilcoxon test with a two-sided significance level of 0.05.

Based on these results, treatment with anastrozole was estimated to yield a 1.5 percent difference in mammographic density compared to the placebo group at 1 year with a standard deviation of 8. However, due to uncertainty around this estimate, a range of power estimates was calculated for varying effect sizes (MD differences of 1 to 2%) and sample sizes of 125 to 250 participants per group, based on a two-sample comparison of means (Table 3-4).

As shown in Table 3-4, for an adequate power of 80%, mammograms from a total of 500 participants (250 participants from each arm) would be required if the difference in mean PD was 2%. As described previously, final accrual (February 2012) was 818 participants at ANZ

BCTG centres. If it had been possible to collect most baseline and first year mammograms from the participants, the project might have yielded an adequately powered, statistically significant result for the difference in mean MD between groups.

Table 3-3 MAP.1 changes in mammographic percent density
Adapted from Cigler et al 2010 [29]

Density	Statistics	Letrozole	Placebo	Unadjusted p-value, between treatments	Adjusted p-value, between treatments[a]
At baseline	N	31	19	NC	NC
	Mean	39.6	40.0		
	SD	20.0	15.9		
	95% CI	(32.3, 47.0)	(32.4, 47.7)		
Change at 12 months	N	30	19	0.67	0.61
	Mean	–1.74	–0.24		
	SD	5.65	9.34		
	95% CI	(–3.85, 0.37)	(–4.47, 4.26)		
	P-value[b]	0.10	0.91		
Change at 24 months	N	27	16	0.69	0.61
	Mean	–0.01	–1.32		
	SD	9.81	14.15		
	95% CI	(–3.89, 3.87)	(–8.86, 6.22)		
	P-value[b]	0.99	0.71		

SD — standard deviation, 95%CI — 95% confidence interval, NC — not computed
[a] Adjusted for age and BMI at baseline
[b] For the comparison of change from baseline within treatment group

Table 3-4 Estimated power to detect a difference in IBIS-II mean PD, treated vs control

# participants per group	Mean PD change between treated and control groups				
	1.00	**1.25**	**1.50**	**1.75**	**2.00**
125	.165	.232	.313	.405	.503
150	.189	.270	.366	.471	.578
175	.213	.307	.416	.532	.645
200	.238	.344	.464	.588	.703
225	.262	.379	.509	.639	.754
250	.286	.414	.552	.685	.797

If the more likely outcome of a smaller 1.5% difference in PD was observed, a total sample size of 500 would result in an insufficient power of 0.55. Smaller sample sizes and smaller mean PD differences would provide even lower power to detect a difference between the IBIS-II

treatment groups if one existed. Even if the two largest accruing IBIS-II centres were specifically targeted— the Calvary Mater Newcastle hospital (142 participants) and Sir Charles Gairdner Hospital in Western Australia (153 participants)— this was likely to yield approximately 125 participants per treatment arm. As per Table 3-4, even with a large MD mean difference between groups of 2% and 125 participants per group, the power to detect a treatment difference at ~12 months post-randomisation was low (~0.5).

These sample size calculations showed comparison of the mean difference in PD between treatment groups was unlikely to yield a useful outcome for this project. Comparisons of means makes good clinical sense; comparison of treatment group means is a popular outcome utilised for MD measurements, as seen for IBIS-I participants [296], and other AI treatment studies [28, 29, 31, 32, 304, 312, 313]. If it can be shown (as for tamoxifen) that a defined decrease in MD at a certain time after starting AI treatment is associated with clinical efficacy, then it is likely worth measuring change in MD over this time period. It would clinically useful to know for instance, that women who experience a 5% or greater decrease in MD twelve months after treatment with an AI are highly likely to remain BC free in 10 years' time. A difference in means measured at two time points also makes economic sense, because additional, unnecessary and expensive collection and measurement of a biomarker (e.g. mammograms, in the case of MD studies) at other time points is not required. Due to the limitations on collection of the IBIS-II MD data, the small expected effect size, as well as changes in mammographic technique (film vs digital mammography), comparisons of mean MD for the treatment groups was not suitable for this project. Another approach was needed.

Because multiple sets of annual mammograms were available for many IBIS-II participants, change in MD over time between the IBIS-II treatment groups could also be examined using longitudinal (repeated measurements) statistical models. Sample size calculations for longitudinal data models incorporate the same parameters used to calculate samples sizes for the

difference in means between two groups— type 1 error rate (α), smallest meaningful difference (Δ), power (1-β), and variability (σ). Two additional parameters are also required— the number of repeated observations per person (n) and the correlation among the repeated observations (ρ) [22].

The correlation (ρ) within the data has opposite effects on the sample size depending upon the outcome being measured. Increasing correlation decreases the sample size needed for studies examining the rate of change between two groups. The opposite is true for sample size calculations for a "time-averaged difference in a response between two groups" [22], where increasing correlation between the groups increases the required sample size by effectively increasing the variance.

Repeated measurements sample size calculations were undertaken in Stata Version 12.1. For a difference in means of 2, standard deviation of 8, for one baseline measurement with three follow up measurements, a sample size of 67 per group was required (double sided p-value of 0.05, power 0.8 with correlations of 0.8 between baseline and follow ups). This is much closer to the actual sample size acquired of 120 total CMN participants, with an average of 4.5 mammographic episodes (follow ups) per participant. A difference in means of 1% PD however, increases the sample size to 268 per group, which is beyond the scope of this project.

3.2.4 IBIS-II mammogram collection

3.2.4.1 *Collection of mammograms via the ANZ BCTG*

As previously described, collection from centres other than the Calvary Mater Newcastle were coordinated only by the ANZ BCTG. The Southern Highlands Cancer Centre completed the NSW reciprocal HREC approval process; fifty-one mammograms comprising 14 mammographic episodes from six participants were collected.

Calvary Mater Newcastle mammograms not received directly from the CMN trial coordinators were forwarded via a shared drive on an ANZ BCTG server for use in this project by ANZ BCTG IBIS-II coordinating staff.

3.2.4.2 *Collection of mammograms via the Calvary Mater Newcastle Hospital*

In order to maximise the potential information from the available sample of IBIS-II participants, the decision was made to collect all available mammographic data for as many follow ups from as many CMN participants as possible. This meant all four standard mammographic Views— the right cranio-caudal (RCC), left CC (LCC), right medio-lateral oblique (RMLO) and LMLO Views— were collected for all available mammographic episodes. Given the inherent variability in most popular (subjective) methods used to measure MD, use of all four Views would likely provide a more consistent representation of each mammographic episode's MD.

The majority of the ~2000 mammograms collected for the project were supplied via the IBIS-II Calvary Mater Newcastle (CMN) Department of Surgical Oncology trial coordinators. Two rounds of mammogram collection were undertaken. The first (Collection 1) was in 2012 after ethics approval was granted in December 2011. A second round of trial mammograms was collected in 2014 (Collection 2). Collection 2 mammograms included mammograms taken in 2013 and 2014, as well as mammograms taken at earlier dates.

IBIS-II digital trial mammograms were forwarded by the CMN trial coordinators on DVDs from the Calvary Mater Newcastle Radiology department. Film-screen and film copies of digital mammograms were collected and digitised during participant IBIS-II trial clinic visits. All Calvary Mater Newcastle mammograms collected from the IBIS-II CMN trial coordinators were forwarded to the ANZ BCTG for the International IBIS-II MD and AI project.

To facilitate a second round of mammogram collection in 2014, the CMN IBIS-II coordinators sent reminders to the IBIS-II participants to bring their mammograms to clinic. They also requested additional follow up mammography DVDs from the CMN Radiology department. Time constraints during the busy clinic typically precluded return of the mammograms during the clinic to the participant. The trial coordinators were often too busy to forward the mammograms for digitisation until after completion of the clinic. However, due to the time needed to digitise the mammograms it was also difficult to return the mammograms to the trial coordinator within the timeframe of participants' clinic visits.

Baseline mammograms were often elusive. Whilst post-randomisation trial mammograms were taken at the CMN Radiology department, many IBIS-II participants undertook pre-enrollment mammograms at other locations. Some mammograms taken at private imaging centres were destroyed due to the time between baseline mammography (e.g. 2008) and the first collection round (Collection 1) in 2012. Hence mammograms from outside facilities were rarely available unless participants brought them with them to a follow up clinic.

In 2012, collection of a set of baseline mammograms from a participant at home was trialled. This was time consuming for the CMN IBIS-II coordinator, and not repeated. However, the second round of data collection in 2014 (Collection 2) yielded many additional mammograms, including a number of baseline mammograms.

The transition from analogue (film-screen) to digital mammography at CMN took place in January 2009, 3.5 years after the trial was activated (all approvals were obtained) at CMN. Approximately half of 85 baseline mammographic episodes and half of 25 six–month episodes are film. This proportion decreased to 13% by 12 months, and 1% at year 2 (1 of 93 total episodes). Forty-five of 120 total CMN participants had at least one film-screen mammographic episode, whilst at least one digital mammographic episode was collected for all

120 participants. Baseline mammograms were frequently taken at locations outside of CMN, however all annual (post-randomisation) mammography was performed on a single mammography machine at the CMN Radiology Department.

A description of the process undertaken to select the scanner and de-identification software used in this project, and the method used to de-identify mammograms is contained in Deidentification Software and Mammogram Deidentification Method.

A review of MD measurement techniques is undertaken in the next chapter, in order to ascertain which assessment techniques best suited this project. All MD measurement assessments in this thesis were undertaken by a single reader (the candidate).

4. MD measurement techniques review

This chapter achieves Aim 1 of the thesis, a review of the methods available to measure mammographic density [prior to 2014], to select techniques to use for this project. Some of the numerous methods available to measure MD are reviewed, including visual assessment, two-dimensional quantitative methods (e.g. Cumulus, Madena), three-dimensional methods (e.g. Volpara, Qantra) as well as other techniques such as MRI, ultrasound, DXA (dual-energy x-ray absorptiometry) and spectroscopy. The chapter concludes with a rationale for the selection of three techniques chosen to measure MD on the CMN IBIS-II mammograms.

4.1 MD measurement technique selection

Selection of the most appropriate acquisition and measurement techniques for the IBIS-II mammograms was determined. Many techniques have been developed to quantitate breast density. Most are based on popular imaging methods, such as mammography (traditional two-dimensional (2D) mammography (x-rays of the breast), and tomography), MRI, ultrasound, DXA/single (S)XA, and PET (positron emission tomography) [372]. Because mammograms were collected retrospectively for this project, and only mammograms were available (not for example ultrasound and/or MRI images, two other popular breast imaging modalities), choice of technique was immediately restricted to those suitable for retrospectively collected mammograms.

This chapter also contains a non-exhaustive description of the more well-known and commonly used methods to measure breast density. These are included because they may be suitable for screening for breast density and/or BC, and therefore for use in longitudinal breast density projects where breast images are prospectively collected. Three dimensional (3D) breast imaging and density estimates are becoming more widespread, and hence are of interest. Please refer to recent [2015 & 2017] reviews for a more comprehensive discussion of available breast and mammographic density measuring techniques [317, 372]. Most longitudinal breast image

data sets, however, are comprised of (2D) mammograms because this is the current modality used by screening programs [373]. Hence this chapter primarily focuses on techniques applicable to this method of breast imaging.

As stated previously, the estimated decrease in MD in response to an aromatase inhibitor was small (~2%); therefore selection of the acquisition and measurement technique with the smallest amount of variability (i.e. most reliable technique) that was capable of detecting change on retrospectively collected mammograms, was core to detecting change in density on the trial mammograms.

Most techniques in common use to assess mammographic density are subjective and/or time-consuming. Visual methods are quick to use, but have larger margins of error due to subjectivity compared to computer-assisted or automated methods.

4.1.1 Wolfe's parenchymal patterns

John Wolfe, a radiologist based at the Hutzel Hospital in Detroit Michigan USA, first described an association between a visual assessment of mammographic tissue composition and BC in 1967 [374]. The "roentgenographic appearance of the parenchyma" was affected by age, parity and possibly by the hormonal influence of natural versus surgical menopause. A prominent duct pattern signalled an increased likelihood of existing breast cancer, and he speculated that the patterns could possibly be related to future breast cancer risk [374, 375]. His discovery was controversial because MD's sole association with BC at that time was its ability to impede BC detection on mammograms.

The Wolfe patterns were a qualitative assessment of breast density, and therefore were subjective. One difficulty with early studies of breast density found by critical reviews [132, 134, 135] was the lack of reproducibility by researchers who tried to implement Wolfe's

parenchymal patterns, leading to conflicting results. Even Wolfe was not entirely consistent in applying the patterns [132].

4.1.2 Tabar's patterns

Laszlo Tabar and colleagues developed a different qualitative method to describe the parenchymal patterns of the breast which uses a more detailed analysis of the anatomical features on a mammogram [84]. This method was also found to be difficult to implement, and provided similar reliability to that Wolfe's parenchymal patterns in the assessment of BC risk [376, 377].

4.1.3 Other patterns

Classification of the breast tissue using fractal patterns and other texture features has been an ongoing challenge, for example [125, 158, 360, 378-380], and has yielded interesting results such as a texture pattern which is a BC risk factor independent of PD [381], texture features which are correlated with PD on digital mammography and digital breast tomosynthesis images [382], and potentially correlated with endogenous hormonal exposure [126]. However, these methods have not been adopted widely as measures of breast cancer risk due to the additional complexity imposed by pattern analyses. A recent [2016] review summarises existing textural classification studies and future directions, and includes a call for large scale longitudinal studies incorporating BC risk [383].

4.2 Percent mammographic density

Wolfe realised his breast parenchymal patterns caused difficulty in the burgeoning field of mammographic density research, and hence developed a more quantitative method for MD assessment. In 1987 he published a study which compared his Wolfe patterns and a new, quantitative method described as measurement of dysplastic involvement [234]. A planimeter

was used to calculate the area/s of density and the whole area of the breast as traced on sheet of plastic overlaid on a mammogram. Percent density was expressed as the sum of the dense areas divided by the area of the whole breast. The percent mammographic density calculated using this technique, like Wolfe's parenchymal patterns, showed an increasing gradient of breast cancer risk as density increased.

4.2.1 "Mammographic Density"

The term 'mammographic density' first appeared in a journal article by Swann, Kopans et al in 1987. The phrase 'mammographic density' was used interchangeably with 'radiographic density' in this publication which found little correlation between Wolfe's parenchymal patterns and the size or compressibility of a breast [329].

Use of Wolfe's parenchymal patterns was common in studies of breast density well into the 1990s. The introduction of visual scales for PD [116, 384, 385] from the early 1990s onwards as well as the advent of computer-assisted techniques in the mid to late 1990s to assess percent mammographic density [128, 289] helped PD to replace Wolfe's patterns as the de facto standard measure of mammographic density. The term 'mammographic density' today generally denotes the percent or area of a breast covered by dense tissue on a mammogram.

4.3 Visual estimation techniques

4.3.1 BI-RADS

The most widely used MD measurement technique is the Breast Imaging-Reporting and Data System (BI-RADS) density categorization, part of the American College of Radiology's BI-RADS standardised image reporting framework [386].

BI-RADS is a system that characterises the breast images created by mammography, breast ultrasound and MRI using standardised classifications. The various components of a BI-RADS

report use the "Breast Imaging Lexicon" (standardised ways to describe imaging findings such as masses, calcifications, asymmetries) which culminate in a final assessment category for the patient. The final assessment categories range from 0 to 6, an ordinal (ranked) numbering system that denotes outcomes from incomplete (0), negative (1) to malignant (6) [117].

The BI-RADS system also incorporates a lexicon to describe the dense mammography tissues: the BI-RADS density categories for breast composition. As mentioned previously, the BI-RADS density categories were initially qualitative [116]. The categories subsequently became quantitative (quartiles of PD: 0-24%, 25-49%, 50-74%, 75%+) in the 4th Edition of BI-RADS [385], and have recently been revised to four new qualitative breast composition categories called A, B, C and D to lessen confusion between the BI-RADS density categories and the overall, final BI-RADS assessment category [119, 385].

Four categories equivalent to the BI-RADS breast composition categories exist on Australian BreastScreen reports, but are not routinely completed.

Although the BI-RADS density categories were created to describe the likelihood cancer was detected on a mammogram, they have been successfully utilised in many epidemiologic studies of BC risk [136], including longitudinal change in BI-RADS density [387, 388]. However, for the purposes of this project, the CMN MD and AI substudy, due to the small anticipated PD change of –2%, the quartile scale is too coarse and not suitable to detect longitudinal change in MD in IBIS-II participants [389].

4.3.2 Six category classification (SCC)

A six category classification (SCC) was created in 1994 [390]. The percent density categories for the SCC (also known as the Boyd scale) are: 0%, 1 to 10%, 11 to 24%, 25 to 49%, 50 to 74% and 75-100%. This classification system has had more limited use than the BI-RADS

density categories, but has proven useful as a research tool [194, 296, 331, 391-393]. A modified version of this scale was utilised in a seminal MD and BC risk paper in the New England Journal of Medicine [201]; this important article raised awareness of the association between MD and BC risk for a broad cross-section of the medical community

Part of the six category classification's usefulness is derived from its sub-division of the first BI-RADS 4th Edition density category (0 to 24% density) into 3 categories. Most post-menopausal women have low density (e.g. Fig 1of [137]), that is less than 25% of the breast area is covered by the dense glandular and connective tissue. Subdivision of the first BI-RADS (4th Edition) breast composition category helps to overcome the loss of information which arises by assigning half to three-quarters of older women to a single category. For instance, if any amount of density is associated with an increase in risk, this information is lost within a single 0 to 24% category. Creation of the first three SCC categories improved the characterisation of BC risk posed by the dense breast tissues for populations with low density.

For the purposes of this project however, the SCC categories were ascertained to be too large for detecting small changes in PD.

4.3.3 Cancer Research UK (CRUK)

A visual assessment scale with 21 categories to help overcome some of the problems (i.e. coarseness) with the BIRADs and SCC density classifications was developed [296]. This scale consists of 21 categories ranging from 0 to 100% density in 5% increments, rounded up to the nearest 5%. That is, 9% density is categorised as 10%, and 52% density is categorised as 55% density. A training DVD for this method was developed by radiologist Ruth Warren to assess participant's eligibility for the IBIS-II breast cancer prevention trial.

The 21 category visual classification method has been validated in several studies published in

peer reviewed journals [296, 394]. The technique is quick to utilise. Although the 5% categories would not be useful to detect small (<5%) changes in PD for this project, substantial (10% or greater) changes in PD might be distinguished. This technique was selected as a potential method for use in this project (the CMN MD and AI substudy).

4.4 Side by side comparison

Side by side comparisons of mammograms to visually estimate longitudinal change in MD has been successfully utilised [262]. This technique was considered for use in this project, however it was not implemented because its use might bias results from other techniques.

4.5 Planimetry

As described in 4.2, manual planimetry can be used to quantify PD on mammograms [234]. Use of a planimeter, whilst accurate, is very time-consuming. Because similar but faster computer-assisted techniques now exist, this technique was not utilised in this project.

4.6 Semi-automated methods for percent density assessment

4.6.1 Cumulus

The Cumulus program [128], developed by Norman Boyd, Martin Yaffe and colleagues at the University of Ontario, is the most widespread, semi-automated method used to assess breast density [395]. It is the defacto standard for measurement of PD [395].

Cumulus users interactively outline the pectoral muscle and other areas to be excluded from the total breast area. The breast edge is delineated through selection of a grey-level threshold and/or manually outlined. A second grey-level threshold which best captures the dense tissue area/s is also selected. The proportion of dense tissue area relative to total breast area is then calculated to yield PD for that mammogram. (See below, Figure 4-1)

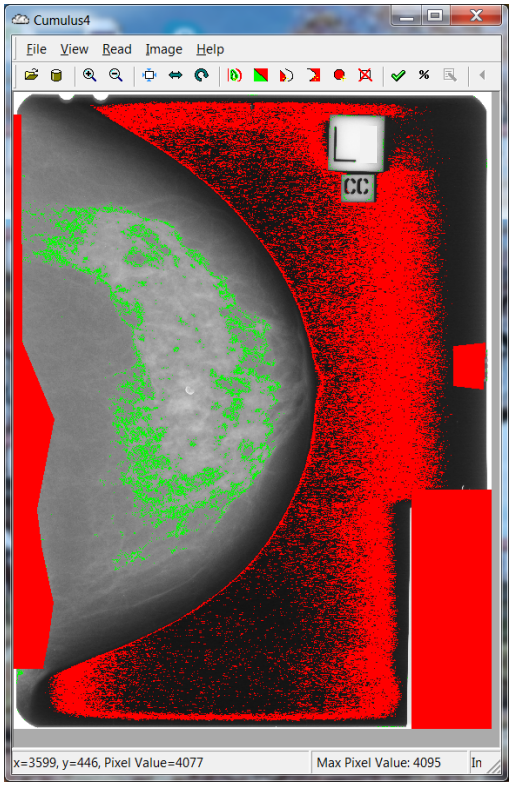


Figure 4-1 Measurement of breast density using Cumulus
The red areas selected by the user delineate the non-breast area from the breast area (raster size 6,743,585). The area surrounded by green pixels was selected by choosing the grey-level which best captures the dense area of the breast (raster size 2,013,137). This breast is approximately 30% dense ($2x10^6/6.7x10^6$). IBIS-II left cranio-caudal (LCC) View mammogram.

Cumulus is well validated. Measurements made by the program have been shown to be very reproducible for trained users, whose correlation coefficients range from 0.95 to 0.99 [182, 396, 397]. Cumulus has been utilised worldwide as an effective tool to quantify density on mammograms and produces consistent results demonstrating increasing gradients of breast cancer risk for increasing percentages of mammographic density.

However, because Cumulus is interactive (i.e. the user determines the areas to outline), measurements with Cumulus are subjective. Even if 100% consistent within their own assessment of a set of mammograms (100% intra-observer reliability), different users of Cumulus can and will vary in their quantification of the density present on mammograms (inter-

observer variability) [personal communication [398] (during training of the Candidate in the use of Cumulus at the University of Melbourne)].

High intra-observer consistency is the most critical aspect for Cumulus use [398]. That is, as long as each user is consistent whilst selecting the breast and dense areas, the relative distributions of mammographic density measurements will be similar for each Cumulus user. This will result in nearly identical gradients of risk for each user's distribution of density measurements, because mammograms will be classified within the same quantile of density even if the absolute value measured for percent density varies among assessors.

To achieve the lowest variability and smallest mean difference, it is advisable to read sets of mammograms from a single patient in random order when assessing PD [399].

Cumulus was designed to assess mammographic density in film-screen mammograms, however Cumulus has successfully been utilised with digital mammograms [18, 400, 401]. Due to the ability of Cumulus to quantify PD precisely, this technique was selected for use in this project.

4.6.2 Madena

Madena is another interactive thresholding computer program, for use on Macintosh computers [289]. Whilst less widely used than Cumulus, Madena is freely available, and has been well-validated [214, 229, 239, 266, 274, 290, 291, 323, 377, 402-405]. It differs somewhat from Cumulus, in that the user manually outlines the breast area— excluding the pectoralis muscle, prominent veins and artefacts— to define the total area of the breast [402]; use of a grey level threshold to capture the outer edge of the breast as shown in Figure 4-1 is not undertaken. A grey level threshold is then selected to capture the dense tissue of the breast. Users manually record density data in spread sheets, whereas Cumulus records these values automatically.

The operating system primarily used at the University of Newcastle and the ANZ BCTG is Windows. Purchase of a separate Macintosh computer to utilise Madena precluded its use in this project.

4.7 Automated techniques – '2D' density assessment

To reduce the variability inherent in subjective measurement tools, many fully automated techniques to measure density on conventional '2D' (two-dimensional) mammograms have been developed [406-409]. These include some of the pattern recognition techniques mentioned previously, as well as programs that automatically measure PD. One of the

A recently developed, fully-automated program using the US National Institute of Health image assessment program, ImageJ, has been validated against the Cumulus program [410]. This assessment tool has successfully measured longitudinal changes in PD for women on TAM vs no TAM to provide further evidence that reductions in MD are associated with reductions in BC risk [297] . The image manipulation techniques utilised to measure PD with the ImageJ program have been documented [410], but a commercially available version of the program was not available.

A fully-automated technique to measure percent and area density on conventional 2D mammograms was developed in Australia via a collaboration between the Commonwealth Scientific and Industrial Research Organisation (CSIRO) and researchers at the University of Melbourne— the AutoDensity program [411]. AutoDensity has a user-friendly interface to select mammograms for measurement, and is easy to use. This measurement technique was selected to trial in this project.

Another automated measurement technique, the LIBRA program (Laboratory for Individualized Breast Radiodensity Assessment) [412, 413], uses DICOM images from both pre-processing

and post-processed digital mammograms. This program is suitable for use on the IBIS-II mammograms, as are other techniques such as STRATUS [317, 414, 415] had they been available at the time techniques for this project were selected in 2013.

4.8 Volumetric techniques – '3D' density assessment

Three-dimensional measurement techniques have been developed to further quantitate the relationship between MD and BC risk. Many techniques generate estimates of the volume of dense tissue using two-dimensional mammographic images, others generate 3-D measurements using volumetric imaging techniques such as MRI. Despite strong criticism of the use of PD as a measurement of BC risk [416], volumetric estimates of MD have not provided consistently better gradients of risk compared to PD.

A number of automated programs which estimate the dense tissue volume from standard (two-dimensional) mammographic images have been developed. Two of these automated programs have achieved commercial success (Volpara[417] and Qantra[418]) and are used clinically.

4.8.1 Standard mammographic form (SMF)

Standard mammographic form (SMF) was one of the earliest programs developed to standardise the appearance of film mammograms due to the variability in exposure times, types of film, etc. [419, 420]. The program calculates the thickness of the breast from its curvature at the margin of the film mammogram. From this, an estimation of the breast volume can be calculated. The density (whiteness) of the tissue at each pixel is used to estimate the amount of dense tissue present in the column of breast tissue present at that pixel. The sum for each pixel is determined for the whole breast, which results in a volume of dense tissue. Percent volumetric density is calculated as the volume of dense tissue in the breast divided by the whole volume of the breast.

Because SMF is it is fully automated, the measurements made by SMF are highly repeatable; an identical volumetric percent density is produced for each mammogram every time the program is run.

Studies utilising SMF as well as percent mammographic density have shown risk gradients for breast cancer with increasing percent volumetric density, but the predictive ability of volumetric breast density from SMF has not been as great as for percent mammographic density [421, 422].

4.8.2 Commercial volumetric programs

Two commercial programs were available in 2013 for use with digital mammography machines and picture archiving and communication systems (PACS):

- Makatina's Volpara (http://www.volparadensity.com/))
- Hologic's Quantra (http://www.hologic.com/en/breast-screening/volumetric-assessment/),

Both programs are approved for use by the FDA (Food and Drug Administration) in the USA with certain digital mammography systems. Both programs provide a fully-automated volumetric density estimate of the breast tissue at the time the mammography is performed. Both commercial programs require access to the 'raw', pre-processed digital mammogram images, which differ from the post-processed images used clinically.

The number of commercially available automated programs has now increased (both 2D and 3D measurement techniques) as reviewed in Destoinis 2017 [317]. Most commercial applications focus on quantifying mammographic density, sometimes emulating the BI-RADS density categories, to meet the new USA reporting legislation requirements. As yet, not many longitudinal studies have been undertaken using the automated measurement programs because they are not yet in widespread clinical use. Automated programs are likely superior for use in longitudinal projects as they are not influenced by user input, and may be able to detect small

changes in density (e.g. such as those expected from AI treatment). Further research is needed to ascertain if longitudinal changes in breast density (whether due to aging, (preventive) hormonal treatment, or other therapy) measured with automated software are associated with changes in BC risk.

4.8.3 Cumulus V

Researchers at the University of Toronto developed Cumulus V [423]. Cumulus V (V for volumetric) uses a 'phantom' (an imaging reference composed of different thicknesses and densities which is imaged alongside the target tissue) on specifically calibrated mammography machines to quantify the volume of dense tissue in the breast.

4.8.4 Magnetic Resonance Imaging (MRI)

Many studies have investigated the feasibility of using magnetic resonance imaging (MRI) to measure breast density [150-156, 158, 424, 425]. The technique is very reliable and repeatable, and unlike mammography, does not involve the use of ionising radiation. The expense of MRI [426] has meant large scale epidemiological MD studies using this imaging modality have not been undertaken. Additionally, standard (contrast-enhanced) MRI breast imaging, such as that utilised in a recent breast density and hormonal exposures project [427], requires use of a potentially harmful contrast agent. The confinement of the body in the magnetic tube is not palatable to numerous people. These issues currently limit its use as a screening tool. However use of non- and pre- contrast MRI techniques to measure breast density [428-430] are available; these provide an alternative method for measuring breast density.

4.8.5 Ultrasound

Like MRI, ultrasound does not involve the use of radiation. Unlike MRI, ultrasound generally is very operator dependent, and it therefore not as consistent a modality as MRI or

mammography for BC screening. Efforts are underway to reduce the amount of dependence upon the skill of the user through use of automated ultrasound systems, such as the commercially available Automated Whole Breast Ultrasound (ABUS) [425, 431-433], and a more experimental ring ultrasound tomography technique [434]. Most ultrasound techniques are quite variable, therefore ultrasound has not been widely used to assess breast density [173]. Whole breast ultrasound tomography, which measures the velocity of sound waves through the breast, is highly correlated with %fibroglandular volume using non-contrast MRI [429] and Cumulus measured MD[435], and may prove useful as a screening tool for breast density.

4.8.6 Tomosynthesis

X-ray tomosynthesis is a specialised type of mammography, and is often used in conjunction with standard mammography systems. It uses a slit opened at regular intervals to emit x-ray beams from a source that travels in an arc over the breast. This produces approximately eight slices which are reconstructed into a three dimensional representation of the breast. Volumetric density is estimated from the reconstructed volumetric representation of the breast [382, 436]. Because multiple x-rays of the breast are taken from different angles, tomosynthesis diminishes the capacity of dense tissue to obscure tumours. This technique may prove superior to standard screening mammography for women with dense breasts [437, 438]. Density estimated on tomosynthesis, digital mammograms and MRI is highly correlated [439]. However, breast density may be underestimated with digital breast tomosynthesis relative to BI-RADS (digital mammography) density [439, 440].

4.8.7 CT, PET and other techniques

Other types of imaging such as computed tomography (CT), positron emission tomography (PET), and DXA/single(S)XA are potentially useful in the assessment of breast density [441, 442]. PET and CT require higher levels of ionising radiation compared to mammography and

are also more expensive. Therefore they are not desirable modalities for use in general BC screening programs. The radiation levels for breast density DXA/SXA are low compared to mammography (30μ Sieverts(Sv) for single breast DXA vs 450μ Sv mammography screening (x4 views)[443]) and are of potential use for breast density assessment [444]. The SXA technique has been shown to vary by about ±2% longitudinally [445], and thus may be suitable for longitudinal studies where (like IBIS-II) the estimated change in MD is likely to be small.

A technique called molecular breast imaging (MBI), developed at the Mayo Clinic in the USA, provides superior detection of cancer in dense breast tissue compared to mammography. The radiation doses utilised in molecular imaging currently far exceed those of mammography, but continued improvements in these techniques may yield an acceptable radiation dose comparable to that of a mammogram [446, 447].

Other potential imaging techniques such as optical imaging [448-450], elastography [451, 452] and electrical impedance [453, 454] have also been utilised to measure the dense tissues of the breast [372]. These are not in widespread use, but this may change as the demand continues to increase for non-ionising methods of assessing breast density.

4.8.8 Volumetric techniques in relation to the project

Because mammographic images were the sole breast imaging modality available for IBIS-II participants, the majority of volumetric density assessment techniques were not available for use in this project. Measurement of volumetric density using a research (non-commercial) version of the Volpara software was considered. As described above, however, Volpara requires the 'raw' (unprocessed) digital image from the mammography machine to estimate volumetric density. Only film-screen and processed digital images were available at baseline for the IBIS-II participants, hence use of Volpara was not pursued.

4.9 Conclusion

Breast density has proven difficult to quantify reliably and repeatably [34]. Visual assessment of BD is widespread. Clinical use of fully automated programs that estimate the volume of BD is rising. PD is typically measured in research environments using semi-automated, interactive computer programs. All methods which require human input are subjective, and hence results are less repeatable and consistent than fully automated methods. Although the fully automated methods are very consistent, many misclassify mammograms and most do not provide gradients of risk as strong as interactive methods.

The type of breast imaging available for IBIS-II participants, namely film and digital mammograms, restricted the choice of MD measurement technique which could be utilised. Furthermore, mammograms were collected retrospectively, and use of fully automated techniques which required access to raw mammography data such as Volpara could not be used.
Visual MD assessment techniques are typically quick and easy to implement, but are hindered by issues of reproducibility and accuracy. The popular BI-RADS and Boyd SCC techniques utilise large categories for MD, and thus were unlikely to be able to discern the small longitudinal changes in MD expected for AI treatment. However the 21-category visual technique utilises 5% PD increments. This technique was selected to test for reliability and compatibility with other techniques, since it had the capacity to detect potentially clinically relevant longitudinal differences in MD quickly and easily.

Semi-automated techniques such as Cumulus and Madena are well established in the MD research community. Cumulus was compatible with the IT infrastructure at the ANZ BCTG and Newcastle BreastScreen facility, and was capable of accurately measuring mammographic MD. Cumulus was selected as a technique to test for reliability and accuracy in this project.

A number of fully automated methods to measure MD on retrospectively collected mammograms are available. A technique utilising ImageJ has been utilised successfully with a number

of mammographic sets. However this technique appeared to be difficult to implement, and was not commercially available. The fully automated AutoDensity program was available via a technology evaluation agreement with the CSIRO, hence this technique was selected to measure density on the CMN IBIS-II mammograms.

In summary, the review of techniques for assessing MD identified three potential methods: visual assessment, Cumulus assessment and AutoDensity assessment. Assessment of the reliability of the visual and Cumulus methods will be presented in the following Chapter.

5. Intra-observer and inter-method reliability

This chapter addresses the second Aim of the thesis: to investigate intra- and inter-technique reliability for MD. The chapter includes results from an analysis of intra-observer (intra-technique) reliability for two MD measurement techniques: visual assessment and Cumulus. The interchangeability of visually assessed and Cumulus assessed MD are compared using an inter-technique agreement method. Additional comparisons are made between film mammograms scanned at 4.3 and 4.7 optical density, and between digital mammograms and film mammograms scanned at 4.7 optical density. A further comparison is made between the average of two MD measurements and single MD measurements.

5.1 Introduction

The modality used to capture and retain mammographic images has undergone a transition in Australia throughout the duration of the IBIS-II breast cancer prevention trial, which commenced accrual in Australia in June 2005. Ongoing improvements in image acquisition, technique and storage are an integral part of medical imaging. John Wolfe— who first published the association between breast cancer risk and mammographic pattern [8]— was an expert in the use of the now defunct xeromammography [455], a popular competitor to film mammography in the 1960s. Film mammography was the worldwide standard for breast imaging from the 1960s [456] until late last decade. Improvements in technology led to the use of a screen in conjunction with film to reduce the radiation dose to the breast in the 1970s ("film-screen" mammography) [455]. Like the conventional film used to capture images in still photography and for movies prior to the invention of digital cameras, mammographic films are developed in dark rooms using a series of chemical exposures. This requires manual handling of the film during the various phases of image development, and increases the overall time needed to acquire the mammograms.

Film-screen mammography has been largely supplanted in the developed world by digital mammography. Digital mammography, invented in the mid-1990s, has advantages over film due to its lower radiation dose [457], superior acquisition speed, and increased contrast and portability of images [456]. Randomised controlled trials have shown no overall difference between detection of breast cancers by digital mammography and film-screen mammography. Detection of lesions is improved for digital mammography in some groups, including women with dense breasts, and those under age 50 years [458]. Digital mammography may be less sensitive than film in women with the opposite characteristics (older age, fatty breasts) [459].

Digitised film-screen, digital, and digitised digital mammograms printed to film ('digital on film', DoF) were utilised in this project. The digital mammograms utilised in this analysis are primarily computed radiography (CR) mammograms. CR images are captured on a reusable phosphorous plate housed in a cassette. The cassette is manually inserted by the radiographer into a separate device to be read by a laser before the image is sent to a computer. CR cassettes are used in the place of film cassettes on film-screen mammography machines, thus providing many of the advantages of digital mammography without complete upgrade of an expensive device. In contrast to CR, direct (or digital) radiography (DR) images are captured directly onto a plate which is an integral part of fully digital mammography machines; DR images are sent immediately from the plate to a computer (manual handling of cassettes does not occur).

Unlike film-screen mammograms, CR and DR mammograms are subject to image post-processing by the software of the mammography machine; this software tends to remove some of the dense tissue from the final (processed) image. Post-processing algorithms vary with the vendor of the mammography machine and over time (e.g. software version), hence the dense tissues of mammograms from the same woman may vary in appearance depending upon which technology and type of mammography machine was utilised.

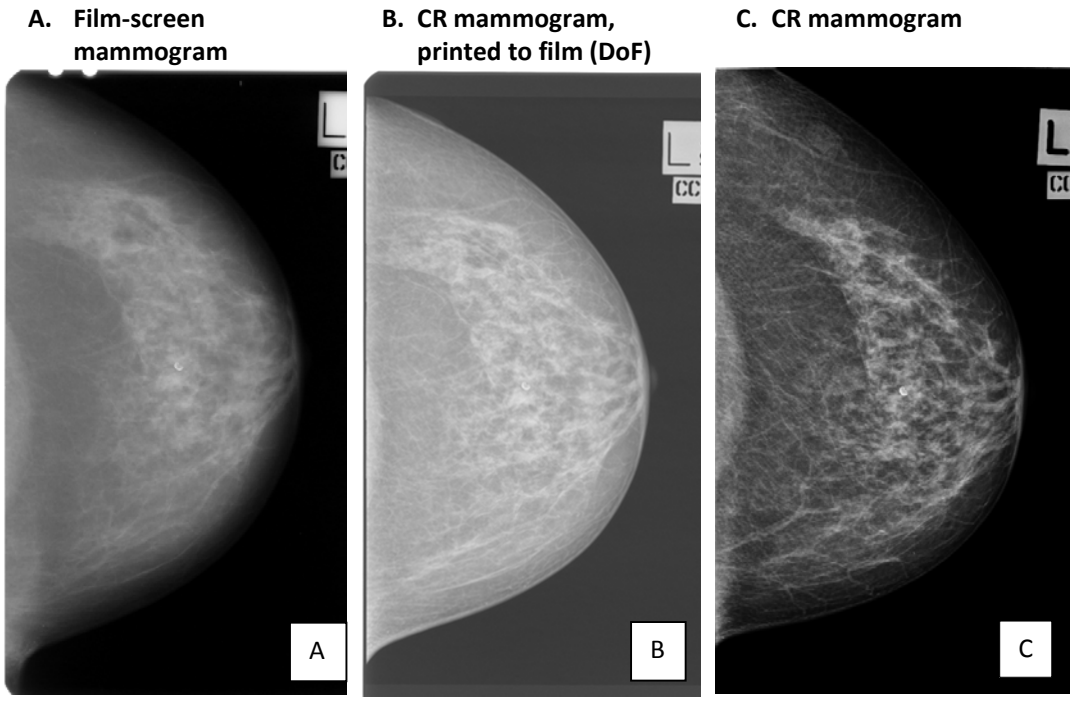


Figure 5-1 Mammograms from an IBIS-II trial participant
A. 2008 digitised film-screen mammogram (left panel); B. 2009 digitised CR mammogram printed to film (digital on film: DoF) (middle panel) C. 2010 computed radiography mammogram (right panel). Density is more diffuse in the film-screen mammogram (A) than in the digital mammograms (B, C). The breast edge is more visible in the digital mammograms, especially for DoF (B). Contrast between the fat (dark) and dense areas of the breast is greatest in the CR image (C).

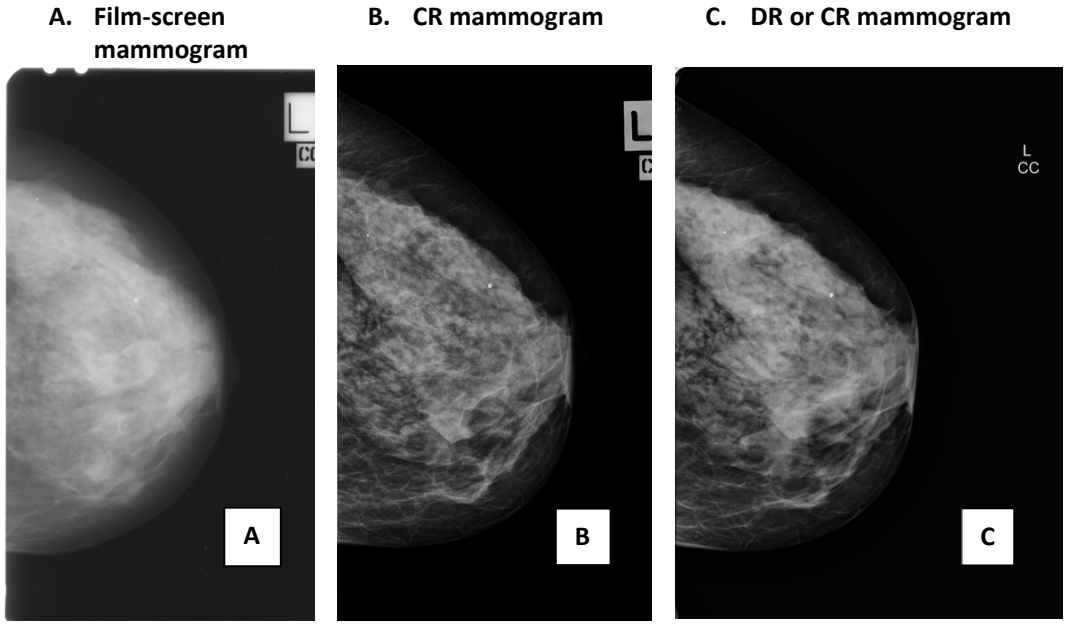


Figure 5-2 Mammograms from an IBIS-II trial participant
A. 2007 digitised film-screen mammogram (left panel); B. 2012 CR mammogram (middle panel) C. 2013 digital mammogram, of CR or DR origin (right panel). The different view label (LCC) for mammogram C indicates that the process that generated this image differs from that of the digital mammogram in panel B.

Film-screen mammograms for this project date from 2005 to 2008. CR mammography started to replace film-screen mammography in 2008 in the lower Hunter (Newcastle) region. Many CR mammograms can be positively identified by their appearance: they are readily distinguished from both film-screen and DR mammograms by the presence of film-screen-like view labels on the digital image (e.g. the LCC labels, Figure 5-1). However, it is not possible to tell whether some digital mammograms are of CR or DR origin solely by viewing the image, e.g. Figure 5-2C.

Many methods to assess mammographic percent density are available, as discussed in the previous chapter. This earlier chapter presented a review of the available techniques to measure mammographic density, and concluded that it would be worthwhile to test three different methods for suitability in this project. Results from Cumulus assessed MD and visually assessed PD to the nearest 5% are presented in this chapter. While MD was also assessed using AutoDensity, the results are not presented here because they are considered confidential information by the CSIRO.

The visual appearance of the DoF mammograms differed from the fully electronic digital mammograms, as did the behaviour of the Cumulus program when selecting the dense area on a DoF mammogram. Film and fully electronic digital mammograms typically had indistinct grey-level differences between the adipose and dense breast tissues, as well as a broad range of graduated grey-levels within the dense tissues. This meant choosing a grey-level threshold which best selected the area of dense tissue was very imprecise (subjective): a large range of grey-levels could be chosen. In contrast, the DoF mammograms had very little grey-level threshold differences between the background (adipose) tissues and the dense tissues, as well as a narrow range of grey-levels within the dense tissues. This made it relatively easy to set a grey-level threshold to select the dense tissues. However it meant that the fully electronic digital mammograms and the DoF were unlikely to be similar representations of the dense

breast tissues when assessed visually or with Cumulus, and thus were excluded from the analysis.

The types of mammograms available for use in this project—film-screen mammograms, digital mammograms printed to film (i.e. copies of digital trial mammograms given to the IBIS-II participants to take home), and fully electronic 'original' digital mammograms— varied over time for trial participants (Figure 5-1, Figure 5-2).

Due to the potential for between–method measurement variation, variability in the measurement method (due to the subjective component), and the likely differences between measurements made on film vs digital mammograms, it was considered important to look at the reliability of MD assessment. Quantification of intra-observer as well as inter-observer reliability when utilising subjective measurement techniques is often reported in the MD literature [189, 394, 399, 460, 461].

This chapter therefore addresses Aim 2 of the thesis: Quantification of intra- and inter-technique measurement reliability.

Several ancillary aims are also addressed in this chapter. The first ancillary aim was to determine if two different acquisition techniques— namely, higher versus lower optical density (OD) scanning— affects MD assessment on the same mammogram. It is unknown if percent density (PD) assessment differs on mammograms scanned at different OD.

Researchers at the University of Toronto advised that Cumulus was compatible only with scans made with an x-ray digitise with at least a 4.0 OD capability. The IBIS-II international trial coordination centre's film-screen mammogram scanning protocol scanned mammograms at 12-bits, 50µ pixel spacing at 4.3OD. However, the Array laser scanner conjointly purchased by the

CMN Department of Surgical Oncology and the ANZ BCTG can scan up to 4.7OD. Higher optical densities correlate to darker greys and blacks. The dense tissues of the breast appear white (lower optical density) and are easily captured by even low (<4) capacity OD scanners. The subcutaneous fat and skin overlying the dense breast tissues appear dark (higher optical density) on film-screen mammograms. Higher OD features such as the breast edge may not be adequately captured on the digitised image when scanning at low OD.

In theory, higher OD scanning will produce a more visible breast edge on the digitised image. User-assessed PD might be lower for film-screen mammograms scanned at 4.7OD compared to 4.3OD, for both Cumulus and visually-assessed PD. This is because the more visible breast edge might increase the total area attributed to the breast, whilst the area of dense tissue remains unaffected. [N.B: the same number of bits (i.e. 12 bits per pixel) then must capture a wider range of greys for 4.7OD vs 4.3OD; this could also possibly affect the assessment of the dense tissue.]

Additionally, mammograms scanned at the highest OD setting (4.7OD) on the Array scanner are likely to produce images with a breast edge similar to the well–defined breast edge of digital mammograms. This is because the higher OD setting should be able to distinguish the dark edge of the breast better than low OD settings. The PD from IBIS-II digital mammograms tends to be lower than that measured from film mammograms. Whilst the lower PD on digital mammograms is likely due to post-processing on the mammography machine, factors such as time between the film and digital mammograms, total area of the breast, and participant age may also influence this discrepancy. Examination of these factors is undertaken in a second (acquisition technique) ancillary aim.

Aim 2 – Intra-observer and inter-method reliability sub-aims

Aim 2 of this thesis was subdivided into separate sub-aims, according to the type of reliability to be tested. This resulted in division of Aim 2 into four sub-aims: intra-observer reliability, inter-method reliability, and x2 acquisition technique reliability as follows:

1.) Intra-observer reliability (**Aim 2.1**): To determine if repeated measurements by an individual on the same mammographic image using the same subjective assessment technique yields similar results. In other words, to determine if measurement 1 is similar to measurement 2 for each subjective assessment technique (Cumulus, visual).
2.) Inter-method reliability (**Aim 2.2**): To determine if measurements made on the same mammographic image by the different assessment techniques (Cumulus and visual) are equivalent ways to assess mammographic density, or if systematic differences between the techniques exist.
3.) Acquisition technique reliability (**Aim 2.3**): Utilising measurements made by the same person using the same technique:
    a. **Aim2.3a**: To determine if MD is similar on film mammograms scanned at 4.3OD vs 4.7OD or if there are systematic differences between measurements made on images scanned with these two optical density settings.
    b. **Aim2.3b**: To determine if measurement of MD differs between film mammograms scanned at 4.7OD and digital mammograms for the same participants

5.2 Methods

As described in Chapter 3, PD and DA are the typical parameters used to represent mammographic density: PD = DA/BA. Hence PD is a parameter derived from a calculation utilising DA and BA. PD may be estimated without formally quantifying the size of the areas for DA and BA (as is done for visual assessment of PD). Alternatively, PD may be calculated if

DA and BA have been quantified (measured). Computer software such as Cumulus typically takes this latter approach. If DA and BA are quantified, then other derived parameters can also be calculated: adipose area (AA) = BA–DA; and percent adipose area (PA) = AA/BA.

Because PD is universal to all common approaches of MD measurement, this chapter primarily addresses the reliability of PD to achieve Aim 2.

To complete Aim 2 of the project, MD was measured using the three techniques selected in Aim 1 (Chapter 4). Digitised film and digital mammograms were assessed by the candidate (sole assessor) visually and rounded up to the nearest 5% PD [296]. The candidate received training in visual assessment using the IBIS-II Mammographic Density training DVD developed by Dr Ruth Warren and CRUK. Digitised film and digital mammograms were assessed for MD by the candidate using the Cumulus technique [128]. The candidate received training in the use of Cumulus from staff at the University of Melbourne prior to assessing the mammograms used in this project. AutoDensity was developed for use on left (L) CC mammograms [411]; for convenience, principally LCC View film and digital mammograms were selected for Aims 2.1 and 2.2.

Mammograms used for this Aim (Aim 2) were those that were obtained by the end of 2012 (Collection 1). (Additional mammograms were obtained for Aims 3 to 5 (Collection 2, in 2014).) Mammograms are also referred to as images or Views throughout this thesis. A mammographic episode consists of one or more Views (typically the four standard Views: RCC, LCC, RMLO, LMLO) taken during a single mammography session. Although four standard Views usually comprise all of the images in a mammography episode, occasionally other Views such as the medio-lateral (ML) view or magnifications were present for digital mammograms. One or more Views were sometimes missing from episodes, or not possible to obtain (e.g. due to unilateral mastectomy).

Table 5-1 Mammogram subsets utilised in Aim 2

Collection 1 mammogram subset	Total partici-pants	Total mmgs	Unique mmgs	4.7 OD film mmg	4.3 OD film mmg	Digital mmg (all)	Digital – non DoF mmg	DoF mmg	Notes
All of Collection 1	112[2]	1712[1]	1333[2]	168[1]	168[1]	1312[1]	1144[1] 1113[2]	117[1]	Collection 1 mmgs used in Aim 2 (this chapter) include duplicated images
Aims 2.1 & 2.2: Intra-observer & inter-technique reliability									
Film	40	130	72	57	57	--	--	--	Film includes 30 RCC Views; digital are all LCC Views.
Digital (no DoF)	44	125	125	--	--	125	125	--	
Aim 2.3a: Film 4.3 v 4.7OD	29	336	168	168	168	--	--	--	All Views utilised All mmgs duplicated at 4.3 & 4.7OD
Aim 2.3b: 4.7OD v digital	25	232	232	116	--	116	98	18	All Views utilised. Views matched to each other during comparison.

[1] Collection 1 mammogram totals, includes duplicated mammograms; [2] Cleaned data, DoF (digital printed to film), and duplicate films excluded

Table 5-2 Mammogram Selection Criteria for each Aim

Aim	Mammogram Selection Criteria
Aims 2.1 & 2.2: Intra-observer & inter-technique reliability	CC Views utilised
Film mammograms	Available LCC mammograms (n=57) were supplemented with 15 RCC View mammograms to better match the number of mammograms available for fully electronic digital mammograms.
Digital mammograms	125 unique LCC View digital mammograms from 37 participants who had both film and digital mammograms were selected.
Aim 2.3a: Film 4.3 v 4.7OD	All available film mammograms with both 4.3 and 4.7 OD scans were utilised. Hence all 4 Views (RCC, LCC, RMLO, LMLO) were selected.
Aim 2.3b: Film v digital	The most recent 4.7OD film mammogram was matched to the oldest digital mammogram for participants with 4.7OD film mammograms. This was done to minimise longitudinal differences in MD due to time.

Mammographic density was measured on different subsets of mammograms from Collection 1, Table 5-1. Collection 1 consisted of all available Views from more than 400 non-unique mammographic episodes representing 112 distinct CMN IBIS-II participants. Non-unique (duplicate) mammograms were collected for the purposes of achieving Aim 2.3 Acquisition technique reliability (e.g. Aim 2.3a: compare PD measured on 4.3 vs 4.7OD images of the same LCC mammogram).

Forty–three of 57 film episodes were scanned by the candidate at both 4.3OD and 4.7OD; 14 film episodes not duplicated both at 4.3OD and 4.7OD were collected by the ANZ BCTG. All available fully electronic LCC digital mammograms for the 37 participants with both film and digital mammograms were selected for the digital mammogram set; the DoF were excluded because selection of the dense area in the Cumulus program differed for DoF vs fully electronic digital mammograms.

The mammogram selection criteria for each Aim in this chapter are shown in Table 5-2. Review of Collection 1 revealed that 57 unique film LCC mammograms were available. All available (fully electronic) LCC digital mammograms for the 37 participants with both film and digital mammograms were selected for the digital mammogram set.

5.2.1 Visual and Cumulus MD measurement

For Aims 2.1, 2.2 and 2.4, the 72 film-screen mammograms were assessed visually and with Cumulus on 9 July 2014 (measurement 1 film) and 12 August 2014 (measurement 2 film). The 125 digital mammograms were assessed using visual and Cumulus methods on 25 September 2014 (measurement 1 digital) and 6 November 2014 (measurement 2 digital). Each assessment was preceded by ‘recalibration’ with the visual assessment technique by retraining on the IBIS-II Mammographic Density training DVD by Dr Ruth Warren/Cancer Research UK (CRUK). The film and digital mammograms were batch read within Cumulus. The 72 film mammograms

were read as one batch and the 125 digital mammograms were read separately in another batch. The assessor was blinded to the percentage calculated by Cumulus during batch reads. The mammograms were first read visually (within the Cumulus program), and subsequently measured using the Cumulus program.

Mammograms collected during Collection 1 were assessed using Cumulus for Aim 2.3 (and Aims 3 to 5) in April 2013. The images were assessed in 16 batches of 100 to 150 mammograms per batch. Batches were composed of all available mammograms (film-screen and digital) for several different participants. Mammograms from each participant were read as a group (sequentially), but in random order: the RCC, LCC, RMLO and LMLO Views for each participant from different episodes were intermixed with each other. Each batch was read during a single sitting (e.g. during 2 consecutive hours on a single day). Thus the 4.3OD and 4.7OD mammograms for each participant were assessed for density in Cumulus within a short time of each other. Visually assessed PD for Aim 2.3 was batch-read using a slightly different paradigm than described for Aims 2.1 and 2.2, because randomly ordered Cumulus batches are time-consuming to prepare. A Bland-Altman plot (not shown) was generated to confirm a significant (>5%) bias did not exist between the two different visual assessment methods.

Method for Aim 2.1 (Intra-observer reliability)

To achieve Aim 2.1— intra-observer (intra-method) reliability for the two subjective measurement techniques (Cumulus vs visual)— two sets of MD measurements were made on 72 cranial-caudal (CC) film-screen mammograms and 125 CC digital mammograms using each measurement technique.

Method for Aim 2.2 (Inter-method reliability for MD)

To achieve Aim 2.2 for PD, the set of repeated measurements (measurement 1 and measurement 2) made on 72 film and 125 digital CC View mammograms in Aim 2.1 were used to compare PD for the Cumulus and visual measurement techniques.

Method for Aim 2.3 (Acquisition reliability)

To achieve Aim 2.3a— acquisition technique reliability for 4.3OD vs 4.7OD film mammograms— 168 film mmgs (i.e. all Views: RCC, LCC, RMLO, LMLO) digitised at both 4.3OD and 4.7 OD were compared for differences in MD measured using the Cumulus and visual techniques. All mammographic Views were utilised because there is no restriction on type of View for the visual and Cumulus techniques, and to increase the sample size used in the comparison.

To achieve Aim 2.3b —Acquisition reliability 4.7OD vs digital— the most recent 4.7OD film mammograms for each participant (n=29) were paired with the oldest fully electronic digital or DoF mammogram for that participant (i.e. the smallest time between the film and digital episodes was selected). Views were paired (e.g. the RCC view of the 4.7OD film was paired with the RCC view of the digital mammogram, the RMLO views of the 4.7OD and digital mammograms were paired, etc.). Because PD measured on DoF mammograms differs from that of original (fully electronic) digital mammograms, separate comparisons are made for the 18 DoF and 98 4.7OD film mammograms for which a fully electronic (original) digital mammogram was available. The average difference in PD and duration between the film and original digital mammograms is tabulated for the 18 DoF and 98 fully electronic digital pairs of mammograms, and also tabulated by time between each fully electronic digital pair (~1, 2 and 3 years difference).

5.3 Statistical Methods

Most analyses in this Chapter contain multiple mammograms from each participant. The effects of measurement error and the differences between the measurement techniques were assumed to be far greater on PD measurement than the influence of the similarities from multiple mammograms from participants. Additionally, longitudinal sets of mammograms inherently

contain repeated measurements for individuals, hence utilising mammograms from different episodes but from the same participant for this Aim reflected the nature of the data for the Primary Aim of this thesis.

The Bland Altman method and Intra-class correlations (ICCs) are the core methodologies used to examine repeatability and reliability of the selected MD measurement techniques. These methods are described in the next two sections.

5.3.1 The Bland Altman reliability method

The Bland Altman method is comprised of graphical and quantitative techniques to assess how well two measurements agree [462]. The Bland Altman "Statistical Method for Assessing Agreement Between Two Methods of Clinical Measurement" was listed as #29 of the most highly cited papers in 2014 [463], and thus is a well-established, popular procedure for assessing agreement among two methods.

Use of the Correlation Coefficient 'r' (or rho: 'ρ') when comparing two methods for agreement is misleading [462]. Two methods may be highly correlated (linearly related), but because the absolute values of the measurements may differ they may not agree very well. If two methods do not agree their measurements cannot be simply interchanged with each other. Many different methods exist for measuring MD. Some are more time consuming to perform than others, and not all are likely to agree. It was necessary to evaluate if the methods selected for use in this project could be readily interchanged with each other.

The Bland Altman method utilises graphs and simple calculations to assess the agreement between two methods. This same procedure can be used to measure agreement of repeated measurements of a single technique, e.g. to calculate how much variability exists when repeating the same measurement technique on same mammograms. A Bland Altman plot is a

scatter plot which graphs the difference between the two measurements on the y-axis and the average of the two measurements on the x-axis. Two measurements that agree should have a very small average difference (bias) between the two measurements, and the variability of the differences around the bias should likewise be small (e.g. the variability should not be clinically significant). The variability of the difference between each pair of measurements is plotted as a line two standard deviations (±2SD) from the bias, and the values ±2SD from the bias are called the limits of agreement (LOA) for the two measurements. If the LOA for two different techniques is within (clinically) acceptable limits, this indicates that the two methods may be used interchangeably.

Ideally, the scatter of the measurement errors (the difference in the two measurements) should be symmetrically and randomly scattered around the bias of the measurements. This shows that the errors are normally distributed, and that the measurement error is the similar for the range of values tested. However, if the scatter of the errors is funnel shaped, this indicates that the measurement error is related to the size of the measurement. This problem typically occurs when the measurement error increases as the size of the measurement increases, which implies that the LOA is smaller for smaller measurements than for larger measurements.

5.3.2 The ICC method

ICCs from the repeated measurements performed for Aim 2.1 were calculated in keeping with the MD literature. ICCs "were devised to deal with the relationship between variables within classes" [464]. A class describes a set of closely related items; a typical example is identical twins. Identical twins can be assigned into two groups randomly. Differences in random assignment into two groups will cause arbitrary (non-systematic) variation in the correlation between the two groups of twins.

The ICC is the average correlation across all possible orderings of pairs into different groups. An ICC can be calculated for repeated measurements using the same technique because the only

difference between the measurements is the random measurement error inherent in the technique. When an ICC is calculated for repeated measurements, the correlation is "essentially a ratio of the variability between the subjects (being measured) to the total variability" [464].

However, ICCs are not suitable for assessing agreement between different measurement techniques because each technique may represent a different class of measurement. If measurements from different techniques are not of the same class they are not 'within' the same class; thus they are not suitable for the ICC: a within ('intra') class correlation. Unlike twins, who can be randomly assigned into either of two groups, measurements made with techniques cannot be randomly assigned into two groups— all measurements from one technique have to be assigned into one group, and measurements from the other must be assigned to the second group. There is a non-random order to group assignment for the two techniques; they are not exchangeable dyads. Such measurements cannot be interchanged because they are of different classes, and therefore should not be compared with an ICC.

In general, a correlation measures "the strength of a relation between two variables, not the agreement between them" [462]. Perfect correlation will exist between two measurements which line up along any straight line, whilst perfect agreement only occurs if the straight line has a slope of 1. A change in scale will not affect correlation, but this does affect agreement. Correlation will be greater for samples with a wider range than for samples from a smaller range. Highly correlated measurements may not agree very well, and the reverse is also possible [464].

The ICCs utilised in this chapter modelled absolute agreement between individual measurements using two-way random effects: the variability for both the rater (observer) and target (mammograms) were included as (two) random effects in the model. This virtual construct of two-way random effects allows the both the observer and the mammograms which were measured to theoretically have been selected randomly from a pool of similar observers

and similar (high risk, IBIS-II participant-like) mammograms. Use of these random effects during statistical modelling of the ICCs permits the result of the ICCs to be generalised, and viewed as a result which arose from a population of similar raters and mammograms.

The interpretation of the ICC would differ if the observer and/or mammograms were viewed as a fixed population which encompassed the entire population/s of interest. The observer and/or mammograms would then be treated as fixed effects. The interpretation of the ICC is not generalisable to a population of similar raters and/or mammograms, but instead only describes the results for a particular observer or set of observers and set of (IBIS-II) participants and their mammograms.

For consistency with the MD literature, correlation coefficients for the repeated measurements of the same technique were also calculated. Correlation coefficients were also generated for the inter-technique analyses, for comparison with the Bland Altman method outcomes.

5.3.3 Statistical methods for each Aim

5.3.3.1 *Intra-observer reliability (Aim 2.1) (intra-technique reliability):*

- Descriptive statistics for each measurement technique were undertaken. Box plots were created for measurement 1 and measurement 2 for PD measured from 72 film and 125 fully electronic digital mammograms. Measures of central tendency (median, quartile 1 (Q1) and quartile 3 (Q3)) for the measurements from each technique were tabulated.
- Bland-Altman plots for Cumulus and visually assessed PD were created. The limits of agreement were expected to be within ±10% absolute percent density (PD) of the average difference (bias) between each pair of measurements (measurement 1, measurement 2) for each mammographic image, for each technique

- Bland-Altman statistics and Intraclass Correlation Coefficients (ICCs) were calculated for measurement 1 and measurement 2 for each technique for each set of film and digital mammograms for PD, DA, BA and AA.
- ICCs and Bland Altman parameters for subsets of these mammograms were also calculated (sensitivity analyses), described as follows:

5.3.3.1.1 Sensitivity analyses for the Aim 2.1 ICC comparisons

Some participants with film mammograms did not have matching digital mammograms. more closely match the numbers of mammograms within each participant in the data set, a second set of ICCs and Bland Altman parameters for a subset of 53 LCC mammograms, one per participant for each Type (film, digital) were calculated.

5.3.3.2 *Inter-method reliability (Aim 2.2)*

Comparison of the repeated measurements of each technique (inter-technique reliability)

- Bland Altman plots of the repeated measurements of each technique were superimposed for PD.

5.3.3.3 *Acquisition technique reliability (Aim 2.3):*

- Aim 2.3a (4.3OD vs 4.7OD): Bland Altman plots of the repeated PD measurements on film mammograms scanned at 4.3OD and 4.7OD were created for each measurement technique (Aim 2.3a). A variance ratio test was undertaken to compare the variability of PD assessed visually on 4.3OD mammograms with the other OD and assessment technique combinations.

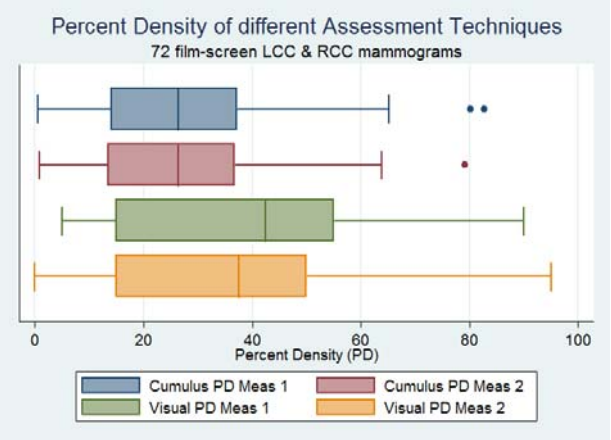
- Aim 2.3b (4.7OD vs digital): The mean, standard deviation, and median difference for PD were tabulated as well as the number of follow ups and years between the most recent 4.7OD film and oldest digital mammograms. Separate tables were made for fully electronic and DoF mammograms. Additional tables were created for mammograms 1 (<1.5years), 2 (1.5 to <2.5) and 3 (2.5 to <3.5)

years apart (original fully electronic mammograms only). The median difference in PD for all paired (within-table) comparisons was tested using a paired signed rank test. Median PD between the subgroups of mammograms 1, 2 and 3 years apart was tested using a median test for unmatched data and a nonparametric trend test. Age at randomisation for participants in the 1, 2, and 3 year groups were tested for equality using a multivariate test for means.

5.4 Results

5.4.1 Aim 2.1 Intra-Technique reliability

The PD distributions of Cumulus measurement 1 and measurement 2 for film mammograms are similar (Figure 5-3). The Visual PD distributions of measurement 1 and measurement 2 are also similar, but differ from the Cumulus PD distributions. Cumulus assessed PD has a lower median (~26%; Quartile 1 to Quartile 3 (Q1-Q3) 14 to 36%) compared to Visually assessed PD (~40%; (Q1-Q3) 15 to ~55%).



PD, 72 film-screen mammograms
Distribution median and IQR

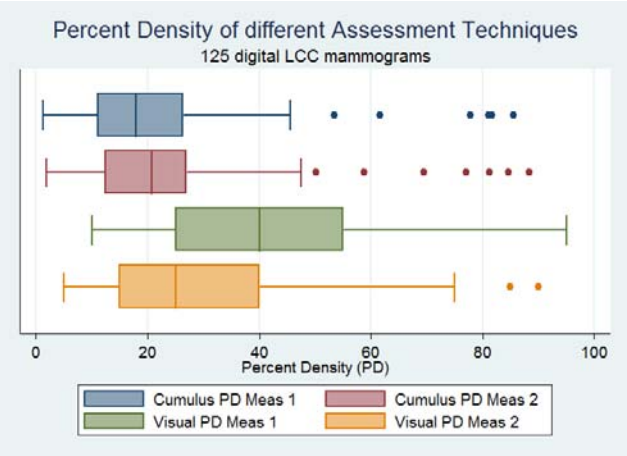
	Median	Q1 to Q3
Cumulus		
Measurement 1	26.4	14 to 37
Measurement 2	26.4	13 to 37
Visual		
Measurement 1	43	15 to 55
Measurement 2	38	15 to 50

Measurement 1 and Measurement 2 are PD estimates made on the same set of mammograms but at different times (at least 30 days apart) by a single reader

Figure 5-3 Distributions of Percent Density for 72 film screen mammograms

The distributions for Cumulus measurement 1 and measurement 2 for digital mammograms are similar as shown in Figure 5-4, but have less overlap than their film-screen counterparts in Figure 5-3. The distributions for measurement 1 and measurement 2 for Visually assessed density are higher and broader than Cumulus assessed PD. Median PD for Cumulus assessed

density is ~20% (Q1-Q3 ~13 to 37%), whilst median PD for visually assessed density is ~32% (Q1-Q3 ~20 to 50%). The two sets of visually assessed PD measurements are less similar than their Cumulus assessed counterparts, and also less similar than their visually assessed film-screen counterparts.



PD, 125 digital mammograms
Distribution median and IQR

	Median	Q1 to Q3
Cumulus		
Measurement 1	18.9	14 to 37
Measurement 2	20.7	13 to 37
Visual		
Measurement 1	40	25 to 55
Measurement 2	25	15 to 40

Measurement 1 and Measurement 2 are PD estimates made on the same set of mammograms but at different times (at least 30 days apart) by a single reader

Figure 5-4 Distributions of Percent Density for 125 digital mammograms

Within-observer (intra-technique) reliability for visually assessed and Cumulus assessed PD was examined using Bland Altman plots, Figure 5-5 to Figure 5-8. Film and digital mammogram were examined separately due to the different distributions of these Types of mammograms. The Bland Altman parameters for these comparisons are tabulated in Table 5-3.

Variability was higher for digital mammogram assessments than for film mammogram assessments; Cumulus assessments were less variable than visual assessments. The limits of agreement (LOA) were ± 13%, Visual PD and ± 8%, Cumulus PD for film-screen mammograms. The limits of agreement were ± 29% Visual PD and ± 10% Cumulus PD for digital mammograms. Bias (mean difference) was <2.5% for measurement 1 and measurement 2 measurements for film-screen mammograms (visual and Cumulus assessed) and for Cumulus assessed digital mammograms. However intra-technique bias was large for visually assessed digital mammograms: nearly 12%.

Film mammogram intra-technique Bland Altman plots

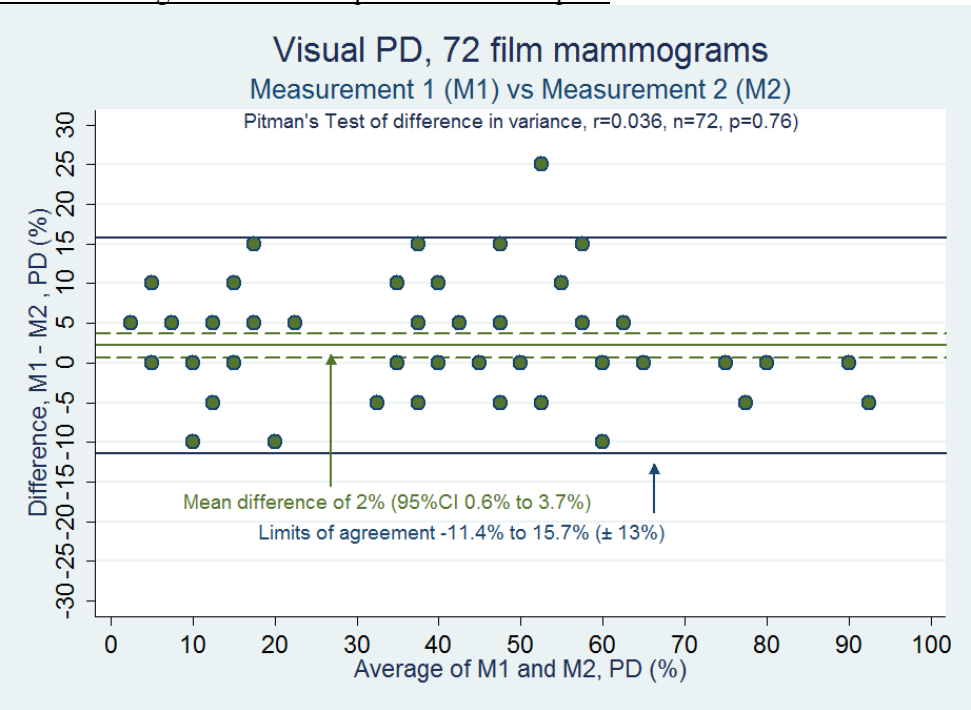


Figure 5-5 Bland-Altman plot, 72 film-screen mmgs, Visually assessed PD
Measurement 1 and Measurement 2 are PD estimates made on the same set of mammograms but at different times (at least 30 days apart) by a single reader

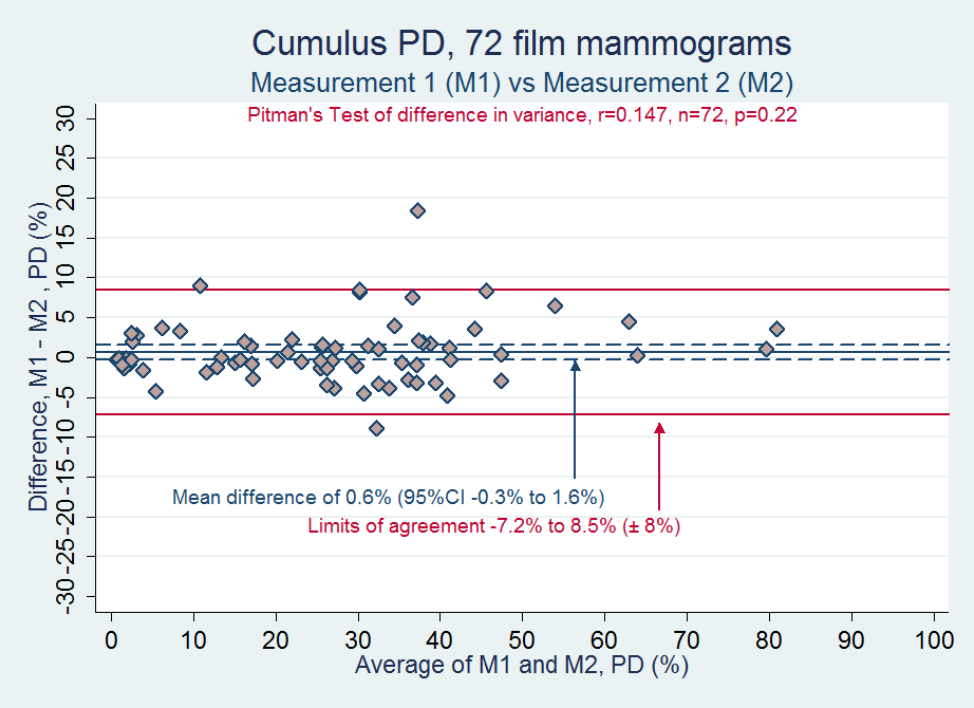


Figure 5-6 Bland Altman plot, 72 film mammograms, Cumulus assessed PD
Measurement 1 and Measurement 2 are PD estimates made on the same set of mammograms but at different times (at least 30 days apart) by a single reader

Digital mammogram intra-technique Bland Altman plots

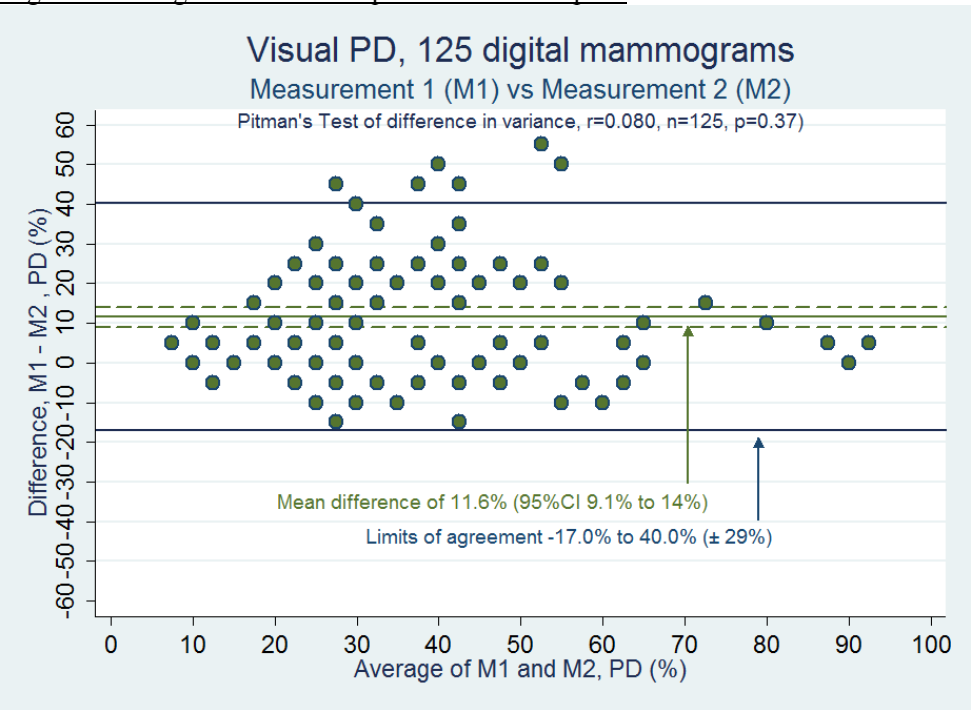


Figure 5-7 Bland-Altman plot, 125 digital mammograms, Visually assessed PD
Please note larger Y-axis scale (± 60) compared to other Bland Altman graphs (±30). Measurement 1 and Measurement 2 are PD estimates made on the same set of mammograms but at different times (at least 30 days apart) by a single reader

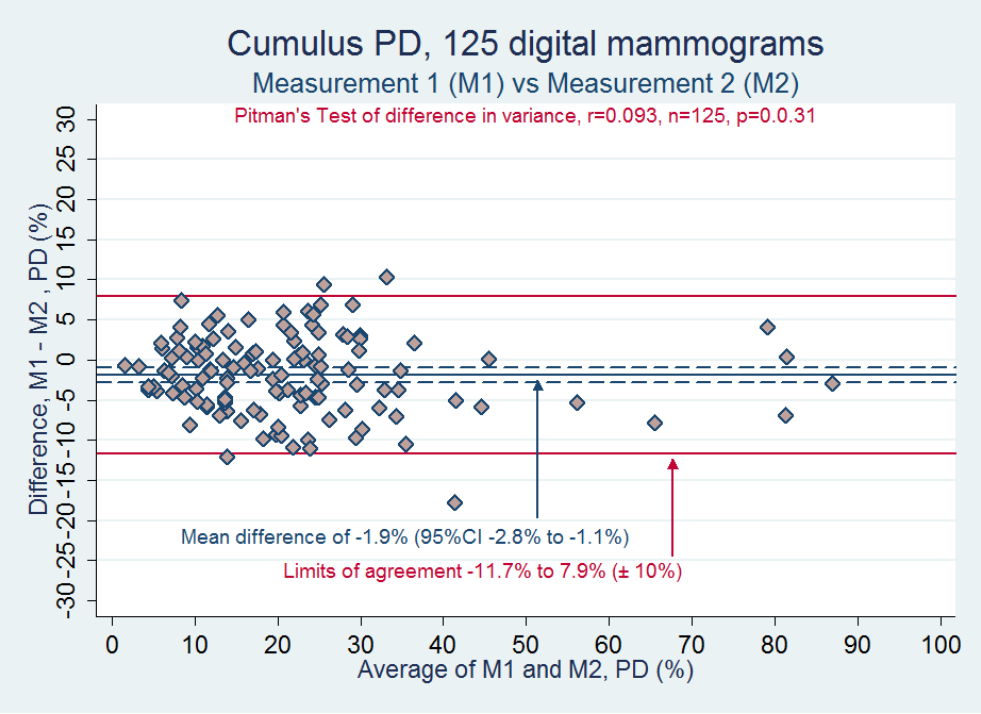


Figure 5-8 Bland-Altman plot, 125 digital mmgs, Cumulus assessed PD
Measurement 1 and Measurement 2 are PD estimates made on the same set of mammograms but at different times (at least 30 days apart) by a single reader

5.4.2 ICCs, film and digital mammograms

ICCs for each technique and mammogram Type were calculated, Table 5-3; ICCs for the subgroups of film and digital mammograms described in the methods section of the chapter are also included in the tables.

Because PD only is measured when visually assessing mammograms, an ICC for only PD can be calculated with this measurement technique. PD, dense area (DA) and BA are all quantified during Cumulus measurement, and thus ICCs for each of these MD parameters were calculated. Additional ICCs for subsets of mammograms, as described in 5.3.3.1.1, are tabulated. Bland Altman plot parameters (bias (95% CI), LOA) were tabulated for all ICC comparisons, Table 5-3.

Intra-technique correlation was generally high for the 72 film mammogram and 125 digital mammogram comparisons for all MD phenotypes (PD, DA, AA) and BA. The results for PD were similar to those for DA and AA: for the 72 film mammograms, visual PD ICC was 0.95 (95%CI 0.92 to 0.97) and Cumulus PD ICC was 0.97 (95% CI 0.96 to 0.98). Intra-technique was less highly correlated for the 125 digital mammograms: visual PD ICC was 0.64 (95%CI 0.21 to 0.81), and Cumulus PD ICC was 0.94 (95%CI 0.91 to 0.96). Restriction of the ICC comparison for film to 53 unique mammograms only did not change the ICCs substantially, Table 5-3; a slight improvement of 0.01 was seen for both the visual and Cumulus techniques.

Similarly as for BA, DA and AA, restriction of the 72 film mammograms to 53 unique mammograms did not markedly affect the Bland Altman parameters for PD, Table 5-3. The bias decreased by a slight 0.2% for visual PD, and was unchanged for Cumulus PD; the LOA increased for Visual PD, but decreased for Cumulus PD. As for the ICC comparisons for the different digital mammogram sets, the Bland Altman parameters for PD, BA, DA and AA for the different digital mammograms changed very little (≤2%).

Table 5-3 Median (Q1 to Q3), Bland Altman parameters and ICCs: Film and Digital mmgs, Visual and Cumulus techniques (PD BA DA AA)

ICCs, film and digital mammograms, Cumulus and Visual, Measurement1 vs Measurement2									
Type of Comparison			Bland Altman parameters			ICC parameters			
N	**Measurement technique**	**Median (Q1 to Q3)[1]**	**Bias (95% CI)**	**Lower LOA**	**Upper LOA**	**ICC**	**95% CI**	**F test[3]**	**Prob > F**
Film-screen mammograms (72 mammograms)									
72	Visual PD	42.5 (15 to	2.2 (0.6 to 3.7)	-11.4	15.7	0.95	0.92 to 0.97	45.3	<0.001
72	Cumulus PD	26.4 (14 to	0.6 (-0.3 to	-7.2	8.5	0.97	0.96 to 0.98	84	<0.001
72	Cumulus	6.3 (5.0 to	-0.0016	-0.19	0.19	0.9996	0.9993 to	4394	<0.001
72	Cumulus	1.6 (0.8 to	0.071	-0.51	0.65	0.97	0.96 to 0.98	78	<0.001
72	Cumulus	4.3 (3.4 to	-0.073	-0.68	0.54	0.996	0.993 to	474	<0.001
Film-screen mammogram – matched 1:1 with digital mmgs within-participant (53 mammograms)									
53	Visual PD	35 (15 to	2.0 (-0.1 to	-13.1	17.1	0.95	0.91 to 0.97	39	<0.001
53	Cumulus PD	25.2 (10 to	0.6 (-0.4 to	-6.2	7.3	0.98	0.97 to 0.99	117	<0.001
53	Cumulus	6.3 (5.3 to	-0.000041	-0.22	0.22	0.9995	0.9991 to	3771	<0.001
53	Cumulus	1.5 (0.8 to	0.077	-0.53	0.68	0.97	0.95 to 0.98	68	<0.001
53	Cumulus	4.6 (3.8 to	-0.077	-0.81	0.56	0.996	0.993 to	515	<0.001
Digital mammograms (125 mammograms)									
125	Visual PD	40 (25 to	11.6 (9.1 to	-17	40	0.64	0.21 to 0.81	6.82	<0.001
125	Cumulus PD	17.9 (11 to	-1.9 (-2.8 to -	-11.7	7.9	0.94	0.91 to 0.96	39.17	<0.001
125	Cumulus	7.2 (6.2 to	-0.01	-0.15	0.13	0.99976	0.9997 to	8412	<0.001
125	Cumulus	1.3 (0.8 to	-0.13	-0.91	0.64	0.94	0.90 to 0.96	34.53	<0.001
125	Cumulus	5.9 (4.5 to	0.12	-0.68	0.93	0.992	0.987 to	261	<0.001
Digital mammograms – matched 1:1 with film mmgs within-participant (53 mammograms)									
53	Visual PD	40 (20 to	9.6 (5.5 to	-20	39	0.65	0.32 to 0.82	6.3	<0.001
53	Cumulus PD	15.8 (10 to	-1.6 (2.8 to	-10.4	7.2	0.95	0.91 to 0.97	45	<0.001
53	Cumulus	6.9 (5.8 to	-0.025	-0.19	0.14	0.9997	0.9995 to	7063	<0.001
53	Cumulus	1.2 (0.7 to	-0.11	-0.74	0.51	0.95	0.91 to 0.97	42	<0.001
53	Cumulus	5.6 (4.4 to	0.088	-0.56	0.74	0.996	0.992 to	483	<0.001

ICCs model Two-way random effects, absolute agreement, for individual measurements.

Measurement 1 and Measurement 2 are MD estimates made on the same set of mammograms but at different times (at least 30 days apart) by a single reader

[1] Data from Measurement 1 used to calculate the Median and Quartiles

[2] Area is presented as (raw) raster size from the Cumulus program $x10^6$; 95% CI not presented for bias due to space restrictions in table

[3] The F test compares the ratio of two variances, to ascertain if the variances are significantly different

5.4.3 Aim 2.2 PD Inter-Technique reliability

Bland Altman plots were prepared for the inter-technique comparison of repeated measurements (measurement 1, measurement 2) for the Cumulus vs Visual techniques. Separate Bland Altman plots were prepared for film-screen and digital mammograms for PD, Figure 5-9 and Figure 5-10.

Film-screen mammogram inter-technique reliability

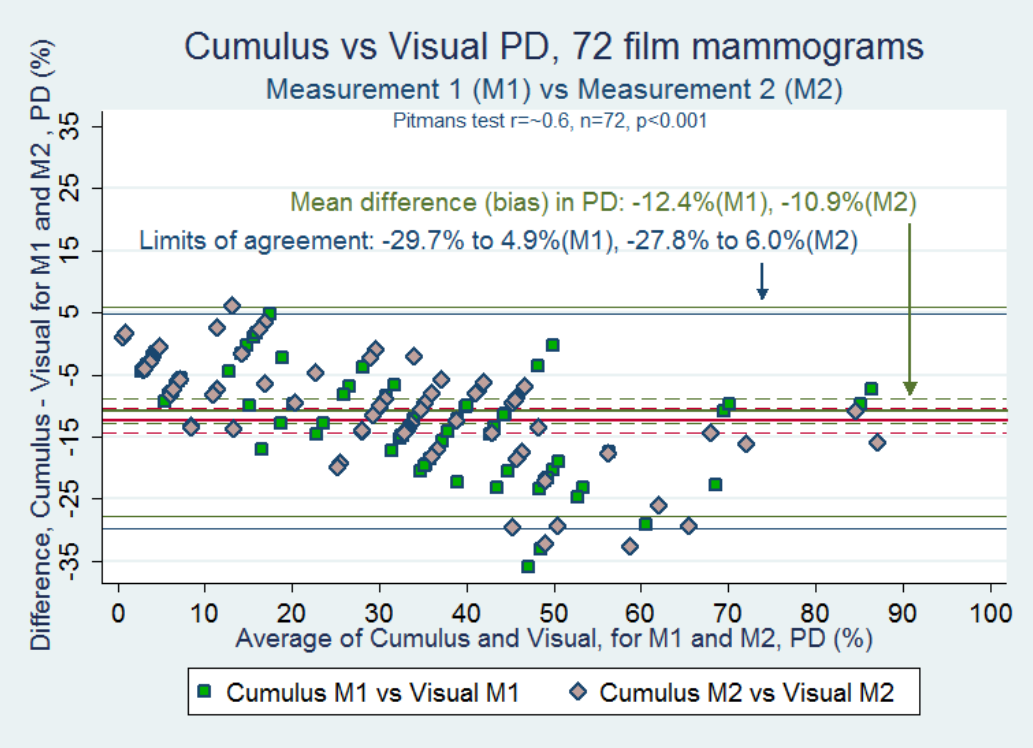


Figure 5-9 Superimposed Bland-Altman plots of Cumulus vs Visual measurements, 72 film mmgs
Measurement 1 and Measurement 2 are PD estimates made on the same set of mammograms but at different times (at least 30 days apart) by a single reader

The bias between each set of measurements in Figure 5-9 is –12% PD. The limits of agreement are approximately ± 17%. The measurements are correlated however: Pitman's test indicates a significant relationship of r = ~0.6. This is evident visually by the downward slope of the plot with increasing average PD. Hence the limits of agreement are likely overestimated. Since a clinically relevant bias of approximately –12 exists, the techniques are not interchangeable. Because of this bias, further analysis was not pursued to more closely ascertain the limits of agreement.

Digital mammogram inter-technique reliability

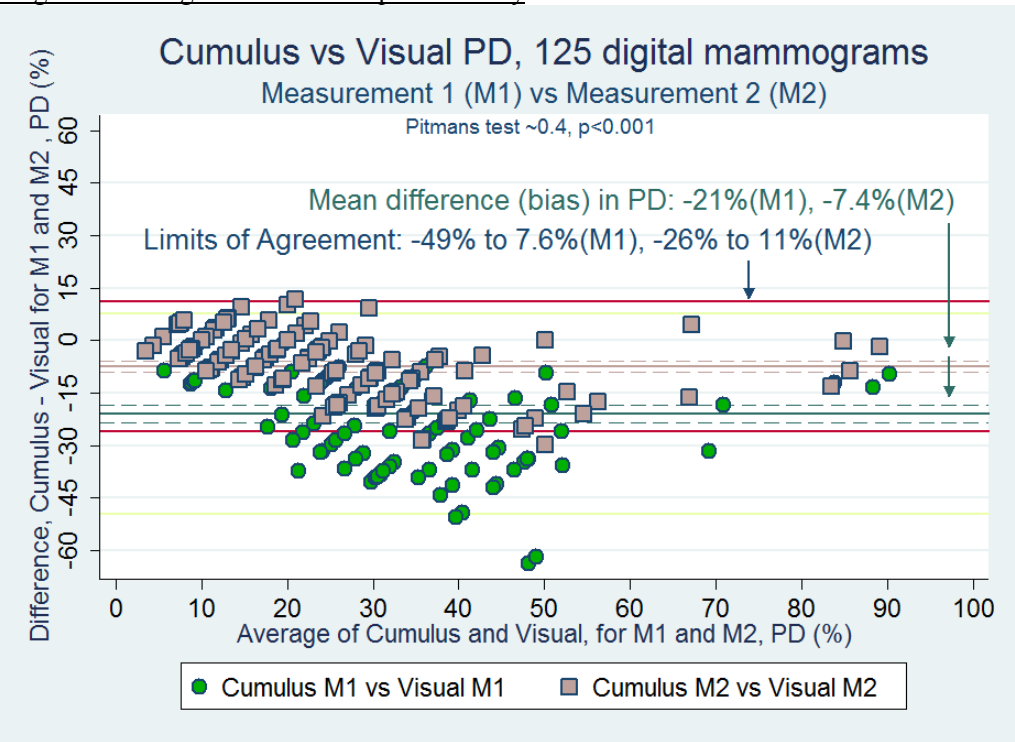


Figure 5-10 Superimposed Bland-Altman plots of Cumulus vs Visual measurements, 125 digital mmgs
Measurement 1 and Measurement 2 are PD estimates made on the same set of mammograms but at different times (at least 30 days apart) by a single reader

The inter-technique Bland Altman plot for digital mammograms (Figure 5-10) resembles the inter-technique Bland Altman plot for film mammograms (Figure 5-9), except the bias and LOA are larger. Pitman's test of correlation is still significant ($p<0.001$) although correlation is lower (~ –0.4) than for film (~ –0.6).

5.4.4 Aim 2.3a PD comparison, film mmgs scanned at 4.3OD vs 4.7OD

PD from film-screen mammograms scanned at 4.3OD (optical density) and 4.7OD were compared, using the visual and Cumulus assessment techniques. The Bland Altman plots (Figure 5-11 and Figure 5-12) show there is very little difference (<1%) between repeated visual and Cumulus measurements made on 4.3 vs 4.7OD film mammograms. The limits of agreement (LOA) are similar (± 9 to 10%) for the visual and Cumulus assessed plots; these LOA are similar to the LOA for film mammograms analysed earlier in this chapter (section

5.4.1). The increase in variability with increasing PD seen for Cumulus assessed PD is not apparent for visually assessed PD.

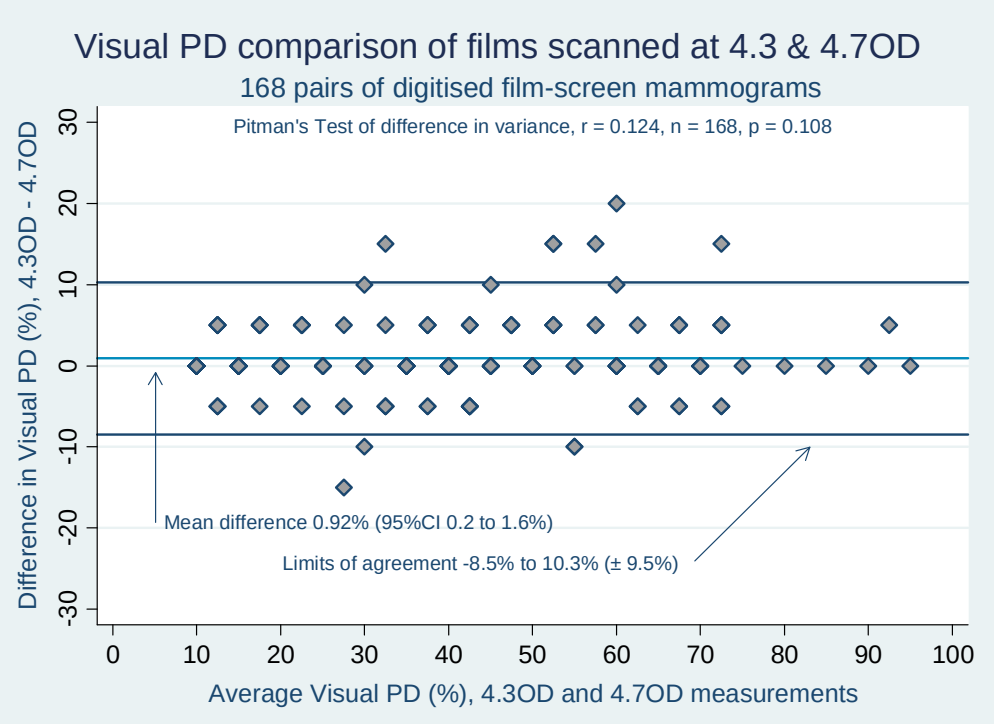


Figure 5-11 Bland Altman plot, visually assessed PD, film mammograms scanned at 4.3 and 4.7OD

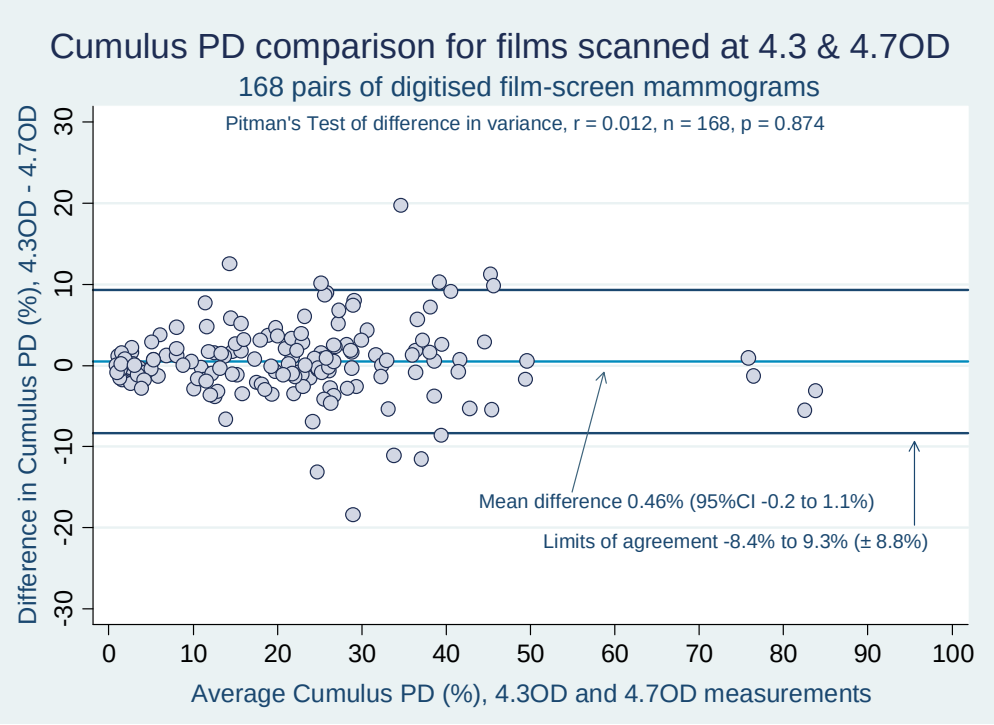


Figure 5-12 Bland Altman plot, Cumulus PD for film mammograms scanned at 4.3 and 4.7OD

The standard deviation (SD) of visually assessed PD from mammograms scanned at 4.3OD was smaller (21.8) than that for the same mammograms scanned at 4.7OD (35.5), Table 5-4. However, these SD were not significantly different ($p = 0.72$, variance ratio test). The SD for the 4.3OD mammograms assessed visually (21.8) was greater than both SDs for 4.3OD (15.7) and 4.7OD (18.6) mammograms assessed with Cumulus. This SD of 21.8 was statistically significantly different from both of the Cumulus assessed SD ($p<0.0001$, variance ratio test).

Table 5-4 Variance ratio tests, 4.3OD and 4.7OD visual PD, 4.3OD and 4.7OD Cumulus PD

Measurement Type	Mean	Standard Deviation	P-value (vs 4.3OD visual PD)
4.3OD Visual PD	39.7	21.8	---
4.7OD Visual PD	38.8	35.5	0.72
4.3OD Cumulus PD	21.4	15.7	<0.0001
4.7OD Cumulus PD	20.9	18.6	<0.0001

5.4.5 Aim 2.3b PD comparison, film 4.7OD vs digital mammograms

The average difference in PD and duration between the film and original digital mammograms is tabulated for all 98 pairs of mammograms (Table 5-5), and also tabulated by time between each pair (1, 2 and 3 years) in Table 5-6 to Table 5-8. A separate comparison was made for the 18 pairs of 4.7OD film and DoF mammograms due to the differences in PD for DoF mammograms (Table 5-9). All mammographic Views were utilised, hence the number of participants is approximately one-quarter the number of mammograms in the analysis.

On average, the difference in PD between the most recent 4.7OD film and oldest digital mammograms was about 8% (SD 8) for all 98 pairs of 4.7OD film and fully electronic digital mammograms, Table 5-5. The median PD difference between the 98 pairs is 5.9%, which is statistically significant ($p<0.001$, Wilcoxon matched-pairs signed-rank test). The mean difference in time between the film and digital mammograms was 2 follow ups, equating to about 2.3 years.

The PD and time differences between 4.7OD film and digital mammogram which differ by one (<1.5 years), two (1.5 to <2.5 years) and three (2.5 to 3.5 years) years are listed below

Table 5-5 PD comparison of paired 4.7OD film and closest digital mammograms, 98 pairs of mmgs

Digitised 4.7OD film-screen mammograms vs fully electronic digital mammograms				
Average age at randomisation 58.6 years (5.8 std), 25 participants				
Metric	**Mean**	**Std dev**	**Median**	**N**
Cumulus PD - 4.7OD (film-screen mmgs) - %	21.6	17.2	20.5	98
Follow up number (4.7OD)	1.5	0.67	1	98
Cumulus PD – Digital - %	13.9	14.7	12.1	98
Follow up number (digital)	3.6	0.97	4	98
Years between 4.7OD and Digital mmgs (year)	2.3	0.77	2.1	98
# Follow-ups between 4.7OD & Digital	2.08	0.81	2	98
Difference between 4.7 and digital PD - %	7.7	7.6	5.9	98

p<0.0001, Wilcoxon matched-pairs signed-rank test, for the comparison PD_4.7 = PD digital

Table 5-6 PD and time differences for ~1 year difference (4.7OD film vs digital mammograms)

Metric	**Mean**	**Std dev**	**Median**	**N**
Cumulus PD - 4.7OD	35.9	23.7	26.3	20
Follow up number (4.7OD)	1.8	1.0	1	20
Cumulus PD - Digital	28.0	26.1	16.0	20
Follow up number (digital)	2.8	1.0	2	20
Years between 4.7OD and Digital mmgs	1.17	0.09	1.1	20
#Follow-ups between 4.7OD & Digital	1	0	1	20
Difference, 4.7 and digital PD	8.0	5.8	8.0	20

Average age at randomisation 61.8 years (3.6 std), 5 participants

Table 5-7 PD and time differences for ~2 year difference (4.7OD film vs digital mammograms)

Metric	**Mean**	**Std dev**	**Median**	**N**
Cumulus PD - 4.7OD	17.7	13.0	14.5	42
Follow up number (4.7OD)	1.4	0.6	1	42
Cumulus PD - Digital	9.6	5.2	9.7	42
Follow up number (digital)	3.4	0.6	3	42
Years between 4.7OD and Digital mmgs	2.1	0.16	2	42
#Follow-ups between 4.7OD & Digital	2	0	2	42
Difference, 4.7 and digital PD	8.0	8.7	6.1	42

Average age at randomisation 57.3 years (6.1 std), 11 participants

Table 5-8 PD and time differences for ~3 year difference (4.7OD film vs digital mammograms)

Metric	**Mean**	**Std dev**	**Median**	**N**
Cumulus PD - 4.7OD	18.9	13.5	21.4	32
Follow up number (4.7OD)	1.5	0.51	1.5	32
Cumulus PD - Digital	11.6	7.3	13.3	32
Follow up number (digital)	4.3	0.84	4.5	32
Years between 4.7OD and Digital mmgs	3.1	0.13	3.01	32
#Follow-ups between 4.7OD & Digital	2.75	0.84	3	32
Difference, 4.7 and digital PD	7.3	7.6	4.2	32

Average age at randomisation 59.2 years (6.8 std), 8 participants

(Table 5-6, Table 5-7, and Table 5-8). Four pairs of mammograms which were more than 3.5 years apart are not tabulated. Because the age of the participants may have an inverse relationship with the decline in PD over time [189], the average age at randomisation for the participants listed in each table is also provided for reference.

The median PD difference for follow up differences of 1, 2 and 3 were statistically different within each group: difference of 1 year, median PD difference 8.0%; differences of 2 years, median PD difference 6.1%; difference of 3 years, median PD difference 4.2% ($p<0.001$ for all comparisons, Wilcoxon matched-pairs signed-rank tests). However, a statistically significant difference in median PD (4.7OD–digital PD) was not found between the groups; i.e. no difference was found between the 1, 2 and 3 years group PD medians of 8%, 6.1%, and 4.2% when they were compared to each other (nonparametric trend test, $p>0.35$). Although the median difference in PD appears to decrease with increasing time between the 4.7OD film and digital mammogram pairs (8% to 6.1% to 4.2%), this trend is not statistically significant ($p=0.48$, nonparametric trend test). The mean age at randomisation for participants in the three subgroups were not statistically different ($p=0.39$, multivariate means test).

Table 5-9 lists the PD and time differences for 18 participants who had a pair of 4.7OD film and DoF mammogram. The median differences of 5.0% PD was statistically significant ($p<0.02$, Wilcoxon matched-pairs signed-rank test) and slightly lower than the median PD difference for fully electronic original digital mammograms (5.9%).

Table 5-9 Digitised 4.7OD Film vs Digital printed to Film (DoF) Mammograms

Metric	Mean	Std dev	Median	N
Cumulus PD - 4.7OD	21.4	10.8	23.8	18
Follow up number (4.7OD)	1.2	0.43	1	18
Cumulus PD – Digital on Film (DoF)	16.9	7.3	17.21	18
Follow up number (DoF)	2.2	0.43	2	18
Years between 4.7OD and DoF mmgs	1.0	0.08	1.0	18
#Follow-ups between 4.7OD & DoF	1	0	1	18
Difference between 4.7 and digital PD	4.5	6.4	5.0	18

Average age at randomisation 61.7 years (5.0 std), 5 participants

5.5 Discussion

This analysis showed that repeated PD measurements utilising either the visual or Cumulus methods varied on average by an amount equal to or greater to the smallest change in PD (10%) associated with a reduction in BC risk [17]. These techniques are likely difficult to utilise for longitudinal measurements where the expected change over time is small (e.g. the expected 1% to 2% PD decrease PD due to anastrozole treatment). Cumulus assessed PD had lower variability (LOA $\leq \pm 10\%$ PD) and smaller biases (<2%) than visually assessed PD (LOA $\geq \pm 13\%$, biases 2 and 12%). This implies Cumulus assessed PD is more reliable than visually assessed PD.

The results also revealed that measurements made with the visual and Cumulus techniques are not interchangeable. The film mammogram inter-technique bias for PD was –12% for both measurement comparisons (M1 vs M1, and M2 vs M2, Figure 5-9). For digital mammograms, the inter-technique bias was –7% for the measurement 1 comparison and –21% for the measurement 2 comparison, Figure 5-10. The M1 and M2 Cumulus and visual assessments were similar for film mammograms (Figure 5-3). The M1 and M2 measurements for digital mammograms made with Cumulus were similar, but the digital mammogram visual assessments were quite dissimilar (Figure 5-4); the dissimilarity of the visual assessments has caused the disparity between the digital mammogram M1 and M2 estimates of inter-technique bias. Both the film and digital mammogram inter-technique biases are approximately equal to or greater than the smallest meaningful clinical difference of 10% PD.

The intra-observer ICCs of 0.95 to 0.97 for PD on film mammogram repeated measurements with the visual and Cumulus assessment techniques are high (Table 5-3), and similar to other film mammogram repeated measurements ICCs reported in the MD literature [189, 192]. The ICCs of 0.94 to 0.95 for Cumulus repeated measurements of digital mammograms are also similar [400]. The ICCs for visual assessments of digital mammograms were still much lower

(~0.6) than for Cumulus (~0.9), even after restriction to make the digital mammogram set more similar to the film set.

Although there numerous ways to calculate random measurement error (precision) [465], including ICCs and least significant change (LSC) [465] as described by Gluer et al, the Bland Altman method used in this thesis uses absolute units to assess precision with 95% confidence via the LOA (limits of agreement). This means the LOA is independent of the MD values measured. Other estimates of precision, including the LSC—defined as a change exceeding $2\sqrt{2}$ times the precision error of a technique—are dependent upon the size of the values measured (e.g. the Coefficient of Variation (CV) for DXA scans), and hence are less desirable than the LOA (also known as the smallest detectable difference (SDD)) for use in daily clinical practice [465]

The descriptive graphs in Figure 5-3 and Figure 5-4 (Aims 2.1 and 2.2) show median percent density is lower for each method of assessment for digital mammograms compared to that for the film-screen mammograms, with the exception of Visual measurement 1. This is not unexpected, due to a combination of factors. Film-screen mammograms from women in the IBIS-II trial predate digital mammograms. Hence the average age of the women contributing digital mammograms to this analysis is likely to be older than for film-screen mammography in this analysis, and PD declines with age. Digital post-processing of the digital mammographic images likely also plays a large role. Breast cancer detection is enhanced for women with dense breasts using digital mammography compared to film (0.59 vs 0.27 sensitivity) [459], which is likely due to removal of density from the image during digital post-processing.

The large intra-observer (intra-method) bias and LOA for visually assessed digital mammograms (Figure 5-7, Figure 5-8) in contrast to the film-only repeated measurements (Figure 5-5, Figure 5-6) was potentially influenced by a number of factors. Second reads of the film mammograms both visually and with Cumulus not included in measurement 1 and

measurement 2 data sets were performed a few days after the first measurements. In keeping with another report in the literature [400], this additional set of measurements ('measurement A') was excluded from further analyses because it occurred < 1 month after measurement 1. This additional reading of the film mammograms but not digital mammograms may have increased the similarity of the film Measurement 1 and Measurement 2 assessments.

The large bias in visual reading of the digital mammograms could also have been influenced by the presence of just one digital mammogram among the set of training mammograms in the IBIS-II density training DVD (the remainder were film-screen). Retraining was undertaken with the IBIS-II mammographic density training DVD prior to reading each set of mammograms (film and digital) visually for this Aim; hence visual MD assessment retraining occurred at least 5 times during 2014. Furthermore, multiple training sessions had been undertaken since 2010, when a copy of the IBIS-II density training DVD was first received. Re-training (recalibration) was also undertaken before each of 15 Cumulus batch reads of the Collection 1 IBIS-II mammograms in 2013. Digital mammograms have an appearance distinct from that of film mammograms. Because reinforcement of PD visual assessment occurred only with film-screen mammograms on the IBIS-II training DVD, this could have reduced the ability to consistently assess PD visually on digital mammograms. The multiple Views from the same episode for film but not digital mammograms may have slightly increased the reliability of the film mammogram assessments

Visual reading of the set of mammograms (film and digital) occurred before assessment of the mammograms by Cumulus for each measurement (measurement 1 and measurement 2). The prior exposure to each set of mammograms visually before assessing the same set of mammograms with Cumulus may have increased the accuracy of the Cumulus assessments.

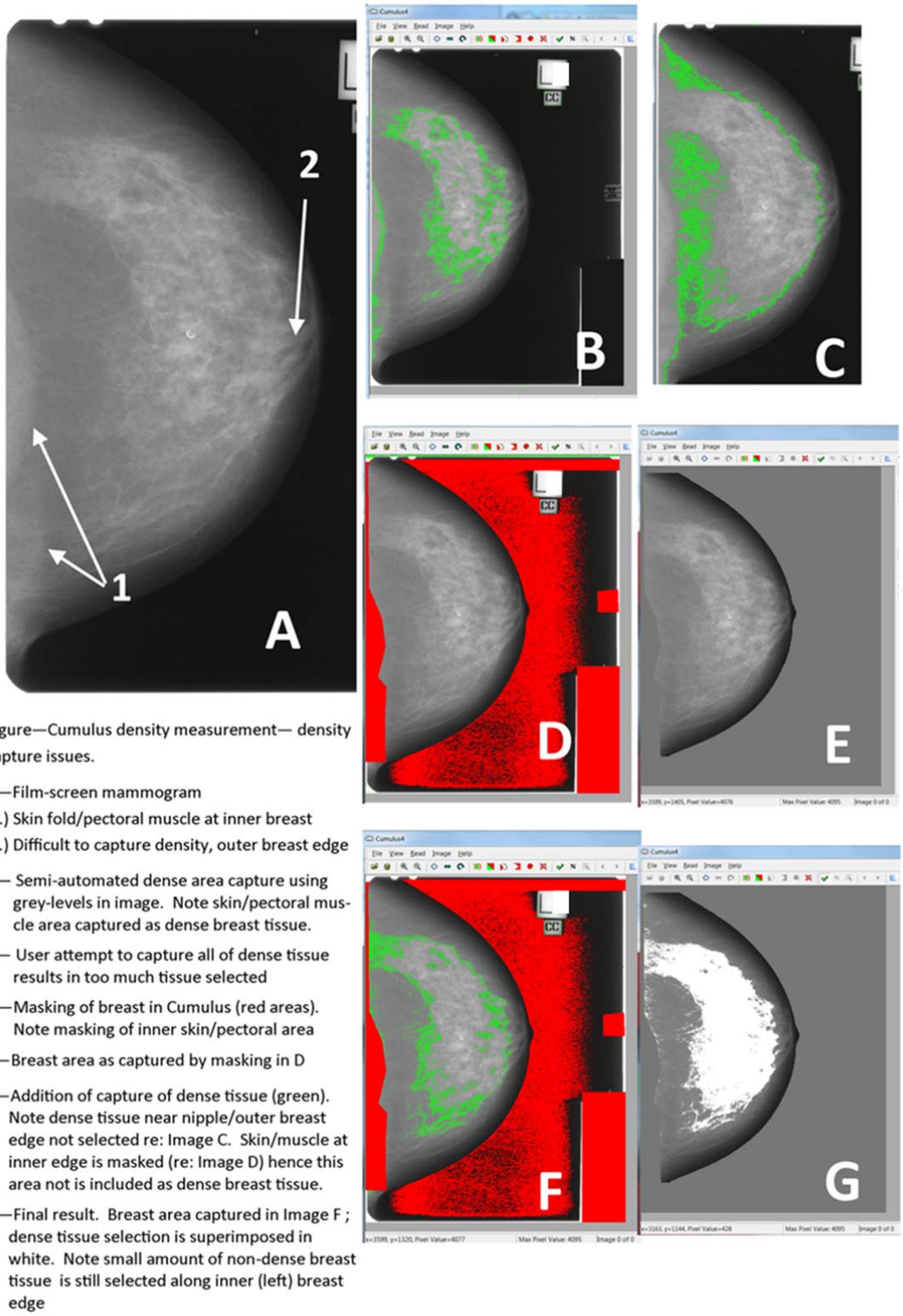
The measurement 1 reads of the digital mammograms on 25 September 2014 occurred relatively soon after the measurements for the film screen mammograms (9 July and 12 August). Because film mammograms have (on average) higher PD than digital mammograms, this potentially

could have influenced the higher visual measurement 1 assessments of the digital mammograms. The appearance of the film mammogram set may have had less influence on the measurement 2 reading of the digital mammogram set which occurred in early November 2014. Alternatively, there may also have been a keen interest in performing the digital mammogram analysis after waiting the required month between reads, and hence proper attention was not paid to the visual assessment of digital mammograms at Measurement 2.

Cumulus was developed for use on film-screen mammograms. This could have also contributed to the increased variability observed in the Cumulus assessments of film-screen vs digital mammograms. Recent reports in the literature state that Cumulus appears to be a robust tool to assess density on digital mammograms [400, 401], hence the increased variability is likely due to the relative unfamiliarity in assessing digital vs film-screen mammograms to complete the Aim of this Chapter.

Regardless of what influenced the substantial bias between the visual PD reads at measurement 1 and 2 for the digital mammogram set, the existence of a bias at least 10% greater for Visually assessed compared to Cumulus assessed PD supports the assertion that Cumulus estimates of density (and that of other semi-automated e.g. Madena and/or quantitative methods e.g. planimetry) are more reliable than estimates of visually assessed density.

Recent data has shown that radiologists who undertook reading of >6,000 mammograms annually are able to accurately categorise breast density into three groups with an accuracy of >90% (ANZCR mammography Synoptic 2002); in contrast radiologists who undertook fewer mammography assessments per year (mean 3,750) accurately assessed breast density 32 to 55% of the time [466]. Hence more frequent assessments of mammograms and/or increased interest/training in mammography assessment may impact radiologists' ability to reliably assess breast density visually [467]. This implies that reliability may improve with further MD assessments of digital mammograms for this project.



Figure—Cumulus density measurement— density capture issues.

A—Film-screen mammogram
1.) Skin fold/pectoral muscle at inner breast
2.) Difficult to capture density, outer breast edge

B— Semi-automated dense area capture using grey-levels in image. Note skin/pectoral muscle area captured as dense breast tissue.

C— User attempt to capture all of dense tissue results in too much tissue selected

D—Masking of breast in Cumulus (red areas). Note masking of inner skin/pectoral area

E—Breast area as captured by masking in D

F—Addition of capture of dense tissue (green). Note dense tissue near nipple/outer breast edge not selected re: Image C. Skin/muscle at inner edge is masked (re: Image D) hence this area not is included as dense breast tissue.

G—Final result. Breast area captured in Image F ; dense tissue selection is superimposed in white. Note small amount of non-dense breast tissue is still selected along inner (left) breast edge

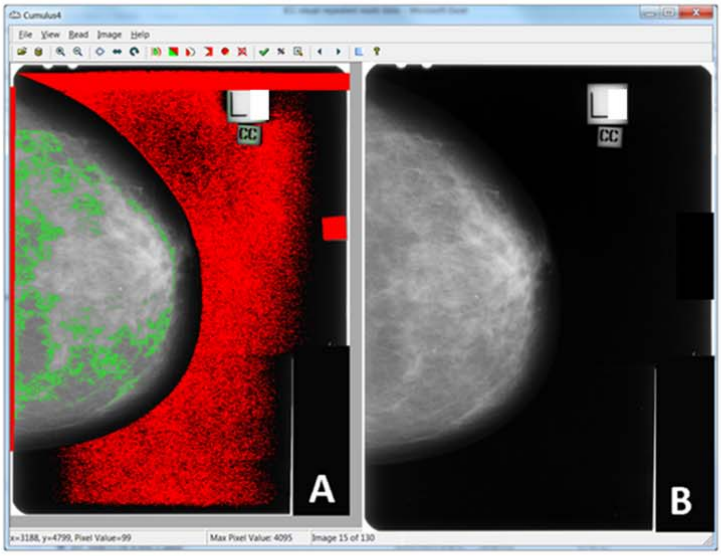
Figure 5-13 Cumulus density measurement and density selection issues

Assessing PD with Cumulus is more time consuming to complete than visually assessed PD (~2 hours/100 mammograms for Cumulus, vs <30 minutes for 100 mammograms visually). However the increase in time taken for Cumulus appears to be offset by the increase in reliability compared to visual assessment.

For Aim 2.2 (PD inter-technique reliability), differences in both breast area and dense area assessment account for the negative relationship with increasing PD seen in Figure 5-9 and Figure 5-10 for the Cumulus vs Visual methods. A mammogram is a two-dimensional representation of a three dimensional structure. Not only are breast tissues visible (skin, subcutaneous fat, glandular/connective tissue), but chest skin and (pectoral) muscle may also appear on the mammographic image (Figure 5-13 A, below). Skin and muscle add to the thickness and x-ray absorption of the tissues viewed on a mammogram, and hence the density of that area of the mammogram. These tissues were typically masked during image assessment with Cumulus to exclude them during measurement (Figure 5-13 D to G). Dense tissue near the outer breast (skin) edge is often difficult to capture in Cumulus without also selecting non-dense tissue in the middle and inner breast because the grey levels are similar (Figure 5-13 C). This is presumably because the tissue adjacent to the outer breast edge contains more adipose tissue (which is translucent to x-rays = darker) and/or is not as thick.

Visual assessment of density is not subject to the same constraints as Cumulus assessed density (Figure 5-14, top and third rows). Visual assessment does not require explicit masking of breast areas to eliminate incorporation of non-dense tissue because visually assessed density is not as affected by the absolute grey levels present in the mammographic image. The eye is able to discern the relative grey levels as dense or non-dense irrespective of the background tissue appearance. As described above, dense tissue in both the area of the breast near the nipple/outer skin area and the area adjacent to the chest can be difficult to accurately capture in the Cumulus program (Figure 5-13, Figure 5-14).

Visual PD, Time 1	Cumulus PD, Time 1	Average PD	Difference in PD
75	46.0	60.5	-29.0



Visual PD, Time 1	Cumulus PD, Time 1	Average PD	Difference in PD
50	51.5	50.8	-1.5

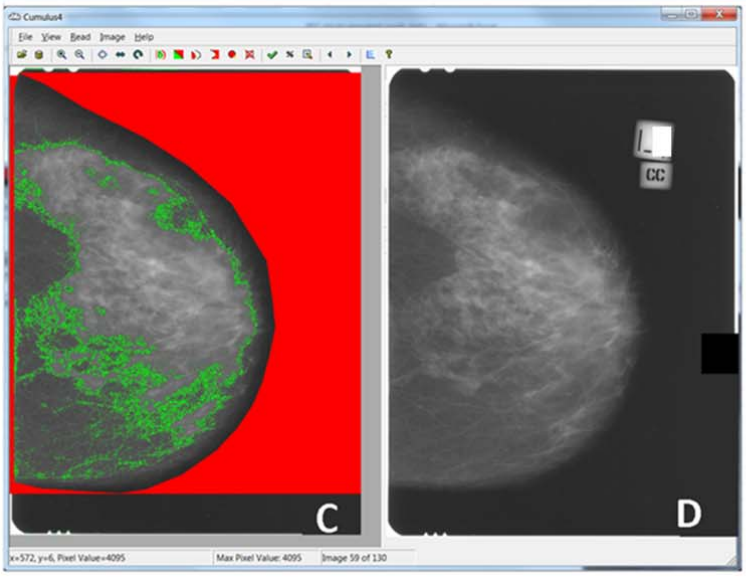


Figure 5-14 Differences in net PD between Cumulus assessed and visually assessed PD
2008 IBIS-II film mammograms, Measurement 1. Cumulus assessed PD is shown in A & C; original mmgs shown in B & D (as used for Visual assessment). A & B are the same mammogram; C & D are the same mammogram. A large difference is found between Visual and Cumulus PD for the upper mammogram (–29.0%) compared to the lower mammogram (1.5%). The dense tissue appears bright in A & B, the background (non-dense) lighter than in C & D. This makes the dense tissue more difficult to capture in Cumulus (green outline, A); note the missed dense tissue on the right hand side (outer edge) of A. The dense tissue in the lower panel (C & D) is not as bright as in A & B, but the non-dense area is darker than in A&B. Differences in image acquisition (e.g. exposure time) may account for the differences in appearance of non-dense tissue on the mammograms. The appearance of the mammograms contributes to the negative correlation on the Visual vs Cumulus PD Bland-Altman plots (Figure 5-9, Figure 5-10) Potentially, additional dense tissue could be captured in Cumulus to better approximate the Visual assessment to correct for these differences for this particular mammogram.

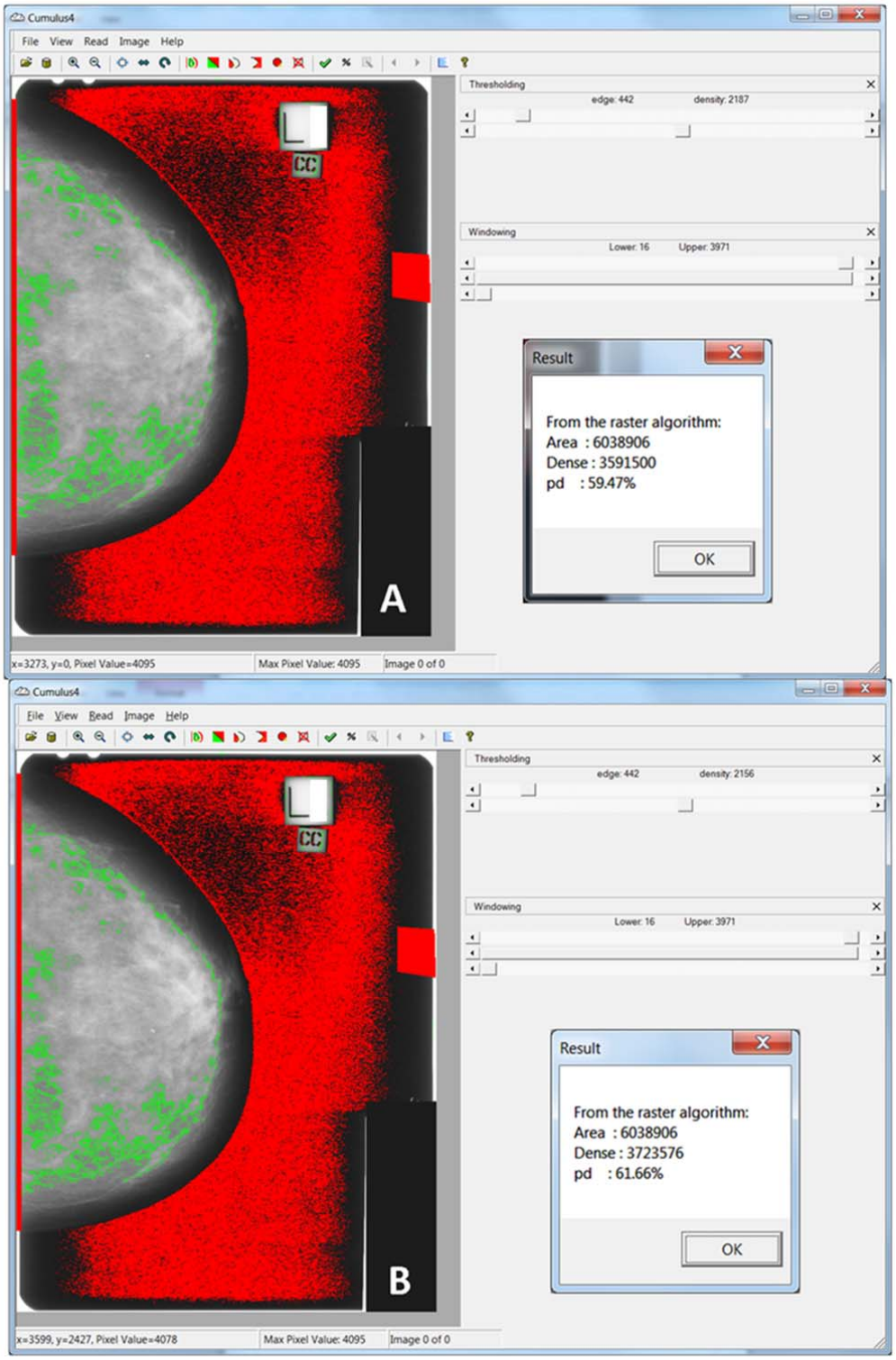


Figure 5-15 Selection of additional dense area in Cumulus to improve similarity to visually assessed PD
The selection of additional dense area in Cumulus (A, B) more closely matches visually assessed PD (75%, top of Figure 5-14). Some of the dense tissue near the outer breast edge is still not selected in Cumulus, which still contributes to the differences in Visual vs Cumulus PD for this mammogram.

As per Figure 5-13 and Figure 5-14, the dense tissue near the nipple not captured by Cumulus is likely to be offset by the unmasked thicker tissue at the inner edge of the breast which has been captured as density. However, in Figure 5-14 A & B, the dense tissue not captured by Cumulus due to concomitant brightness of the background tissue contributes to the lower PD calculated by the Cumulus program compared to visually assessed PD. The difference in PD between Cumulus and visual PD for this particular mammogram could potentially be reduced by additional capture of dense tissue in Cumulus (Figure 5-15).

The results from the film mammogram 4.3OD vs 4.7OD comparison imply that percent density assessed at 4.3OD may be slightly higher than PD assessed on the same film-screen mammograms scanned at 4.7OD (as shown in the Bland Altman graph biases of ~0.5% in Cumulus, ~1% in visual), Figure 5-11and Figure 5-12. However, repeated measurements 1 month apart using both the Cumulus and visual assessment techniques on film-screen mammograms in Aim 2.1 (Intra-technique reliability) yielded PD biases of 1% and 2% respectively. The slight increase in PD observed in this analysis for film mammograms digitised at 4.3OD may be just due to the inherent variability of these subjective assessments.

Even if the slight increase in PD observed for the 4.3OD scans is attributable to differences in appearance of 4.3OD and 4.7OD scans, the slight difference in PD (~0.5 to ~1%) is unlikely to have a clinical impact on the assessment of absolute PD for these women. However, since the reduction in PD due to treatment with anastrozole is expected to be small (~1 to 2%) relative to the variability in repeated assessments (± ~10%), use of a standardised scanning procedure (e.g. either 4.3OD or 4.7OD) may assist with improving the likelihood a reduction in PD due to AI treatment will be detectable if it is present.

The variability (limits of agreement (LOA)) in Aim 2.3a of ±9 to 10% is similar to those in Aim 2.1, which were ±8% for Cumulus and ±13% for Visual (Table 5-3). The Cumulus LOA for Aim 2.3a, ±9%, may be slightly higher than that in the repeated measurements in Aim 2.1

because mammograms from all four views (RCC, LCC, RMLO, LMLO) were included in this analysis. PD measured on CC vs MLO views has been shown to produce slightly different gradients of BC risk [468]. Aim 2.1 included only CC view mammograms.

The LOA for the visually assessed digitised film-screen mammograms for this Aim may have been lower (±9.5%) than in Aim 2.1 (±13%) because visual assessments for each participant's mammograms occurred only minutes apart, instead of being separated by 1 month of time. Comparison of the bias and LOA between the (CC view) mammograms assessed visually using Cumulus/full size DICOM files (Aim 2.1) and Windows Explore/quarter-size PNG files (Aim 2.3a) yielded a bias of ~2%, and LOA of ± 14.5% (Bland Altman plots not shown). These assessments occurred approximately nine months apart, and are not substantially different from the bias (2%) and LOA (± 13%) observed in Aim 2.1. Visual assessment of mammograms in either Cumulus or Windows Explorer seems to be interchangeable. However, to reduce potential variability, one technique or the other should probably be used when trying to assess small changes in PD over time (i.e. longitudinal PD), since the LOA between the two methods was slightly higher (14.5%) compared to repeated visual assessments in Cumulus alone (13%).

The standard deviation (SD) of visually assessed PD from mammograms scanned at 4.3OD was smaller (21.8) than that for the same mammograms scanned at 4.7OD (35.5), Table 5-4. However, these SDs were not significantly different ($p = 0.72$, variance ratio test). The standard deviation (SD) for the 4.3OD mammograms assessed visually (21.8) was greater than both SDs for 4.3OD (15.7) and 4.7OD (18.6) mammograms assessed with Cumulus, this SD of 21.8 was statistically significantly different from both of the Cumulus assessed SDs ($p<0.0001$, variance ratio test). This supports the observation that Cumulus assessed PD is less variable than visually assessed PD, and that use of 4.3OD vs 4.7OD mammograms does not significantly affect the absolute value of PD assigned using either assessment method.

The assumption that PD would be similar when measured in both film-screen and digital mammograms was further disproved by the results of Aim 2.3b (PD comparison of 4.7OD film and digital mammograms). The descriptive distributions of PD for film vs digital mammograms (Figure 5-3 and Figure 5-4) showed median PD is lower for digital mammograms than film. This difference could have resulted from the natural decline over time in PD, since the film mammograms preceded the digital mammograms. The results from Aim 2.3b show more specifically there are statistically significant differences between PD assessed on film screen vs digital mammograms from the same IBIS-II trial participant, Table 5-5. Furthermore, the average PD difference between film and digital mammograms is more than would be expected due to an annual decrease in PD of 1% (Table 5-6, Table 5-7, Table 5-8). The decrease in PD is likely due to the inherent differences in the presentation of PD on film-screen and (processed) digital mammograms.

A limitation for the results of this chapter is that the sets of mammograms used in some analyses contained multiple potential violations of the requirement for independent samples. Restriction to a single view per episode in Aim 2.1 did not substantially change the ICC results (Table 5-3); trends in difference for the bias and LOA were inconsistent (Table 5-3).

The variability for most results in this chapter is likely smaller than might have occurred if each participant had only been sampled once. However, as mentioned in Section 5.3 Statistical Methods, the use of multiple images from the same mammogram, mammographic episode and/or participant may not be as problematic for the Aims of this chapter as for some research projects. This is because the variability of the subjective nature of the visual and Cumulus techniques is likely far greater than the opposing effect of what are effectively repeated measurements within the datasets. The use of repeated mammograms for some participants during completion of Aims 2.1 to 2.3 is consistent with the nature of the repeated measurements which will be utilised for Aim 4 and the Primary Aim of this thesis, and hence reflects the likely variation present in the dataset during assessment of individual Views. The entire set of analyses

in this chapter could potentially be repeated using the larger mammographic data set available from both Collection 1 and Collection 2 rounds of mammogram collection; these data may provide results less impacted by the non-independence of the Collection 1 data and provide a slightly different insight into the repeatability of Cumulus vs visual assessments for film and digital mammograms.

Another limitation for this project was the use of one observer to perform all measurements for this Aim. Inclusion of additional observers would provide a more thorough insight into the subjective nature of the Cumulus and visual assessment methods on this set of IBIS-II trial mammograms. However, use of single observers is common during studies involving measurement of MD [188, 296, 312], in part due to the introduction of other factors such as inter-observer differences in MD assessment.

5.6 Chapter summary

Percent density (PD) was challenging to measure reliably and repeatedly on film and digital mammograms. Ideally, the longitudinal (repeated) measurements made on IBIS-II participant mammograms to achieve the Primary Aim of this project would vary less than the average change in PD expected to occur from trial hormonal treatment. For IBIS-II participants treated with anastrozole, the expected average reduction in PD compared to that of controls is approximately 1 to 2% [29]. The most reliable outcome for repeated measurements was observed using the Cumulus assessment technique of 72 film-screen mammograms: Bland Altman limits of agreement (LOA) were ± 8%, with a mean difference (bias) of 0.6% (95%CI 0.1 to 1.5%). The Intra Class Correlation (ICC) for this set of repeated measurements was 0.97 (95%CI 0.96 to 0.98). This ICC compares favourably with repeated measurements performed by experts in the use of Cumulus [184].

However, ICCs were generally lower and Bland Altman LOA and biases greater with visual assessment of film mammograms, and Cumulus and visual assessment of digital mammograms. Digital mammograms assessed with Cumulus compared well with that for film screen mammograms: Bland Altman LOA was ± 10%, with a mean difference (bias) of ~2%; the ICC was 0.94 (95%CI 0.91 to 0.96). Visual assessment of film mammograms had LOA of ± 13%, with a bias of 2%, and an ICC of 0.95 (95%CI 0.92 to 0.97). Visual assessment performed much more poorly on digital mammograms: LOA ±29%, bias ~12 %, ICC 0.64. Cumulus provides a more reliable and repeatable method for assessing density on both film and digital mammograms. Assessing PD with Cumulus is more time consuming to complete than visually assessed PD (~2 hours/100 mammograms for Cumulus, vs <30 minutes for 100 mammograms visually). Cumulus assessment is much more time consuming to perform than both visual and fully automated assessment techniques, and hence is not practical for clinical settings (e.g. BreastScreen). Variability with Cumulus assessed PD is still high, and is approximately equivalent to the smallest meaningful change in PD (10%) associated with a clinical difference in breast cancer outcomes.

Due to the differences in PD observed with the Visual and Cumulus methods, these assessment techniques are not interchangeable. The lower average PD observed for digital mammograms compared to film mammograms for the same IBIS-II participants is greater than expected due to the estimated average annual change in PD of 1%; digital mammogram PD may be lower due to inherent differences in acquisition, as well as image post-processing by mammography machine software. PD assessed visually on 4.3OD and 4.7OD scans of the same mammogram did not differ significantly, and could be interchanged if required.

Because MD measurements made with the Cumulus program were less variable than MD measurements made visually, longitudinal assessment of IBIS-II mammograms for annual change in MD (Aims 4 and 5) as well as the participants baseline characteristics analysis (Aim 3) was undertaken with the Cumulus program. With adequate training, visually assessed

density may be useful in settings where Cumulus assessed PD is impractical. Ready access to freely available, fully automated techniques to measure MD on film, processed digital and raw (unprocessed) digital mammograms is desirable. Due to differences in MD assessment which are likely to result if two or more different assessment techniques are utilised, use of standardised processing and measurement techniques is recommended when undertaking longitudinal measurements of density. This is challenging in the multi-centre, international setting typical in breast cancer clinical trials.

6. Descriptive statistics and MD associations of baseline characteristics

This chapter describes the aims, general methods, and statistical methods utilised for Aim 3. Results of the analysis follow, leading with a discussion of the mammographic technical factors which affected quantification and modelling of MD. Next, measures of central tendency are tabulated for the baseline characteristics of the CMN IBIS-II participants, including MD. Simple regression of the MD parameters with baseline covariates is undertaken, followed by multivariable (multiple) linear regression of the MD parameters and baseline covariates. The chapter concludes with a discussion of the results.

6.1 Aim

The Aim for this chapter (Aim 3) was to undertake descriptive statistics of the baseline characteristics of the CMN IBIS-II participants who contributed mammograms to this project, and to quantify the relationship between baseline MD and the participants' baseline characteristics.

6.2 Methods

6.2.1 Participants and mammograms

Eligibility for the IBIS-II CMN and AI substudy was described in Chapter 3. Participants undertook trial mammography approximately annually; trial mammographic episodes usually coincided with the IBIS-II annual follow up visit. Mammograms taken up to 1 year prior to randomisation were utilised as the baseline measurement of MD. Baseline mammograms were taken at a variety of institutions, including the Calvary Mater Newcastle. All trial follow up mammograms were taken on a single mammography machine in the Radiology Department at the Calvary Mater Newcastle hospital (CMN). The transition from film-screen to computed radiography (CR) digital mammograms at CMN occurred after 17 December 2008.

Standard mammographic Views of the breasts, namely the right cranio-caudal (RCC), left CC (LCC), right medio-lateral oblique (RMLO), and left MLO (LMLO) Views were assessed for each woman for each mammography episode (time point), including all baseline episodes. This chapter contains some analyses utilising individual Views and pairs of Views (averaged), however, the Aim for this chapter was undertaken primarily with MD represented as the average of the four standard mammographic Views per episode. The average of all four Views was employed to reduce MD measurement error— which on average was ±8% PD for film-screen mammograms and ±10% PD for digital mammograms using Cumulus (Aim 2, Chapter 5). Use of average MD per episode also helped simplify statistical analyses, and enabled independence assumptions for regression models to be met.

6.2.2 Assessment of mammographic density

Participant mammograms were assessed for density in Cumulus in random order for each participant [399]. Hence, for a woman with baseline mammograms and three follow ups, 16 mammograms were assessed for density in Cumulus in random order. For instance, the LCC View from the last follow up could be read first, followed by the RMLO View from the baseline follow up, etc. Cumulus assessments for mammograms from Collection 1 were performed on all available mammograms— film, digital printed to film (DoF), and (fully electronic) digital— whilst Collection 2 Cumulus assessments were performed only on film and digital mammograms.

Only measurements from film-screen and fully electronic 'original' digital mammograms were utilised for this and subsequent analyses. Measurements from digital mammograms printed to film (DoF) were not utilised because they were fewer in number than their fully electronic counterparts, and did not appear to provide a representation of breast density which was consistent with other digital mammograms; this introduced additional variability into the longitudinal MD measurements. Hence measurements from DoF mammograms were omitted.

6.2.3 Measures

As described previously, percent (mammographic) density is the proportion of the breast covered by mammographically dense tissues. The Cumulus program is used to assess the total breast area (BA) and the area of dense tissue (DA) on the breast, which outputs DA and BA as a raster size (number of pixels). The percent density (PD) parameter value for these Cumulus measurements was derived by dividing area of the breast selected as dense divided by the total breast area: PD = DA / BA. PD is a dimensionless parameter, because it is one area, DA, divided by another, BA. Comparisons of one PD to another are hence independent of the pixel spacing in the images.

DA and BA can also be used to estimate the adipose area (AA) and percent adipose area of tissue on the breast, as follows: AA = BA–DA, and PA = AA/BA. Although less commonly reported than PD or DA, AA and PA may be independently [186] and inversely associated with BC risk [469]. Because DA, BA and AA are all measures of square area, the size of the pixel spacing of the mammograms affects the estimates of these MD parameters.

PD, DA, AA and PA can be considered different MD attributes, components, or 'phenotypes'[469]. Three of these MD attributes (PD, DA, AA) were considered as potential longitudinal MD outcomes of interest for this project, and therefore utilised as outcomes for the Aim of this chapter. PA was not utilised because the associations for PA would likely be identical to those for AA. BA was also utilised in baseline analyses for the current Aim because it was necessary to assess the longitudinal behaviour of BA in future Aims; BA should remain relatively constant longitudinally, since both AI treatment and time (age) were not expected to affect it.

The covariates considered during the analysis for Aim 3 are listed in Table 6-1, below.

Table 6-1 Covariates (confounders/modifiers of mammographic density) and MD technical factors

Covariate	Description, rationale for covariate
Age	Age at randomisation. Density is expected to decrease as age increases. Increasing age is also associated with a slower rate of change in MD.
Weight	Increasing weight is associated with decreasing percent density (PD).
Height	Increasing height is associated with increasing density.
Body Mass Index	BMI = (weight in kilograms)/(height in metres)^2. PD decreases as BMI increases.
Age at menarche	Density tends to increase with increasing age at menarche, although BC risk is decreased with increasing age at menarche.
Parity (yes/no)	In general, increasing parity is associated with decreasing mammographic density, as well as a decrease in breast cancer (BC) risk.
Age at first birth	Age at first pregnancy ≥ 28 weeks. Lower age at first birth is associated with decreased BC risk and lower MD. Late age at first birth (30 to 35+) may be associated with an increase in MD, and may be associated with a decrease in age-related MD decline in the general population. The effect of age at first birth was examined in the model by use of a continuous variable (in years) and as a categorical parameter described in a later section of this chapter.
Previous oral contraceptive use	The relationship with oral contraceptive use and MD is not well established. Modelled as both a dichotomous parameter (never/ex-user) and continuously (months of total use).
Age at menopause	Later age at menopause may be associated with higher MD, and is associated with higher BC risk.
Previous HRT use	Whilst current HRT use is associated with both an increase in MD and BC risk, the relationship with MD and previous HRT use is less well characterised. Modelled as a categorical variable (dichotomous), and as a continuous parameter: total months of use.
Family history of BC and/or OC	The number of first and second degree relatives, age at which breast cancer or ovarian cancer was diagnosed, and whether the BC was bilateral was recorded for IBIS-II participants at baseline. Number of relatives is likely to be positively associated with MD at baseline because increasing family history is associated with higher BC risk. Number of relatives may or may not be associated with change in MD over time. BC risk will be modelled as total number of relatives with BC and/or OC (as both a continuous and categorical parameter). A derived parameter, weighted number relatives (first degree relative = 1, second degree relative = 0.5) was also used to better approximate the increase in BC risk due to first compared to second degree relatives (this is because on average, second degree relatives share half the number of genes that first degree relatives share with an individual).
Smoking status	Never, current or ex-smoker. Smoking likely has a weak, inverse relationship with MD through reduction of (lifetime) estrogen exposure. Modelled as a three category categorical variable.
IBIS-I participation (yes/no)	IBIS-I participants were randomised to receive 5 years of tamoxifen or placebo. Former IBIS-I participants were eligible to enter the IBIS-II trial from June 2007. Active trial treatment in IBIS-I may (inversely) affect subsequent baseline and/or longitudinal MD in IBIS-II, because tamoxifen treatment tends to lower MD.
'Technical' mammographic parameters	
Film vs Digital mammograms	Density tends to be higher on film compared to digital mammograms, due to the propensity of post-acquisition processing to remove density on digital mammograms.

Covariate	Description, rationale for covariate
Post-processing software version and/or hardware configuration	Different software versions, and/or different digital mammography systems may utilise different algorithms to remove the dense tissue on the final, processed image. This can cause the appearance of increased or decreased mammographic density for individual participants, compared to how other software versions and/or digital mammography systems may render the appearance of the density on their mammograms. This parameter is modelled as a categorical variable.
CC vs MLO Views	The cranio-caudal standard mammographic View contains less cross sectional area of the breast than the medio-lateral oblique standard mammographic View. These differences in cross-sectional area likely affect the amount of dense area measured on the CC vs MLO mammograms, and may also affect the relative percentage of the breast covered by dense tissues.
Breast laterality (right vs left side)	The left breast tends to be slightly larger than the right breast in many women. As for the differences in BA between CC vs MLO Views, this (smaller) difference in cross sectional area between right and left mammograms may also affect the amount of DA and PD measured.

6.3 Statistical Methods

All analyses were undertaken using Stata v12.1. P-values <0.05 were considered significant. Back transformation of coefficients from models with transformed outcome parameters was not performed to minimise reporting errors. Coefficients from regression models utilising untransformed MD thus are presented in % (PD) or mm^2 (DA, BA and AA). Coefficients from models utilising square root and log transformed MD are presented in square root or log transformed % and mm^2.

Pixel spacing varied for different mammographic Versions. This meant that the raster (total number of pixels) for DA and BA from mammograms with different spacing of pixels could not be directly compared, as they represented different sized areas. The DA and BA rasters were converted to area in mm^2 using the following formula:

Equation 6-1

$$Area\ in\ mm^2 = \frac{total\ \#\ pixels\ (raster\ size)}{\#\ pixels/mm^2} = total\ \#\ pixels \times (pixel\ spacing)^2$$

Box plots of all mammograms collected during Collection 1 and Collection 2 (2130 film and digital mammograms from 541 episodes) were prepared to compare the differences in raw vs

converted area of BA and DA. Comparisons were made for BA and DA between different mammographic Versions using a Kruskal-Wallis test on 2126 mammograms from 540 episodes (one episode collected at year 8 was omitted). A non-parametric trend test [470] was also utilised to assess trend across mammogram Versions for adjusted BA.

An average value was calculated for PD, DA, BA and AA for each baseline episode. For example, average baseline PD for each participant was calculated using the PD measurements for the RCC, LCC, RMLO and LMLO views at baseline for that participant; this is equal to (RCC + LCC+ RMLO+ LMLO)/4. This was done to provide a better estimate of the MD parameters for each episode, given the inherent variability of the subjective measurements made using Cumulus as noted during Chapter 5 (Aim 2, reliability analysis).

6.3.1 Variable checking

Checking of each parameter was undertaken at baseline. Categorical variables were expressed as frequencies and percentages, and these values examined for plausibility. Histograms and scatter plots were created to check for missing values, potential outliers and assess normality. MD parameters were also checked longitudinally using scatter plots and line plots.

PD, DA, AA and BA all were right-skewed. Ladder of powers graphs indicated the distributions for all four MD parameters improved with square root transformation. Log transformation also improved normality of the distributions. Due to the strong right skew of PD and DA, square root (sqrt) and natural logarithm (ln) transformations of these MD parameters were assessed during statistical modelling. BA and AA were less strongly right skewed, and were modelled in both original units and with a square root transformation.

Four baseline episodes had average PD <1%. Hence the square root of average PD for these four baseline episodes increased compared to the raw PD value— unlike episodes with average PD>1. However, the relative order of the measurements was maintained. Given the inherent

variability of the semi-automated PD measurements in Cumulus, this slight loss of precision of less than 1% was considered negligible. All averages of DA, BA and AA per episode were values greater than 1, hence square root transformation was consistent for all episodes.

One participant was notable with very dense breasts (PD≥75%); this was unusual compared to other CMN participants. Visual comparison of PD and the Cumulus assessments confirmed that the MD measurements for this participant were valid.

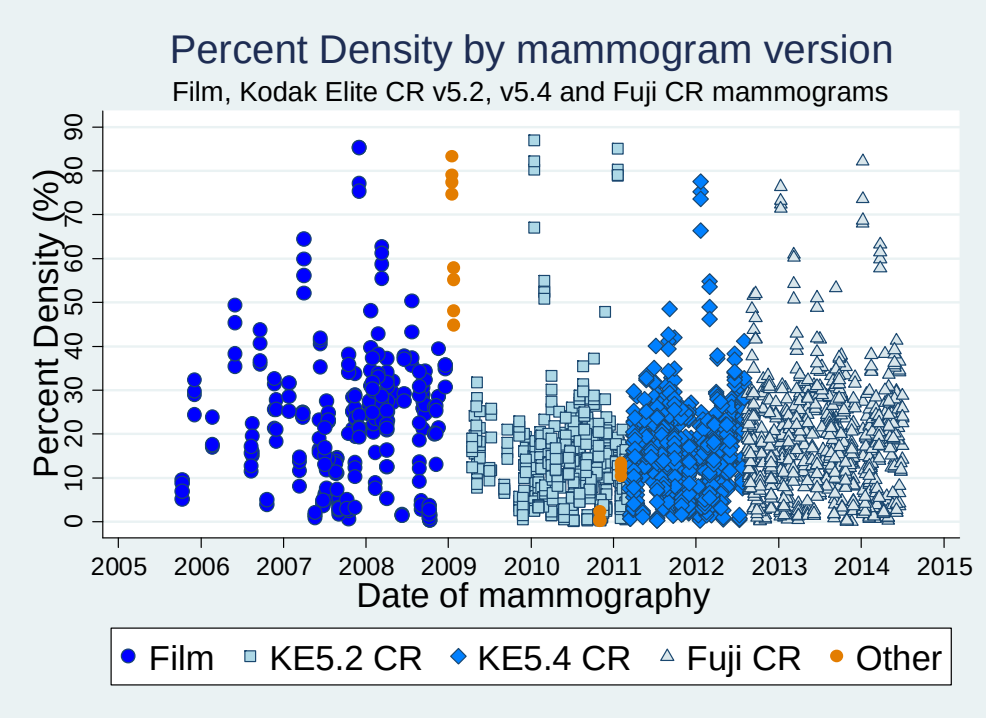
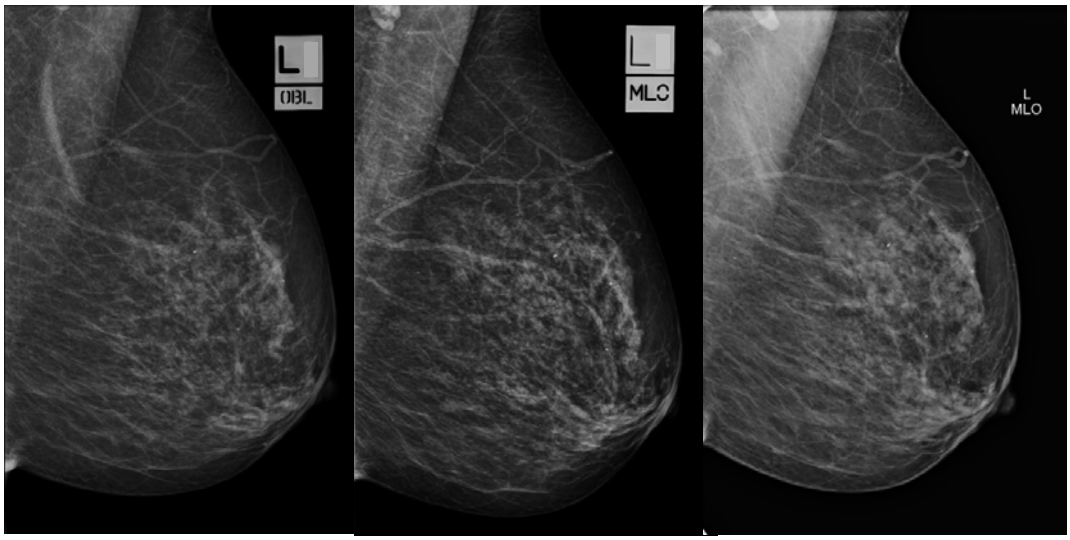


Figure 6-1 PD over time, by mammogram Version over time.
The trend towards increasing PD over time for digital mammograms is due to changes in software and hardware configuration.
CR – computed radiography (digital) mammograms; KE – Kodak Elite mammography machine
5.2 – software version 5.2; 5.4– software version 5.4; Fuji– Fuji mammography machine

A trend towards increasing PD over time was noted for participants with digital mammograms, Figure 6-1. Because the mammograms were read in random order during Cumulus assessment, this trend was not noted earlier in the project. The pattern of increasing PD over time was unexpected, as typically PD will decrease over time with age— until perhaps age 70 when this trend may reverse [471]. The analysis for Aim 2 had revealed that PD tended to be different between mammogram Types (film vs digital): PD on film mammograms tends to be higher than

for digital mammograms. However a different factor appeared to be affecting PD within the digital mammographic Type.

Subsequent inspection of the DICOM header file for digital mammograms revealed three different configurations of software and hardware which were utilised during the period from April 2009 to July 2014 at the CMN Radiology department. When reviewed visually, mammograms captured using the older Kodak Elite CR software (v5.2) were generally darker with less dense tissue than the later version of the Kodak Elite CR software (v5.4). Mammograms taken most recently with the Fuji CR machine tended to be brighter with more dense tissue than v5.4 of the Kodak CR machine, Figure 6-2. Hence, the post-processing performed on the mammographic image appears to have retained more of the dense tissue over time, although an increase in MD over time due to other factors could not be excluded.



2010 mammogram Kodak Elite CR v5.2 (KE52) | **2011 mammogram** Kodak Elite CR v5.4 (KE54) | **2012 mammogram** Fuji CR (Fuji)

Figure 6-2 Sequential annual LMLO CR mammograms for one participant
The figure depicts the visual differences in MD for the different CR software and hardware versions of the CMN mammography machine.

Figure 6-1 is tinted with different colours for each mammographic Version (film, KE52, KE54, Fuji). Differentiation of PD by mammographic Version showed the pattern of increasing PD over time is related to changes in the software version of the Kodak Elite mammography

machines, as well as the change from the Kodak Elite CR machine to a Fuji CR mammography machine. As a result of the discovery of the relationship between PD and mammogram Version, mammogram Version was added as technical mammographic covariate for this analysis.

6.3.2 Covariate checking

Two unusually high values were found during checking of covariates. One participant had an extremely long duration of oral contraceptive use; another had an extremely long duration of HRT use. Comparison of the time span between age at menarche and age at menopause revealed that the reported long duration of oral contraceptive use was plausible. Review of the span of time between menopause and randomisation for the latter patient revealed an early age at menopause, which explained the long duration of HRT for this participant. No other unusual covariate values or distributions were noted.

Because some covariates were not applicable for some participants, categorical parameters were created for age at first birth (AFB), oral contraceptive (OC) duration, and HRT duration. To examine the possible dose-effects of these parameters on MD, the quartile boundaries for HRT and oral contraceptive duration were used to create a five-category covariate; never-user participants were coded as 0 in the first (reference) category. A six category parameter was created for age at first birth, with the first three categories utilising the CMN participants' age at first birth quartile boundaries of 20, 22, and 25. The fourth quartile was subdivided into two categories, for women aged 26 to 29 and age 30+; this was done to emulate the expected age related dose-response effects of first birth on PD [296]. The eight non-parous women were provided with a separate category.

Categorical parameters with multiple categories are problematic in analyses when the sample size is small relative to the number of categories. This is because few or none of the individuals in the sample may be present in some categories. Review of the breast cancer and

mammographic density literature was undertaken to examine which dichotomisations of AFB, duration of OC use and HRT duration were useful. AFB is sometimes dichotomised (e.g. age <26 vs 26+, age <20 vs 20+) and paired with number of children, e.g. 1 to 2 vs 3+ to create four or more categories for analysis [191, 203]. HRT and OC use are often divided into never vs ever, or never vs current vs past [218, 356]. The relationship between MD and AFB was also modelled through dichotomisation of AFB as <20 vs 20+, <26 vs 26+ and <30 vs 30+ [218], with a separate category for nulliparous women. HRT and OC use were divided into ever vs never use. The category with the youngest parous women was utilised as the reference group for all age at first birth categorical parameters, whilst never use for OC and HRT was used as the reference category for all OC and HRT categorical parameters.

Determination of one established BC risk factor, age at menopause, was complicated for 27 participants for whom age at hysterectomy only was known. Natural menopause was experienced by 57 participants whilst 36 participants underwent ovarian oblation, typically in conjunction with hysterectomy. The remaining 27 participants underwent hysterectomy without a known age at natural menopause. Age at natural menopause was estimated for this latter group of women using multiple imputation based on the age at natural menopause experienced by their 57 IBIS-II peers. The imputed values for age at natural menopause were modelled as a multivariable linear relationship with age at natural menopause as the dependent variable and the IBIS-II baseline covariates related to age at natural menopause—smoking status, parous status, age at menarche, age at first birth, oral contraceptive use, age at randomisation and BMI [472-474]— as independent predictors of age at natural menopause. Twenty-five imputations were used during estimation of the missing age at menopause values.

The estimated age at natural menopause was compared to age at hysterectomy for the 27 participants by averaging the 25 MI ages for each participant. Six participants were found to have had hysterectomy at an age older than their imputed age at menopause. Two of these participants commenced HRT the same year as their hysterectomy, and hence were unlikely to

experience menopausal symptoms. The average of the twenty-five MI ages for these participants were within 2 years of their age at hysterectomy. The other four participants previously participated in the IBIS-I trial. As for HRT use, randomisation to active IBIS-I trial treatment may have masked any menopausal symptoms. Hence it is possible that these six participants may have experienced menopause more closely to their imputed age than their age at hysterectomy. The imputed ages for the remaining 21participants who underwent hysterectomy without ovarian oblation are likely higher than their actual age at menopause; this is because women who undergo hysterectomy may become menopausal 2 to 4 years earlier than their counterparts who do not undergo hysterectomy [475].

Regression modelling with the imputed values for age at menopause:

In Aim 3, an average of the 25 imputations for age at menopause was calculated for each of the 27 women for whom this parameter was imputed; this value was used as the age at menopause for the regression models. Although use of the average of the imputations likely underestimates the variability of this parameter (e.g. the standard error for age at menopause would tend to be smaller than it should be), this allowed use of R^2 (the coefficient of (multiple) determination) as output by the statistical program for consistency with all regression models.

To assess the impact of the averaged MI values, regression coefficients for the parsimonious multivariable regression model of square root transformed PD with the properly imputed age at menopause ('MI model') were compared with a model fitted with the average of the 25 MI ages for the 27 CMN participants ('non-MI model').

6.3.3 Baseline characteristics

Baseline characteristics for the 120 CMN IBIS-II participants were reported and compared visually to all IBIS-II study participants; formal statistical testing was not undertaken because the CMN MD and AI substudy participants are a subset of the overall IBIS-II sample.

Summaries of measures of central tendency were prepared for important CMN IBIS-II trial baseline characteristics (covariates) including the MD parameters PD, DA, BA and AA. Median, 25th percentile (Q1), and 75th percentile (Q3) were reported for consistency with the baseline covariate measures reported for the International IBIS-II participant data.

6.3.4 Bivariable analyses

The relationship between each of the MD outcome parameters (PD, DA, AA and BA) and each of the explanatory variables at baseline was examined. Relationships for continuous covariates were examined using scatter plots to check linearity, correlation coefficients to measure the strength of association, and simple linear regression. Relationships between the MD outcomes and categorical explanatory variables were examined with summary statistics and graphs by each level of the categorical variable, and simple regression.

The relationship between the MD parameters and covariates were checked graphically using scatter plots (for continuous covariates) and box plots (for categorical covariates); this output is not shown. Graphical comparisons including scatterplots with best linear fit lines (for continuous covariates) and boxplots (for categorical covariates) were performed. Covariates were also examined graphically by mammogram Type for any unexpected relationships between the explanatory variables and the MD parameters. Collinearity testing of continuous and categorical covariates was also performed to check for unexpected relationships between the MD parameters and the covariates. These relationships were compared against the expected relationships reviewed in thesis Chapter 2 (Literature Review).

Collinearity testing was undertaken using the Stata collin and coldiag2 functions for continuous and categorical parameters respectively. Collin provides output including variance inflation factor (VIF), tolerance [$1/VIF = 1/(1-R^2)$], and condition index. VIF >3 (equivalent to a tolerance <33%) and condition indices >30 were used to identify problems with collinearity of

continuous parameters. Condition indices >30 and variance-decomposition proportions >0.3 output by the function coldiag2 were used to identify problematic categorical parameters.

With the exception of the non-dichotomous categorisations for age at first birth, oral contraceptive duration and HRT duration (described in 6.3.4.1, below) simple regression of each baseline covariate with the MD parameter at baseline for the 85 participants with baseline mammograms was undertaken; e.g. baseline age at randomisation (independent variable) was compared against PD for the baseline episodes (dependent variable). Regression was performed for all 85 baseline episodes (film + digital) and by Type: film (n=42 episodes), digital (n=43 episodes). Regression by mammogram Version (film n=42, KE52 n=28, KE54 n=12) was also performed for the strongest modifiers of MD— age (at randomisation) and BMI— to more closely examine the relationship of these confounders and MD within digital Version. Use of Fuji CR mammograms commenced in August 2014, hence baseline data for this Version is not available. Except where noted in the tables (e.g. for models by mammographic Type), each participant with baseline mammography contributed one MD measurement (the average of all four Views from the baseline episode) to each regression model.

The coefficients (β) for simple regression were tabulated for untransformed per-episode averages of baseline PD, DA, BA and AA and also for square root transformed values of these MD parameters. R^2— an estimate of model fit [476]— was also reported for each simple regression model. Model residuals were checked against fitted values (MD predicted by the model) for heteroskedasticity and influential points. As a sensitivity analysis, the one participant with very high PD (~80%) was typically omitted from PD regression models; three women with high PD (≥50%) were also frequently omitted from models of PD. These results were compared with the regression results for all participants with baseline mammograms, to assess the influence of these higher PD participants on the coefficients and p-values. Results from log transformed MD models are not presented due to heteroskedasticity of the residuals.

6.3.4.1 *Simple regression, different categorisations for AFB, OC and HRT duration*

The simple regression analyses described above did not include the non-dichotomous categorisations for age at first birth, oral contraceptive duration and HRT duration. Only the dichotomisation for OC and HRT use, and continuous representation of age at first birth were utilised. This was done because, like the other parameters in Table 6-9, these parameters did not require additional calculations to create them; the values for these parameters were taken directly from the IBIS-II database To see which categorical representation best described the relationship between age at first birth, OC use and HRT with the MD parameters, simple regression for the different categorisations of each parameter was performed. The best relationship for each parameter was determined by examining coefficient significance and R^2 value. The parameter categorisation with the smallest number of categories which maximised both significance (i.e. smaller p-value) and R^2 was selected. If no categorisation yielded any significant coefficients, then the simplest (most parsimonious) representation of that parameter was selected (i.e. the one with the fewest categories).

6.3.5 Assessment of correlation and collinearity among explanatory variables

Inter-covariate relationships were assessed using graphical techniques and collinearity testing as described for testing of correlation/collinearity between covariates and the MD parameters (e.g. box plots, scatter plots, review of VIF).

6.3.6 Multivariable regression, avg baseline MD parameters vs baseline covariates

Full multivariable regression models of MD were created utilising the best categorisation for AFB, HRT and OC use along with other baseline covariates: age at randomisation, BMI, age at menarche, age at menopause (imputed average), smoking status, previous IBIS-I participation, and the technical mammography covariate mammographic View. Height and weight are collinear with BMI, as is mammographic Type with mammographic View thus these parameters could not be utilised in multivariable models simultaneously. The simple regression relationship

between family history and PD and DA (modelled as number of relatives and also as weighted number of relatives) appeared to be non-linear for weighted number of relatives (Table 6-9). The option of a dichotomous (yes/no) representation of the relationship of MD and number of relatives was not possible, because all CMN participants had at least one relative with BC and/or ovarian cancer. Due to the non-linear relationship, number of relatives was not included in the multivariable MD models.

Models were reduced from full to parsimonious using both backwards elimination and forward stepwise regression. For backwards elimination, the covariate with the highest p-value was removed from the full model, and the resulting model retested until only covariates with $p \leq 0.1$ remained. The upper limit of 0.1 was selected for retention of covariates in the multivariable model to be conservative, in particular because of the small number of study participants. R^2 for each model was also reviewed to assess for the percentage of the variability in the dependent variable explained by the covariates in the model. Forward stepwise regression reversed the process: single covariates were added to a simple regression model until only covariates with $p \leq 0.1$ remained. Higher order relationships such as Age^2 (age* age interaction) which are sometimes significantly associated with MD, and other interactions between covariates were not prospectively explored during stepwise regression of these models.

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), which are indices of overall (global) model fit [477] were evaluated for the model at each step, to ensure more parsimonious models corresponded to smaller AIC and BIC. The AIC and BIC account for the complexity of a model (i.e. number of covariates in the model) [478]. Covariates for which BIC decreased <3 upon removal from the model were retained regardless of p-value. Covariates which decreased the AIC by more than 1 (but not the BIC by more than 3) upon removal from a model were noted. AIC and BIC are a better assessment of fit across models with differing numbers of covariates than (non-adjusted) R^2, as well as -2LL (-2* Log

Likelihood, which is mainly used to compare nested models. AIC and BIC are therefore better than other model fit measures at identifying both fit and parsimony.

Age at randomisation was retained in the models because age (along with BMI) is a well-recognised strong confounder of MD. PD and DA models which utilised digital mammograms were also adjusted for mammogram Version. Model coefficients for AA are not tabulated because both BA and AA showed a similar relationship with the covariates in the simple and multivariable regression models.

Regression diagnostics were undertaken by checking for residual heteroskedasticity with scatter plots and histograms. A set of sensitivity analyses was undertaken to review MD differences in CC vs MLO Views, and right vs left mammograms, as described next.

6.3.7 Differences in MD measurements, by View, Type and Version

Typically, one MD measurement made on a single View from a mammographic episode is utilised as representative of MD for that episode. Potential MD differences for measurements made on right vs left mammograms, and MD differences for measurements made on CC vs MLO Views are not frequently reported. The rationale and methods for this analysis, as well as results for the analysis and a discussion, are described in an appendix (Differences in MD measurements, by View, Type and Version).

6.3.8 Simple regression of baseline PD with different View combinations

As stated above, typically, one MD measurement made on a single View from a mammographic episode is utilised as representative of MD for that episode. Potentially, the average of two, three or four Views may enhance the ability to observe significant expected, but less strong relationships between MD and other covariates (such as age at menarche) in small samples. Differences in the association of BC risk factors with average MD measured on different combinations of right vs left mammograms, and CC vs MLO Views are not frequently reported.

Simple regression was performed with PD represented as the average of different combinations of mammographic Views. The different View combinations tested for untransformed PD (%) were each single View (RCC, LCC, RMLO, LMLO), the average of right and left CC Views, the average of right and left MLO Views, the average of both right side Views, the average of both left side Views and the average of all four Views (the MD representation usually used in this thesis). Simple regression models were also fitted for natural log and square root transformed PD using the average of all four Views, as well as the RCC View and the average of the right and left MLO Views.

Coefficients from the different PD simple regression models were examined to ascertain if they were similar (within +/- 10 to 20%) to the coefficients from the PD models with all four Views. Coefficient p-values for the models were also compared for significance ($p<0.05$).

6.4 Results

6.4.1 Dense and Breast Area measurements on mammograms

CR mammograms from the Kodak Elite system utilised a 48.5micrometres (µm) pixel spacing (horizontally and vertically), whilst the electronic images from films digitised on the Array scanner and images acquired from the Fuji CR system had 50µm pixel spacing (horizontally and vertically). This equates to 20 pixels per mm for a 50µm pixel spacing, and 20.62 pixels per mm for a 48.5µm pixel spacing [177, 479]. Using the formula in Equation 6-1 on page 163, the approximate number of pixels per square millimetre (mm^2) are 425 pixels for a 48.5µm pixel spacing and 400 pixels for a 50µm pixel spacing. Average DA and BA converted to mm^2 using pixel densities of 425 pixels/mm^2 and 400/mm^2 are listed in columns 3 and 4, Table 6-2. The approximate area difference in mm^2 and % difference in area for average DA and BA for pixel spacings of 48.5µm vs 50µm are listed in columns 5 and 6, respectively.

Table 6-2 Estimation of the difference in average[1] breast BA and DA for differing pixel sizes

MD parameter	Average # of pixels (raster size)	Area in mm^2 48.5µm pixel spacing (425 pixels/mm^2)	Area in mm^2 50µm pixel spacing (400 pixels/mm^2)	Area Difference (mm^2) 50µm – 48.5µm pixel spacing[2]	% increase 50µm vs 48.5µm (ref)[3]
DA	1,286,000	3026	3205	179	5.9%
BA	7,772,000	18,287	19,430	1143	6.3%

[1] Collection 1 mammograms only (2012, n=1333 unique mammograms (Table 5-1))
[2] Area difference (column 5) = column 4 – column 3
[3] % increase in area 50µm vs 49.5µm (column 6) = column 4/column 3

The 'average' breast during Collection 1 (2012) had dense and breast area rasters of 1,286,000 and 7,772,000. The approximate average breast DA and BA areas differ by 179mm^2 (1.8cm^2) and 1143mm^2 (11.43cm^2), respectively, for 48.5µm vs 50µm pixel spacings (Table 6-2, column 5). The difference in the number of pixels per mm^2 for a 48.5µm vs 50µm pixel spacing (400 / 425 = 94.1%, a ~6% relative difference) roughly corresponds to the % increase in approximate area for average DA and BA of ~6% for the 50µm spacing vs 48.5µm spacing (column 6). Converted (adjusted) Breast and Dense area to mm^2 using Equation 6-1 (Figure 6-3, right column) shows a slight change in distribution compared to unconverted, 'raw' area in number of pixels (left column). The relative distribution for BA (upper row, Figure 6-3) amongst mammogram Versions is more affected by the conversion than DA (lower row, Figure 6-3). Whilst the raw BA distribution for film mammograms (blue) is visibly lower compared to the raw digital mammogram distributions, this difference almost disappears after conversion from number of pixels to area in mm^2. In comparison, both the raw and converted DA distributions for film mammograms (blue) is higher than all digital distributions. An increase across DA digital Versions for KE52 (pink) to KE54 (green) to Fuji (yellow) is clearly visible on both the raw and the converted DA graphs.

Statistical comparisons of the BA and DA distributions, Table 6-3, using a Kruskal-Wallis test confirmed that different raw mammogram Version distributions for BA are significantly different from each other, and the BA mammogram Version areas converted to mm^2 are not significantly different. The trend test for BA converted to mm^2 was not significant but was close to significance at the 0.05 level (p=0.059).

Figure 6-3 BA and DA by mammogram Version, 2130 mammograms from 540 episodes
Distribution before (left) and after conversion from pixel raster to area in mm^2. Left column, raw area in # pixels. Right column, area converted to mm^2 (Equation 6-1, page 163)

The conversion to mm^2 has assisted with stability of BA across mammogram Versions, because significant differences between mammogram Versions are no longer present as they were for BA measured in raw units (pixel raster). However, the trend test for BA was marginally non-significant, which is supported by the visible increase in BA with change of mammographic Version over time (Figure 6-3). DA for the different Versions differed significantly both before and after conversion, indicating substantial differences in DA were associated with mammographic Version, even after conversion of pixel raster to common units. Unlike BA, conversion of DA to mm^2 appears to have enhanced the differences in area distribution between the mammographic Versions, Figure 6-3.

Table 6-3- Statistical comparisons† of differences between BA, DA for mammogram Versions

Parameter	Comparison of area for mammogram Versions film, KE5.2, KE5.4 & Fuji*	
	Raw (unadjusted) area (# pixels)	Adjusted area (mm^2)
BA	p<0.003	p=0.184‡
DA	p<0.001	p<0.001

†Kruskal-Wallis test *16 'other' mmgs not tested. ‡p for trend = 0.059

6.4.2 Baseline Characteristics, CMN IBIS-II participants

Due to the high quality of the IBIS-II trial data, complete data was recorded for the four well established confounders of MD available for trial participants: age, body mass index (BMI), age at first birth (AFB), and age at menarche. As mentioned in 6.2 Methods, determination of another established BC risk covariate, age at menopause, was complicated for 27 participants for whom age at hysterectomy only was known; an imputed age was estimated for these participants. Age at natural menopause and/or age at ovarian oblation was known for the remaining 93 participants.

Information for all 120 CMN participants was available for all baseline exploratory covariates: number of relatives with BC and/or ovarian cancer, smoking status, duration of hormone replacement therapy (HRT), duration of oral contraceptives (OC), and previous participation in the IBIS-I trial. The median age at randomisation for the one-hundred and twenty IBIS-II participants from the Calvary Mater Newcastle hospital was 61.8 years, Table 6-4. Median body mass index (BMI) was 28.5 kg/m^2.

Table 6-4 Summary of 120 IBIS-II CMN Participant Characteristics at Baseline (Randomisation)

Parameter	N	Median or %	25th to 75th percentile (Q1 to Q3)	Min	Max
Age (years)	120	61.8	57.0 to 65.5	44.9	71.3
Height (cm)	120	162	157 to 166	141	178
Weight (kg)	120	75	66 to 86	44	136
Body Mass Index (BMI, kg/m^2)	120	28.5	25.9 to 31.8	18.3	49.4
<25 (N, %)	26	22%			
25-30 (N, %)	53	44%			
>30 (N, %)	41	34%			
Age at menarche (years)	120	13	12 to 14	9	18
Age at menopause (years)	120	49[1]	46 to 52[1]		
Natural menopause	57	50	48 to 52	25	60
Hysterectomy age	64	42.5	37 to 50.5	28	62
Ovarian oblation age	40	45	38.5 to 51.5	28	62
Nulliparous (N, %)	8	7%			
Age at first birth (years)	112	22	20 to 25	16	35
Previous HRT Use (N, median (%))	74	60 (62%)	18 to 108	1	361
Smoking Status	120				
Never (N, %)	69	58%			
Current (N, %)	11	9%			
Ex-smoker (N, %)	40	33%			
Previous Oral Contraceptive use (months)	112	90	48 to 120	1	528
Previous participant in IBIS-I (N, %)	68	57%			

N– number of participants; SD– standard deviation; Q1– first quartile; Q3– third quartile.
[1] Values shown for age at menopause include mean imputed age at menopause for 27 participants

Median age at randomisation for the CMN participants (~62) was slightly older than the median age for the entire cohort of IBIS-II participants (~59.5 years, n=3864), Table 3-1 [26]. Median height was comparable for the CMN and entire IBIS-II cohort (~162cm), however median weight (74.5 vs 72 kg) and the proportion of participants with BMI ≥ 25 were higher for CMN participants. Age at menarche and menopause were similar, however age at first birth (22 vs 24 years) and the proportion of non-parous participants (7% vs 17%) was lower for the CMN cohort. Previous HRT use was higher for CMN participants (62% vs 47% of the entire cohort).

All IBIS-II participants are at approximately double the risk of BC compared to their same-aged peers. Although not every entry criteria for the IBIS-II trial utilised family history of breast or ovarian cancer to assess BC risk, all 120 CMN IBIS-II participants who contributed mammograms to this project had at least one first or second degree relative with breast or ovarian cancer (OC). The majority of the 120 CMN participants had one (n=41) or two (n=46) relatives affected by BC or OC; twenty-four had three, and nine had four or five. Total number and measures of central tendency for the number of relatives with BC and OC, stratified by familial relationship, for the CMN participants are listed in Table 6-5.

Table 6-5 Number of relatives affected by breast or ovarian cancer, and age of diagnosis

Family History of Breast or Ovarian cancer	N	Total #	Average # per participant (SD)	Min	Max	Mean age at diagnosis (SD)
Breast Cancer – 1st degree relatives	112	146	1.3 (0.6)	1	3	51.5 (12.5)
Breast Cancer – 2nd degree relatives	61	93	1.5 (0.8)	1	4	61.2 (19.0)
Ovarian Cancer – 1st degree relatives	4	4	1	1	1	55 (13.5)
Ovarian Cancer – 2nd degree relatives	2	2	1	1	1	82 (24.0)

N–number of participants; Total #–total number of 1st and 2nd degree relatives; SD–standard deviation

Of the one-hundred and twenty IBIS-II participants from the Calvary Mater Newcastle hospital who contributed mammograms to the CMN MD and AI study, 339 mammograms were collected from 85 mammographic episodes from 85 participants for Aim 3. Baseline mammograms were taken at a variety of institutions, including the Calvary Mater Newcastle.

Characteristics of the baseline mammograms are tabulated in Table 6-6. Baseline mammography occurred ≤ 3months prior to randomisation for the majority of women. Approximately half of the baseline mammograms were film mammograms.

Median (Q1 to Q3) age at randomisation and BMI differed only slightly for the 85 CMN IBIS-II participants with baseline mammograms compared to the entire cohort of 120 participants; for the 85 participants with baseline mammograms, median age at randomisation was 61.5 (56.2 to 65.4), whilst median BMI was 28.6 (26.4 to 31.6). Distributions for other covariates were also similar.

Table 6-6 Baseline breast density parameters for the 85 CMN IBIS-II Participants with Baseline mammograms[1]*

Parameter	Mean	SD	Median	25th to 75th percentiles (Q1 to Q3)	Min	Max
Months, baseline mammogram to randomisation**	-3.5	3.2	-2.3	-5.5 to -1.0	-11.7	2.5
Film (42 women)	-4.5	3.2	-4.3	-6.7 to -1.9	-11.7	-0.1
Digital (43 women)	-2.5	3.0	-1.3	-3.3 to -0.7	-10.0	2.5
Percent Density (PD, %)	18.9	14.1	17.7	8.4 to 26.6	0.26	80.8
Film (42 women)	23.7	15.8	24.2	11.9 to 29.5	0.87	80.8
Digital (43 women)	14.2	10.2	14.3	6.6 to 18.9	0.26	51.5
Dense Area (mm^2)	3346	2660	2755	1438 to 4759	57	13215
Film (42 women)	4030	2986	3535	1673 to 5850	123	13215
Digital (43 women)	2677	2127	2135	1054 to 4045	57	8797
Breast Area (mm^2)	18724	6918	16832	14055 to 21853	7455	42592
Film (42 women)	17960	6554	16460	13347 to 20336	7823	38480
Digital (43 women)	19471	7253	17216	14437 to 23523	7455	42592
Adipose Area (mm^2)	15378	6887	13720	10617 to 19278	3213	37781
Percent Adipose (%)	81.1	14.1	82.3	73.4 to 91.6	19.2	99.7

[1] The measures of central tendency for PD, DA, BA, AA and PA were calculated for the average of each MD parameter for each baseline episode (e.g. n=85 for calculations for all 85 participants). Episodes versions comprise x42 film, x28 KE52, x12 KE54 and x1 unknown digital version.
* The baseline characteristics such as randomisation age and BMI for participants with baseline mammograms do not differ substantially from the baseline characteristics of all participants.
** All baseline mammography occurred up to 12 months prior to or on the date of randomisation, bar 2 participants whose baseline mammograms were taken 10 days and 2.5 months post-randomisation

When the number of relatives was weighted according to closeness of the relationship with the participant (i.e. first or second degree relative), most participants (98 of 120) had an approx.-imate BC risk equivalent of 1 to 2 first degree relatives, Table 6-8. With the exception of the 0.5 category, the decrease in frequency of participants in each category shows an almost linear

relationship with weighted number of relatives, Table 6-8 (graph not shown); in contrast, the relationship between (non-weighted) number of relatives and frequency is less linear, Table 6-7.

Table 6-7 Number of relatives with a history of BC and/or ovarian cancer

Number of relatives	1	2	3	4	5
All 120 CMN participants	41	46	24	6	3
Only participants with baseline mammograms (n=85)	33	31	14	6	1

Table 6-8 Weighted number of relatives with a history of BC and/or ovarian cancer*

Number of relatives	0.5	1	1.5	2.0	2.5	3.0	3.5	4
All 120 CMN participants	3	39	32	27	10	7	--	2
Only participants with baseline mammograms (n=85)	3	30	21	17	8	5	--	1

* 0.5 for each second degree relative, and 1.0 for each first degree relative

6.4.3 Baseline PD, DA, BA and AA analysis

6.4.3.1 *Assessment of correlation and collinearity among outcomes and covariates*

The expected relationships between the MD parameters and potential covariates described in Chapter 2 of this thesis generally held true. For example, age at randomisation, weight, BMI, age at first birth, were inversely associated with PD and DA. An inverse relationship between duration of HRT (in months) and baseline PD and DA was noted; this coincided with a positive relationship between duration of HRT and baseline BA and AA. Duration of HRT was also significantly and positively related to randomisation age (older women have taken HRT for longer). Collinearity testing did not indicate significant collinearity of HRT duration with other continuous parameters.

BA and AA differed by smoking status: BA and AA for never smokers was smaller than BA and AA for current smokers, who in turn had smaller than BA and AA than ex-smokers (never < current < ex-smokers). PD/DA was lower for ex- and never smokers, and higher for current smokers. The interrelationship of smoking with height, weight and age largely explain the apparent relationships of the MD parameters with smoking status. BMI was highest for ex-smokers, lowest for current smokers and in-between for never smokers, which was potentially

due to the influence of smoking on appetite and hence weight. Current smokers also tended to be younger and taller than their non- or ex-smoking counterparts.

6.4.3.2 *Assessment of correlation and collinearity among explanatory variables*

As for smoking status described above, many of the covariates were interrelated; however most interrelationships between covariates were weak. For instance, a line of best linear fit overlaid on scatterplots of randomisation age and BMI, randomisation age and BMI, and randomisation age and age at first birth revealed a weak inverse relationship between randomisation age and these other covariates. However when the scatter plots were viewed without the guidance of linear best fit they showed little relationship between these (continuous) covariates. During graphical comparison of continuous parameters with categorical parameters, BMI was also found to be lower for non-parous participants compared to BMI for parous participants.
Use of the collin function yielded VIFs of <1.5, tolerances >0.65 and condition indices <30 (for all dimensions except the last) for the following set of continuous parameters: randomisation Age, BMI, age at menarche, age at first birth, menopause age, length of HRT in months, length of oral contraceptives in months, and weighted number of relatives with OC/BC. When the categorical parameters smoking status, mammogram Version, CC vs MLO, right vs left mammogram, and follow up in months, were added to the collin test, VIFs were <1.4, tolerances >72%, and condition indices <30 for all but the last dimension. This indicated significant collinearity was not present for these sets of parameters.

Use of the coldiag2 test for collinearity amongst categorical parameters (parous status, weighted number of relatives, smoking status, mammogram Version, CC vs MLO, right vs left) yielded a maximum condition index of 28. This condition index of 28 had variance-decomposition proportions >0.3 for the constant (0.94) and mammogram Version (0.75). The other condition indices were <13. This indicated that significant collinearity was not present for this set of categorical parameters.

6.4.4 Simple regression of average baseline MD parameters vs baseline covariates

The results of the simple regression for each baseline covariate and MD parameter (PD, DA, BA, AA), and with square root transformations of each MD parameter are tabulated in Table 6-9, below.

Age like BMI is typically a strong confounder of breast density which has an inverse relationship with PD and DA. A significant negative relationship was found between age at randomisation and PD and DA assessed from digital mammograms, however the relationship of age with PD and DA was non-significantly positive for film mammograms. Omission of all three women with high PD (≥50%)— two with baseline film mammograms— strengthened the all mammogram inverse relationship with PD to –0.44, p=0.054, R^2 5%. The film mammogram coefficient increased non-significantly to –0.25.

Two strong confounders of MD, BMI and weight, were consistently and significantly associated with PD, BA and AA, both for the all mammograms relationship [i.e. both Types of mammograms—film and digital] and within each Type (all expected). The association was negative for PD because PD tends to decrease as weight increases, and positive for BA and AA because breast size tends to increase with weight and height. DA did not show a significant relationship with BMI or weight.

As a single parameter, weight explained about 30 to 60% of the total variability of BA and AA for film, digital and all mammograms (R^2 values ranged from 0.31 to 0.58). The explanatory relationship was stronger for digital mammograms than for film mammograms (50 to 60% for digital vs 30 to 40% for film). The percentage of variability in PD explained by weight was stronger for film mammograms however (16 to 20%) than for digital (7%).None of the relationships tested (PD, DA, BA & AA, for all mammograms and by Type) were significantly associated with height. The all mammogram relationship as well as the within-Type

Table 6-9 Simple regression, Baseline Average MD parameters with established confounders & potential covariates

Covariate	N	PD (%)		PD sqrt		DA (mm^2)		DA sqrt		BA (mm^2)		BA sqrt		AA (mm^2)		AA sqrt	
		β	R^2	β	R^2	β	R^2	β	R^2	β	R^2	β	R^2	β	R^2	β	R^2
Randomisation Age years (yr) All mmgs	85	-0.25[1]	0.01	-0.05	0.02	-51[1]	0.01	-0.6	0.17	26	<0.001	0.17	0.002	76	0.004	0.3	0.004
Film-screen	42	0.14[1]	0.002	0.002	<0.001	43[1]	0.01	0.14	0.001	105	0.01	0.4	0.006	62	0.002	0.2	0.001
Digital (CR)	43	**-0.6***	0.13	**-0.09***	0.15	**-127***	0.13	**-1.3***	0.13	-8	<0.001	0.11	0.001	120	0.01	0.6	0.02
KE52	28	**-0.7***	0.23	**-0.10***	0.21	**-125***	0.17	**-1.3***	0.16	134	0.01	0.6	0.02	259	0.05	1.1	0.06
KE54[2]	12	-0.3	0.03	**-0.04***	0.05	-85	0.07	-0.9	0.08	-272	0.04	-0.7	0.03	-187	0.03	-0.5	0.02
Fuji (none)	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
BMI (kg/m^2)	85	**-0.9****	0.11	**-0.13****	0.15	-41	0.01	-0.7	0.03	**932*****	0.48	**3.2*****	0.47	**973*****	0.53	**3.7*****	0.50
Film-screen[a]	42	**-1.4***	0.14	**-0.19****	0.21	-120	0.03	-1.3	0.06	**975*****	0.40	**3.5*****	0.42	**1096*****	0.48	**4.3*****	0.47
Digital[a]	43	**-0.6***	0.10	**-0.09***	0.13	13	0.001	-0.2	0.003	**900*****	0.54	**3.0*****	0.52	**887*****	0.60	**3.2*****	0.58
KE52	28	**-0.6***	0.18	**-0.13****	0.25	-25	0.01	-0.8	0.008	**887*****	0.48	**3.2*****	0.48	**912*****	0.55	**3.4*****	0.54
KE54[2]	12	-0.5	0.14	-0.05	0.10	52	0.03	0.4	0.02	**924****	0.62	**3.0****	0.59	**871*****	0.75	**3.1*****	0.72
Height (cm)	85	-0.2	0.01	-0.01	0.001	-8	<0.001	0.12	0.001	132	0.02	0.5	0.02	140	0.02	0.6	0.03
Film-screen	42	-0.5	0.04	-0.05	0.03	-43	0.01	-0.4	0.01	124	0.02	0.4	0.01	167	0.02	0.7	0.04
Digital	43	0.2	0.01	0.04	0.04	41	0.02	0.7	0.06	124	0.01	0.5	0.02	83	0.02	0.4	0.01
Weight (kg)	85	**-0.3****	0.11	**-0.04****	0.13	-14	0.01	-0.2	0.01	**302*****	0.45	**1.0*****	0.44	**316*****	0.49	**1.2*****	0.48
Film-screen[a]	42	**-0.4****	0.16	**-0.06****	0.20	-40	0.04	-0.4	0.06	**257*****	0.31	**0.9*****	0.32	**297*****	0.39	**1.2*****	0.39
Digital[a]	43	**-0.2**	0.07	**-0.02**	0.07	12	0.01	0.07	0.003	**331*****	0.56	**1.1*****	0.54	**319*****	0.59	**1.2*****	0.58
Menarche (years)	85	0.7	0.01	0.10	0.01	166	0.01	1.5	0.01	-117	<0.001	-0.3	<0.001	-283	0.005	-0.9	0.003
Film-screen	42	0.9	0.01	0.16	0.02	346	0.04	3	0.04	203	0.003	0.8	0.004	-143	0.001	-0.09	<0.001
Digital	43	-0.9	0.02	-0.11	0.01	-213	<0.001	-2	0.02	-243	0.003	-0.7	0.002	-30	<0.001	0.08	<0.001
Nulliparous – yes (parous ref)	85	-5	0.007	-0.5	<0.001	-1249	0.01	-11	0.01	-3284	0.01	-12	0.01	-2035	0.005	-7.6	0.005
Age at First Birth (yr)	80	0.6	0.02	0.06	0.02	**153***	0.05	**1.2**	0.05	226	0.02	0.7	0.01	73	0.002	0.2	0.001
Film-screen	42	-0.1	<0.001	-0.01	0.001	10	<0.001	-0.2	0.001	19	<0.001	-0.11	<0.001	9	<0.001	-0.05	<0.001
Digital	38	**0.8***	0.12	**0.11***	0.10	**235****	0.22	**2.1****	0.18	413	0.06	1.4	0.06	177	0.01	0.6	0.02
Menopause (years)																	
Natural+ovariectomy[7]	67	**0.7****	0.12	**0.10****	0.16	**135****	0.13	**1.4*****	0.18	-28	<0.001	-0.06	<0.001	-163	0.03	-0.6	0.03
Film-screen	32	0.5	0.02	0.04	0.01	82	0.02	0.5	0.02	-22	0.001	-0.003	<0.001	-104	0.01	-0.4	0.008
Digital	35	**0.5****	0.19	**0.09****	0.24	**118****	0.23	**1.4****	0.30	51	0.003	0.2	0.004	-66	0.006	-0.2	0.004

Table 6-9 continued, Simple regression baseline coefficients

Covariate	N	PD (%)		PD sqrt		DA (mm^2)		DA sqrt		BA (mm^2)		BA sqrt		AA (mm^2)		AA sqrt	
		β	R^2	β	R^2	β	R^2	β	R^2	β	R^2	β	R^2	β	R^2	β	R^2
Reported+imputed[7]	85	**0.5**	N/A	**0.07***	N/A	**93**	N/A	**0.98***	N/A	8.5	N/A	0.06	N/A	-84	N/A	-0.32	N/A
Film-screen	42	0.2	N/A	0.005	N/A	30	N/A	0.12	N/A	52	N/A	0.23	N/A	21	N/A	0.06	N/A
Digital	43	**0.4**	N/A	**0.08***	N/A	**95**	N/A	**1.2***	N/A	33	N/A	0.14	N/A	-62	N/A	-0.18	N/A
HRT ever v never (ref)	85	-1.0	0.001	-0.09	0.001	-91	<0.001	-0.77	<0.001	232	<0.001	0.9	<0.001	322	<0.001	1.4	0.001
HRT ever (months)	54	-0.04	0.04	**-0.01**	0.07	-5.6	0.02	-0.06	0.04	14	0.03	0.06	0.03	**20**	0.05	**0.08**	0.06
Film-screen	26	0.04	0.03	0.01	0.04	11.2	0.05	0.1	0.07	8	0.01	0.04	0.01	-3	0.001	0.006	<0.001
Digital	28	**-0.06****	0.24	**-0.01****	0.27	**-10.8***	0.22	**-0.11****	0.23	14	0.03	0.06	0.04	25	0.10	**0.10**	0.11
OC ever vs never (ref)	85	3.2	0.003	0.36	0.003	377	0.001	4.6	0.002	1255	0.002	4.1	0.002	878	<0.001	2	<0.001
OC ever use (months)	80	0.004	<0.001	<0.001	0.002	2	0.003	0.02	0.007	10	0.01	0.03	0.01	8	0.01	0.03	0.01
Film-screen	40	0.006	0.001	<0.001	0.003	-0.3	<0.001	0.005	<0.001	-4	0.003	-0.01	0.002	-3	0.002	-0.1	0.001
Digital	40	0.02	0.01	0.003	0.02	8	0.07	0.08	0.07	**31**	0.09	**0.1**	0.09	23	0.05	0.08	0.05
Smoking- never (ref)	44	ref.	0.01	ref.	0.01	ref.	0.02	ref.	0.02	ref.	0.01	ref.	0.01	ref.	0.01	ref.	0.01
Current	9	2		0.39		1176		9.3		1840		6.1		664		3.4	
Ex-smoker	32	-2		-0.22		-100		-0.6		1333		5.6		1432		6.8	
IBIS-I yes (vs no ref)	85	-2.4	0.007	-0.22	0.005	-123	<0.001	-1.6	0.001	606	0.002	2	0.002	729	0.003	3.5	0.004
Number of relatives with BC or OC	85	1.6	0.01	0.12	0.005	222	0.007	1.7	0.005	403	0.003	1.4	0.003	180	0.001	0.09	<0.001
Weighted # relatives with BC or OC as a continuous parameter: First degree = 1; Second degree = 0.5 (min 0.5 to 4.0 max)																	
All mammograms	85	0.5	0.02	0.05	0.009	55	0.005	0.42	0.004	-120	0.004	-0.4	0.003	-175	0.008	-0.8	0.01
Film-screen	42	**1.3[3]**	0.08	0.12	0.06	125[3]	0.02	0.9	0.02	-539	**0.08**	**-1.8**	0.07	**-665*[3]**	0.11	**-2.7***	0.12
Digital	43	0.3	0.01	0.03	0.006	70	0.01	0.6	0.01	210	0.01	0.8	0.01	140	0.005	0.5	0.005
Weighted # relatives with BC or OC as a categorical parameter																	
One 1st degree (ref)	30	ref.	0.06	ref.	0.07	ref.	0.05	ref.	0.05	ref.	0.13	ref.	0.13	ref.	0.18	ref.	0.15
One 2nd degree	3	-14		-2		-1981		-16		**12894****		**43****		**14875*****		**52**	
One 1st + one 2nd	21	-3		-0.8		-891		-11		975		7		**3219**		11	
Two 1st	17	-1		-0.2		-446		-5		-1640		-3		158		-1.4	
One 2nd + two 1st or more relatives	14	3		0.1		525		2		417		6		1244		3	
Mammogram Type – film (ref)	42	ref.	0.12	ref.	0.10	ref.	0.07	ref.	0.06	ref.	0.01	ref.	0.01	ref.	0.04	ref.	0.05
Digital	43	**-9.5**	--	**-1.1****	--	**-1353***		**-11***		1511		5.2	--	**2864**		**12***	--

Table 6-9 continued, Simple regression baseline coefficients

Covariate	N	PD (%)		PD sqrt		DA (mm^2)		DA sqrt		BA (mm^2)		BA sqrt		AA (mm^2)		AA sqrt	
		β	R^2	β	R^2	β	R^2	β	R^2	β	R^2	β	R^2	β	R^2	β	R^2
Mammogram Version[5]- film (ref)	42	ref.	0.13	ref.	0.10	ref.	0.08	ref.	0.07	ref.	0.02	ref.	0.02	ref.	0.05	ref.	0.16
KE52	28	**-10.3**	--	**-1.1****	**--**	**-1571***	--	**-13***	--	1521	--	5	--	**3092**	--	**13***	--
KE54	12	**-9.0**	--	**-0.9**	--	-1040	--	-8	--	2270	--	7	--	3308	--	13	--
Fuji	0	--	--	-0.7	--	--	--	-8	--	--	--	-1.3	--	--	--	-0.5	--
Comparison of individual mammograms: 4 per baseline episode (not MD averages of each follow up)																	
MLO mmgs[6] v CC (ref)	339	-1.06	0.001	**--**	**--**	-85	<0.001	**--**	**--**	731	0.003	**--**	**--**	817	0.003	**--**	**--**
L. mmgs[6] v right (ref)	339	0.02	<0.001	--	--	42	<0.001	--	--	229	<0.001	--	--	188	<0.001	--	--
Mmg View[6], RCC (ref)	85	ref.	0.001	--	--	ref.	<0.001	--	--	ref.	0.003	--	--	ref.	0.004	--	--
LCC	85	0.006	--	**--**	**--**	0.8	--	**--**	**--**	274	--	**--**	**--**	273	--	**--**	**--**
RMLO	85	-1.1	--	--	--	-126	--	--	--	775	--	--	--	900	--	--	--
LMLO	85	-1.0	--	--	--	-43	--	--	--	964	--	--	--	1007	--	--	--

y, yr years; ref. denotes the reference category; **Bold** indicates the coefficient (β)p-value is less than or equal to 0.1; * p<0.05; ** p<0.01; *** p<0.001;

[a] R^2 is similar (±1%) if the participant with PD≥75% (n=1) is omitted from the models. R^2 was slightly more affected (by up to ±3%) if participants with PD≥50% (n=3) are omitted from the models; e.g. the PD and weight model R^2 changed slightly: the film-only model R^2 decreased –3 to 17% and digital only model R^2 decreased –2 to 5%

[1] If a single influential episode with PD≥75% is excluded (outlier on residual vs fitted plot), the magnitude of the coefficient (β) increases, R^2 increases and the sign is negative.

[2] Baseline KE54 are few. Hence sub-group comparisons are subsequently made for mammogram Type only; NB: three digital baseline mammograms were of Version 'other'

[3] Relationship decreases (smaller coefficient (0.7) & R^2, p=0.24) when the influential participant with PD≥75% is omitted

[4] Weighted number of relatives encompasses equivalent risk combinations. Women with one 2nd degree relative not used as reference category due to small cell size (n=3)

[5] Three mammograms of Version 'other' omitted from the comparison

[6] All baseline mammogram measurements utilised (4 views/episode), not the averages of all four Views of PD, DA, BA, & AA for each episode

[7] Natural menopause and ovariectomy age reported at baseline. Age at menopause imputed for women with date known for only for hysterectomy.

relationships for film and digital mammograms with height were non-significantly but consistently associated with larger overall BA as well as AA (i.e. the regression coefficients (βs) for BA and AA were positive).

The overall (combined film+digital) relationship between PD (%) and height was non-significantly negative (–0.2 % for each 1 cm increase in height). The all mammograms relationship was comprised of a non-significant negative relationship between PD and film mammograms (–0.5) and a non-significant positive relationship for digital mammograms (0.2). Omission of the three women with high baseline PD attenuated the relationship for film mammograms with PD and DA (the coefficient decreased in size to –0.13 and R^2 reduced to 0.006).

The all mammogram relationship between menarche and PD & DA was non-significantly positive but within-Type relationships differed. Film showed a non-significant positive relationship, whilst the relationship for digital mammograms was non-significantly negative. R^2 for PD and DA ranged from ~1 to 4%, whilst age at menarche provided little explanation for the variability in BA and AA (R^2 0.5% or less).

Age at first birth (n=80) was significantly (p=0.031) and positively associated with digital baseline PD, but non-significantly and inversely associated with film mammograms. When the three women with high baseline PD (≥50%) were omitted from the simple regression equations, the association was non-significant but positive for film mammograms and marginally non-significant for all mammograms (p=0.065, β=0.61). R^2 for all mammograms also increased to ~5% (n=77 women). The relationship for DA strengthened with omission of the high PD women: the coefficient for DA (mm^2) became 165, p=0.014 with an R^2 of 8% for the all mammograms regression. The relationship for PD and DA remained non-significant and inverse for nulliparous women, with or without women with high PD in the model. A positive relationship was seen for both PD and DA with menopausal age, for all mammograms and

within each Type. The relationship was stronger for women with known age at menopause (natural or via ovariectomy) than for imputed + known age at menopause, and both stronger and significant for digital mammograms vs film mammograms.

Ever use of HRT was not significantly associated with higher or lower PD or DA in this dataset, and its explanatory strength for the variability in any of the MD parameters was very low (R^2 ~0.001(~0.1%)). Univariable HRT duration had a stronger explanatory relationship with most MD parameters (R^2 from 2 to 27%) and was significantly associated with lower PD and DA for digital mammograms. HRT duration was also associated with an non-significant increase in AA (p=0.092) for all mammograms. Omission of the three women with high baseline PD increased the explanatory power of HRT duration for both all and film-only mammograms, with increases of R^2 to 5% and 8% respectively. The coefficient for all mmgs decreased slightly to –0.033, but p-value also increased to 0.097.

Past OC use had a very low explanatory relationship with all MD parameters (0.2% or less) as well as a non-significant and positive association. OC duration tended to have a positive non-significant association with the MD parameters, but similarly had very low R^2 values for all regression relationships except for digital mammogram-only models. Whilst still non significant, DA, BA and AA had coefficients of ~0.5 and R^2 of 5 to 7% for digital mmgs. Omission of the three women with high MD changed the coefficients slightly but non-significantly; for instance DA (mm^2) had a non-significant coefficient of 0.8 and R^2 of 0.001 (film mammograms). In this dataset, current smokers tended to have non-significantly higher PD and DA but non-significantly higher BA and AA than never smokers. Former smokers had non-significantly lower PD and DA but non-significantly higher BA and AA than never smokers. The explanatory relationship of smoking and the MD parameters was low (R^2 ~1%).

Neither total number of relatives nor weighted total number of relatives (1 for first degree and 0.5 for second degree) was associated significantly with PD or DA, although R^2 was around 8%

for PD and the (non-significant) associations for all, film and digital mammograms were positive. The relationship of weighted number of relatives as a continuous parameter with PD in film mammograms was attenuated if the woman with the highest PD (~80%) or all three women with high PD ≥50% were omitted from the regression model.

One of the assumptions for linear regression is homoscedasticity of the residuals. The right skewed distributions of the untransformed MD parameters, e.g. high PD for only a few individuals in the dataset, created right skewed residual distributions. Log transformation of PD and DA produced residual plots with a left skew, whilst square root transformation of PD, DA, BA and AA produced the best approximation of a normal distribution for both the dependent variable and model residuals.

The relationship between the distribution of the dependent variable and the distribution of the regression residuals is demonstrated for PD in Figure 6-4. Frequency histograms of the distributions for untransformed baseline PD (in %, ranging from 0.3% to 81%, Table 6-6, n=85), square root transformed baseline PD (ranging in value from 0.5 to 9) and natural log transformed PD (ranging in value from –1.4 to 4.4) are shown in the top row of the figure. The lower row of Figure 6-4 displays scatter plots of the residuals (y-axis) vs the PD predicted values (x-axis) from simple regression models of BMI and the different transformations of PD. To meet model assumptions of homoscedasticity, the residuals should be evenly scattered around the value of 0 on the y-axis. Residuals from the square root transformation of PD (lower row, middle column) are approximately evenly scattered around 0, however the residuals for untransformed PD (lower row, left column) show a right skew (the residual values >20%, which were generated for participants with observed baseline PD >40%). The distribution of the residuals from the regression model of untransformed PD hence mirrors the distribution of the dependent variable of the model. Similarly, the left skew of the residuals from the regression model with natural log transformed PD reflects the left skew of the dependent variable (right column, Figure 6-4).

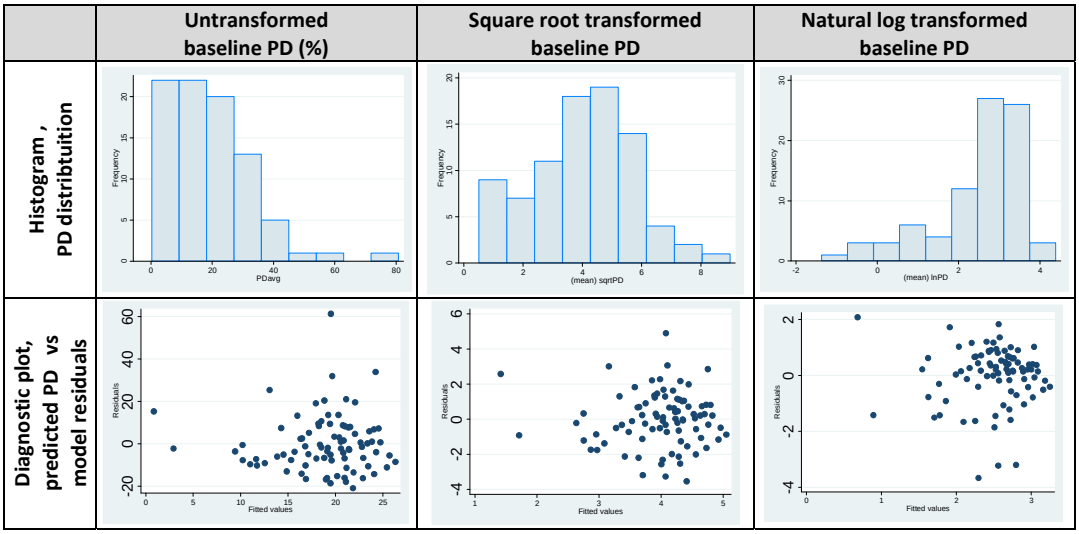


Figure 6-4 PD distribution histograms (upper row), and simple regression residuals vs predicted (fitted) plots for untransformed PD, square root transformed PD, and natural log transformed PD
The distribution of the dependent variable in the regression models (upper row) is reflected in the distribution of the residuals (y-axis) vs PD predicted values (x-axis) in the scatter plots of the lower row. Whilst the residuals are scattered evenly and approximately normally around 0 for the square root transformation of PD (middle column), the residuals are skewed to the right and left, respectively, for the regression models of untransformed and natural log transformed PD (left and right columns).

6.4.5 Simple regression, different categorisations AFB & OC, HRT duration

The coefficients for the simple regression models utilising the re-categorisations of AFB, OC duration and HRT duration are presented in Table 6-10. Significant differences in DA were found between younger AFB vs older AFB for the AFB ≥ 26 (p=0.027) and AFB ≥ 30 (p=0.004) categories; the p-values for the same relationships for square root transformed DA were very similar to those for untransformed DA: p=0.031 for AFB ≥ 26, and p=0.009 for AFB ≥ 30 (compared to younger AFB). There was some evidence (p=0.078) of a difference in square root transformed PD for AFB ≥ 30 vs AFB < 30. A possible dose-response trend for increasing PD (p=0.092) and DA (p=0.046) with increasing AFB was noted for the quintile division of AFB (parous participants only); square root transformed PD (p=0.097) and DA (p=0.04) reflected the trend noted for the untransformed MD models. The PD model p-value for the trend decreased (p=0.072) with omission of the woman with the highest PD (~80% PD, ≤22 AFB), but increased to p=0.174 with omission of three women with PD ≥50% (one of

whom had AFB ≥ 30). P-values for square root transformed PD models with these participants omitted were similar to those for PD, however p-values for both DA and square root transformed DA models for PD <70% and PD <50% only were lower (p= ~0.03). The discrepancy between the PD and DA models is because some participants with high DA had mid-range PD (~20 to 40%). No trend was detected for AFB and BA or AA.

Table 6-10 Simple regression coefficients (β), Baseline avg MD with categorical AFB, OC and HRT

Covariate	N	PD (%)	PD sqrt	DA (mm^2)	DA sqrt	BA (mm^2)	BA sqrt	AA (mm^2)	AA sqrt
		β	β	β	β	β	β	β	β
Age at First Birth (AFB), years									
AFB <20y (ref)	18	ref	ref	ref	ref	ref	ref	ref	ref
≥20 years	62	1	-0.14	185	-0.8	1272	4.5	1087	3
Non-parous	5	-4	-0.61	-1105	-11	-2298	-9	-1193	-5
AFB <26y (ref)	69	ref	ref	ref	ref	ref	ref	ref	ref
≥26 years	11	6	0.85†	**1898***	**16***	2309	6	411	0.9
Non-parous	5	-4	-0.38	-988	-8.5	-2966	-12	-1978	-7.5
AFB <30y (ref)	74	ref	ref	ref	ref	ref	ref	ref	ref
≥30 years	6	**10**	**1.3**	**3200****	**25****	3722	13	522	3
Non-parous	5	-4	-0.40	-1009	-9	-3005	-11	-2000	-7
AFB ≤ 20y (ref)	23	ref	ref	ref	ref	ref	ref	ref	ref
21 to 22	20	3.0	0.2	582	3.3	-266	0.09	-848	-3
23 to 25	26	3.6	0.3	358	2.7	324	0.9	-33	-1.5
26 to 29	5	2.9	0.5	614	6.7	612	-1	-2	-3
≥ 30	6	12.2[1]	1.4[2]	**3525**[3]**	**27.6***[3]	3806	13	281	1.4
Non-parous	5	0.1	-0.2	-684	-6.5	-2921	-11	-2236	-9
Oral Contraceptives (OC), months									
OC Never (ref)	5	ref	ref	ref	ref	ref	ref	ref	ref
Past user	80	3	0.36	377	5	1255	4	878	2
OC Never (ref)	5	ref	ref	ref	ref	ref	ref	ref	ref
48 months	20	5.2	0.53	475	5	-550	-2	-1026	-6
90 months	17	0.5	-0.03	-247	-2	-457	-2	-211	-1.5
120 months	27	3.6	0.46	535	7.0	2171	7	1636	5
≥120 months	16	2.7	0.37	649	7.5	3786	13	3137	11
Hormone Replacement Therapy (HRT), months									
HRT Never (ref)	31	ref	ref	ref	ref	ref	ref	ref	ref
Past user	54	-1	-0.09	-91	-0.8	232	0.9	323	1.4
HRT never (ref)	31	ref	ref	ref	ref	ref	ref	ref	ref
18 months	14	0.2	0.17	4	2	573	1	569	1.3
60 months	18	2.3	0.28	630	5	112	0.5	-517	-2.2
108 months	11	-4.3	-0.48	-1224†	-10	-3739†	-13	-2514	-8.4
109+ months	11	-4.5	-0.66	-257	-4	3964†	15	4221	17

Bold indicates the coefficient (β) has **p≤0.1**; ref. reference category; y years;
* p<0.05; ** p<0.01; *** p<0.001; † p<0.2†;
[1] p=0.092 for trend (parous only, n=80); p=0.072 for PD<70% (n=79); p=0.174 for PD<50% (n=77);
[2] p=0.097 for trend (parous only); p=0.08 for PD<70%; p=0.18 for PD<50%;
[3] p<0.05 for trend (parous only)

The AFB categorisation which was most parsimonious, had the lowest p-value for one or more categories and highest R^2 was AFB divided at age 30 years. This parameter was selected for use in the multivariable models of MD.

Non-linear trends predominated within the quintile (dose-response) divisions of OC and HRT use. Thus the binary categories of HRT ever vs never use and OC ever vs never use were chosen for utilisation in the MD multivariable linear regression models as the most parsimonious representations.

6.4.6 Multivariable regression, avg baseline MD parameters vs baseline covariates

During collinearity testing prior to fitting the simple regression models, it was noted that many covariates were interrelated. Relationships were noted between smoking status, age and BMI, for example. To account for these interrelationships, multivariable (multiple) linear regression was undertaken. Full models and age- and mammogram Version-adjusted parsimonious models containing parameters with $p \leq 0.10$ were estimated for the MD parameters. The results from the full and parsimonious (parsimon.) models for all 85 baseline mammograms are listed in Table 6-11 below.

Coefficients for the multivariable models generally showed the expected sign (e.g. negative for randomisation age and BMI in models with PD and DA as the dependent variable). Use of square root transformed PD provided a significant value for age at randomisation for the square root PD full ($p=0.044$) and parsimonious ($p=0.031$) models, whilst the coefficient for this covariate in the untransformed PD models was not significant (e.g. $p=0.133$ for the PD parsimonious model). With the exception of age at randomisation, coefficients which were significant in the untransformed models were also significant in the square root transformed MD models. No additional covariates became significant with use of transformed MD as the

Table 6-11 Multivariable linear regression coefficients, Baseline Average MD with confounders of breast density and other potential covariates

Covariate	PD %		PD square root		DA mm^2		DA square root		BA mm^2		BA square root	
	Full	Parsimon	Full	Parsimon	Full	Parsimon	Full	Parsimon	Full	Parsimon	Full	Parsimon
ALL MMGS	R^2	R^2	R^2	R^2	R^2	R^2	R^2	R^2	R^2	R^2	R^2	R^2
85 participants	31%	24%	37%	32%	27%	26%	29%	23%	60%	59%	59%	58%
Age at Rand.	-0.33	-0.35	**-0.07***	**-0.06***	-47	-31	-0.75	-0.53	**204**	**198**	**0.79***	**0.77***
BMI (kg/m^2)	**-0.89****	**-0.88****	**-0.13*****	**-0.13*****	-12	--	-0.52	--	**1086*****	**1066*****	**3.7*****	**3.7*****
Menarche (years)	0.33	--	0.06	--	142	--	1.2	--	128	--	0.57	--
Menopause (years)	0.32	--	**0.07**	**0.07**	**84**	**83**	**0.90***	**0.82***	92	--	0.34	--
Age at First Birth												
<30 years	ref.	--	ref.	--	ref.	--	ref.	--	ref.	--	ref.	--
≥ 30 years	8.3	--	0.97	--	**3293****	**3265****	**24***	**24***	**5879****	**5825****	**20***	**20***
Non-parous	-3.1	--	-0.54	--	-82	219	-4.5	-1.7	2909	1530	8.2	4.3
OC use (months)												
Never users	ref.	--	ref.	--	ref.	--	ref.	--	ref.	--	ref.	--
Ever users	-0.71	--	-0.33	--	-574	--	-6.9	--	1152	--	2.8	--
HRT (months)												
Never users	ref.	--	ref.	--	ref.	--	ref.	--	ref.	--	ref.	--
Ever users	2.2	--	0.42	--	245	--	3.8	--	**-2136**	**-2241***	**-7.2**	**-7.8**
Smoking- never	ref.	--	ref.	--	ref.	--	ref.	--	ref.	--	ref.	--
Current	2.5	--	0.40	--	**1539**	**1581**	12	--	**3687**	**3407**	**12**	**11**
Ex-smoker	1.1	--	0.15	--	-120	-145	0.0	--	-1812	-1725	-5.0	-4.8
IBIS-1 participant no (ref)	-1.2	--	-0.001	--	69	--	1.0	--	78	--	-0.56	--
Mammogram Version[5]- film	ref.	--	ref.	--	ref.	--	ref.	--	ref.	--	ref.	--
KE52	**-9.7****	**-10****	**-0.99****	**-0.96****	**-1547***	**-1629***	**-12***	**-12***	1176	1060	4.3	3.7
KE54	-6.7	**-7.4**	-0.54	-0.64	-528	-626	-3.3	-6.1	1399	1332	3.9	3.7
Intercept	**28**	**31*****	**6**	**6**	4977	**4141*****	**73***	**63*****	**12382**	**14922*****	**114*****	**121*****

Bold indicates the coefficient (β) is significant (p≤0.1); Parsimon. column lists significant parameters' β (age-adjusted, except BA)
ref. denotes the reference category; * p<0.05; ** p<0.01; *** p<0.001; ‡ p<0.15; † p<0.2;
R^2 values from regression models with average imputed values for menopause age

dependent variable in the models. As expected, BMI was not significant in the DA models. As for square root PD, age at menopause is associated with a significant increase in DA.

Whilst HRT duration indicated a possible positive dose-response relationship with BA in simple regression (Table 6-10), after controlling for other factors such as BMI and smoking status, HRT use was associated with a significant decrease in BA. Removal of smoking status from the parsimonious model did not substantially change the coefficient for HRT use (–2071, p=0.074); the other coefficients in the model also decreased slightly.

Residuals for the multivariable parsimonious models of PD, DA and BA were right skewed for untransformed MD models (heteroskedastic), but not as skewed as the residuals for the simple regression models with untransformed PD, DA and BA as the dependent variable. Residuals from the square root transformed multivariable parsimonious MD models were approximately normally distributed (homoscedastic).

6.4.7 Comparison of parsimonious MI and non-MI models

Comparison of the coefficients from the square root transformed PD MI and non-MI models revealed that all parsimonious parameters in the non-MI model continued to meet the pre-specified criterion for retention in a parsimonious model ($p \leq 0.1$). The coefficients for age at randomisation and BMI changed less than 3% (e.g. from –0.065 to –0.063 for randomisation age), and the coefficients for mammogram Version changed less than 10%. The coefficient for menopause age was the most affected, changing 35% from 0.069 in the non-MI model to 0.051 in the properly imputed (MI) model. The SE for both coefficients was 0.028. Hence the relative change in the age at menopause coefficient was still well within ± one SE limits, despite an apparent overestimation in magnitude for the non-MI model coefficient.

6.4.8 Differences in MD measurements, by View, Type and Version

As described in Differences in MD measurements, by View, Type and Version, additional comparisons were made to characterise the differences in MD measured on CC vs MLO and right vs left mammograms, by mammogram Type and Version. These comparisons showed the CC mammograms had significantly higher PD (~1%, p=0.0009) but significantly lower BA (–8cm^2, p<0.0001) and AA (–3cm^2, p<0.0001) on average than PD, BA and AA measured on MLO mammograms (Table 10-6). BA and AA were also significantly lower on average for right mammograms compared to left mammograms (–4cm^2, p=0.002 (BA) and –3cm^2, p=0.004 (AA), Table 10-6) however PD did not differ significantly for right vs left Views (p=0.88). DA was non-significantly lower (–0.7cm^2, p=0.33) for right compared to left mammograms, and non-significantly higher (0.5cm^2, p=0.41) for CC vs MLO Views (p=0.41), Table 10-6. Wilcoxon's paired signed-rank test was utilised for all comparisons.

6.4.9 Simple regression of baseline PD with different View combinations

The PD simple regression coefficients for different (single) Views, the average of pairs of Views, and the average of all four Views are presented in an appendix (Simple regression of baseline PD with different combinations of mammographic Views, Table 10-8). With the exception of models fitted with LMLO View mammograms, coefficients from the untransformed PD single View models (RCC, LCC, RMLO) were generally similar (± 10 to 20%) to those for the all View models of untransformed PD. The LMLO View model coefficients tended to be smaller in magnitude and therefore non-significant when the coefficients from models with other View combinations were significant. Some of the left side model coefficients were similarly slightly smaller in magnitude and non-significant compared to their four View counterparts. The coefficients for the three different View combinations for log transformed PD were similar in both magnitude and significance, as were the coefficients for models utilising a square root transformation of PD. The relationships for both the weaker modifiers of MD, such as menarche and age at first birth, as well as stronger modifiers such as

age, BMI and weight, did not appear to change with use of one, two or four Views in the PD models.

6.5 Discussion

The calculated Breast Area for baseline mammograms (n=85 participants, averaged from 4 views from 85 baseline episodes, Table 6-6) of 18,724 mm^2 (~187cm^2) is higher than average breast size reported elsewhere. Slightly younger and thinner cohorts of participants (average age ~51to57 years, average BMI ~26) had digitised CC view breast sizes of median 123.7 cm^2 (Q1-Q3: 94.7 cm^2 to 65.8 cm^2) [177], 74.4 cm^2 (average, SD 42.1 cm^2) [479], and ~75 cm^2 (average) [192].

Reasons for the larger estimated BA for CMN IBIS-II participants are not clear, as the method to calculate area on the IBIS-II mammograms is similar to others [177, 479]. The candidate may have over-estimated/been generous whilst outlining the breast during Cumulus evaluation. The report listed above with the largest BA used xeromammograms, whose 'positive' image has a well-defined breast edge [177] . Perhaps the 4.3OD and 4.7OD setting used whilst scanning the breast x-rays made the 'true' edge of the breast more visible than the film-screen scans used in earlier reports. Median BA for the CMN IBIS-II participants is not too different from median BA measured during 82 serial screening CC view mammograms, 145.7 cm^2 (Q1-Q3: 109 cm^2 to 198 cm^2) [480]. Breast measurements were made along the chest wall, and from the chest wall to the nipple. BA at the 5th percentile was ~77 cm^2 whilst it was ~264 cm^2 at the 95th percentile. Age and BMI were not reported for this study, so it is not possible to take these MD confounders into account. In the US, screening starts from age 40, so it is highly likely the women in this study were younger (leading to perhaps, firmer less compressible breasts, with possibly lower BMI (also likely to decrease BA)) than the CMN IBIS-II participants.

Both CC and MLO views were used to calculate median BA for the IBIS-II participants, however MLO views tend to have a larger BA than CC views [481]. Restriction of the IBIS-II

mammograms to solely baseline CC film mammograms (as the average of the RCC and LCC views per episode, n=42 episodes) yields a median BA ~20 cm^2 smaller: 168 cm^2 (Q1-Q3: 127 cm^2 to 205 cm^2). Given that the majority of CMN IBIS-II participants were over age 60 years (e.g. they had more compressible breasts) in combination with the slightly higher BMI (~29), the larger BA observed compared to other reports is not implausible.

The median DA of ~33.5 cm^2 is also higher than earlier reports: 26.4cm^2 (Q1-Q3 0.3 cm^2 to 138.0 cm^2) [177], 20cm^2 (average, 0.8cm^2 SD) [479] and 40cm^2 (average, 26 cm^2 SD) [192], whilst median PD (~19%) is fairly typical for women aged ~60 [144, 189]. This implies that the calculated areas for BA, AA and DA are overestimated, compared to other studies. The calculated # of pixels/mm^2 is 400/mm^2 for scans using a 50μm pixel spacing (20 'pixels' per linear mm); this means each 'pixel' is about 0.0025 (1/400) mm^2 in size. However, the expert team which developed the Cumulus program report that one pixel was 0.0676mm^2 for scans made at "50μm spatial resolution" [137]; 0.0676mm^2 per 'pixel' is larger than 0.0025 mm^2/'pixel', so use of this latter conversion factor would have (further) increased estimated DA, BA and AA.

If area for the IBIS-II mammograms has been overestimated, the (external) generalisability of this project's results for 'absolute' measures (DA, BA, AA) will be affected, however results for PD (dimensionless) will be unaffected. Rates of change, such as those discussed in Chapters 7 and 8 for the absolute measures may also be affected, but less so.

Age is typically a strong confounder of density which has an inverse relationship with PD and DA [137]. A significant negative relationship was found between age at randomisation and digital mammograms for PD and DA during simple regression, however the relationship of age with PD and DA was non-significantly positive for film mammograms, Table 6-9. This latter unexpected relationship may have two possible explanations. Median age at randomisation was 61.8 years, hence over half of the women in this dataset were aged over 60 at randomisation.

The rate at which PD decreases per annum tends to slow post-menopause [144, 188, 189]; PD may even increase in later life [145, 471]. This likely attenuated the typical effects of age upon PD and DA for women in this dataset.

Additionally, the result for film mammograms was partly due to one participant whose PD was ~80% at baseline, very unlike the remainder of participants whose baseline PD was <60%. Removal of this participant from the simple regression model for film mammograms yielded a non-significant coefficient of –0.2% PD per one year increase in age at randomisation. Omission of all three women with high PD (≥50%)— two of whom had baseline film mammograms— strengthened the all mammogram model's inverse relationship with PD to –0.44, p=0.054, R^2 5%. The film mammogram coefficient similarly increased non-significantly in absolute value to –0.25. These three participants with high baseline PD also were influential in other bivariable relationships with the MD parameters, as described in subsequent paragraphs. The relationship between age at randomisation with PD and DA remained inverse for the multivariable regression models, Table 6-11, and was significant for the square root transformed models of PD (e.g. p=0.028 for the square root PD parsimonious model). The decreases in PD and DA with age for the CMN IBIS-II participants (about –0.5% to –1% , and –0.5cm^2 (50mm^2) to –1cm^2 (100mm^2) per year, respectively), are similar to others [188, 192, 252], given that the decline in PD and DA tends to slow for women aged 60+ years [144]. Age was non-significantly and positively associated with BA and AA during bivariable regression, Table 6-9, and positively associated with BA as well as positively and significantly with square root transformed BA during multivariable regression, Table 6-11.

Two other strong confounders of MD, BMI and weight [174], were consistently and significantly associated with PD, BA and AA, both for the all mammograms relationship [i.e. both Types of mammograms—film and digital] and within each Type (all expected), Table 6-9. The association was negative for PD because PD tends to decrease as weight increases, and positive for BA and AA because breast size tends to increase with weight and height [178]. The

relationship for AA from the CMN IBIS-II population was consistent with another report. Non-dense area (AA) increased by ~3cm^2 per kg (300mm^2/kg) [197], but this was much larger than the univariable association between non-dense area (AA) and weight in a slightly younger cohort (mean 50.5 years (6.2 SD)) of 804 high risk women, 0.08cm^2/kg (8mm^2/kg). Similarly, the large association between BA and AA and BMI (~900mm^2/(kg/m^2) = 9cm^2/(kg/m^2)), is also much higher than for the younger, high risk cohort (0.24cm^2/(kg/m^2)). DA did not show a significant relationship with BMI or weight (Table 6-9) which was also consistent with some [192] but not all [178, 197] of the MD literature, perhaps because the IBIS-II participants are a high risk population. The multivariable regression BMI relationships with PD and BA (Table 6-11) were consistent in sign and magnitude with their simple regression counterparts (Table 6-9).

As a single parameter (Table 6-9), weight explained about 30 to 60% of the total variability of BA and AA for film, digital and all mammograms (R^2 values ranged from 0.31 to 0.58). The explanatory relationship was stronger for digital mammograms than for film mammograms (50 to 60% for digital vs 30 to 40% for film). The percentage of variability in PD explained by weight was stronger for film mammograms (16 to 20%) however than for digital (7%); this implies the R^2 differences are not (solely) due to removal of dense area due to digital mammogram post processing. The percentage of variability for percent adipose (PA) explained by weight was also higher for film only (14%) vs digital only (7%) models (data not tabulated). R^2 for BMI also showed a similar trend (higher for digital only models of BA and AA, but lower for digital only compared to film only for PD and PA). The difference in film vs digital R^2 for BA and AA vs PD and PA might have resulted because BA and AA are measured in mm^2 and PD and PA are dimensionless parameters which are independent of the size of the breast. Removal of the participant with very high baseline PD ≥75% did not change R^2 of any MD and BMI or MD and weight model by more than ±1%, however omission of all three participants with high PD (≥50%) changed R^2 for the weight and BMI PD models slightly more (±3%). Participants with high PD do not always have high DA relative to other participants, and

participants with the lowest PD do not always have the highest weight or BMI. These complex relationships between PD, BA and AA with weight and BMI may have also influenced the differing relationships for film vs digital mammograms for weight and BMI. The opposing relationships between the film vs digital R^2 for PD vs BA and AA might be also be spurious due to the small number of participants utilised in this analysis (n=85).

Height was expected to be positively associated with PD and DA, but to have no relationship or possibly an inverse association with BA and AA [174, 192]. This is because density, but not necessarily breast size, tends to increase with height. None of the relationships tested (PD, DA, BA & AA, for all mammograms and by film or digital Type) were significantly associated with height, Table 6-9).

The inconsistent relationship between PD and height for film vs digital mammograms (Table 6-9) disappeared when weight was controlled for during regression (multivariable models, data not tabulated because BMI was selected for use in the multivariable models). Similar confounding during simple regression appeared to be present for the parameters age at menarche, age at first birth, and months of HRT use. The relationship of MD with height and most other confounders of MD is less strong than the relationship for MD with age, BMI and weight. It was not surprising to observe inconsistent and non-significant relationships between lesser confounders such as height and the MD parameters in this dataset, especially when the distribution of baseline PD and DA differed between film and digital mammograms.

Adjusted R^2 adjusts the original R^2 for the number of parameters in the model; this provides a more complete estimation of the amount of variance of the dependent variable (e.g. PD) explained by the independent variable/s. Although not tabulated, most of the adjusted R^2 for the regression models were negative (equivalent to 0), emphasising the lack of association between most covariates and the average MD parameters at baseline.

The MD literature suggests that many significant univariable relationships between PD and BC risk factors become attenuated and/or non-significant after adjustment for other factors in multivariable analyses. However, with the exception of the non-significant associations with HRT and OC use and ex-smoking status, the direction of the relationship between the PD and DA simple regression model coefficients (Table 6-9 and Table 6-10) and the PD and DA multivariable regression models (Table 6-11) were generally the same. The magnitude of the coefficients changed between the simple regression and multivariable regression models, however (as expected).

Later age at menarche is associated with higher PD and DA compared to earlier age at menarche [194] (although later age at menarche is associated with lower BC risk). The relationship between older age at menarche and higher PD may be attenuated for women over age 60 [202]. The all mammogram relationship between menarche and PD & DA was non-significantly positive but within-Type relationships again differed. Film showed a non-significant positive relationship, whilst the relationship for digital mammograms was non-significantly negative.

Both nulliparity and older age at first birth (AFB) are associated with higher PD and DA [149, 192, 203]; the association with between PD and DA with age at first birth appears to be modified (inversely) by the number of children [191], perhaps particularly so for premenopausal women [203]. The association between PD and AFB may be stronger in women over aged 60 [202]. For this project, AFB was significantly and positively associated with digital baseline PD during bivariable regression, but non-significantly and inversely associated with film mammograms, Table 6-9. Although AFB ≥ 30 was significantly associated with PD in simple regression models, after adjustment for age at randomisation, BMI and age at menopause, AFB ≥ 30 did not meet retention thresholds ($p \leq 0.1$) for retention.

Nulliparous women tend to have higher PD, DA and lower AA than parous women [177], including older first time mothers (e.g. age 26+ [191] and/or 30+ [192]). The non-parous

women with baseline mammograms (5 participants) in this sample of IBIS-II participants (85 women) had on average 5% non-significantly lower PD than their parous counterparts during simple regression (Table 6-9), and –3% non-significantly lower PD than their counterparts with age at first birth <30 years during multivariable regression (Table 6-11). However, PD was inversely associated with increasing parity in the much larger sample of high-risk participants in the IBIS-I trial (>100 non-parous women, >800 total participants) [192, 296], so the association of lower PD for non-parous women observed for this sample of high-risk IBIS-II participants is not likely due to their high-risk status. The average age at first birth for the 80 parous women with baseline mammograms was 22.6 (sd 4) years of age; this low average age at first birth is associated with lower PD in the general population [202], so the unexpected association of higher average PD for parous vs non parous participants is likely due to the small number of women in this (baseline) dataset. The relationship for PD and DA remained non-significant and inverse for nulliparous women relative to parous women, with or without the women with high PD in the regression relationships. Hence the three women with high PD did not substantially alter the observed relationship between PD and parous status, unlike the bivariable relationships observed for MD with age and height.

Later age at menopause is associated with both higher MD [177] and increased BC risk [5]. A positive relationship was seen for both PD and DA with menopausal age, for all mammograms and within each Type. The relationship was stronger for women with known age at menopause (natural or via ovariectomy) than for all women (imputed + known age at menopause), and both stronger and significant for digital mammograms vs film mammograms.

Past HRT use does not have a well-defined relationship with MD. PD may be higher for premenopausal former HRT users, but not former post-menopausal users compared to never users [203], but was associated with lower PD, DA and higher BA in the high risk IBIS-I population [192]. Current HRT is positively associated with PD in most populations, e.g. [177], especially for post-menopausal women [202, 203]. Combination HRT with estrogen and

progestin, in particular, is associated with both higher PD and BC risk [246, 482]. Past use of HRT may be associated with higher PD and DA compared to never use [253]. Ever use of HRT was not significantly associated with higher or lower PD or DA in this dataset, and the proportion of variability explained for any MD parameter was very low (R^2 ~0.001, which equates to about 0.1% of the variability explained).

Past oral contraceptive use is not a well-defined MD confounder, however OC use and duration may be associated with differences in MD [253]. Past OC use had a non-significant and positive association, as well as a very low explanatory relationship with all MD parameters (R^2 of 0.2% or less). OC duration tended to have a positive non-significant association with the MD parameters, but similarly had very low R^2 values for all regression relationships except for digital mammograms.

Smoking status has an inverse dose-response association with PD [483]; however, both current and past smoking have a dose-response association with increased BC risk, which is modified (positively) by alcohol intake [484, 485]. PD tends to be lowest in current smokers, but also lower in former smokers compared to never smokers [192]; this trend might only occur for premenopausal, but not post-menopausal, women [203]. In this dataset, current smokers tended to have both non-significantly higher PD and DA in the simple regression models (unexpected) and non-significantly higher BA and AA than never smokers, Table 6-9. Past smokers however, had non-significantly lower PD and DA compared to never smokers (expected) and non-significantly higher BA and AA, Table 6-9. The bivariable relationships between smoking and the MD parameters were very weak (R^2 ~1%).

The multivariable, positive relationship between current smoking status and PD, DA and BA was unexpected. Although the p-values for these relationships were not significant ($p<0.05$), the p-values met the predefined criterion of $p\leq0.1$ for retention in parsimonious models of untransformed DA and both untransformed and transformed BA (e.g. DA p-value=0.075).

Given that the p-value for current smoking status in the multivariable square root transformed DA model was >0.1, the slightly stronger relationship between current smoking status and untransformed DA is likely due to distortion of the linear relationship caused by the right-skew of the untransformed DA distribution. The unexpected higher PD, DA and BA associated with current smoking status could be due to the small sample size, or (for unknown reasons) due to the high risk status of the women in the sample. Given that none of the MD and current smoking relationships met the specified $p<0.05$ criteria for significance, these possible relationships are likely due to chance alone.

Smoking status is not a widely utilised confounder of breast density, probably because the effects of smoking on MD seem to be primarily mediated through BMI (i.e. via differences in weight) and are also possibly modified by age. It is possible that tobacco utilisation may have other effects upon MD as for the confounding which likely exists between smoking and BC risk e.g. smoking appears to suppresses estrogen production, which then reduces overall BC risk (and probably MD as well) in smokers despite the BC risk increase caused by carcinogens in the tobacco smoke. However the effects of smoking on MD are likely too minimal to impact MD differentially in this small dataset— large (hundreds) to very large (thousands) numbers of people are needed to ascertain the relationship between a risk factor and an outcome if the overall effect is small. The number of current smokers with baseline mammograms (9 of 85 women) and within the total sample of CMN IBIS-II participants (11 of 120 total women) is small, within this small sample of high-risk women. The relationship between smoking status and changes in longitudinal MD— if examined during modelling of longitudinal MD (Aim 4 and Aim 5)— will be need to be interpreted with caution due the small sample size.

Family history of BC appears to be associated with higher PD, as suggested by data from a large (>140,000 woman) general screening population sample [202]; a stronger family history (i.e. more relatives with BC) was also associated with higher PD in another large general screening population [233]. This relationship, however, is not consistent in all studies. The relationship

between PD and BC risk appears to be modified by familial risk status in some populations (i.e. Asian and Caucasian), as shown in an analysis of approximately 4,000 women from four case-control studies [237]; although mean PD did not differ between those with and without a (first degree) family history of BC in that study, a 10% increase in PD was associated with higher risk of BC in women with a family history of BC (32% increase in risk) compared to women without a family history of BC (13% increase in risk of BC per 10% increase in PD). Additionally, higher PD does not appear to be associated with BRCA status, but (as for the general population) within the population of women with BRCA mutations higher PD is associated with higher BC risk [244]. The inverse relationship between higher predicted Tyrer-Cuzick BC risk and lower PD in the high risk IBIS-I population [296] is intriguing.

In this project (the CMN MD and AI study), number of relatives and weighted number of relatives (1 for 1st degree and 0.5 for second degree) was used as a surrogate for BC risk. All CMN IBIS-II participants had at least one first or second degree relative affected by BC and/or ovarian cancer. An increase in the number of relatives with BC and/or ovarian cancer was associated with non-significantly higher PD and DA for the CMN IBIS-II participants during simple regression analyses, whether modelled as a continuous number representing the total number of relatives, or total number weighted by first or second degree status (reflecting the (average) genetic similarity between the participant and her first vs second degree relatives), Table 6-9.

As noted in Table 6-10 for the re-categorised parameters of AFB, OC and HRT, non-linear trends predominated within the quintile (dose-response) divisions of OC and HRT use. This implied a linear relationship did not exist between duration of OC or HRT and the baseline MD parameters for this dataset. No significant associations were noted for the dichotomous divisions of OC or HRT use. Because the relationship was expected (consistent with the literature) and it had higher R^2 and adjusted R^2 values, the AFB parameter with the division at ≥30 years was selected for use in the multivariable baseline regression model.

The relationship between BA and BC risk factors is not well described in the MD literature. The simple regression coefficients for menarche, HRT use, ex-smoking and non-parous status (Table 6-9) did differ in sign compared to the multivariable regression models of BA (Table 6-11). However, the p-values for the BA regression coefficients which differed in sign between the simple and multivariable models were not significant, thus it is not surprising that some of the simple regression coefficients subsequently differed in sign in the multivariable regression model.

Few unexpected relationships were noted amongst the MD parameters (PD, DA, BA & AA) and MD covariates via graphical (descriptive) analyses and quantitative collinearity checks. Most of the unexpected relationships during regression analyses became attenuated or were reversed with exclusion of the one participant with very high PD (≥75%) and/or exclusion of all three participants with high PD (≥50%). The three participants with high PD often appeared to be outliers in graphical analyses and residual checking because they were few in number. Although high MD (≥50%) for post-menopausal women is less common in the general post-menopausal population compared to lower MD <50% (e.g. Figure 2-9 [137]— where most women aged 50+ with PD≥50% also appear to be outliers), the high MD for measured for these women is real. Hence the data for these women was retained. Due to the relatively small sample size in this pilot study (the CMN MD and AI study), women with PD ≥ 50% are few in this sample of IBIS-II participants. A larger sample size would likely improve stability of the coefficients estimated in the regression models (i.e. reduce bias as well as improve efficiency (smaller SE)) and provide better estimates of the relationships between the covariates and MD for the (high risk) IBIS-II population.

For both the simple (Table 6-9) and multivariable regression models (Table 6-11), digital mammograms were associated with both lower PD and DA than film mammograms. In particular, KE52 digital mmgs had significantly lower (square root transformed) PD and (untransformed and square root transformed) DA than film mammograms. BA was also higher

for KE52 mammograms compared to film, as was square root transformed AA (significantly so in the simple regression model, Table 6-9). Because PD and DA differed significantly by mammogram Type and Version at baseline, they are also likely to differ significantly longitudinally. Therefore these parameters needed to be accounted for during mixed modelling of longitudinal MD (Aims 4 and 5).

Compared to log transformation, square root transformation of PD yielded the most evenly distributed residuals, Figure 6-4. The residual diagnostic plots had less heteroskedasticity, though this effect appeared to still be present on a number of the simple regression diagnostic plots. The square root transformation makes the regression results more difficult to translate back into 'real' units (e.g. %, mm^2) however. In contrast, regression results from log transformation are interpreted as a % change in the original units (if the coefficients are not too large), which is straightforward to perform. Mixed modelling of both untransformed and square root transformed PD, DA and BA was undertaken in Aim 4 (next chapter) to compare the model outcomes for both the square root transformed and untransformed MD parameters. Square root transformation of BD measures is common transformation applied during BD analysis, e.g. [486-488].

One of the limitations of this study is the population of IBIS-II participants is likely not representative of the general female population because these women have at least a 1.5-fold risk of BC compared to their same-age counterparts from the general population. Most of the relationships between the BC risk factors and MD described in Chapter 1 and in this chapter are for women from general screening populations. Hence unexpected relationships such as the inverse relationship noted between women at high familial risk and PD in the IBIS-I population [296] may also be present for IBIS-II participants. Therefore the relationships modelled in Aim 3 (this chapter) may only apply to the population of (high risk) IBIS-II participants, but possibly may be generalisable to other populations of high risk women. Median PD for the CMN participants (19.3%) was comparable to median PD (18.7%, Q1 to Q3: 8 to 30%) for the subset

of post-menopausal women in the Minnesota Breast Cancer Family Study (n=1,284; median age 62 years (Q1 to Q3: 55 to 58y)) [144]; most women in the Minnesota Breast Cancer Family Study were at medium to high-risk of BC which is comparable to the IBIS-II participants. The median ages of these groups of women were also similar (~age 62). The similarities in PD (for women of similar age) between the high risk women in the CMN IBIS-II group and the Minnesota BC Family Study lends support to extrapolation of the results from this CMN MD and AI substudy to other similar (e.g. high-risk) populations.

Typically only a single View from a mammography episode is assessed for density, although pairs of Views such as both CC mammograms are sometimes utilised. The simple regression results (collated in Simple regression of baseline PD with different combinations of mammographic Views) suggest models utilising any single View or combination of Views are roughly equivalent. The regression coefficients were generally consistent in magnitude, sign and significance for single Views, averages of pairs of Views, and the average for all four Views within each transformation of PD (identity (%), square root, natural logarithm).

6.6 Conclusion

The results of this analysis of baseline covariates and baseline MD were generally consistent with the reports in the MD literature, although many of the relationships noted during simple regression modelling were affected by a few participants with high (≥50%) PD. PD and DA tended to show an inverse relationship with age at randomisation and BMI, whilst the relationship of PD and DA with age at menarche, age at menopause and age at first birth (≥30 years) tended to be positive. The relationship between BA and potential covariates is not well described in the MD literature; however the strongest relationship between BA was with BMI ($p<0.001$) which is expected.

A larger sample size of IBIS-II participants would help elucidate the discrepant relationship for PD and DA noted for non-parous women vs parous women in the dataset, as well as help clarify the influence of smoking status on MD (in high risk women). A larger sample size may also assist with the relationship between MD and BC risk status (e.g. family history of BC and/or ovarian cancer) in populations of (post-menopausal) high-risk women.

Square root transformed PD and DA provided regression models with the least heteroskedastic residuals. This implies that square root transformations of PD and DA should be utilised during MD longitudinal modelling. However, care will need to be taken during back transformation of the regression coefficients, which is needed to provide useful quantification of the longitudinal change in MD over time for this sample of IBIS-II participants.

The results of the technical analyses as well as the baseline covariate analyses in this chapter imply that mammographic Version is an important factor which should be accounted for during modelling of the CMN IBIS-II MD data. The KE52 mammographic Version in particular, differed significantly from the film Version for all models of PD and DA, e.g. Table 6-11.

The simple baseline regression models utilising one, or the average of two or four (standard) mammographic Views were similar (Table 10-8), which implies that modelling with one View or the average of four Views may yield similar results during longitudinal MD modelling. A small but significant difference exists between PD measured on CC vs MLO mammograms; models utilising different Views should take this relationship into account. Models utilising BA and AA with different mammographic Views should account for both the CC vs MLO and right vs left differences in mammograms (Table 10-5).

7. AIM 4: Aggregate CMN IBIS-II MD longitudinal change

This chapter contains the methods, results and discussion for Aim 4 of the thesis— the model for MD change over time for CMN IBIS-II participants with the treated and control groups aggregated. The first part of this chapter describes the Aims of the chapter and the methods used in this analysis. Descriptive analyses of longitudinal PD follow. Next, the results of the analysis are presented, followed by a discussion of the results.

7.1 Aim

Aim 4 of this thesis was to develop an adequate model for mean change in MD over time for the aggregate (treated + control) groups— a "blinded" longitudinal analysis. Treatment group allocation for the IBIS-II participants will remain blinded until at least January 2022, when the last participant completes ten years of follow up. Currently, only the IBIS-II statistician has access to the unblinded IBIS-II trial data.

In order to complete the Primary Aim of this thesis (Aim 5) — comparison of longitudinal change in MD for treated vs control IBIS-II participants— a statistical model to analyse the longitudinal IBIS-II MD data needed to be developed. One longitudinal modelling strategy is to develop a good model for mean change over time, before adding in important covariates and the primary parameter/s of interest such as treatment group [36]. The Aim of this chapter was to define a good model for mean MD change over time for the IBIS-II treated and control groups, which includes all important covariates except treatment group.

7.2 Introduction

The effect of anastrozole treatment for participants in the IBIS-II trial was expected to exert its influence on MD by reducing PD as well as DA relative to control participants. Most of the

change in MD due to anastrozole treatment was expected to occur within the first 12 months of follow up (Figure 7-1, below), as occurred for tamoxifen in the IBIS-I trial [296].

Whilst PD continued to decline (more slowly) between years 1 to 5 for the IBIS-I tamoxifen treated participants relative to the control treated participants, this may not be true for anastrozole treated participants in the IBIS-II trial. This is because the IBIS-I group of high risk women was a mix of premenopausal participants (higher PD) as well as post-menopausal women (lower PD after the menopausal transition). All of the women in the IBIS-II trial are

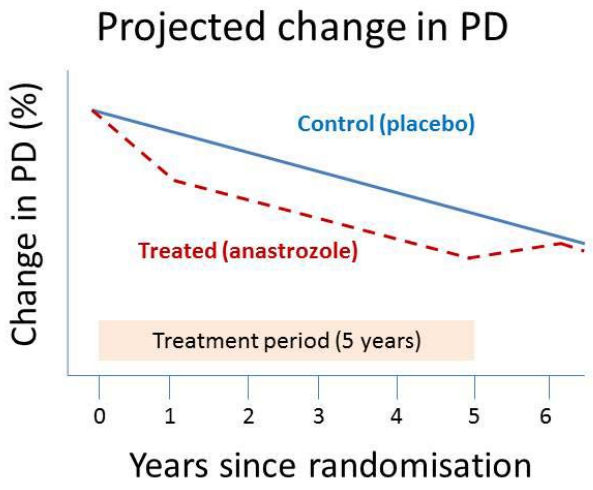


Figure 7-1 Theoretical change in PD due to trial treatment, IBIS-II participants

postmenopausal, a requirement for treatment with aromatase inhibitors such as anastrozole. Post-menopausal women on average are older than premenopausal women; the rate of annual decline in PD tends to diminish with increasing age [144, 189]. Given the subtle effect expected for anastrozole in post-menopausal women, a negligible decline was expected between years 1 to 5 for anastrozole treated participants relative to IBIS-II control participants. The decline in PD for IBIS-II treated participants was expected to rebound to control levels after year 5, as seen in women treated with other hormonal medications including AI [29, 261, 290].

7.3 Methods

7.3.1 Participants and mammograms

One hundred and twenty CMN IBIS-II participants contributed mammograms for Aim 4 (this chapter) and Aim 5 (Primary Aim, unblinded treated vs control analysis, Chapter 8). Only mammograms from participants at the CMN IBIS-II centre were utilised for Aim 4 (this chapter) and the Primary Aim (Chapter 8). Although fifty-one digital mammograms comprising 14 episodes from six IBIS-II participants were collected from the Southern Highlands Cancer Centre in New South Wales, (as a result of the de-identification process) mammographic Version could not be determined for these mammograms. Two set of baseline mammograms and two sets of follow up episodes for three participants were also unsuitable for measurement in Cumulus due to issues such as poor image quality and missing images. This reduced the total number of participants from the Southern Highlands Cancer Centre to four. Because mammographic Version for each episode could not be determined, and the number of participants was small compared to the potential issues which might ensue (e.g. increase in variability because exact mammogram Version could not be controlled for), data from the Southern Highlands IBIS-II centre was not included in analyses for this thesis.

CMN IBIS-II participant baseline mammograms were taken at a variety of institutions, including the Calvary Mater Newcastle. All CMN trial follow up mammograms were taken on a single mammography machine in the Radiology Department at the Calvary Mater Newcastle hospital.

As outlined in Chapter 3, mammograms for the IBIS-II Study were obtained at baseline and then annually thereafter. Mammograms were considered baseline episodes if they occurred up to 1 year prior to randomisation. While follow-up visits, including mammography, for IBIS-II were on a specified schedule, there was some variation around the timing of the visits and examinations including mammography. Mammographic episodes occurring more closely to the

6 month follow up than the 12 month follow up were associated with the 6 month follow up data (6 months ± 3months). Episodes taken between 9 and 18 months were associated with the 12 month follow up data. Thereafter episodes occurring within six months of an annual visit (i.e. ± 6 months) were associated with the follow up data for that annual visit.

Mammograms were acquired in two rounds of collection. The first (Collection 1) was in 2012 after ethics approval was granted in December 2011. Collection 2 was undertaken in 2014; mammograms in the second Collection included mammograms taken in 2013 and 2014, as well as mammograms taken at earlier dates. Mammograms from the two collection rounds were assessed for density separately: in 2012 for Collection 1, and 2016 for Collection 2. As described in previous chapters, participant mammograms were measured in Cumulus in random order, grouped by participant. Hence all mammograms for a particular participant acquired during each Collection were read in random order, but sequentially prior to reading mammograms from the next participant.

Standard mammographic Views of the breasts, namely the right cranio-caudal (RCC), left CC (LCC), right medio-lateral oblique (RMLO), and left MLO (LMLO) views were collected for each participant for all available trial mammography episodes, as described in the Methods chapter (Chapter 3), and Chapters 5 and 6. Multiple Views were collected per episode to help reduce PD measurement error, which on average was ±8% PD for film-screen mammograms and ±10% PD for digital mammograms using Cumulus (Aim 2, Chapter 5).

Compliance with trial treatment (anastrozole or placebo) was also recorded for each episode, as follows: full compliance; a small deviation from the protocol (e.g. <2 weeks without treatment); a treatment holiday; or if the participant stopped completely. If available, the reasons for the deviation, treatment holiday or permanent cease was also recorded.

7.3.2 Measures

The mammographic density attributes percent density (PD), dense area (DA) and breast area (BA) were selected for use in the longitudinal models in Aim 4 (this chapter). As described in previously, BA and DA are directly measured using the Cumulus program. PD is a derived parameter calculated as follows: PD = DA/BA. Whilst other MD attributes such as adipose area (AA = BA–DA) and percent adipose (PA = AA/BA) are sometimes reported in the MD literature, these attributes were not utilised for longitudinal analyses in this thesis. The change in PA was likely to be very small due to treatment with anastrozole. The results from Aim 3 of this thesis (Chapter 6) showed that AA largely had the same associations as BA; as for PA, AA would likely undergo only small changes due to anastrozole treatment— changes which are better characterised using PD and DA. A graphical distribution for AA, and median for AA and PA only are reported in this chapter for consistency with other reports in the MD literature.

The effect of anastrozole on longitudinal PD and DA is the focus of Aim 4 (this chapter) and the thesis Primary Aim (Aim 5), hence most of the analyses emphasise these two parameters. BA is also reported as part of this Aim, because BA has a direct influence upon the interpretation of change in DA. Whilst changes in BA are less likely to affect assessment of PD because PD will change proportionally with BA (as described in Aim 2), a reduction in DA but not BA implies that only DA has changed. As described in Aim 4 (previous chapter, section 6.4.1), adjustments were made to DA and BA as output by Cumulus (raster area) in order to account for differences in pixel spacing between mammogram Versions by converting raster size to area in mm^2; hence all MD attributes which have a dimension (area), namely DA, BA and AA, are reported in mm^2.

The covariates[3] [489] utilised in this analysis are the same as those described in Aim 3 (Chapter 6) and utilised in the full multivariable models in Aim 3— age at randomisation (years), BMI

[3] The regression covariates are often referred to as confounders (of mammographic density) in this thesis. However in the context of randomised controlled trials, the predictors in linear regression are used to improve precision. They are not related to treatment assignment due to the randomisation

(kg/m^2), age at menarche (years), age at first birth (3 categories: <30 years, ≥30 years, non parous), smoking status (never, current, former), menopause age (years), HRT use (never/ever), oral contraceptive use (never/ever), and previous IBIS-I participation (yes/no). As described in Chapter 6, the menopause age parameter utilised an imputed age at natural menopause for 27 of the 120 participants for whom age at hysterectomy only was known; age at natural menopause was imputed for the 27 participants based on the age at natural menopause experienced by their 57 IBIS-II peers. All other covariate values were taken directly from the IBIS-II data supplied by the ANZ BCTG. Besides the age at hysterectomy-only reported for 27 participants, due to the high quality of the IBIS-II trial data, full information was available for all remaining covariates for the 120 CMN participants (no data was missing or unknown).

7.4 Statistical Methods

All analyses for Aim 4 (this chapter) were undertaken using Stata v12.1. P-values < 0.05 were considered significant. Coefficients from regression models utilising untransformed MD as the outcome parameter (dependent variable) are presented in % (PD) or mm^2 (DA, BA and AA). Coefficients from models utilising square root MD as the outcome parameter are presented in square root or log transformed % and mm^2. Back transformation of coefficients was performed only for the parsimonious models of PD, DA and BA, to minimise the potential for reporting error. Back transformed values for the coefficients were calculated by squaring the sum of the coefficient and the (main effects) intercept and subtracting the square of the intercept, Equation 7-1 :

Equation 7-1

$$Back\ transformed\ coefficient = (\text{coefficient} + \text{intercept})^2 - (\text{intercept})^2$$

Use of the square of the intercept was derived from Singer and Willet [490]— in chapter 4 of that book, the square of a (main effects) intercept from a model with a square root transformed

process, although they are associated with the outcome. Covariates are therefore not considered to be confounders in the context of randomised trials. Gelman, 2006 [472].

outcome parameter was utilised to obtain the value of the intercept on the outcome parameter's original scale. Because all coefficients in a regression model provide estimates in relation to the intercept of the model, subtracting the square of the intercept from the square of the sum of the intercept and the coefficient provides an approximation of the back transformed coefficient value. This approach helps to overcome difficulties with the squaring of coefficients with values less than 1, which become smaller when squared. However, the values of the back transformed coefficients are highly dependent upon the value of the intercept used during back transformation. The value of the intercept is the estimated value for the outcome parameter when all coefficients in the model are equal to 0; hence the intercept will vary if centreing of the predictor (independent) variables is used, and it will also vary if the values used to centre the parameters change. Because the intercept estimates for PD and DA for models utilising film mammograms (e.g. ~30% PD) was higher than the observed median baseline PD and DA (PD 18%, DA 2800mm^2, Table 6-6, Chapter 6), values closer to the baseline values for the CMN participants were also substituted for the intercept in Equation 7-1. A range of back transformed coefficient values are tabulated for the MD change over time coefficients using different 'intercept' reference values for Equation 7-1. Two of the selected 'intercept' values, 20% PD and 27cm^2 (2700mm^2) for DA, are close to the median baseline values for the CMN participants (Table 6-6), and are identical to the reference values used in another MD longitudinal study, which used a variant of Equation 7-1 to calculate back transformed coefficient values from models with square root transformed MD as the outcome parameter [188].

This chapter refers to two related groups of longitudinal models for IBIS-II MD. MD was initially modelled using the four standard mammographic Views individually as the lowest level units of measurement in the longitudinal model—i.e. all four Views grouped within an episode were incorporated into the statistical model. These initial results are annotated in Three-Level Unconditional Means and Unconditional Growth Models and Three-Level Full and Parsimonious Models of PD, DA and BA. To simplify the statistical model, the density measurements from all four Views were averaged together to create a composite measurement

of PD, DA and BA for each mammographic episode; this approach was also utilised in Aim 3 (baseline characteristics analysis, Chapter 6) which utilised average MD for the baseline episodes to meet the independence assumptions of simple and multivariable linear regression. Use of average MD per episode also helps to stabilise the longitudinal MD measurements, which are inherently variable due to the subjective nature of the semi-automated method utilised to measure MD (re: Aim 2, reliability analysis, Chapter 5). This simplified approach also permitted modelling of a covariance structure for the data, which was not possible to perform in the statistical program with all four mammographic Views as the primary unit of analysis. These simplified, average PD, DA and BA models were utilised as the final models for this Aim, as well as for the Primary Aim (next chapter).

Due to the strong right skew of PD and DA and the right skew of BA, square root transformations of these MD parameters were assessed during statistical modelling for comparison with models of untransformed MD measurements. This was done to ascertain if square root transformation would improve normality of the residual distribution. Preliminary modelling using Collection 1 mammograms showed natural log transformed PD and DA produced residuals with a left skew. This also occurred during simple regression with natural log transformed baseline MD in Aim 3 (Chapter 6, re: Figure 6-5). Natural log transformation of PD was modelled for comparison with untransformed and square root transformed PD during unconditional means and unconditional growth mixed modelling, however, due to non-normality of the residuals, further modelling using natural log transformations of PD and DA was not undertaken for Aim 4 (this chapter) or the Primary Aim (next chapter).

The majority of change due to active trial treatment with anastrozole was expected to occur between randomisation and the first annual (12 month) follow up [296]. MD change over time was analysed both categorically using follow up number (in months), and as a continuous parameter (years since randomisation, converted from days). Due to the differences in MD measured on film vs digital mammograms, analyses were also undertaken separately for film

and digital mammograms, with further adjustment by hardware and/or software Version of digital mammograms. The mammograms were then combined in additional analyses which included an indicator for different Versions of mammograms. Four digital mammographic episodes of unknown Version ('other') were not included in the indicator for different mammogram Versions, due to the small number (4) of these episodes.

To account for correlation amongst the repeated mammographic density measurements in subjects over time, a linear mixed model was utilised to model the longitudinal changes in mammographic density in Calvary Mater Newcastle IBIS-II participants [490]. The steps undertaken during fitting of the longitudinal mixed model are described in more detail in sections 7.4.3 to 7.4.7.

7.4.1 Covariate checking and centreing

The majority of pre-analysis covariate checking was undertaken prior to fitting the baseline simple and multivariable regression analyses in Aim 3 (Chapter 6), where collinearity between the response and predictor variables, as well as the predictors was undertaken and the distributions checked for outliers. Most of the independent predictor variables were found to be interrelated, although the interrelations did not exceed acceptability limits (e.g. the variance inflation factor (VIF) for the interrelationships were below 3).

Although most of the continuous and discrete covariates had skewed distributions (BMI, randomisation age, age at menopause, and age at first birth), they were utilised as per Aim 3 as continuous covariates. This was done to further avoid portioning the data into small categories which represented only a few or no participants during multivariable modelling, and to avoid the loss of information which results when a continuous covariate is transformed into a categorical covariate. Use of original units for all continuous parameters aided with ease of interpretation of model coefficients since back-transformation is not required. Additionally, it is

not a requirement that the dependent and independent variables in regression models have normal distributions; only the residuals are required to be normally distributed to meet regression assumptions. However, as noted in Aim 3 (last chapter) transformation of the dependent (outcome) variable, PD, using a square root transformation provided an acceptable normal residual distribution whilst the residuals from models of untransformed PD were skewed to the right. This is because the residual distribution generally reflects the distribution of the outcome variable.

To see if the residual plots could be further improved (i.e. in terms of normality of the distribution), BMI was selected for transformation; this was because, of the two strongest confounders of MD, BMI was more strongly skewed than age. Use of square root and natural logarithm transformed BMI improved the distribution of BMI, producing a distribution which more closely approximated the normal distribution. Natural logarithm and square root transformed BMI was substituted for untransformed BMI in the mixed linear models (sensitivity testing). The resulting residuals were graphed using histograms, and compared to residual histograms from models with untransformed BMI.

To provide a meaningful value for the intercept, as well as assist with back transformation of coefficients from square root transformed PD and DA models, continuous covariates utilised in final models were centred at a noteworthy value. Parameter centreing only affects the intercept value of a regression model, not coefficient values (β). Intercept values of regression models with centred covariates represent the mean response for an average participant with characteristics matching the centred values of the continuous covariates and the reference values (ref.) of each categorical covariate. Without centreing, the model intercepts represent mean MD for participants with values of 0 for continuous parameters. For example, a BMI of 0 is outside the range of BMI for the women sampled in this project; this is problematic for model interpretation. Furthermore, 0 BMI is not a possible value for a real person, further reducing the appeal of uncentred BMI for this analysis.

Randomisation age and menopause age were centred at age 50. Age 50 was chosen because menopause typically occurs at around age 50, and thus 50 is the likely age for high-risk women to consider use of an AI for BC prevention. Menarche was centred at age 13 (median for the sample population and historically). BMI was centred at 25 kg/m^2 (at the borderline between healthy and overweight, and a more appealing number than 24 or 24.9). Hence for a model with (centred) age at randomisation, BMI, menopause age, and AFB (<30y reference category) as the independent variables, the intercept represents the mean MD response for participants who were age 50 at randomisation, have a BMI of 25, experienced menopause at age 50, and gave birth to their first child at age 29 or under. Mammogram Version was also modelled as a categorical parameter; the model intercept is also representative of participants with film mammograms in all mammograms models (film + digital) and film only models. For digital mammogram only models, the intercept is representative of participants with KE52 (Kodak Elite, software Version 5.2) mammograms at baseline.

The back transformed coefficient values tabulated for the parsimonious models of PD, DA and BA therefore are also representative of participants whose characteristics match the values used to centre the continuous parameters, and whose characteristics also match the reference categories of the categorical parameters.

7.4.2 Descriptive Analyses

Descriptive analyses were undertaken initially to examine the 'raw' (as measured) distributions of the data for individual mammograms (not average MD per episode) and the timing and distribution of individual mammograms with the (ideal) annual trial follow up dates. Measures of central tendency were calculated for the episode averages of PD, DA, BA, adipose area (AA) and percent adipose (PA); median, Q1 (25^{th} percentile), Q3 (75^{th} percentile) were tabulated for PD and DA, and median tabulated for BA, AA and PA. These measures were calculated for all mammograms, film-only, digital-only, and for each digital mammogram Version. The

longitudinal average PD per episode response of participants was examined using Stata's "panel-data" line plot facility to search for participants whose longitudinal response appeared unusual, particularly in regard to additional variability in the measurements which may have been introduced by measuring mammograms for the two Collections separately in 2012 and 2016. Participants who appeared to have more than 5 to 10% changes in PD over time not attributed to other causes (e.g. the film to digital transition, the transition between different digital mammogram Versions) had their entire set of mammograms remeasured in Cumulus, using the per-participant random order strategy utilised during earlier MD assessment.

7.4.3 Longitudinal modelling of MD

Linear mixed regression (regression which utilises fixed and random effects) was undertaken to model average longitudinal change in MD for CMN IBIS-II participants (both treated and control). A mixed linear model was chosen for this dataset because it accounts for correlations amongst observations which result from repeated measurements, e.g. sequential measurements of PD on IBIS-II participant trial mammograms. Mixed models can also accommodate data which is irregularly spaced (time-unstructured) and unbalanced (varying numbers of follow ups per person). The model has both fixed and random effects, and hence is considered to contain 'mixed' effects. Interpretation of the coefficients for the fixed effects component of the model is identical to that for standard regression. The fixed effects are the average response of the data; they represent the mean, structural portion of a mixed model. Inclusion of random effect components tends to reduce the variability otherwise attributed to the fixed part of the model, for example by specifically modelling heterogeneity in PD among different individuals who have multiple measures of time within the data. This increases the precision of the estimate of the fixed components by reducing the standard error (SE) associated with the fixed effects.

The random effects (RE) component of the mixed model utilises groups to create a hierarchy (levels) within the data; these groups provide a structure to account for similarities amongst members of groups within the data. For instance, data from a national (cross sectional) survey

might be grouped by city, then region, then state. A study observing secondary students might group the students by classrooms, then schools, then school districts within a state, and then finally by state. Each group constitutes a 'level' within the model. Random effects can be modelled for each level of the model, and estimate the variability in the outcome associated with that level (group) of the data.

Within a multilevel model there is generally at least one random effect associated with each level / hierarchy of the data. There may be multiple random effects at each level, depending on the hypothesised nature of the relationship among outcome and explanatory variables. Typically, a random intercept for each RE level within the data is always modelled. If enough variability in change over time exists amongst units within a level, a RE for rate of change is usually modelled at that level.

The random effects components of the model can provide insight into relationships within and between the different levels or grouping within the data. A relationship sometimes exists between the initial status of units (represented by the RE intercept) and RE rate of change for these units; these RE thus co-vary with each other (for example, units / groups / clusters with high values of the outcome initially may have a steeper decline over time than units with lower initial values). This co-varying relationship is modelled as a RE covariance component, and it provides quantification of the relationship (covariance) between initial status and the rate of change between units. The RE covariance is typically converted to a correlation coefficient, which standardises the relationship between the RE components. For example, the correlation could quantify the relationship between initial status of participants in the IBIS-II trial (e.g. PD ≥25% vs PD <25% at baseline) and differences in change over time (e.g. the rate of change in PD during the first year of anastrozole treatment). Higher baseline PD might be associated with a greater annual rate of decline in PD during the first year of treatment, whilst lower PD might be associated with slower declines in PD; hence the correlation would be positive. The reverse could also be true (higher PD could be associated with a lower rate of change), and thus the

correlation (and covariance) between initial status and the rate of change for the IBIS-II participants would be negative.

The analysis for Aim 4 was performed following the guidelines of 'Applied Longitudinal Data Analysis' by Singer and Willett [490] and other texts [36, 491, 492]. The steps undertaken to model the data are described in more detail in sections 7.4.4 to 7.4.7 below.

As mentioned previously, a three level model was originally fitted for the data. However, Stata does not permit modelling of covariance structures other than the default 'independent' covariance structure for models with repeated times at the base (lowest) level of the model. Due to this issue with covariance structure modelling for the three level model, a two level model with the average of the four mammographic Views as the lowest level was selected. The upper level of the final hierarchical data structure was participant; this allowed the model to account for similarities of the longitudinal MD measurements for each participant. Ordering of episodes (repeated measurements of average PD, DA and BA) for each participant was provided by follow up number and mammographic date.

Fixed effects (e.g. from covariates) were retained in the model if they yielded a p-value of 0.10 or less. This more conservative threshold value of 0.1 was utilised due to the small sample of women in the dataset. Thresholds of 0.10, 0.15 or 0.20 are often used during stepwise reduction in multivariable regression to help ensure important covariates are not eliminated. Covariates with p-values >0.1 but <0.2 were closely examined for strong changes in BIC and AIC upon removal.

Forward selection and backwards elimination for models with 'all mammograms' (both film and digital) was undertaken to create parsimonious models. Backwards elimination only was primarily utilised for film only and digital only models; this was because forward and backwards selection always resulted in identical parsimonious models for the all mammogram

models, and also resulted in identical models for every film only and digital only model on which both approaches were utilised in preliminary analyses. The Bayesian Information Criterion (BIC) and Akaike information criterion (AIC) as output by Stata were reviewed during stepwise regression to compare models. As described in Aim 3 (last chapter), changes in BIC ≥ 3 were considered significant, whilst changes in AIC ≥ 2 were considered significant. The AIC takes into account the log likelihood of the regression model and the number of parameters (complexity) in the model, whereas the BIC considers both the sample size (number of observations) and number of parameters to estimate complexity whilst accounting for the log likelihood of the model. To create more parsimonious models, selection of covariates based on the BIC outcome instead of AIC outcome was undertaken; covariates which increased BIC ≥ 3 were removed from the model.

Each parameter was also modelled in a bivariable relationship with MD ('simple' mixed regression), including the alternative HRT, OC and age at first birth categories tested in Chapter 6; the simple mixed regression coefficients were compared to their parsimonious and full mixed model counterparts for any unusual differences (e.g. coefficient values which differed by more than 1 or 2 SE).

P-values for mixed model RE estimates are not generated during model estimation by Stata. Random effects were considered to be significant if the SE for that component was less than or approximately equal to half to the estimate for that random component. Only significant RE were retained in models.

Time was modelled both as a categorical parameter (follow ups at 0, 6, 12, 24, 36, 48, 60, 72, and 84 months— to match the IBIS-II trial follow up schedule) and continuously as days since randomisation converted to years. The continuous time parameter was centred at the date of randomisation for each participant.

As described in section 7.2 Introduction, the MD change over time due to trial treatment was not expected to be linear over the five year treatment period and after treatment cease. However, the initial change of anastrozole treatment on MD could have been shorter in duration than predicted (e.g. over the first 3 to 6 instead of 12 months), or longer (over the first 24 to 36 months of treatment) or non-linear over the treatment period (e.g. decreasing in magnitude over time). Hence the shape of the average trajectory of the participants was closely examined by reviewing the coefficients for categorical time vs those for continuous time. The best linear fit over time for the change in PD and DA was initially determined by systematic examination of the categorical time coefficients vs linear time coefficients from baseline through year 7. Linear splines of continuous time were created by selecting cut points which suited the mean growth curve of the IBIS-II participants [36]. Cut points allow the linear spline—a piecewise linear model— to change slope at the cut point (also known as a knot [491]).

Models of both untransformed and square root transformed PD and DA were checked for the cut points which best suited the data. The first cut point for continuous time was selected which maximised the slope of change over time from baseline to 0.5 through 2.5 years post-randomisation in 0.25 year (3 month) increments. Additional cut points starting from year 1 through 5.5 years post-randomisation in 0.5 year increments were also compared for differences in slope (rate of change over time).

The full model included both mammogram Types (film, digital); the continuous covariates randomisation age (years), BMI (kg/m^2), age at menarche (years), menopause age (years); the categorical covariates age at first birth ($\leq$30 years, >30 years, non-parous), HRT use (yes/no); oral contraceptives use (yes/no), previous IBIS-participant (yes/no) and mammogram Version (film, KE5.2, KE5.4, and Fuji). Age at randomisation was retained in all final models, regardless of significance or coefficient sign, because age is a well-recognised, strong modifier of MD.

Parameter estimates during covariate (fixed effect) model comparisons were obtained using maximum likelihood (ML) estimation. Interactions between covariates were not examined because there was no biological or clinical rationale for these. However, sensitivity analyses involving interactions between two covariate and time were undertaken. A quadratic term for time was tested in some non-multiply imputed parsimonious models of PD because a quadratic term for time was significant in another longitudinal study of PD [189].
Models were fitted for all mammograms, and separately for film only and digital only mammograms. All mammogram and digital only mammogram PD and DA models were also adjusted for mammogram Version as a (categorical) fixed effect.

Statistical models were checked for residual heteroskedasticity, including at different levels of the models. The parsimonious models of PD, DA and BA were examined for functional form, by graphically comparing the observed MD values with the linear prediction and fitted values generated by the model.

Sections 7.4.4 to 7.4.7, below, describe important steps undertaken to model the data using the hierarchical mixed linear model. The data were first modelled 'unconditionally'[490]. The representation of time was altered to a more suitable configuration as a continuous spline 'cut' (allowed to bend) at suitable time points. Forward stepwise and backwards stepwise modelling was undertaken to select a set of parsimonious covariates. Different covariance structures were tested. The models used for post-estimation testing of residual heteroskedasticity and model fit are described. Subgroup and sensitivity analyses are listed in section 7.4.8.

7.4.4 Unconditional means models for longitudinal PD, DA, BA

As per the procedure suggested in Singer and Willet's text [490], an unconditional means model was fitted to the data. The unconditional means model of MD utilised MD as the dependent variable and only a single RE term for participant as an explanatory variable (a two–level

"variance component" model). Use of the RE term for participant allowed each participant's mammographic episodes to be grouped within each participant; as described previously, this allowed the model to account for the repeated measurements made for each participant. This model provided a baseline against which later models with additional covariates were compared. Three quantities are estimated by this two–level model: a (fixed effect) intercept (a constant) and two random effects: the variance (variability) of the measurements between each person (inter-person variability) and the variability over time within each participant in the model (intra-person variability). Hence this model partitions the variability in the MD measurements into two sources— intra- and inter-person variability. The model is considered an 'unconditional' model because it does not contain any explanatory variables in the fixed effect (FE) part of the model, the main 'structural' part of the model.

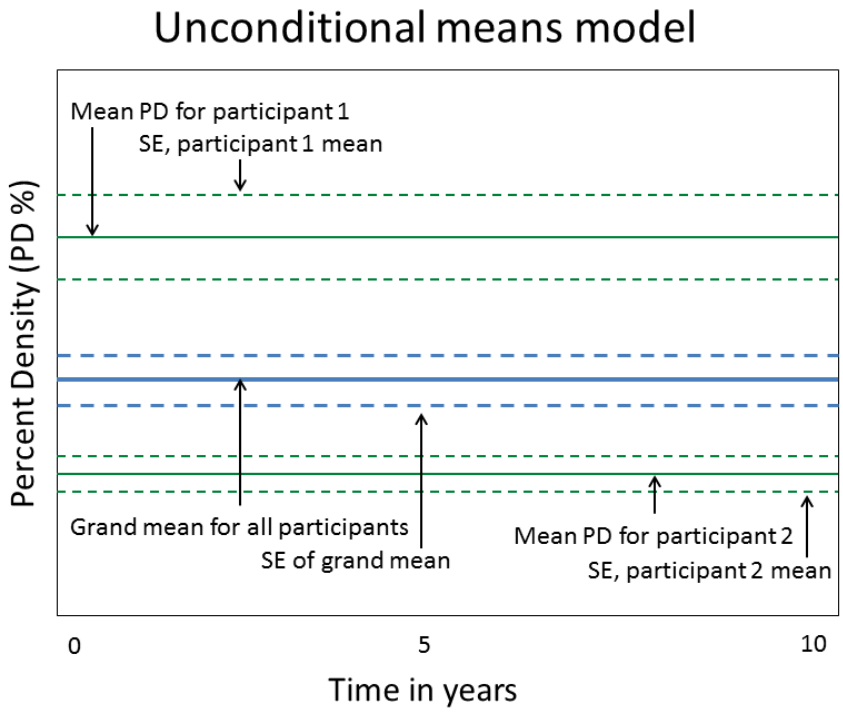


Figure 7-2 Unconditional means model, example participant and grand mean trajectories
Each participant has their own mean MD and SE. The grand mean is comprised of the average of the participant's means. Trajectories are flat because time is not included in the model.

The lowest level (level 1) of the hierarchy is the MD measurement for each episode (the average PD, DA or BA for the four mammographic Views per episode is the MD measurement for that

episode). The set of MD measurements for each participant is grouped within each participant; participant is hence the second level (level 2) utilised in the unconditional means model.

Each participant has their own average trajectory (growth curve). Because time is not a parameter in an unconditional means model, each person's average trajectory is flat, Figure 7-2. Each participants' MD measurements vary around their own mean by their own SE. The average of all participants' unconditional means is the 'grand mean' for the model. Hence the grand mean is the mean of the individual participant means. The amount the individual participant means vary around the grand mean is the SE of the grand mean. The intercept of the grand mean trajectory forms sole the 'fixed effects' (FE) estimate for an unconditional MD means model. Thus the 'grand mean' is the population average true mean for all participants (the highest level units in the model) in an unconditional means model. For models in this project, the FE denote the average MD response of the 120 CMN IBIS-II participants. The RE variance estimates represent the average variability of the participants for each RE at each level of the model.

As described above, the total variability in the model is partitioned into the two RE estimated by the model (the within-person variance and the between-person variance). The within-person (intra-person) random effect is estimated by the model as the average of the within-person variability. For example, if participants' measurements vary only slightly around their individual means, the model will have low average within-person (level 1) variability.

The between-person (inter-person) variability is the estimate of how much the participants' means vary (on average) around the grand mean. If all participants are similar, the between-person (level 2) variance estimate will be relatively small; if participants' mean estimates vary greatly the between-person variance estimate will be relatively high.

An ICC for the proportion of the total variance attributable to within-person differences (variability) was calculated by dividing the between-person variance by the sum of the between-person variance and the within-person variance. This ICC provides information on the residual autocorrelation in the model, which is the amount that the density parameters (e.g. PD) measured for each person's follow up visits are correlated with the density measurements in their other follow up visits.

The model coefficients, ICCs and model fit estimates from the PD, DA and BA models were tabulated for comparison with results from later models. The model coefficients for the unconditional means models are reported as output by the statistical program— in untransformed and square root transformed units. The mixed models were configured in Stata to report RE estimates as variances as per Singer and Willet (instead of the default setting of standard deviations[4]); hence RE estimates in this thesis are reported as variances instead of standard deviations.

7.4.5 Unconditional growth models for longitudinal PD, DA, BA

The unconditional growth model is comprised of two levels as per the unconditional means model, with time added to the model as both as a fixed effect and as a random effect. The addition of time as a fixed effect in the model allows the estimated average growth curve for all participants to change over time— the trajectory does not have to be flat. The inclusion of a RE for time allows each participant to have a unique growth curve which varies from the slope of the grand mean growth curve, Figure 7-3. The 'growth' in the name of the model refers to the 'growth curve' of the dependent variable, i.e. the change over time of the outcome being modelled. No additional (FE) explanatory variables are added to the model, hence this growth model is also 'unconditional'.

[4] Variance is the square of the standard deviation; Variance = (standard deviation)^2

The model now contains two fixed effects, a constant and a slope (change over time) which comprise the 'mean (average) structure' for the model. The FE intercept (constant) represents the grand mean (averages) of all participants' density measurements at time 0. The FE time parameter coefficients represent the average change over time for all participants. Categorical time was modelled as months since randomisation (0, 6, 12, 24, 36, 48, 60, 72, and 84 months). In separate unconditional growth models, continuous time was modelled as days since randomisation converted to years (days/365.25). Continuous time therefore provided a more precise date of mammography than categorical time, because the latter rounded continuous time to the nearest follow up visit. Categorical time, however, permitted non-linear growth curves to be fitted. Time modelled as a single continuous trajectory forced the growth curves to conform to a single, straight line; the growth trajectories could not bend like the growth curves for models with categorical time.

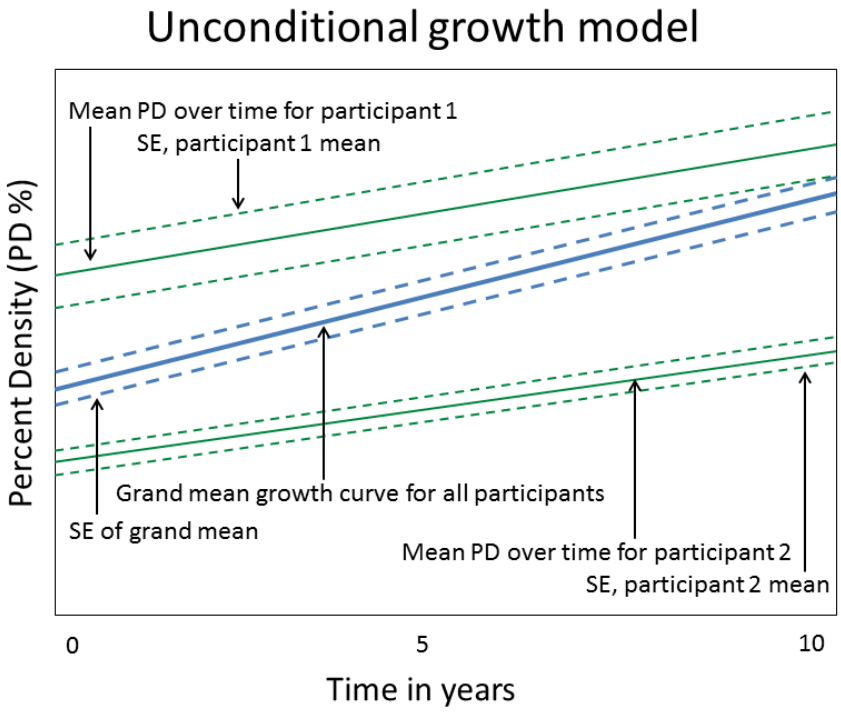


Figure 7-3 Unconditional growth model, example participant and grand mean growth curves
Each participant has their own mean growth curve SE, the slope of which can vary from other participants and the grand mean growth curve. The grand mean is comprised of the average of the participants' mean growth curves.

The addition of time to an unconditional means model typically explains some of the within- and between-person variability in the model, which is a characteristic effect of variables which change over time such as ambient temperature and income. In contrast, parameters which do not change over time (constants) such as sex and race primarily influence between person variability in mixed models. Parameters such as BMI which can change over time but are measured only once (e.g. IBIS-II baseline BMI) are considered to be constants, and hence primarily affect between person variability.

Two RE at the level of participant (level 2) are now included in the model: a random slope for each participant and a random intercept for each participant. As for the unconditional means model, the random intercept allows each participant to have a different baseline PD from the grand mean (FE intercept). Similarly, the random slope for each participant allows each participant to have a different slope from the grand mean trajectory (the FE time parameter).

A between person (RE) covariance was also fitted for the between person RE of the model, due to the potential for an association between initial status (baseline PD) and change over time. To express the covariance as a (standardised) correlation coefficient (ρ), the covariance was divided by the square root of the product of the variance components (initial status [RE intercept] and change over time [RE slope for time]),

Equation 7-2

$$\rho = \frac{Between\ person\ (BP)\ covariance}{\sqrt{(BP\ intercept * BP\ slope)}}$$

The reduction of within person variability due to the addition of time in the model was estimated by calculating the difference of the within person variance (WP var) between the unconditional means and growth models as a proportion of WP variance of the unconditional growth model,

Equation 7-3

$$\frac{WP\ var.unconditional\ means\ model - WP\ var.\ unconditional\ growth\ model}{WP\ var.unconditional\ means\ model}$$

Statistics of model fit—log likelihood (LL), AIC and BIC [493] [478]— were obtained for each unconditional growth model. The Deviance statistic (–2LL, for multilevel models) [490], was also calculated to assess the improvement in fit by the addition of time to the model.

7.4.6 Covariance structure selection

Simplification of the model with average MD at the lowest level of the model—rather than View-specific MD— permitted variance-covariance structure testing of the density measurement residuals. The residuals are the (within-person, level 1) deviations of the MD measurements from mean growth curve of each participant. Reduced maximum likelihood (REML) was used during covariance structure model comparisons [490].

All available linear mixed model covariance structures in Stata (independent, exchangeable, auto regressive (AR), moving average (MA), unstructured, banded, toeplitz, exponential) were trialled for models of the square root transformed MD parameters (PD, DA and BA); these MD models were configured with the parsimonious set of covariates selected during ML estimation and the default covariance structure (independent covariance, i.e. no covariance between mammographic episodes).

Parsimony, lower BIC scores and theoretical fit of the covariance structure to the data were considered during comparisons of the covariance structures which converged properly[5].

[5] The BIC cutoffs for model selection of 0-2, 2-6, 6-10 and >10 [weak, positive, strong and very strong evidence, respectively] are for models with independent observations. This is not the case for mixed longitudinal models [Jones 2011]. Singer and Willett [2003] recommend use of –2(log likelihood), the Deviance statistic, for multilevel models, rather than the AIC or BIC when assessing model fit.

Because log likelihood, AIC and BIC model statistics are not available after multiple imputation estimation, as for the regression models in Aim 3 (previous chapter), an average of the 25 multiply imputed estimations for the unknown age at menopause was utilised during covariance structure selection.

A sequential integer time/ordering variable for the AR, moving average (MA), toeplitz, unstructured and banded structures was created by cutting continuous time in years at 0.2 years after each annual follow up (i.e. 1.2 years, 2.2 years, etc.). Hence all 6 month follow ups were allocated into the 1 year category for this variable. 0 years for this variable was defined as the time between –1 and 0.25 years post-randomisation. Duplicate observations within categories were manually reordered to retain all mammographic episodes, however 2 participants with both –1 year baseline mammograms and ~2 month mammograms had these dates categorised as –1 and 0 years respectively. The exponential covariance structure utilised the 'uncut' (no splines) continuous time parameter (episode date converted to years, centred at date of randomisation).

During covariance structure testing, mammogram Version lost significance as a separate RE level within the model. Because each mammogram Version was still likely to induce randomly varying effects for each participant, separate indicator variables for each mammogram Version were created. Separate indicator variables were required, because support is limited for factor variable notation (an automated method of creating indicator variables from a categorical parameter) in Stata for random effects. All possible different combinations of the four indicator variables were modelled as RE: all four simultaneously, different sets of three, different pairs, and each RE slope for Version individually.

For generalised mixed linear models, the ratio of the FE default SE to robust SE indicates potential problems with the covariance structure specification [494]. If this ratio exceeds the

interval from 3/4 to 4/3 the covariance structure may be misspecified. These ratios were evaluated for covariance structures which converged properly.

Robust SE were obtained, since this allows the model SE estimations to likely be more correct (consistent and less biased) and less prone to misspecification problems and departures from assumptions (e.g. non-normality of residuals).

After a covariance structure was selected, forward selection for significant covariates was repeated using ML estimation to ascertain if the parsimonious models fitted using the default (independent) covariance structure were still correct. Robust SE were utilised during this second round of covariate selection, and were also utilised in final models to improve SE estimation in the event the models were incorrectly specified. Regression post estimation VIF (variance inflation factor) testing was performed as part of forward selection of significant covariates for the square root transformed MD parsimonious models (modelled only with 'first' (baseline or earliest) mammograms) for each participant), to cross check that age at randomisation and age at menopause were not significantly collinear (i.e. VIF was not >3).

7.4.7 Post estimation

As described previously, multiple imputation (MI) post-estimation commands in Stata are not as numerous as those for models which are not imputed. Post-estimation linear prediction of multiply-imputed mixed linear models is limited to the 'linear' predictions for the FE portion of the model (xb, the mean structural linear prediction) and its standard errors. Predictions for the RE portion of the model are not available for multiply imputed mixed models, hence 'fitted' predictions of both the linear (FE) predictions and RE effects cannot be obtained.

In order to see if it was feasible to use non-imputed models, properly imputed models of PD were compared against models of PD that utilised an average of the 25 MI imputed values for age at menopause for the 27 participants with unknown age at menopause. The model

coefficients and SE as well as graphs of the linear predictions (xb) were compared. Few differences were found. Hence it was judged acceptable to use the non-MI models to undertake post-estimation checking of the parsimonious aggregate models of MD change for the CMN IBIS-II participant data.

Use of non-MI models allowed full access to the range of post-estimation commands for mixed models in Stata. Fitted MD values for each participant were estimated in Stata by adding a 'Best Linear Unbiased Prediction' (BLUP) for each level of random effects to the mean structural prediction (FE 'xb' estimates) for each participant. These linear (FE) and fitted (FE+RE) estimates were utilised to create a number of graphs to check the parsimonious models against the measured (observed) values for MD. Because the axes of the predicted vs observed graphs were inadvertently reversed (the observed values are plotted on the y axis instead of the x axis), a line of equality (y=x going through the origin) is shown for reference, instead of the more typical linear prediction reference line.

The residuals from the linear (xb) and fitted estimates were visually assessed for normality. Additional checking of normality using skewness and kurtosis tests were undertaken. Plots of the observed, linear and fitted predictions for low, mid and higher PD participants were visually inspected for fit.

Models were checked for heteroskedasticity, normality of residuals, and linearity assumptions. The parsimonious models of PD, DA and BA were examined for functional form, by graphically comparing the MD values input into the model with the linear prediction and fitted values generated by the model.

7.4.8 Subgroup, sensitivity and exploratory analyses

To examine the possible influence on model outcomes, participants with unusual values for MD

and other covariates were omitted from the parsimonious models. The resulting coefficients and SE were examined for any substantial differences from the original models. Three participants with high PD (≥50%) were omitted, as well as a participant with an unusually long duration of HRT for the dataset (361 months) and a participant with unusually high OC duration (528 months). Women with imputed age at menopause were also omitted, and the resulting models compared with those for all 120 participants for coefficient size and significance. Additional comparisons were also made with another PD model which utilised 'as reported' values from the IBIS-II database for age at menopause for the 93 participants for whom these values were known, and the age at hysterectomy for the 27 participants for whom age at menopause was not known. MI was not required for this 'as reported' PD model, since age at hysterectomy was substituted for age at menopause for the 27 participants for whom age at menopause was not known.

Given that other longitudinal MD studies have shown an easing reduction in PD change over time with increasing age, e.g. [144, 189], further modelling was done according to age group at baseline: ≤60 years, and >60 years, which approximated median age of the 120 CMN participants at baseline (~62). Baseline to year 1 square root transformed PD change was tabulated for women ≤60 years vs >60 years for models fitted with all mammograms, film only, digital only, KE52 only and KE54 mammograms only; these different digital mammogram Versions were modelled separately to assess if PD baseline to year 1 growth differed for the different digital Versions. Models for Fuji mammograms only were not fitted, because all baseline and 6 month mammograms were taken at CMN prior to use of this mammographic Version.

Models by age (≤ or > 60 years) for the film only, KE52 only and KE54 only subsets of mammograms were modelled with the default (independent) covariance structure due to small sample size; this allowed the models to converge because there were too few within person

repeated measurements to calculate 'rho' (the within person correlation estimate) for these subgroup models. An additional set of all mammogram models were fitted for women ≤55 vs >55 years at baseline, and for women aged ≤65 vs >65 at baseline, for comparison with the original model and the models divided at age 60. The number of distinct episodes and participants available from baseline to 1 year post-randomisation for each age group subset was also tabulated, to provide an indication of the samples sizes and repeated measurements per participant likely to contribute to each model.

The interaction between age at randomisation and time was also modelled as an interaction between time and age as a continuous variable. This allowed full use of the data set within one model; separate models for younger and older women, as described above, were not required. An interaction between time and mammogram Version was also modelled because it seemed likely that the average slope for PD and DA over time might vary for each mammogram Version, e.g. Figure 7-7. The time and mammogram Version interaction model was fitted both with and without RE terms for film and Fuji mammograms, to compare if model coefficients differed substantially (>1 or 2 SE) with and without the RE for Version. These interactions (between age and time, and Version and time) were exploratory only, and were not considered for retention in the final aggregate longitudinal models. Retention of an interaction term with time in the aggregate models would create three-way interactions with treatment group during modelling for the Primary Aim; the sample size was too small to have adequate power to examine all of these effects simultaneously.

Because of potential differences in longitudinal MD due to mammogram Version, models for each digital mammogram Version only were used to control for longitudinal differences in change due to mammogram Version. Models for film, KE52, KE54 were modelled with both an exponential and default (independent) covariance structure for compatibility with the Fuji only model. This is because the model with Fuji mammograms only would not converge when

fitted with an exponential covariance structure; hence only results for the default (independent) covariance structure are tabulated for the model with Fuji mammograms only.

Some studies assessing MD change in response to AI treatment have excluded women with low PD (e.g. <5%) [29, 313], presumably to enhance the likelihood of observing a difference between treatment groups. Other MD and AI studies have shown that change in PD differs for women according to baseline PD [31, 312]. To assess if any differences in rate of MD change existed for women with higher (or lower) initial PD, participants with first episode (baseline or earliest episode) PD up to 10%, from 10 to <25% and ≥25% were modelled separately. The frequency of participants per age group at randomisation (<55, 55 to 59, 60 to 64, and 65+) by initial PD were tabulated for reference.

All sub-group models (e.g. age ≤60 vs > 60 years) utilised a square root transformed PD non-MI parsimonious model, with an exponential covariance structure except as noted above for certain models fitted by mammogram Version. Hence all sub-group models are adjusted for age at randomisation, BMI, age at menopause, parous status and mammogram Version. Although only annual change in PD from baseline to year 1 is reported for some tables, all episodes (baseline to year 7) were modelled to improve stability of the adjustment by mammogram Version, the coefficients for which can vary substantially if only a few episodes from each Version are modelled.

7.5 Results

7.5.1 Mammogram collection

A total of 2,130 film and (original, fully electronic) digital mammograms from 541 episodes were collected during two rounds of collection, representing 120 Calvary Mater Newcastle participants. Five-hundred of the episodes (1971 mammograms) were from follow up 0 (baseline) to the cease of trial treatment at follow up 5 (60 months). Forty-one episodes (159 mammograms) were obtained from follow-up visits between five and eight year post

randomisation. Only one mammographic episode was collected for the year 8 (96 month) follow up. This single episode was omitted from further analyses because it was unlikely to be representative of year 8 for all CMN participants. Hence 2126 mammograms, comprising 540 mammographic episodes, were utilised for Aim 4.

Due to timing of mammographic episodes, mammograms for two participants which were taken after randomisation were designated as baseline mammograms for these participants. One episode occurred 10 days post-randomisation, the other 2.5 months post-randomisation. It is unlikely that measureable changes in MD occurred 10 days after randomisation, however it is possible that discernible changes in density may have occurred 2.5 months post-randomisation [258]. As discussed further below, the principal models for this Aim were undertaken with time modelled as the number of days from randomisation, not with time categorised as the month of trial follow up (0, 6, 12, 24, etc.). The results for Aim 4 were unlikely to be affected.

Mammograms dated prior to January 2009 were film-screen mammograms. The final film mammogram for this data set was taken on 17 December 2008. Mammograms from January 2009 were digital post-processed CR (computed radiography) mammograms. The majority of digital mammograms were one of three CR "versions": Kodak Elite CR software version 5.2 ("KE52", 15 April 2009 to 1 March 2011); Kodak Elite CR software version 5.4 ("KE54", 10 March 2011 to 31 July 2012); Fuji CR ("Fuji",14 August 2012 to 1 July 2014). Four digital episodes of unknown Version were assigned to Version category 'Other'.

Table 7-1 Number of mammograms and mammographic episodes by mammogram Type and Version

Mmg Type	# Episodes (# Mmgs)	% of total	# Pts	Episodes per Participant					
				Mean	SD	Median	Q1–Q3	Min	Max
Film	63 (248)	12%	45	1.4	0.7	1	1–2	1	3
Digital (all)	477 (1,878)	88%	120	4.0	1.0	4	3–5	1	6
Total	540 (2,126)	100%	120	4.5	1.4	4.5	4–5	1	8
Digital Mammogram Version:									
KE5.2	123 (482)	23%	94	1.3	0.5	1	1–2	1	3
KE5.4	151 (596)	28%	114	1.3	0.5	1	1–2	1	2
Fuji	199 (784)	37%	113	1.8	0.4	2	2–2	1	3
Other	4 (16)	1%	4	1	0	1	1–1	1	1

Mmg mammogram; Pts participants; SD standard deviation; Q1 1st quartile; Q3 3rd quartile

The majority of mammograms collected were digital mammograms, n=1878 (Table 7-1, above); 248 unique film-screen mammograms from 63 episodes were also collected. The total number of mammography episodes per participant ranged from 1 to 8, with a mean of 4.5 follow ups per participant (Table 7-1). Four sequential, annual follow ups equated to about 3 years between the earliest and latest episodes.

The number of mammographic episodes collected for each trial follow up and mammogram Version are listed in Table 10-7.

7.5.2 Descriptive analyses

7.5.2.1 *Graphs of average PD over time for all participants*

Average raw (unmodified by covariates) PD for 120 IBIS-II participants is graphed in Figure 7-4 and Figure 7-5. Raw, average PD (per episode) for all 120 participants from 541 episodes

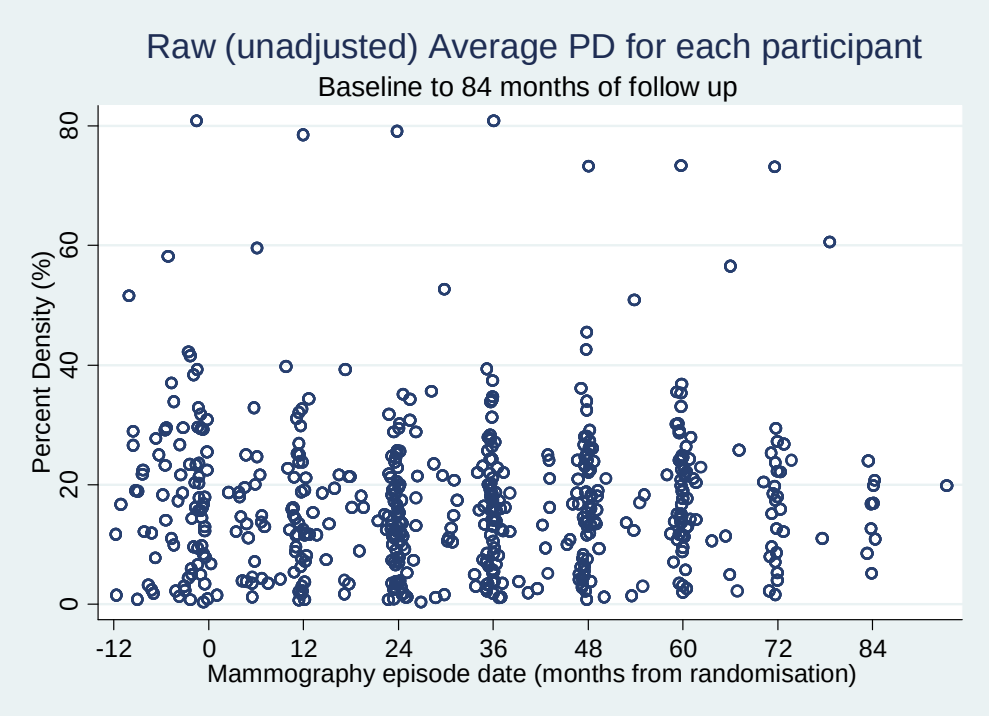


Figure 7-4 Average percent density (PD, %) per participant, 0 month (baseline) to 84 months follow up
Raw (as measured) average PD from 541 episodes (~2130/4 mammograms) from 120 Calvary Mater Newcastle IBIS-II participants. The episodes show a trend to cluster around the ideal follow up time points at 12-month (yearly) intervals from baseline (year 0) through 84 months (7 years). Outlying right-most episode was taken at 96 months (year 8); this data point was omitted from further analyses

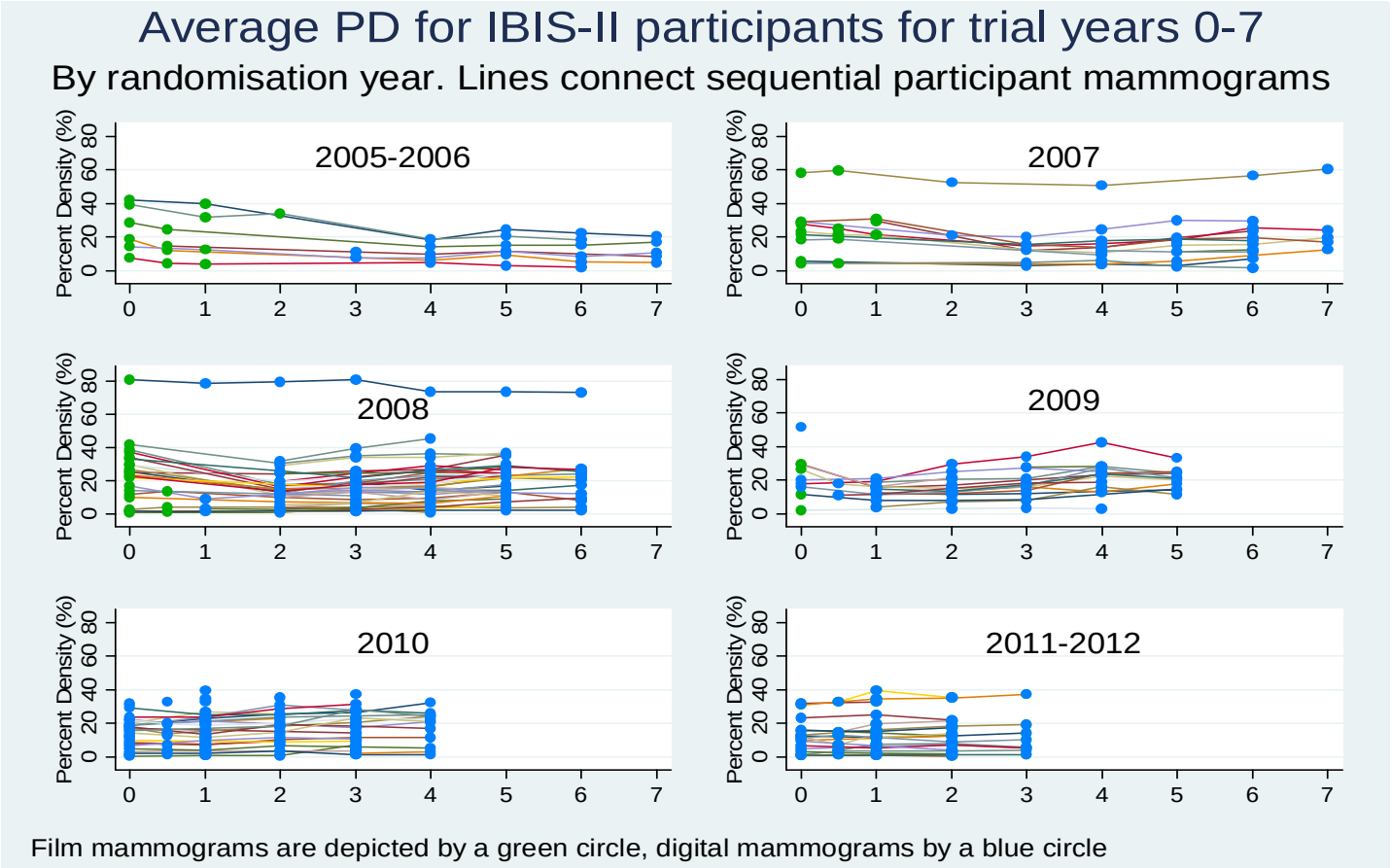


Figure 7-5 Average percent density (PD) over time, 120 IBIS-II CMN participants, by mammogram Type

Graphs by randomisation year 2006-2012. Green circles represent film mammograms, blue circles are digital mammograms. There are 540 episodes from 120 participants. An expected drop in PD is seen from the transition from film to digital mammograms. An unexpected increase in PD over time for digital mammograms is also shown.

(2130 mammograms) is graphed against time (months since randomisation) in Figure 7-4. Episodes are clustered around the yearly follow up time points, and appear to be most numerous for the 24 through 60 month follow ups. Whilst the expected trend toward decreasing density (due to ageing and trial treatment) is apparent from baseline (month 0) through 24 months, it is not clear in this figure if longitudinal data from same participants are represented over these time points.

In Figure 7-5, measurements of average, raw PD from the same participant are connected by a coloured line overlaid with green and blue dots to illustrate longitudinal MD by mammogram type (film vs digital). Participants contributing only a single episode to the data set are represented by a dot. The participants are grouped by randomisation year in order to match the different mammographic Types (film, digital) over time. Green circles represent film mammograms, blue circles represent digital mammograms.A marked drop between the film (green dots) to digital (blue dots) transition is visible on these raw longitudinal PD trajectories. Digital mammogram PD also appears to increase over time for many participants (blue dotted lines). This latter effect is influenced by the change from KE52 to KE54 to Fuji CR mammograms, as described in Aim 3 (Chapter 6). Changes to post-processing of the digital mammograms on the Kodak Elite (KE) and Fuji machines appear to have caused more density to be retained over time (Figure 6-2 and Figure 6-1).

CMN IBIS-II participants reported they were fully compliant at 551 of 622 total follow up visits (with or without associated mammograms). 21 visits were recorded as a protocol deviation (e.g. "missed 7 days while on holidays") and 24 were recorded as treatment holidays (e.g. 'stopped for 5 weeks due to hot flashes'). Cease of trial treatment was noted on 25 visits; 13 of these were due to side effects such as hot flushes and arthralgia, whilst 12 were due to other reasons including one multiple sclerosis diagnosis, one DCIS diagnosis, and two breast cancer diagnoses. Treatment compliance status for the 540 follow ups with associated mammograms

(episodes) are listed in Table 7-2. Compliance with trial treatment was not applicable for baseline mammograms, and year 6 and 7 mammograms. Of the 415 follow ups collected during the trial treatment period, full compliance was present for (346/415=) 83% of the episodes

Table 7-2 Treatment compliance reported at trial follow ups with mammograms (n=540)

Compliance Status	Follow up number (years post-randomisation)									
	0	.5	1	2	3	4	5	6	7	Total
Not applicable	85	--	--	--	--	--	--	30	10	125
Full[1]	--	20	52	81	75	64	50	--	--	342
Deviation[2]	--	2	1	4	0	0	1	--	--	8
Tx Holiday[3]	--	3	4	1	2	2	0	--	--	12
Ceased Permanently	--	0	3	7	14	16	13	--	--	53
Total	85	25	60	93	91	82	64	30	10	540

[1] Full compliance: missed ≤3 consecutive pills or ≤14 pills total per year)
[2] Deviation: missed >3 consecutive pills or > 14 pills total per year
[3] Tx (treatment) holiday: a planned break from trial medication (e.g. due to side effects)

7.5.2.2 *Distributions of PD, DA, BA, and AA*

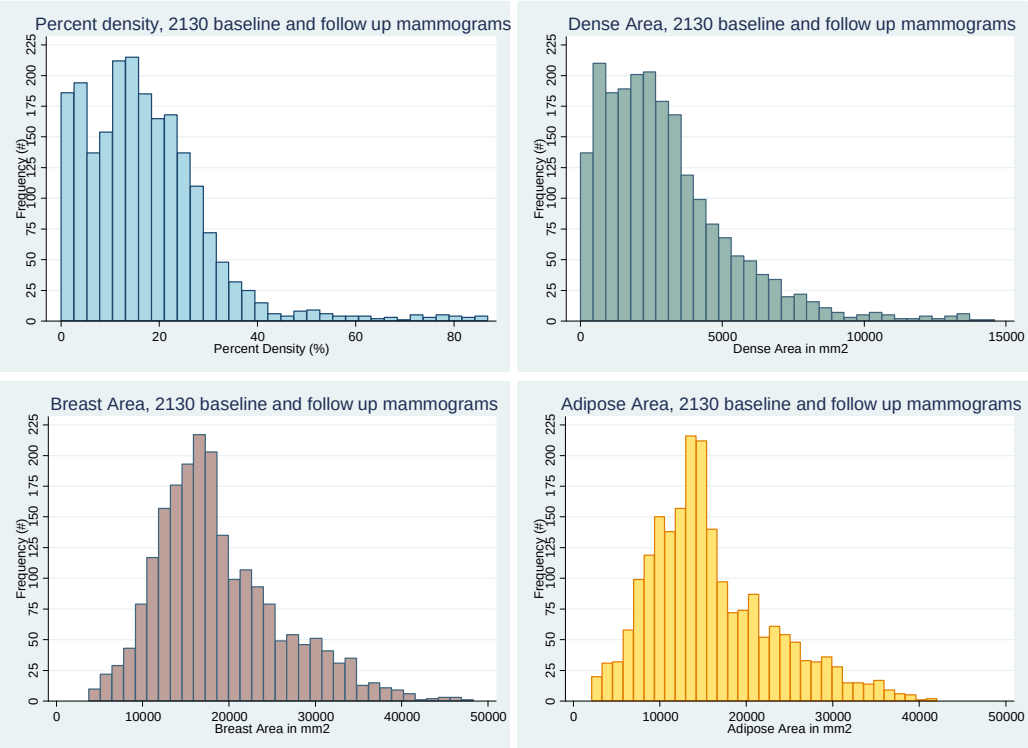


Figure 7-6 Histograms of PD, DA, BA and AA baseline and follow up mammograms
120 IBIS-II CMN participants. A right skew is present in all histograms, which is more pronounced for the PD and DA distributions (upper row) than for BA and AA (lower row).

collected for this analysis. A "per-protocol" treatment sensitivity analysis of 117 participants and 447 episodes (including baseline mammograms, deviations and treatment holidays) up to three months after trial treatment ceased is presented in Chapter 8 (Table 8-16).

PD and DA had a pronounced right skew, whilst BA and AA were more normally distributed (Figure 7-6). Film mammograms had the highest median PD and DA, whilst KE52 mammograms had the lowest median PD and DA (Figure 7-7, Table 7-3). Median PD and DA increased for KE54 mammograms, and the medians for these MD parameters were higher again for the Fuji mammograms.

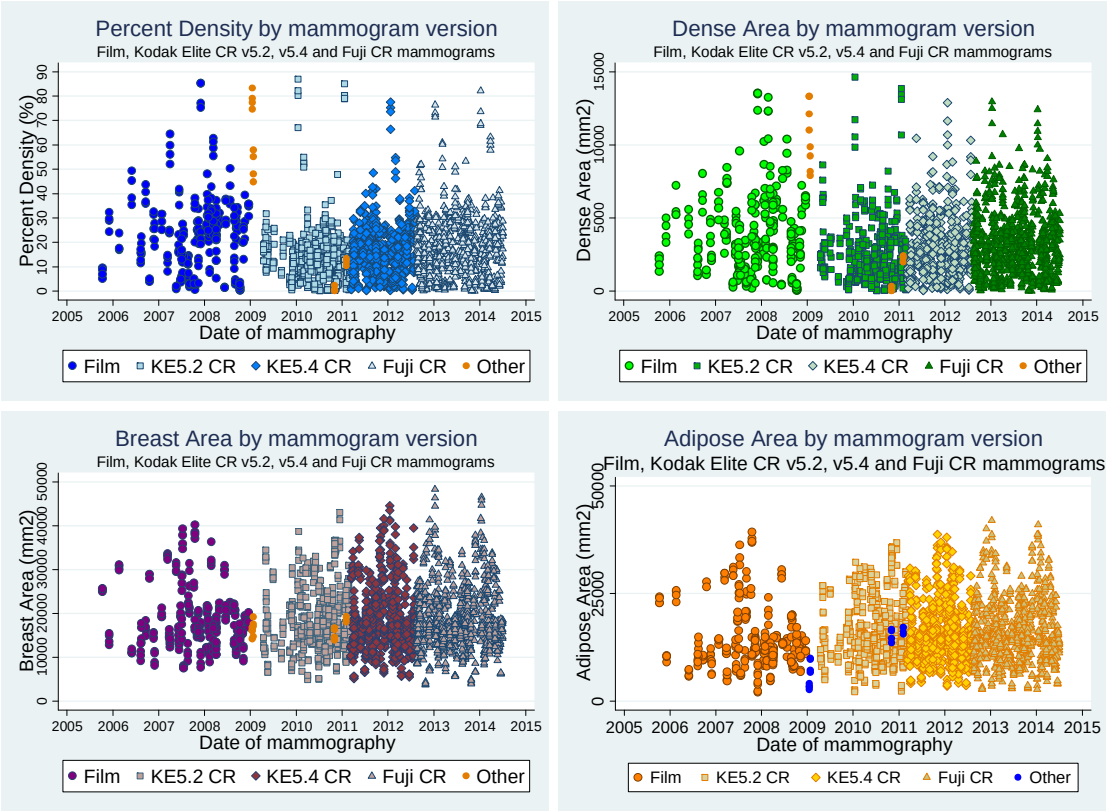


Figure 7-7 PD, DA, BA and AA by mammogram version over time.
PD for film mammograms is more evenly dispersed over the possible range of values for PD (0 to 100%) than PD for digital mammograms; the distribution range for DA is also correspondingly higher for film mammograms than for the digital mammograms. Digital mammograms show a trend towards increasing PD and DA over time which is likely due to changes in software and hardware configuration. An increase in BA over time due to changes in digital mammography version or age related factors (e.g. increased compressibility of the breast) may also be present, whilst the opposite may be true for AA.

Median PD differed by about 2% between CR mammography Versions (Kodak Elite v5.2, Kodak Elite v5.4, Fuji), double the expected change in longitudinal PD of approximately (–)1% per annum. The increasing trend in PD over time also opposes the anticipated decrease in PD over time. As stated previously, this implied mammogram Version would need to be taken into account when analysing longitudinal MD in this dataset.

Table 7-3 Number of episodes, MD measures of central tendency, by mammogram Type and Version

Parameter	N	PD (%)	DA (mm^2)	BA (mm^2)	AA (mm^2)	PA (mm^2)
		Median (25^{th}-75^{th}percentiles)		Median		
All mammograms	540	15.6 (8-23)	2576 (1400-4032)	17354	14377	84.4
Film mammograms†	63	22.4† (12-31)	3383† (1615-5539)	16493††	12524†	77.6†
Digital (CR)‡	477	15.0 (8-22)	2456 (1337-3853)	17384	14642	85.0
Kodak Elite v5.2	123	13.0‡ (7-18)	2119‡ (1143-3031)	17376	14892	87.0‡
Kodak Elite v5.4	151	15.7‡ (8-21)	2553‡ (1239-3868)	17222	14481	84.3‡
Fuji	199	17.3‡ (9-24)	2879‡ (1634-4151)	17455	14622	82.7‡
Other	4	31.8 (6-65)	5532 (1194-10618)	16496	11833	68.2

† PD and PA differ for film vs digital mammograms, p<0.001; Significant differences between film and digital mammograms are also found for DA (p=0.009) and AA (p=0.02); Wilcoxon's rank-sum test
†† BA does not differ for film vs digital mmgs, p=0.45, Wilcoxon's rank-sum (Mann Whitney) test
‡ KE52, KE54 and Fuji CR mammograms differ from each other significantly for PD, DA, PA, p<0.01, Kruskal-Wallis test (p>0.75 for BA, AA); The Cuzick non-parametric trend test was significant for trend for PD, DA, and PA measured on these CR mammograms, p<0.003

7.5.2.3 *Review of the variability in PD measurements*

Line plots of average PD were inspected for unusual (>5 to 10%) differences in average PD for sequential mammograms for participants. Mammograms from 13 participants were re-measured with the Cumulus program to reduce within-individual longitudinal differences in density which likely resulted from assessing Collection 1 and Collection 2 mammograms at different times. Substantial longitudinal variation in PD between and within Versions remained after PD re-measurement for the 13 participants (Figure 7-8). This may have occurred due to the inherent variability of the subjective assessment technique, or potentially due to differences in image acquisition and/or post-processing (other than differences due to Versions), or other factors such as the positioning of the breast during mammography [416], which may substantially change the appearance of the dense tissues on mammograms (i.e. >10%PD change).

PD and DA had a pronounced right skew, whilst BA and AA were more normally distributed (Figure 7-6). Film mammograms had the highest median PD and DA, whilst KE52 mammograms had the lowest median PD and DA (Figure 7-7, Table 7-3). Median PD and DA increased for KE54 mammograms, and the medians for these MD parameters were higher again for the Fuji mammograms.

Median PD differed by about 2% between CR mammography Versions (Kodak Elite v5.2, Kodak Elite v5.4, Fuji), double the expected change in longitudinal PD of approximately (–)1% per annum. The increasing trend in PD over time also opposes the anticipated decrease in PD over time. As stated previously, this implied mammogram Version would need to be taken into account when analysing longitudinal MD in this dataset.

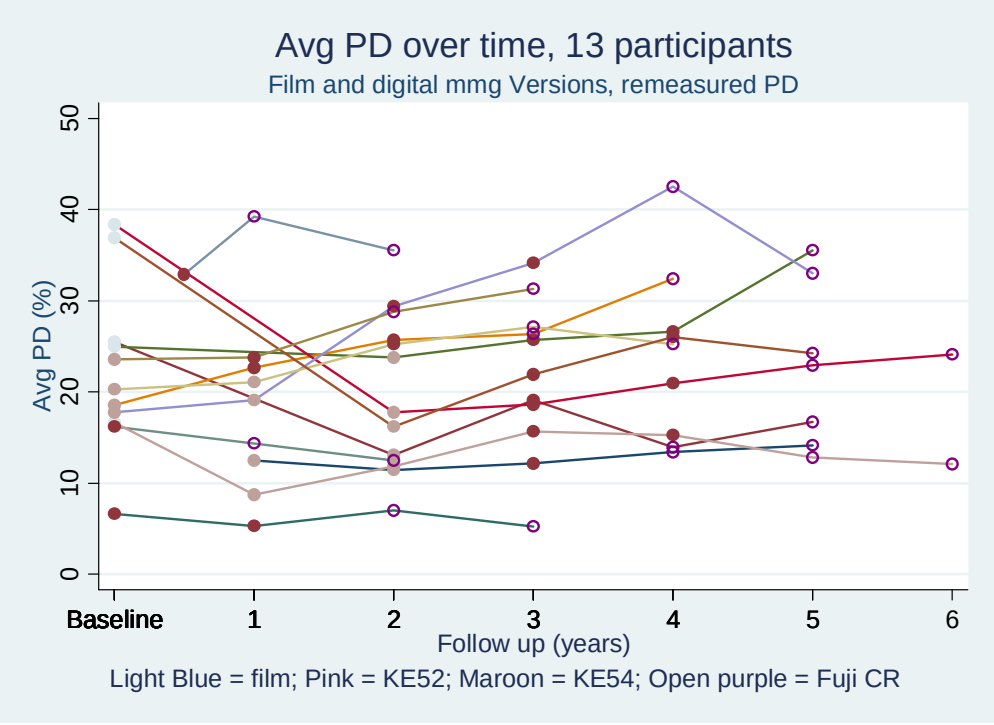


Figure 7-8 Avg PD over time by digital (CR) mammogram Version
Average (unadjusted) PD tended to increase over time for digital mammograms and was related to the software and hardware version of the CR mammography machine used.

The analysis for Aim 3 (Baseline characteristics, Chapter 6) was not re-performed with the updated measurements for the 13 participants because longitudinal data was not required for

Aim 3. The subjective nature of the Cumulus assessments means that any reasonably consistent set of measurements—as shown in Aim 2 (Chapter 5, reliability analysis)— is likely to be as valid as another set of similar measurements.

7.5.3 Three level mixed model results

The individual mammographic density measurements (one per mammographic View) were originally utilised as the observations (units) which formed the bottom level of the data hierarchy. The model development strategy used for the two level mixed model described in sections 7.4.3 to 7.4.7 was also used to develop the three level mixed model, with the exception of section 7.4.6 (covariance structure selection); the covariance structure could not be modified for the three level model due to (within episode, or within Version) repeated times at the lowest level of the model. The model coefficients for the set of PD, DA and BA three level mixed models are listed in two appendices (Three-Level Unconditional Means and Unconditional Growth Models, Three-Level Full and Parsimonious Models of PD, DA and BA). The three levels modelled are: participant (level 3), (episodes or) mammographic Version (level 2), and mammographic Views (level 1).

The results for the three level model are similar to the results for the two level model. Both the two and three level all mammograms models show a decrease in annual PD and DA change for baseline to year 1, an increase from years 1 to 5, and negligible change for years 5 to 7. However, the rate of annual change is larger in magnitude for the 2 level model from baseline to year 1, but smaller for years 1 to 5 (i.e. MD does not increase as quickly). This may imply that use of average MD per episode (in the two level model) helped to stabilised the longitudinal measurements. The film only model rates of change are almost identical for both sets of models. Larger differences between the two and three level models are noted for PD and DA for the digital mammogram only models. The magnitude of annual rate of change for baseline to year 1 and years 1 to 5 is larger in the three level digital only models. Whilst the two level baseline to year 1 all mammogram model rates of change for PD and DA are intermediate between the rates of

change for the film and digital only models, evidence of potential confounding is present for the three level model. The annual rates of decline for PD and DA from baseline to year 1 for the three level film only and digital only models are greater in magnitude than the three level all mammogram models, which should be in between those for film and digital.

The three level model has higher "within person" variance, approximated by summing both the level-1 (residual, within Version) variance and the level-2 between mammogram variance. This is likely due to the use of all four Views instead of average MD per episode. The estimated variability of the between person intercept for the three level model is also higher than the between person intercept for the two level model. However, total between person variability for the two level model (approximated by the sum of the variances for the between person intercept as well as the random slope variances for film and Fuji mammograms) is higher than the sum of the three level variances for baseline to year 1 random slope and intercept. The higher between person variability for the two level model may mirror the improved efficacy of the model to more accurately allocate sources of random variability. For example, although the coefficient estimates for the parsimonious model of square root transformed PD are similar for age at randomisation for the two and three level models (–0.04), the p-value for the two level model is less than 0.01, whilst the p-value for the three level models is between 0.05 and 0.1. The SE for age at randomisation in the two level model are therefore smaller. This appears to indicate that the two level model has improved main effects efficiency (precision) via a decrease in (FE) SE variability, by ascribing that variability to RE sources.

7.5.4 Unconditional means models for longitudinal PD, DA, BA

The unconditional means model of MD contained only participant group as an explanatory variable in the model. A graph of the estimates for the untransformed PD unconditional growth model is shown in Figure 7-9. The fixed effect and random effect components are all constants; hence all participants' estimated trajectories are flat, which therefore results in a flat trajectory

for the Grand Mean of the model (red solid line). The observed (measured) trajectory for one participant is also included on the graph as a fine, dotted green line.

Coefficients for the different PD, DA and BA unconditional means models are listed in Table 7-4. Each participant's true mean PD (or DA or BA) varied around their own mean PD on average by the within person variance, and each participant's own mean PD varied on average around the PD 'Grand Mean' by the between-person variance. For example, the grand mean for PD was 16.7%, PD between person variance was 131, & the PD within person variance was 17.

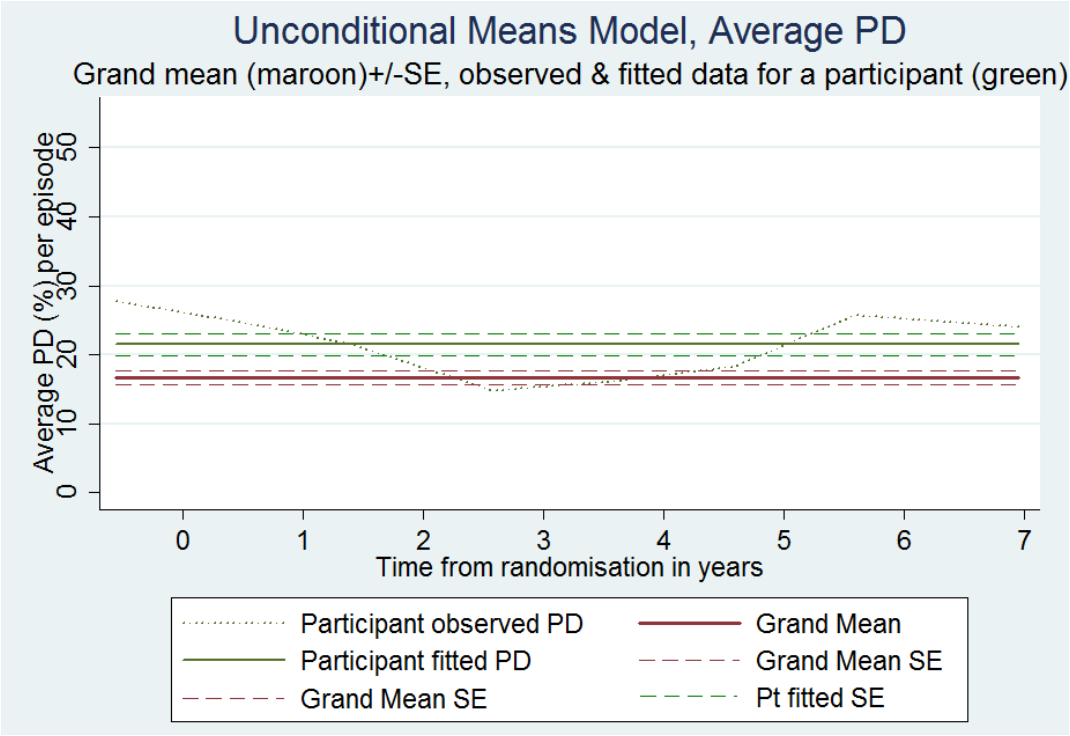


Figure 7-9 Unconditional means model: Grand mean ± SE w/participant observed & fitted trajectories
The PD grand mean for the unconditional means model is the mean of all PD measurements for all participants (16.7%). The SE limits for the grand mean (1.1%) is shown as dashed lines above and below the grand mean. The observed trajectory for one participant is shown with green dots. The participant's fitted trajectory (grand mean + the best linear prediction from the between-person RE for the participant) with SE limits is shown as a thin solid green line with dashed green lines above the red line of the grand mean of the model. The modelled lines are flat because PD is the sole parameter in the model (e.g. time and other parameters have not yet been added to the model). The SE for the fitted participant values (green dashed lines) may be smaller than expected when compared to the observed PD (dotted green line) for the participant because they are for best linear unbiased predictions (BLUPs).

The calculated ICCs in Table 7-4 for PD and DA range from 0.89 to 0.90, hence the correlation between follow ups for each participant is high, and only a small proportion of the variability of the density measurements is due to within person differences in MD. This likely implies that the change in MD over time is small per person compared to the between person differences in MD, although more generally, it could indicate between person variability was just very high compared to within person variability. The ICC for BA is high (0.98), which likely implies BA tends to be stable over time. This inference is supported by the increase of the ICC to 0.99 after adjustment of the BA model with FE and RE for mammogram Version (results not shown); this adjustment appears to have further increased the stability of BA over time. Unconditional means model ICCs for film mammogram only were ~0.98 for PD and DA, and 0.98 for BA; ICCs for

Table 7-4 Mixed modelling coefficients, unconditional means models for PD, DA and BA

Covariate	PD %	PD square root	PD natural log	DA mm^2	DA square root	BA mm^2	BA square root
			ALL MMGS, 541 episodes[1], 120 participants				
FE Intercept	**16.7*** **	**3.78*** **	**2.44*** **	**2920*** **	**50.0*** **	**19006*** **	**135*** **
			RE variance estimates				
Between person intercept (initial status)	**131**	**2.0**	**0.89**	**$4.1x10^6$**	**357**	**$52x10^6$**	**664**
Within person	**17**	**0.24**	**0.10**	**$0.53x10^6$**	**39**	**$1.3x10^6$**	**14**
ICC	**0.89**	**0.89**	**0.90**	**0.89**	**0.90**	**0.98**	**0.98**
			Model fit estimates				
Log Likelihood	-1750	-594	-356	-4545	-1979	-4883	-1812
AIC	3505	1195	717	9095	3965	9773	3631
BIC	3518	1208	730	9108	3978	9786	3643

Bold indicates p≤0.1; * p<0.05; ** p<0.01; *** p<0.001; † p<0.2

[1]The unconditional means model coefficients differed slightly for MD averages of all 2130 mammograms, e.g. average PD for 2130 individual mammograms yielded an intercept of 16.7%, with a between-person variance of 134 and within-person variance of 25. This equated to an ICC of 0.84.

digital mammograms only models were ~0.92 for PD and DA, and 0.98 for BA (results not shown). The lower PD and DA but not BA ICCs for the digital only models reflect the differences in MD caused by differences in post-processing of the digital images. All ICCs are high, hence the density measurements are also highly auto-correlated within participants over time. However this likely implied that the aggregate (treated + control) change in PD and DA due to trial treatment also was small. All RE variance estimates were significant. This meant

additional covariates could be added to the model to explain (reduce) the variability of the between- and within-person variances.

7.5.5 Unconditional growth models for longitudinal PD, DA, BA

Time as a fixed effect in the model allowed the estimated average MD for all participants to change over time— the trajectory is no longer flat. The inclusion of time as a RE in the model allowed each participant to have a unique growth curve which varied from the slope of the grand mean growth curve. Figure 7-10 illustrates this for a different participant from that shown in Figure 7-9. The coefficient estimates for the PD, DA and BA models are tabulated in Table 7-5.

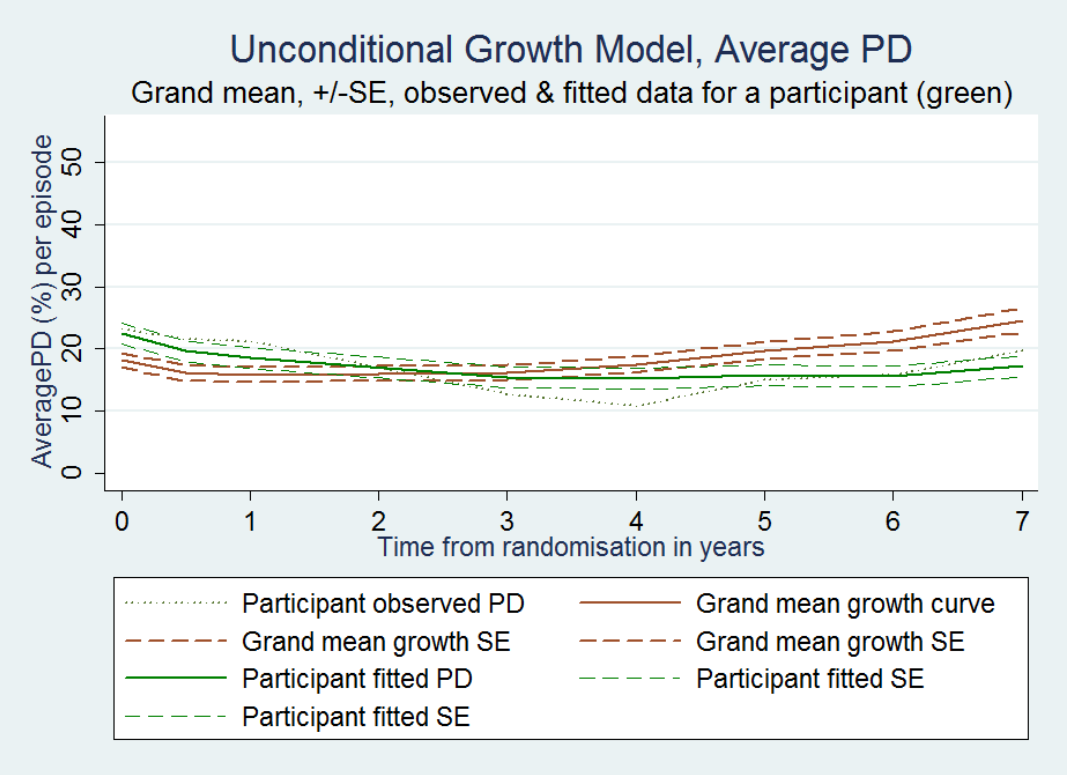


Figure 7-10 Unconditional growth model, Grand Mean growth curve with observed & fitted PD for a participant

The average unconditional growth curve for all participants (± SE) is shown in maroon, whilst the observed (as measured) PD and fitted PD (± SE) are shown for a participant in green. This participant differs from the participant whose mean trajectory is shown in Figure 7-9. Time is modelled as a categorical variable. The modelled growth curves now change over time (because time has been added as a parameter to the model). Approximately half (0.47) of the within person variability is explained by the addition of time to the model (Table 7-5), implying the model has improved fit compared to the Unconditional means model (which does not include time as a parameter).

The model now contained two fixed effects, a constant and a slope (change over time) which comprised the 'mean (average) structure' for the model, shown in red in Figure 7-10. The FE intercept (constant) represented the Grand Mean (averages) of all participants' baseline density measurements. The intercepts were 18.2% (categorical time) and 16.1% (continuous time) for PD, Table 7-5. The FE time parameter coefficients represented the average change over time for all participants. More specifically, the categorical time coefficients estimated the mean change

Table 7-5 Mixed modelling coefficients, unconditional growth (time only) models for PD, DA and BA

Covariate	PD %	PD square root	PD natural log	DA mm^2	DA square root	BA mm^2	BA square root
ALL MMGS, 541 episodes, 120 participants							
Categorical time FE estimates							
Months since randomisation	0 (ref)	0 (ref)	0 (ref)	0 (ref)	0 (ref)	0 (ref)	0 (ref)
6	**-2.1*** **	**-0.20***	-0.03	**-341***	**-2.3***	**515***	**1.7***
12	**-2.3*** **	**-0.26*** **	**-0.11** **	**-411*** **	**-3.3*** **	**235**	**0.9**
24	**-3.0*** **	**-0.31*** **	**-0.11** **	**-559*** **	**-4.0*** **	217†	0.9
36	**-2.0** **	**-0.18***	-0.01	**-299** **	**-1.7**	**541** **	**2.0** **
48	-0.60	-0.03	0.04	-144	-0.24	400	**1.4***
60	**1.7***	**0.25***	**0.20** **	**295***	**0.7** **	**633***	**2.1***
72	**3.2** **	**0.41** **	**0.26** **	**478***	**5.3** **	**699***	**2.3***
84	**6.8*** **	**0.80*** **	**0.47*** **	**1026***	**10** **	**1015***	**3.3***
Intercept	**18.2*** **	**3.9*** **	**2.4*** **	**3170*** **	**48*** **	**18626*** **	**134**
Continuous time FE estimates							
Change per year	**0.28**	**0.04***	**0.04***	**46**	**0.64***	**106**	**0.36**
Intercept	**16.1*** **	**3.7*** **	**2.4*** **	**2817*** **	**48*** **	**18723**	**134**
Continuous time RE variance estimates[1]							
Between person							
Time (years)	**1.5**	**0.02**	**0.01**	**38648**	**3.2**	**105316**	**1.3**
Intercept	**149**	**2.3**	**1.1**	**$4.8x10^6$**	**417**	**$51x10^6$**	**651**
Covariance	**-5.2**	**-0.08**	**-0.05**	**-181832**	**-16**	**147065**	**1.1**
Within person	**11.4**	**0.16**	**0.06**	**$0.38x10^6$**	**27**	**$0.82x10^6$**	**9.2**
Correlation coefficient	**-0.35**	**-0.37**	**-0.48**	**-0.42**	**-0.44**	0.06	0.04
Continuous time estimates of the within-person (WP) reduction in variability (Unconditional growth model compared to unconditional means model, re Equation 7-3)							
Change in WP variability	0.49	0.50	0.67	0.39	0.44	0.59	0.52
Continuous time model fit estimates[1]							
Log Likelihood	-1711	-561	-322	-4512	-1947	-4829	-1765
AIC	3435	1133	657	9036	3907	9671	3542
BIC	3460	1159	682	9062	3233	9696	3568

Bold indicates p≤0.1; * p<0.05; ** p<0.01; *** p<0.001; † p<0.2
Numbers shown are the values output by the statistical program
[1]The estimates for categorical time models are similar to those for continuous time. The continuous time estimates are shown for comparability with later tables.

in MD at that time point relative to baseline. The continuous time coefficient estimated the mean linear change for all participants from baseline to year 7. For PD, the average continuous change per year was estimated as 0.28% (an average *increase* over time from baseline to year 7), whilst at 6, 12, 24 and 36 months post-randomisation the average (unconditional) change in PD for the categorical time variable was estimated as a significant –2 to –3% *decrease* from baseline. PD also showed an average increase from baseline at 60, 72 and 84 months of 1.7%, 3.2% and 6.8%. These latter increases account for the slight but positive increase in annual PD per year for the coefficient for continuous time (0.28%).

Two RE were now included in the model: a random slope for each participant and a random intercept for each participant. The random intercept allowed each participant to have a different baseline PD from the grand mean (FE intercept). Similarly, the random slope allowed each participant to have their own growth curve which differed from that of the grand mean growth curve shown in red in Figure 7-10.

As described in section 7.4 Statistical Methods, the fitted intercept for the participant shown in Figure 7-10 (solid green line) was estimated by the model by adding a RE 'Best Unbiased Linear Prediction' (BLUP) for that participant to the Grand Mean (FE intercept). Similarly, the random slope for the participant was modelled by estimating a RE BLUP for each category of time which was added to the mean FE for each time category to derive a 'fitted' growth curve for the participant.

As per Table 7-5, change over time appeared to significantly decline from baseline to 24 months for PD and DA, but then increase through 84 months. The model of untransformed BA showed a non-monotonic increase in area from baseline of about 500 mm^2 for most follow ups. This equated to an annual increase of 106mm^2 for each year of the trial, in part driven by the modelled increase in BA of more than 500 mm^2 at the 60, 72 and 84 month follow ups.

A between person covariance was also fitted to the RE of the model. As described previously, this covariance quantified the population covariance between initial status (baseline PD) and change over time. PD and DA both showed significant negative covariances, which implied women with higher baseline PD tended to decrease in PD over time more than women with lower initial PD. As for covariance, PD and DA also showed a significant negative correlation over time. In contrast, the (more stable) BA had small (<0.7) non-significant positive correlations between initial BA and change over time.

As per Equation 7-3 in the Statistical Methods section, change in within person variability for PD% after time was added to the model was (17-11.4)/11.4 = 0.49. This meant about half of the unconditional within-person variability for PD was explained by the addition of time to the model. Within person variability reduction estimates for DA were slightly smaller than for PD: 0.39 and 0.44 for untransformed and square root transformed DA respectively. BA within person variability was also approximately reduced by half with the addition of time to the model.

The RE for change over time (time) and initial status (intercept) were significant for all of the MD parameters. This implied that participants differed from each other enough not only in average offset from the Grand Mean growth curve (as indicated by the significant variance estimate for the within person intercept), but their individual growth curves also differed from each other sufficiently to produce significant RE for change over time. Within person variability was also significant, implying further covariates could be added to explain within person variability. The statistics for model fit—log likelihood (LL), AIC and BIC— were smaller (closer to 0) for the unconditional growth model vs the unconditional means model. For example, the BIC decrease for the square root transformed PD unconditional means model to the unconditional growth model provided very strong (>10 units) evidence of an improvement in fit: the BIC decreased from 1208 to 1159, a 49 unit difference. This implied model fit was greatly improved with the addition of time to the model.

7.5.6 Modelling time

Initial models with one continuous parameter for time from years 0 to 7 were found to be inadequate for PD and DA. This was due to the within person drop in average PD from film to digital mammograms (e.g. PD trajectory decline from green dots (film) to blue dots (digital), Figure 7-5), as well as the increase in average PD over time which was related to change in digital mammogram software and hardware configuration (e.g. Figure 7-7, Figure 7-10). A single linear trajectory for continuous time was not able to accommodate the u-shape which often resulted from the film to digital transition and/or initial decline in breast density due to trial treatment. Use of categorical time rather than continuous time was unsatisfactory (not parsimonious) due to the number of categories (nine). Hence a linear spline with two or more segments that could change slope at one or more cut points (knots) was fitted to the model. Potential cut points for continuous time were initially determined by examining the magnitude of change in PD and DA at each categorical time point. Two times of best fit (1.0 and 1.25 years) maximised the slope of change in MD over time from baseline compared to other cut points between 0.5 and 2.5 years examined by modelling 0.25 year (3 month) increments.

The largest change in PD from baseline to a later time point was observed when a cut point at 1 year post-randomisation was utilised in the model. No further decreases in PD were found from 1.0 years to 5 years of follow up, however a slight increase in linear change over time was noted. Hence a second linear relationship between 1 and 5 years post-randomisation was modelled. The remaining time during years 5 to 7 was modelled with an additional linear segment, to reflect the theoretical rebound in PD and DA after the cease of anastrozole treatment at year 5 (Figure 7-1).

No evidence was found of a significant higher order effect of time (e.g. quadratic ($time^2$), cubic ($time^3$) relationship) in addition to the significant linear relationships modelled for the 0 to 1 year or 1 to 5 year time periods. Modelling of the change over time between 1.0 or 1.25 to 5

years post-randomisation in 0.5 year increments showed that a single segment from 1.0 (or 1.25) to five years was adequate. RE slopes at the level of person (i.e. between person RE) for the first two time segments were significant for models of PD and DA, and were retained in the model along with the RE between person constant. The random slope variance for the third segment (years 5 to 7) was not significant.

Models for BA showed that a single continuous growth curve (slope/line) from baseline to year 7 was sufficient, however BA was modelled with all three segments of continuous time for consistency with PD and DA.

7.5.7 Forward and backward covariate selection

Strong (6 to 10) or very strong (>10) changes in the BIC [493] during FE parameter addition/removal was noted for the covariates BMI, CC vs MLO (in three level models) and Version. Strong changes in the BIC were not observed for other covariates. Because of the small sample size, the BIC and AIC were sometimes at odds during stepwise comparison of models. The removal or addition of one covariate in the models often yielded differences in AIC greater than 2, but would change the BIC by less than 3. To create more parsimonious models, selection of covariates based on the BIC outcome was undertaken.

As stated previously, age at randomisation was retained in all final models, regardless of significance or coefficient sign, because age is a well-recognised, strong modifier of MD. Interestingly, removal of age at randomisation from most PD models increased AIC >2 (worse model), whilst decreasing BIC (improved model).

7.5.8 Covariance structure selection

Substitution of the exchangeable and unstructured covariance structures for the default structure (independent) resulted in non-convergence of the mixed model. The banded structure (a special

case of the unstructured covariance specification) was modelled with a single band of off-diagonal covariance estimates. The banded structure model converged, but the SE for the RE parameter estimates were not calculated. This may have occurred because the SE were too small (or too large) to be estimated by the statistical program. This problem showed that the banded covariance structure did not suit the model.

Of the models which converged properly, the BIC scores were similar for the AR, exchangeable, MA, and toeplitz structures (within ±3 of each other). Models with single covariance coefficients for the AR, MA, banded and toeplitz structures (e.g. toeplitz 1, first off-diagonal band covariance only) had lower BIC than those with higher numbers of (higher-order) covariance coefficients— e.g. toeplitz 2, which estimates both the first and second off diagonal bands in the variance-covariance matrix; all other off diagonal bands are set to 0.

The RE for time (the random slopes modelled for each participant) became non-significant when the AR, exponential and other covariance structures were incorporated into the model. Therefore time was omitted as a RE at the level of participant from the model. This meant that growth curves for all participants now paralleled the Grand Mean growth curve (represented by the coefficients for time in the FE portion of the model). Individual participants' MD trajectories were no longer allowed to differ from the Grand Mean growth curve. The lack of significance for the RE for time indicated that the (between person) MD trajectories for the participants did not differ substantially from one another; this potentially indicated the lack of a strong treatment effect between the treated and control groups, which might be detected as between person heterogeneity over time by the model.

Another potential reason the non-significance of the random slope time parameter was the frequent change from one mammographic Version to the next; these changes in mammographic Version may have caused differences in MD growth for participants which did not coincide with the cut points (change points) of the continuous spline for time. The indicator parameters

for each mammogram Version were used to model (account for) potential random effects of Version on the model at the level of participant. All possible different combinations of the four parameters were modelled as RE: all four simultaneously, different sets of three, different pairs, and individually. The KE52 and KE54 RE were consistently found to be not significant, but the RE slopes for film and Fuji Version were consistently significant, whether modelled individually or with other mammogram Version RE. Inclusion of the RE slopes for film and Fuji mammogram Version reduced the BIC score substantially (>35) despite the increase in complexity due to two additional parameters. Hence these RE slopes appeared to be quite beneficial in accounting for between person heterogeneity. Re-introduction of the between person RE for time to the model with the RE for film and Fuji Version did not achieve significance for the RE slope for time. Further simplification of the model by utilising a single random effect (i.e. a single indicator) for both film and Fuji mammograms increased the BIC by more than 3 (indicating worse fit). The final model therefore includes separate RE for Film and Fuji mammograms at the level of participant.

Table 7-6 Covariance structure comparison of naïve vs robust Standard Errors

Covariate	Ratio of default SE to robust SE, for different covariance structures				
	Independent	Auto Regressive 1	Exponential	Moving Average 1	Toeplitz 1
Age at Rand. (yrs)	1.23	1.24	1.23	1.24	1.24
BMI (kg/m^2)	1.09	1.09	1.09	1.09	1.09
Menopause (yrs)	1.04	1.04	1.04	1.04	1.04
Age First Birth <30y	--	--	--	--	--
≥ 30 years	**1.33⁺**	**1.35⁺**	**1.35⁺**	**1.35⁺**	**1.35⁺**
Non-parous	**1.57⁺**	**1.59⁺**	**1.59⁺**	**1.60⁺**	**1.60⁺**
Mammogram Version[5]- film	--	--	--	--	--
KE52	0.80	0.80	0.82	0.82	0.82
KE54	0.87	0.84	0.87	0.84	0.84
Fuji	0.92	0.88	0.91	0.89	0.89
Intercept	1.20	1.21	1.21	1.22	1.22
Baseline to Year 1	0.95	1.00	1.00	1.00	1.00
Years 1 to 5	1.04	1.06	1.07	1.05	1.05

⁺ Values outside of the acceptable range of 3/4 to 4/3

The ratio of the FE default SE to robust SE for each model was inspected to see if it exceeded the interval from 3/4 to 4/3 [494], Table 7-6. With the exception of the parameter which had small cell sizes in two categories (age at first birth ≥30 years and non-parous women), the naïve

SE to robust SE ratios of the FE for the AR, exponential, MA, toeplitz as well as the default covariance structure (independent) were within the 3/4 to 4/3 limits. The SE ratio for age at first birth ≥ 30 years was 1.33 to 1.35, and 1.57 to 1.59 for age at first birth = non-parous. The FE SE ratios for other covariates for all the covariance structures were also quite similar to each other (e.g. 1.09 for BMI).

The toeplitz, MA and AR structures assume equal spacing between all occasions (follow ups). Figure 7-4**Error! Reference source not found.** showed fairly regular spacing for most but not all participants. The exponential covariance is a generalisation of the AR model which allows for unequal spacing of occasions
as well as non-integer values for time. The exponential structure consistently had non-significantly but slightly higher BIC scores (~2) than the BIC scores for AR(1) models. Due to irregular spacing of some mammographic episodes as well as gaps in follow ups, the exponential variance-covariance structure was selected as the preferred residual structure for the data.

As described in section 7.4.6 Covariance structure selection, forward selection for significant covariates was repeated using ML estimation. The set of parsimonious covariates remained the same before and after covariance structure testing for square root transformed PD and DA, however square root transformed BA gained three significant covariates in addition to BMI: age at menarche, AFB and smoking status.

VIF testing of 'first mammogram' multivariable parsimonious square root transformed PD and DA regression models did not find evidence of collinearity between age at randomisation and age at menopause, nor between any of the model covariates. Removal (separately) of the two age related parameters (age at randomisation and age at first birth) from (non-MI) parsimonious square root transformed PD model yielded models similar to the (non-MI) original model with both age related parameters; coefficient values and SE did not greatly differ between these

models except for the age at first birth parameter due to the small cell sizes of two of its categories.

Differences in the BIC, AIC and LL, however, were noted between the PD and DA models fitted with only one of the age related covariates, as well as these models and the original models fitted with both age related covariates. The BIC was lowest (not significantly, by 1 unit) for the PD model which retained age at menopause but not age at randomisation; log likelihood and AIC values were lowest (in absolute value) for the original model with both coefficients. The AIC was lower significantly (–3 units) for the original PD model compared the model with age at menopause alone. The PD model with age at randomisation alone showed significantly worse BIC values (+8) and AIC values (+11) compared to the original parsimonious model with both covariates. A slightly different trend was noted for (non-MI) square root transformed models of DA. Whilst the coefficients did not change very much between the models, BIC was significantly lower (–4) for model with age at menopause only compared to the original model with both age related covariates but the AIC and LL were almost identical. The model with age at randomisation alone had significantly higher BIC (+8), AIC (+12) and larger magnitude LL (+7) than the original (non-MI) parsimonious square root transformed DA model. Age at randomisation was significant for the PD square root transformed models, but not for the DA models. Although an a priori decision was made to retain age (at randomisation) in all models of PD and DA, the retention of both age related may not have been necessary, especially for the DA model; use of the Deviance statistic (–2LL), the method favoured by Singer and Willet [490], may have indicated age at menopause alone provided the most parsimonious model with adequate fit.

7.5.9 Full and parsimonious models of PD, DA and BA

The ‘full’ models contain a mix of significant ($p<0.05$) and non-significant parameters, Table 7-7. Age at randomisation, BMI, age at menopause and age at first birth ≥ 30 years had p-

Table 7-7 Longitudinal MD full and parsimonious multivariable mixed linear model regression coefficients

Covariate	PD %		PD square root[a]		DA mm^2		DA square root[a]		BA mm^2		BA square root[a]	
	Full	Parsimon	Full	Parsimon	Full	Parsimon	Full	Parsimon	Full	Parsimon	Full	Parsimon
ALL MAMMOGRAMS 120 participants 540 episodes (follow ups)												
Age at Rand. (yrs)	**-0.29**	**-0.26**	**-0.05***	**-0.04****	-44†	-31	**-0.58**	-0.4‡	118	--	0.5†	--
BMI (kg/m^2)	**-0.81*****	**-0.82*****	**-0.11*****	**-0.12*****	-18	--	-0.34	--	**990*****	**965*****	**3.6*****	**3.4*****
Menarche (yrs)	0.23	--	0.03	--	113†	--	1.0†	--	471†	**500**	1.8‡	**2.0**
Menopause (yrs)	**0.38****	**0.36****	**0.06****	**0.06****	**81****	**75****	**0.92****	**0.85****	44	--	0.2	--
Age First Birth <30y	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
≥ 30 years	**8.2***	**8.3***	**1.0****	**1.1****	**2659****	**2603****	**22*****	**22****	**4621***	**4376***	**15***	**14***
Non-parous	-1.3	-0.97	-0.10	-0.05	398	492†	3.8	5.4	**3580**	3252‡	13	12
OC use - Never	ref.	--	ref.	--	ref.	--	ref.	--	ref.	--	ref.	--
Ever users	-2.0	--	-0.32	--	-563	--	-5.7	--	243	--	--	--
HRT – Never	ref.	--	ref.	--	ref.	--	ref.	--	ref.	--	ref.	--
Ever users	0.03	--	0.05	--	-85	--	-0.6	--	-946	--	-4	--
Smoking- never	ref.	--	ref.	--	ref.	--	ref.	--	ref.	ref.	ref.	ref.
Current	1.2	--	0.27	--	**653**	--	**7.9***	--	**4150***	**3769***	**14***	**13***
Ex-smoker	0.64	--	0.12	--	-17	--	0.6	--	-727	-862	-2.1	-2.8
IBIS-1 – No	ref.	--	ref.	--	ref.	--	ref.	--	ref.	ref.	ref.	ref.
Yes	0.49	--	0.19	--	162	--	2.6	--	-568	--	-2.2	--
Mammogram Version- film	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
KE52	**-6.6*****	**-6.6*****	**-0.76*****	**-0.76*****	**-1233*****	**-1234*****	**-10*****	**-10*****	-62†	-63‡	-0.11†	-0.12†
KE54	**-4.5*****	**-4.5*****	**-0.48*****	**-0.48*****	**-859*****	**-864*****	**-6.6*****	**-6.6*****	**38**	**36**	**0.25**	**0.24**
Fuji	**-2.7***	**-2.7***	**-0.29**	**-0.29**	**-638***	**-648***	**-4.5***	**-4.6***	**-267**	**-269**	**-0.82**	**-0.84**
Intercept	**31*****	**27*****	**5.7*****	**5.2*****	**5290*****	**3967*****	**74*****	**60*****	**13929*****	**13996*****	**119*****	**118*****
Annual change in MD (All mammograms)												
Baseline to Year 1	**-1.06****	**-1.06****	**-0.13****	**-0.13****	**-196****	**-194****	**-1.6****	**-1.6****	194	197	0.66	0.68
Years 1 to 5	**0.36***	**0.37***	**0.04**	**0.04**	**98***	**100***	**0.79***	**0.81***	144	145	0.48	0.49
Years 5 to 7	-0.17	-0.17	-0.02	-0.02	-4.32	-3.78	0.05	0.05	312	311	1.09	1.09

Covariate	PD %		PD square root[a]		DA mm^2		DA square root[a]		BA mm^2		BA square root[a]	
	Full	Parsimon	Full	Parsimon	Full	Parsimon	Full	Parsimon	Full	Parsimon	Full	Parsimon
Random effects – estimates for between- and within-person change												
Between person variance												
Film mmgs	**5.4**	**5.4**	**0.5**	**0.5**	**1023**	**1027**	**7.0**	**7.1**	**1607**	**1613**	**4.8**	**4.9**
Fuji mmgs	**2.7**	**2.7**	**0.4**	**0.4**	**524**	**523**	**5.2**	**5.2**	**749**	**753**	**2.5**	**2.5**
Intercept	**9.4**	**9.4**	**1.1**	**1.1**	**1739**	**1773**	**16.1**	**16.6**	**4523**	**4595**	**16.3**	**16.6**
Within person correlation (rho)	**0.6**	**0.6**	**0.5**	**0.5**	**0.3**	**0.3**	**0.25**	**0.25**	**0.6**	**0.6**	**0.6**	**0.6**
Within person variance	**2.8**	**2.8**	**0.3**	**0.3**	**410**	**410**	**3.5**	**3.5**	**1166**	**1158**	**4.1**	**4.0**
Statistics of model fit[b]												
Log–likelihood	-1512.1	-1512.5	-374.6	-375.8	-4331.2	-4333.6	-1780.7	-1784.6	-4746.5	-4748.1	-1688.7	-1691.0
AIC	3070.3	3059.0	795.3	785.5	8708.4	8699.1	3607.5	3601.2	9538.9	9532.3	3423.5	3418.0
BIC	3169.0	3132.0	894.0	858.5	8807.1	8767.8	3706.2	3669.9	9637.7	9609.5	3522.2	3495.3
FILM mammograms only: Annual change in MD 63 episodes , 45 participants												
Baseline to Year 1	**-1.8***	**-1.8***	**-0.2***	**-0.2***	**-313***	**-315***	**-2.2**	**-2.2**	**---[1]**	281[2]	**---[2]**	**1.5[3]**
DIGITAL mammograms only: Annual change in MD 434 episodes, 120 participants												
Baseline to Year 1	**-0.72**	**-0.72**	**-0.09***	**-0.10***	**-153***	**-149***	**-1.3***	**-1.5**	136	139	0.47	0.49
Years 1 to 5	**0.52****	**0.52****	**0.07****	**0.07****	**115****	**119****	**1.0****	**1.1**	123‡	124‡	0.39†	0.39‡
Years 5 to 7	-0.06	-0.06	-0.004	-0.005	7.2	8.0	0.18	0.17	281†	282†	0.97‡	0.97†

Bold p≤0.1; * p<0.05; ** p<0.01; *** p<0.001; ‡ p<0.15; † p<0.2;

Parsimon. parsimonious model; y yrs years; ref. reference category; -- not applicable, p>0.1 parameter removed during stepwise selection

[a] Coefficient values of square root transformed MD variables have not been back–transformed to original units, e.g. DA square root coefficients' scale is in mm (not mm^2).

[b] The LL, AIC, BIC statistics of model fit were generated by models which utilise an average of the imputed value for menopause age

[1] BA full model would not converge; BA parsimonious model converged with BMI menarche only in the model – rho very small (0.000025)

[2] Square root transformed BA full model would not converge; Sqrt BA parsimonious model converged with removal of all covariates except BMI and baseline–Year 1 time.

Table 7-8 Coefficients for parsimonious untransformed and back transformed models of PD, DA, BA

Covariate	PD % (from Table 7-7)[1]	PD % Back trans-formed sqrt	DA mm^2 (from Table 7-7)[1]	DA Back trans-formed sqrt	BA mm^2 (from Table 7-7)[1]	BA mm^2 Back trans-formed sqrt
ALL MAMMOGRAMS 120 participants 540 episodes (follow ups)						
Age at Rand. (yrs)	**-0.26**	**-0.41****	-31	-48	--	--
BMI (kg/m^2)	**-0.82*****	**-1.23*****	--	--	**965*****	**814**
Menarche (yrs)	--	--	--	--	**500**	**476**
Menopause (yrs)	**0.36****	**0.63****	**75****	**103****	--	--
Age First Birth <30y	ref.	ref.	ref.	ref.	ref.	ref.
≥ 30 years	**8.3***	**12.65****	**2603****	**3124*****	**4376***	**3500**
Non-parous	-0.97	-0.52	492†	677	3252‡	2976
Smoking- never	--	--	--	--	ref.	--
Current	--	--	--	--	**3769***	**3237**
Ex-smoker	--	--	--	--	-862	-653
Mammogram Version- film	ref.	ref.	ref.	ref.	ref.	ref.
KE52	**-6.6*****	**-7.33*****	**-1234*****	**-1100*****	**-63‡**	**-28†**
KE54	**-4.5*****	**-4.76*****	**-864*****	**-748*****	**36**	**57**
Fuji	**-2.7***	**-2.93**	**-648***	**-531***	**-269**	**-198**
Intercept	**27*****	**27*****	**3967*****	**3600*****	**13996*****	**13924*****
Annual change in MD (All mammograms)						
Baseline to Year 1	**-1.06****	**-1.34****	**-194****	**-189****	197	161
Years 1 to 5	**0.37***	**0.42**	**100***	**98***	145	116
Years 5 to 7	-0.17	-0.21	-3.78	-6.0	311	258
FILM mammograms only: Annual change in MD 63 episodes , 45 participants						
Baseline to Year 1	**-1.8***	**-2.2***	**-315***	**-259***	281[2]	**363**
DIGITAL mammograms only: Annual change in MD 434 episodes, 120 participants						
Baseline to Year 1	**-0.72**	**-0.87***	**-149***	**-146**	139	116
Years 1 to 5	**0.52****	**0.63****	**119****	**110**	124‡	92‡
Years 5 to 7	-0.06	-0.04	8.0	17	282†	230†

The PD and DA back transformed values are representative of CMN IBIS-II participants randomised at age 50, BMI 25 kg/m^2, who underwent menopause at age 50, first gave birth under age 30.
Bold indicates p≤0.1; * p<0.05; ** p<0.01; *** p<0.001; ‡ p<0.15; † p<0.2; y yrs years;
ref. reference category; -- not applicable (p>0.1 parameter removed during stepwise selection)
[1] The first column of the paired PD, DA and BA model coefficients presents the coefficient values from the untransformed parsimonious models tabulated in Table 7-7; the second column of each pair presents back-transformed values from the square root models for PD, DA and BA from Table 7-7
[2] Sqrt BA model converged with removal of all covariates except BMI and baseline–Year 1 time.

values ≤0.1 in the PD models (full and parsimonious); the latter three covariates were significant (p<0.05). Age at menopause and age at first birth ≥ 30 years were significant (p<0.05) covariates for the DA models (full and parsimonious). Current smoking status had a p-value ≤0.1 in the DA full models only. PD and DA were also significantly modified by the 'technical' mammographic parameter mammogram Version in all models; the effect of Version on BA was less consistent (non-significant p-values). BMI, age at first birth ≥30 years and current smoking

status were significantly associated with BA in all models, whilst age at menarche was retained in parsimonious models of BA because the p-value for this covariate was <0.1.

All PD and DA models estimated a decrease in the rate of annual change for PD and DA between baseline to year 1, Table 7-7. The PD and DA all mammograms models first year annual decline was significant ($p<0.01$), as was the first year rate of change for the PD film only and square root transformed PD digital only models ($p<0.05$). The film only and digital only first year decline in DA annual change was only significant for untransformed DA; the film only and digital only DA parsimonious square root transformed model baseline to year 1 rate of change was not significant ($0.05< p \leq 0.1$).

However, for years 1 to 5 for the all mammogram and digital only models a significant ($p<0.05$) increase in annual change for PD and DA is seen. (Film mammograms were not taken during years 1 to 7, therefore coefficients for the film only models from year 1 onwards are not available.) A significant annual change in PD and DA is not estimated for years 5 to 7 for any model. The rate of annual change in BA is not significant ($p\geq0.05$) for all models—for any time period, indicating the BA is relatively stable across time (and therefore across all mammogram Versions).

Back-transformed parsimonious model coefficients are presented in Table 7-8. The back transformed coefficient values for PD, DA and BA (second column for each pair of models for PD, DA and BA, Table 7-8) were calculated as per Equation 7-1 using the square of the intercept for each square root transformed MD model from Table 7-7. Because the square of the intercept values for PD (27%) and DA (3600 mm^2) are higher than median PD and DA for baseline mammograms in this data set (PD 18%, DA 2800mm^2, Table 6-6 in Chapter 6), the back transformed coefficient values calculated for the square root transformed PD DA models are greater in magnitude (absolute value) than would be if the median PD and DA values for CMN IBIS-II participants were utilised as the reference (intercept) values.

Back transformed annual rates of change for PD and DA for a range of 'intercept' reference values are shown in Table 7-9, for the MD change over time coefficients of the square root transformed parsimonious models (Table 7-7) for all mammograms, film only, and digital only mammogram models of PD and DA. The absolute value (magnitude) for the rate of annual MD change increases as the value of the PD and DA reference value ('intercept') increases. Because the 20% PD and 2700 mm^2 DA reference values are closer to the median PD and DA values for CMN participants (PD 18%, DA 2800mm^2, Table 6-6 in Chapter 6), the MD annual rates of change calculated for these reference values, Table 7-9, are likely more representative of the average annual MD change for the 120 CMN IBIS-II participants.

Table 7-9 Back transformed PD, DA coefficients calculated for a range of 'intercept' reference values

Covariate	Percent Density estimates			Dense Area estimates		
	15% PD	20% PD	25% PD	2000 mm^2 (20cm^2)	2700 mm^2 (27cm^2)	3500 mm^2 (35cm^2)
Intercept	15	20	25	2000	2700	3500
Annual change in MD						
ALL MAMMOGRAMS 120 participants 540 episodes (follow ups)						
Baseline to Year 1	-1.01	-1.16	-1.28	-142	-164	-182
Years 1 to 5	0.29	0.34	0.40	71	85	101
Years 5 to 7	-0.18	-0.20	-0.20	3	5	11
FILM mammograms only 63 episodes , 45 participants						
Baseline to Year 1	-1.53	-1.77	-1.96	-194	-224	-251
DIGITAL mammograms only 434 episodes, 120 participants						
Baseline to Year 1	-0.79	-0.90	-0.99	-134	-154	-171
Years 1 to 5	0.52	0.61	0.70	98	115	136
Years 5 to 7	-0.06	-0.06	-0.05	13	18	25

The back transformed values are representative of CMN IBIS-II participants randomised at age 50, BMI 25 kg/m^2, who underwent menopause at age 50, and first gave birth under age 30.

Comparison of the univariable ('simple') mixed regression coefficients (not tabulated) to their full and parsimonious multivariable counterparts showed they were generally similar (±10%) for significant covariates (e.g. BMI, age at menopause). However univariable non-significant covariates were often more than 10% different than their multivariable counterparts in the full models. For PD models, age at randomisation showed evidence of confounding between univariable and multivariable models because the coefficients differed by more than one SE. However, all multivariable model coefficients including age at randomisation still fell within the

95% CI of their univariable model counterparts, therefore no substantial collinearity issues appeared to present in the parsimonious models.

7.5.10 Post estimation plots for PD and DA parsimonious models

Figure 7-11 plots the estimated fitted values (RE + FE) of the parsimonious untransformed PD (%) model growth curves (Table 7-7), overlaid on a scatter plot of the fitted values for the parsimonious model for untransformed PD (%). Comparisons of the fitted scattered PD data in Figure 7-11 with the observed scattered PD data in Figure 7-4 demonstrates the close emulation of the observed PD values by the mixed model —even for the model of untransformed PD which does not meet all regression model assumptions (re: skewed residuals, Figure 6-5 in Chapter 6, and fitted values with a 'floor' upper row of Figure 7-12, below).

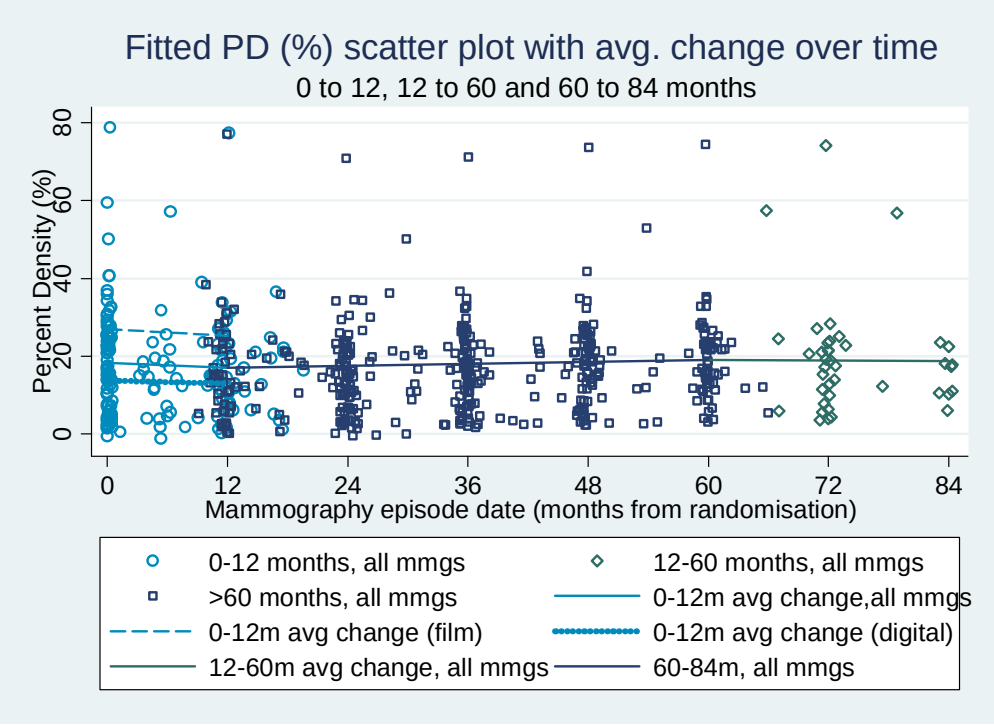


Figure 7-11 Fitted values for PD (%) with estimated growth curves for 0-12, 12-60 and 60-84 months

The fitted growth curves for all mammograms, and for each Type (film, digital) in Figure 7-11 display a small decrease in PD for the period of 0 to 12 months (baseline to 1 year). The growth curve subsequently rises slightly over the period from 12 to 60 months (years 1 to 5),

and then declines again slightly from 60 to 84 months (years 5 to 7).

The model coefficients and SE as well as graphs of the linear predictions were compared for the MI vs non-MI PD models (data not shown) and few differences were found. The coefficients of the MI and non-MI models were all well within ±2 SE of each other. A comparison of the MI vs non-MI linear (FE) predictions for parsimonious models of PD showed few visual differences, Figure 7-12, although approximately one-fifth of the values for age at menopause (116 episodes from 27 participants) have changed (slightly). The MI graphs (left column Figure 7-12) and the non-MI graphs (right column, Figure 7-12) appeared quite similar, for both untransformed PD (upper row) and square root transformed PD (bottom row). The graphs for

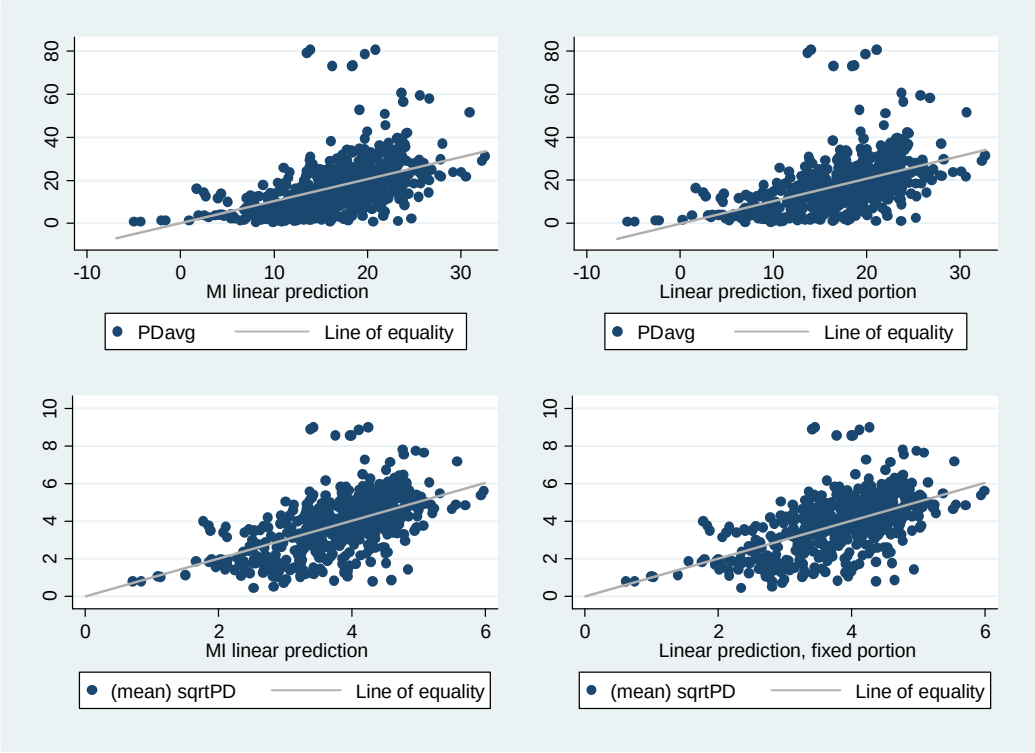


Figure 7-12 MI and non-MI linear predictions vs measured values for PD (%) and sqrt PD
The plots for the multiply imputed models (left column) look very similar to their non-MI counterparts in the right column. Untransformed PD (upper row) shows a floor at 0%, whilst square root transformed PD (sqrtPD, bottom row) is more symmetric. The lines of equality reveal the models do not work as well for women with very high PD (outliers at 50 to 80% (upper graphs) and high values of transformed PD (values of 6 to 8 on the lower row graphs)). Predicted PD (x-axis) appears to be higher than observed PD (y-axis) for some women with observed square root transformed PD values of 0 to 2 (y-axis, lower row)

transformed PD show better fit than the untransformed PD graphs because they do not show a 'floor' effect. Because the linear prediction plots for untransformed PD (Figure 7-12, upper row) and untransformed DA (graph not shown) both showed evidence of a floor effect, subsequent post estimation checks are primarily shown only for square root transformed models of PD and DA.

The residuals of the linear (FE) prediction for the parsimonious square root transformed PD model are quite similar for the non-MI model (upper plot) vs MI model (lower plot), Figure 7-13. These plots are very similar to the square root transformed PD plots in the lowest row of Figure 7-12; the scatter plot in Figure 7-12 has effectively been rotated to create Figure 7-13. Because the MI and non-MI coefficient values were similar, and substantial differences were not detected for the MI vs non-MI predicted values the non-MI models were judged as adequate representations of the MI models. Hence subsequent post-estimation plots were created using non-imputed data which utilised an average of the 25 MI imputed values for age at menopause for 27 participants with unknown age at menopause. This allowed full access to the range of

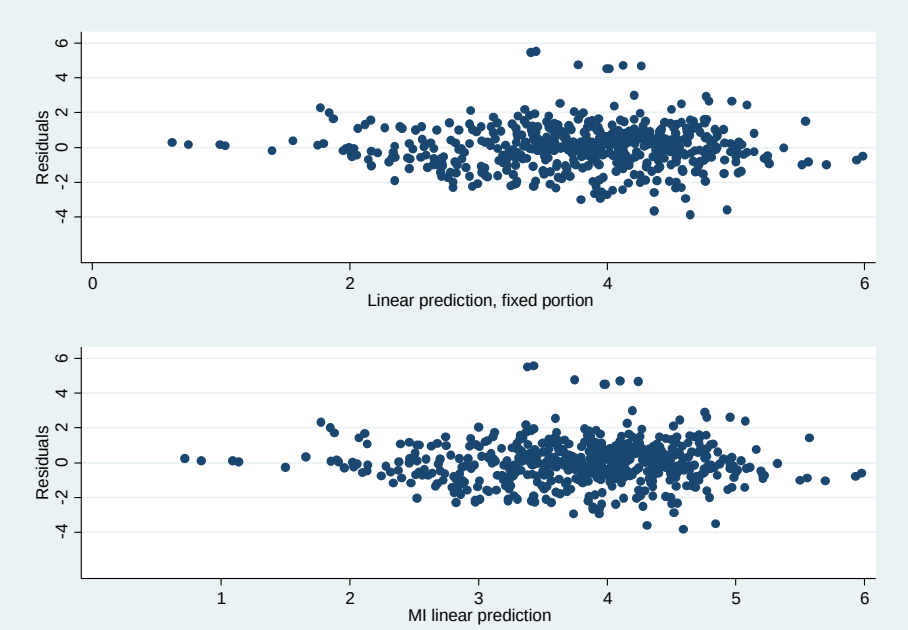


Figure 7-13 Similar residual plots for non-MI (upper) & MI (lower) linear predictions, transformed PD

post-estimation commands for mixed models in Stata, particularly for fitted values of the models (i.e. predictions for both the FE and RE parts of the mixed models).

More so than histograms of the untransformed PD best linear unbiased predictions and residuals (not shown), the residuals BLUPs for each RE level of the (non-MI) parsimonious model of square root transformed PD model showed reasonably normal distributions, Figure 7-14. They are approximately symmetric and normally distributed. As noted in the figure caption, the large peak at 0 for the film mammogram RE constant (BLUP RE for id: mmgv7) is due to the binary indicator used to model the film mammogram RE slope. A film RE is modelled only for episodes with film mammograms, else the value for the film RE is 0 (for digital episodes).

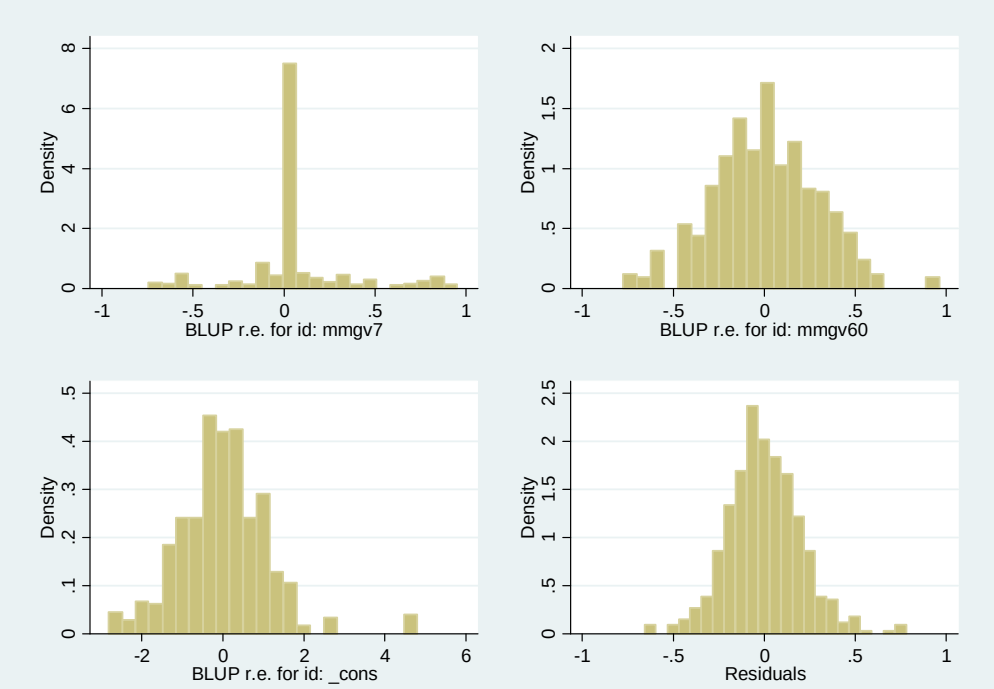


Figure 7-14 Square root transformed PD mixed model RE residuals
BLUP Best Linear Unbiased Prediction; mmgv = mammogram Version (7=film, 60=Fuji)
Between person RE residuals: upper left film; upper right Fuji; lower left RE constant; Within person RE residuals: lower right. The large peak at 0 for the film residuals is due to few film mammograms in the dataset. The other residuals are symmetric and approximately normally distributed. The right-tail of the residual plots is due to a few women with high PD (≥50%).

Further normality assumptions of the square root transformed PD model BLUP and residual plots were examined graphically with a method recommended by the Singer and Willet text

[490], Figure 7-15. The graphs in Figure 7-15 plot the same residual data shown in Figure 7-14, but as normal probability plots (left column) and as scatter plots of the standardised residuals vs participant id (right column). Again, the normal probability plot for the film mammogram intercept residuals for the between person RE are enriched for values of 0 (due to few film mammograms in the dataset), otherwise no particular issues with non-normality of the data are noted.

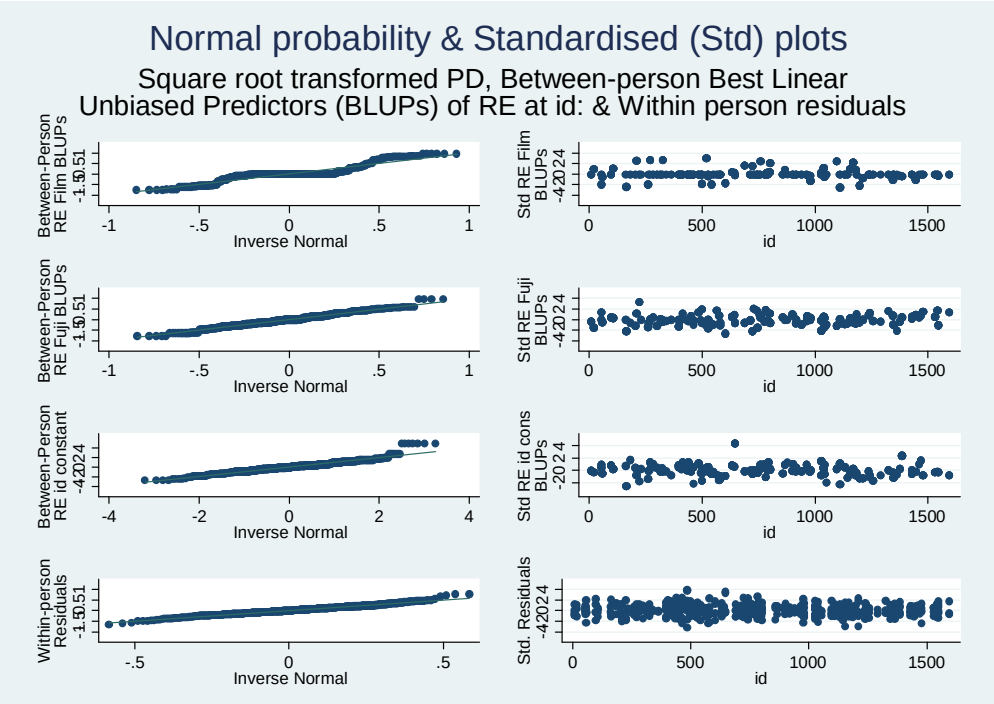


Figure 7-15 Examining normality assumptions, all mammograms model of square root transformed PD
The figure's left column presents normal probability plots for the random-effect residuals; the right column presents standardised residuals for each participant (by id). The top pair and second row of plots depict the between–person Film and Fuji residuals respectively, the third row pair of plots depict the between–person RE constant. The bottom pair of graphs plots the within–person residuals. As for the residual histograms, enrichment at 0 is seen for the Film mammogram residuals, with a less marked effect for the Fuji mammogram residuals. The outliers at value of 4 in the right column are due to the participant with very high PD (~80%). The residuals for all RE are otherwise normally distributed.

The BLUPs and residuals plots for the mixed model of untransformed PD (%), Figure 7-16, are very similar to those for square root transformed PD, Figure 7-15. However, square root transformation has constrained more of the standardised residuals to fall within the desirable limits of ± 2 to 3 (right column, Figure 7-15) and only a few of the standardised residuals have a value of 4 or more (outliers from participants with high PD). The high PD of some participants has caused outliers of with values at approximately 6 on the standardised residual plots of

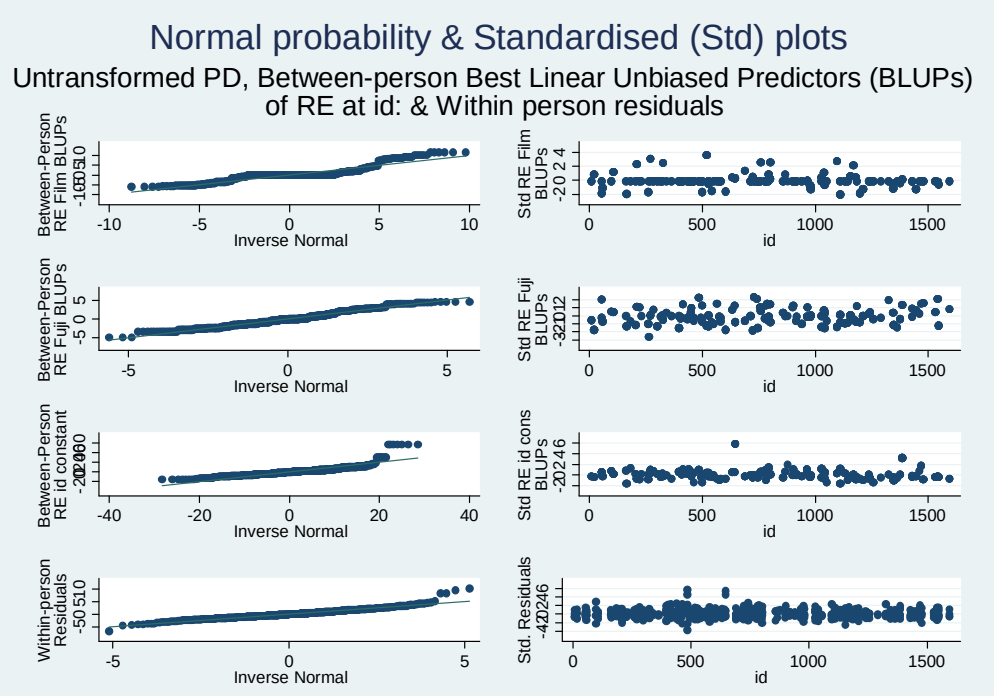


Figure 7-16 Examining RE normality assumptions, all mammograms model of untransformed PD (%)

untransformed PD (right column, Figure 7-16). The diagnostic evidence from this chapter as well as the last chapter (Chapter 6, baseline MD analysis) supports use of square root transformed PD instead of untransformed PD in the mixed linear regression models.

Graphs of the measured, linear (FE) prediction and fitted (FE + RE) model predictions for square root transformed PD were made from randomly selected participants for whom five or more mammographic episodes were present in the dataset: Figure 7-17, Figure 7-18 and Figure 7-19. Four participants each were selected from the following ranges of PD: low PD (<20%), mid-range PD (20 to 30%) and higher PD (>30%). In general, participants with lower PD have linear and fitted values which are closer to their measured (observed) values for PD (Figure 7-17 and Figure 7-18) than for participants with higher PD (Figure 7-19). The linear prediction for each participant (in blue) mirrors the growth curve of the measured values for PD (in red) for both low, mid-range and high PD participants.

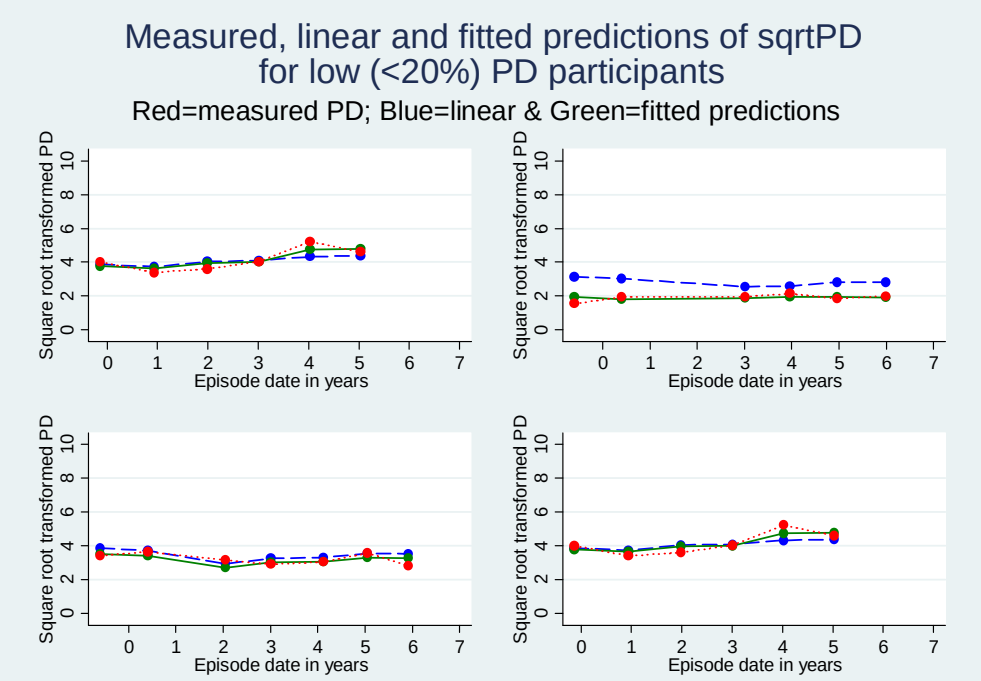


Figure 7-17 Observed, linear and fitted predictions for low PD (<20%) participants
Square root transformed PD fitted (FE+RE) predictions for participants (green) with linear prediction line (blue) and original measured PD (square root transformed) in red.

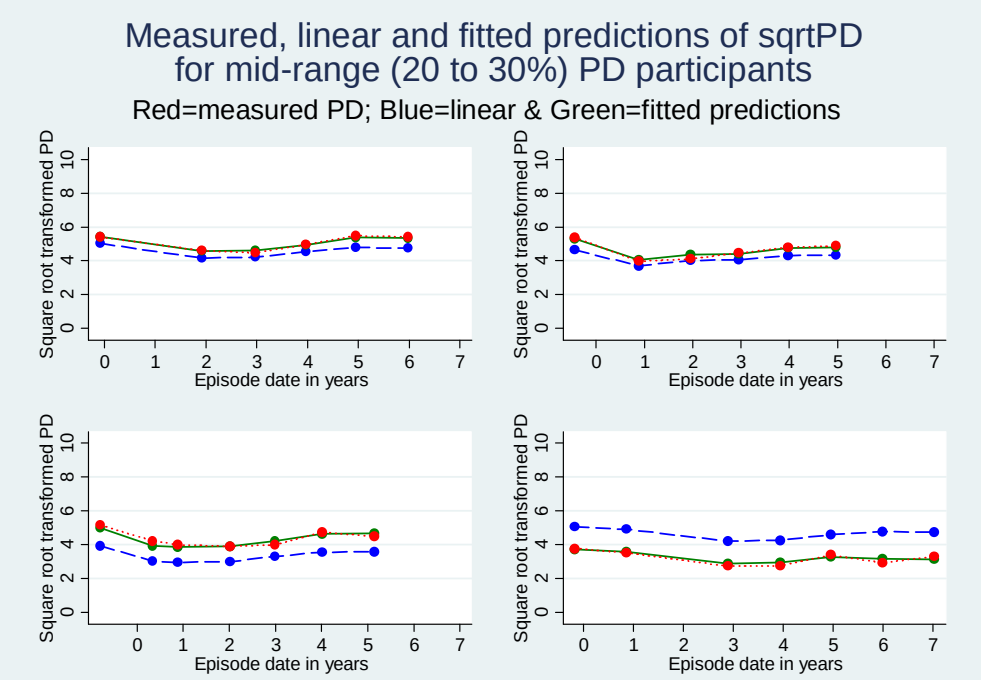


Figure 7-18 Observed, linear and fitted predictions for mid-range PD (20 to 30%) participants
Square root transformed PD fitted (FE+RE) predictions for participants (green) with linear prediction line (blue) and original measured PD (square root transformed) in red.

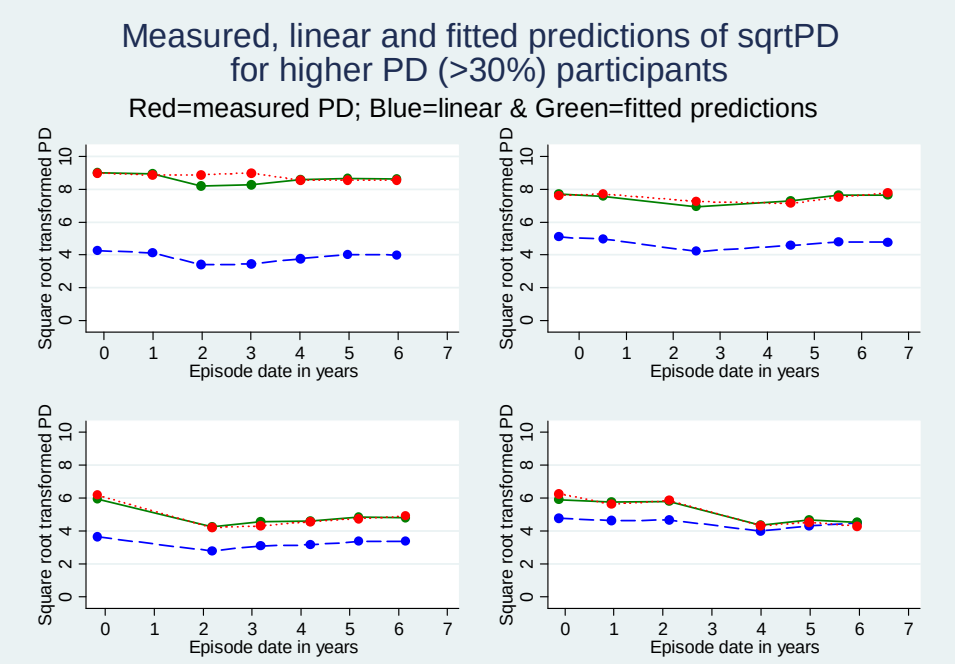


Figure 7-19 Observed, linear and fitted predictions for higher PD (>30%) participants
Square root transformed PD fitted (FE+RE) predictions for participants (green) with linear prediction line (blue) and original measured PD (square root transformed) in red.

The RE residuals of the parsimonious model for square root transformed DA also had approximately symmetric and normally-distributed residuals, Figure 7-20. The few film mammograms in the dataset are again reflected in a large peak at 0 for the film mammogram RE residual (mmgv7).

As for PD, further normality assumptions of the square root transformed DA model residual plots revealed few unexpected departures from normality, Figure 7-21. Histograms, normal probability plots, and standardised scatter plots for the BLUPs and residuals from the untransformed parsimonious model of DA (not shown) showed reasonably normal distributions. As for the square root transformation of PD, the diagnostic plots from the parsimonious model of square root transformed DA showed fewer outliers (e.g. fewer standardised residual values > 3) than did the graphs for the model of untransformed DA. Hence the mixed models of square root transformed PD and DA were both selected for use in the Primary Aim (Chapter 8, unblinded analysis of treated vs control IBIS-II participants), as well as for the subgroup, sensitivity and exploratory analysis for Aim 4 (this chapter, next section).

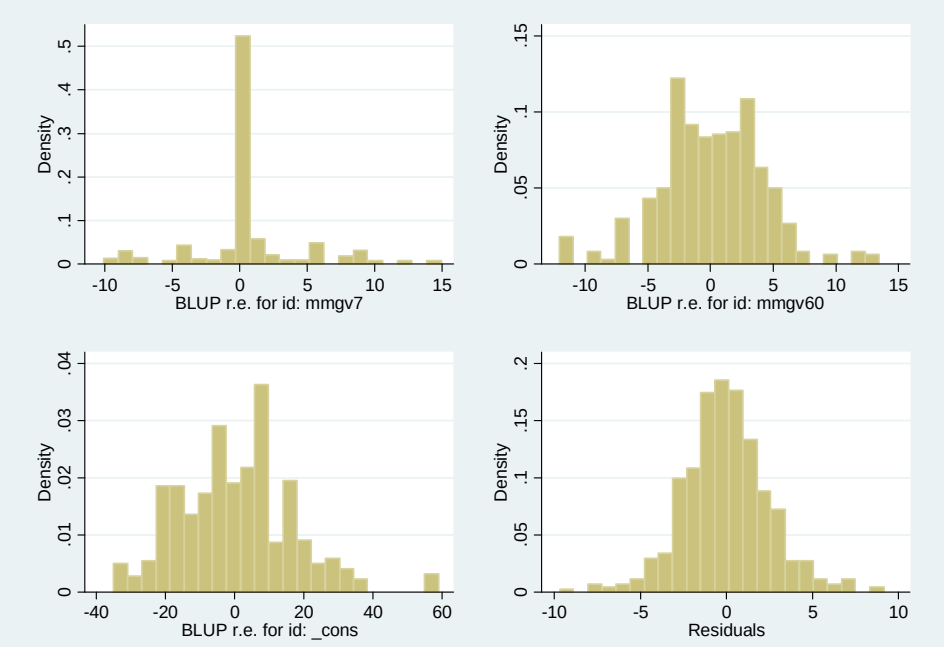


Figure 7-20 Square root transformed DA mixed model RE residuals

Between-person RE residuals: upper left mmg Version (film); upper right mmg Version (Fuji); lower left RE constant; Lower right: Within–person RE residuals. The large peak at 0 for the Film mmg residuals is due to few film mammograms in the dataset. The other residuals are symmetric and approximately normally distributed. The right-tail of the residual plots is due to a few women with high PD (≥50%).

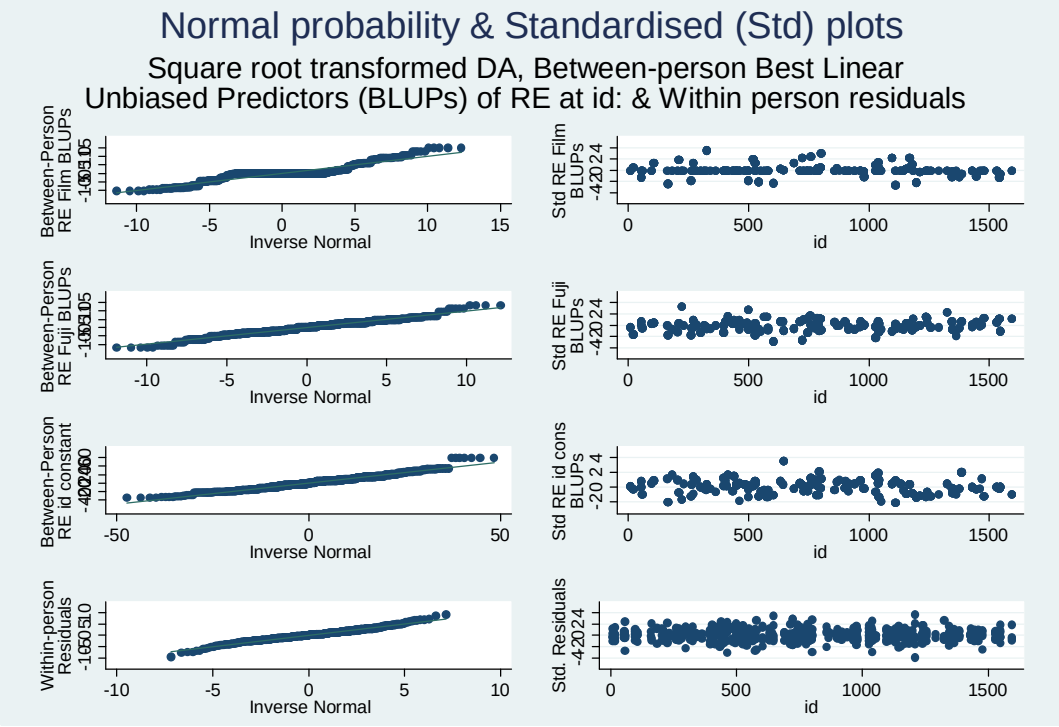


Figure 7-21 Examining normality assumptions, all mammograms model of square root transformed DA

The left column of the figure presents normal probability plots for the transformed DA random-effect residuals; the right column presents standardised residuals for each participant (by id). The top pair and second row of plots depict the between–person film and Fuji RE residuals respectively, the third row pair of plots depict the between–person RE constant. The bottom pair of graphs plots the within–person

residuals. As for the residual histograms, enrichment at 0 is seen for the Film mammogram residuals, with a less marked effect for the Fuji mammogram residuals. The outliers at value of 4 in the right column are due to the participant with very high PD (~80%). The residuals for all RE are otherwise normally distributed.

7.5.11 Subgroup, sensitivity and exploratory analyses

No unusual coefficient changes were noted for models with and without participants with high PD (≥50%), a participant with unusually long duration of HRT compared to others in the dataset (30 years), and a participant with an unusually long duration of oral contraceptives (>40 years). The PD model fitted with 93 participants (424 episodes) for whom age at menopause was known did not differ substantially from the multiply imputed model with all 120 participants and 540 episodes. The coefficient for age at menopause was 0.0746 (SE 0.0178, $p<0.001$) for the subgroup of 93 participants with known age at menopause compared to 0.0557 (SE 0.0193, $p=0.004$) for the multiply imputed model with all 120 participants including 27 participants for whom age at natural menopause was imputed. Only one other coefficient differed by >1 SE between these models: first year PD change was –0.10 ($p=0.018$, SE 0.04) for the subgroup of 93 participants compared to –0.13 ($p=0.002$, SE 0.04) for the imputed model with all 120 participants.

Substitution of the MI covariate for age at menopause with the 'as reported' age at menopause or hysterectomy from the IBIS-II database yielded (non-MI) PD model coefficients similar to the MI model; the coefficient for (rounded) first year annual change in the 'as reported' age at menopause or hysterectomy model was identical (–0.13, SE 0.04, $p=0.002$) to that of the MI model, although the coefficient for age at menopause differed slightly (0.05, SE0.02, $p=0.004$). The (rounded) coefficients, SE and p-values were also identical for PD annual change for years 1 to 5 and years 5 to 7. The most marked difference was found in the RE variance estimates, which were lower in the 'as reported' age at menopause or hysterectomy model for film and Fuji mammograms (0.3 and 0.1, respectively, compared to 0.5 and 0.4 respectively in the MI model) whilst the estimate for the between person variance intercept was slightly higher (1.3, vs

1.1 in the MI model). Estimated within person variance was also lower in the 'as reported' age at menopause or hysterectomy model (0.09, vs 0.3 in the MI model). Rounded RE variance estimates for the non-MI model (with an average age at menopause for the 27 participants) were identical to the RE estimates of the 'as reported' age at menopause or hysterectomy model. Comparison of the three different parameterizations for age at hysterectomy (MI, non-MI, and 'as reported' age at menopause or hysterectomy) imply that (not surprisingly) the MI introduced extra variability into the model, as reflected in the higher within person and film and Fuji RE variance estimates. Some of the between person variability in the non-MI and 'as reported' models has likely been shifted to the RE within person portion of the model as a result of the MI performed for some participants.

Differences in randomisation age under 60 vs over 60 years of age were associated with differences in change in PD during the first year of trial treatment, Table 7-10. A greater PD change in the first year of treatment was associated with younger women (≤60 years at randomisation) for all, film-only and digital-only mammograms. The p-values for these rates of first year change in PD were all <0.1. The baseline to year 1 PD decrease for film only mammograms was larger than that for all digital mammograms, however the KE52 only decline was slightly larger than that for film only.

Table 7-10 PD change from randomisation to year 1, by age at Randomisation (≤60 vs >60 years)

	All mmgs	Film only[2]	Digital only	KE52 only[1]	KE54 only[1]
Baseline to Year 1		**≤ 60 years at Age at Randomisation**			
PD change	**–0.20****	**–0.29%****	**–0.18%***	**–0.31*****	0.015
Number of participants	39	18	27	15	9
Number of episodes	63	28	35	21	11
Baseline to Year 1		**> 60 years at Age at Randomisation**			
PD change, 0 to 1yr	–0.05%	0.12%	–0.02%	0.01	–0.03
Number of participants	60	27	48	22	20
Number of episodes	83	31	52	25	21

Bold indicates p≤0.1; * p<0.05; ** p<0.01; *** p<0.001; † p<0.2; mmgs mammograms;
[1] Film, KE52 and KE54 only models fitted with default covariance structure (independent) due to small sub-group sample sizes to enable model to converge

A dose-response effect of age was also noted for models with all mammograms with age at randomisation divisions at ≤55 years vs <55 years, ≤60 years vs >60 years, ≤62 years vs >62

years and ≤65 years vs > 65 years, Table 7-11. Younger age at randomisation appears to be associated with a more substantial rate of decline for annual PD from baseline to year 1, relative to older age at randomisation. Most models include only a few repeated measurements for each subset of mammograms from baseline to year 1 (Table 7-11), hence these age-related dose-response results are based on very small sample sizes.

Table 7-11 PD change from randomisation to year 1, by ages 55, 60, 62 and 65 at randomisation

Age, ≤ or > comparison	Age 55	Age 60	Age 62	Age 65
Baseline to Year 1		**Younger participants**		
PD change[1]	**–0.24****	**–0.20****	**–0.19*****	**–0.16%*****
Number of participants[2]	19	39	53	72
Number of episodes[2]	31	63	83	110
Baseline to Year 1		**Older participants**		
PD change[1]	**–0.09%**	–0.05%	–0.02	0.05%
Number of participants[2]	80	60	46	27
Number of episodes[2]	115	83	63	36

Bold indicates p≤0.1; * p<0.05; ** p<0.01; *** p<0.001; † p<0.2

[1] Coefficient for baseline to year 1 time segment

[2] Number of participants and episodes for baseline to Year 1

Selected model coefficients—for the parameters age at randomisation, MD change over time, the age and time interactions, and the intercept— are presented in Table 7-12 for parsimonious models of PD and DA which were fitted with an interaction between age and time. The age and baseline to year 1 interaction was significant for both PD (p=0.032) and DA (p=0.048), as well as for years 5 to 7 (PD, p=0.016; DA, p=0.018). The age and baseline to year 1 interaction coefficient for PD, for instance, is interpreted as a significant increase in (square root transformed) PD of 0.014 per year increase in age at randomisation. For a participant age 50 years of age at randomisation, the average annual baseline to year 1 decline in PD is –0.26/year; in comparison, a participant randomised at age 60 has a relative baseline to year 1 decline in average annual PD which is (10 years * 0.014/year =) 0.14 higher, which equates to –0.12/year. The significant interactions between age and time for PD and DA during baseline to year 1 and years 5 to 7 also suggest, as per the data in Table 7-10 and Table 7-11 that the rate at which PD and DA decline decreases as age increases.

Table 7-12 Selected model coefficients, PD, DA continuous Age at Randomisation interaction models

Covariate	Continuous age at randomisation interaction models			
	Percent Density[1]		Dense Area[1]	
	Coefficient	SE	Coefficient	SE
Age at randomisation (y)	**-0.056****	0.016	**-0.53***	0.26
Baseline to Year 1	**-0.256****	0.078	**-3.1****	1.0
Year 1 to Year 5	0.033	0.033	**0.81**	0.45
Year 5 to Year 7	**-0.195***	0.085	-**2.8***	1.2
Baseline to Year 1 x interaction	**0.014***	0.0064	**0.16***	0.08
Year 1 to Year 5 x interaction	0.001	0.0022	0.0008	0.03
Year 5 to Year 7 x interaction	**0.020***	0.0083	**0.33***	0.14
Intercept	**5.4*****	0.25	**61.6*****	3.6

Bold indicates p≤0.1; * p<0.05; ** p<0.01; *** p<0.001; † p<0.2

[1] Square root transformed PD and DA, modelled with all 540 episodes and continuous time (3 segments)

A dose-response effect with increasing PD for the initial (first) episode at either baseline (n=85) or the earliest episode (n=35) was noted, Table 7-13; the rate at which square root transformed PD declines increases (in absolute value) with higher initial PD (baseline or earliest episode). As for most other sub-group models, the estimated annual PD decline from baseline to year 1 is derived from data with only a few repeated measurements from baseline to year 1 for any participants. However, the observed dose-response effect of higher PD associated with greater PD change over time is expected, and may potentially be replicated within larger samples of the IBIS-II population.

Table 7-13 PD[1] change from randomisation to year 1, by first episode PD <10%, 10 to <25%, & ≥25%

PD (%) for first episode (baseline or earliest episode)	<10%	10 to <25%	≥25%
Baseline to Year 1 only			
PD rate of change[2]	**–0.11**	**–0.12**[4]	**–0.24****
Number of participants[3]	31	42	26
Number of episodes[3]	45	63	38
Baseline to Year 7			
Number of participants	37	54	29
Number of episodes	150	246	144

Bold indicates p≤0.1; * p<0.05; ** p<0.01; *** p<0.001; † p<0.2

[1] Square root transformed PD;

[2] coefficient for baseline to year 1 time segment;

[3] Number of participants and episodes for baseline to Year 1 only

[4] p=0.053

The numbers of participants were too small to properly assess the age x time interaction within each PD stratum (<10%, 10 to <25%, & ≥25%), Table 7-14, however the PD stratum with the

most participants and episodes, PD 10% to <25%, showed a marginally non-significant (at the 5% level) baseline to year 1 interaction between age and time (p=0.05), Table 7-15.

Table 7-14 Number of participants per PD <10%, 10 to <25%, & ≥25%, by age at randomisation

Age at randomisation	<10%	10 to <25%	≥25%	Total
<55 years	4	14	4	22
55 to 59 years	10	9	6	25
60 to 64 years	11	16	12	39
65 + years	12	15	7	34
Total	37	54	29	120

Table 7-15 PD[1] change and age x time interaction, by first episode PD <10%, 10 to <25%, & ≥25%

PD (%) for first episode (baseline or earliest episode)	<10%	10 to <25%	≥25%
# participants/# episodes	37/150	54/246	29/144
Age at randomisation (years)	-0.02	0.02	-0.004
Baseline to Year 1	-0.23†	**-0.25**	-0.24†
Year 1 to Year 5	-0.07	0.04	-0.06
Year 5 to Year 7	-0.21	**-0.31**	-0.13
Baseline to Year 1 x interaction	0.01	**0.02**[3]	0.0009
Year 1 to Year 5 x interaction	0.004	-0.0004	0.007†
Year 5 to Year 7 x interaction	**0.04***	**0.02**	0.009
Intercept	**2.5*****	**4.8*****	**6.1*****

Bold indicates p≤0.1; * p<0.05; ** p<0.01; *** p<0.001; † p<0.2
[1] Square root transformed PD
[2] Coefficient for baseline to year 1 time segment
[3] p=0.05

Although sample sizes within any of the strata for mammogram Version are also small (Table 7-16), longitudinal PD change modelled for each mammogram Version separately (Table 7-17) demonstrated significant decreases (p<0.05) in square root transformed PD for the first year of trial treatment for film and KE52 Version mammograms. The PD rate of annual change for years 1 to 5 was not significant for KE52 and Fuji Version mammograms. In contrast, PD annual change for KE54 mammograms increased non-significantly for the first year of trial treatment (0.05/year), whilst the PD annual change from years 1 to 5 for KE54 was a significant 0.20/year (p<0.001). Most of the unusual changes in PD over time for models of digital mammograms (e.g. reduced rated of PD change for baseline to year 1, and increase in annual PD for years 1 to 5) appeared to have resulted from measurements made on KE54 mammograms. Interestingly, PD decreased non-significantly for KE54 mammograms for years

5 to 7 at an annual rate of –0.3/year, whilst PD did not change significantly (0.05/year) for Fuji mammograms during this time period.

Table 7-16 Distinct number of participants, total episodes per time segment, by mammogram Version

Mammogram Version	0 to 1.0 year	>1.0 to 5.0 years	>5.0 to 7.0 years
	Distinct number of participants/Total episodes[1]		
Film only	45/59	4/4	none
KE52 only	37/46	66/77	none
KE54 only	29/32	90/115	4/4
Fuji only	5/5	90/135	42/59

4 episodes from mammograms of 'other' digital Version not tabulated

[1] Due to differences between the date in years and categorical time (follow up number), total number of follow ups for baseline to year 1.0, >1.0 to 5.0 years and >5.0 years (Table 7-16, above) does not exactly match the number of episodes per follow up in Table 10-7 (Number of episodes (participants) at each follow up, by mammogram Version)

Table 7-17 Square root transformed PD annual change coefficients (%), by mammogram Version

Mammogram Version	N participants /episodes	0 to 1 year	1 to 5 years	5 to 7 years	Intercept (FE) ***	Between person variance	Within person variance
Film only[1]	45/63	**-0.20***	--[3]	--[3]	**5.6**	**2.2**	**0.05**
Film only, default covariance	"	**-0.20***	--[3]	--[3]	**5.6**	**2.2**	**0.05**
KE52 only[1]	94/123	**-0.15***	-0.04	--	**4.5**	--[4]	--[4]
KE52 only, default covariance	"	**-0.14***	-0.05	--[3]	**4.5**	**1.3**	**0.03**
KE54 only[1]	114/151	0.05	**0.20*****	-0.30	**4.3**	1.2	0.08
KE54 only, default covariance	"	0.05	**0.20*****	-0.30	**4.3**	**1.3**	**0.02**
Fuji only, default covariance[2]	113/199	--[3]	0.01	0.05	**4.7**	**1.5**	**0.05**
KE52 + Fuji only[1]	117/322	**-0.13***	0.03	0.002	**4.4**	**1.3**	**0.05**
KE54 + Fuji only[1]	119/350	**-0.09***	**0.10*****	0.02	**4.6**	**1.2**	**0.06**
KE52 + KE54 + Fuji[1,5]	119/473	**-0.09**	**0.07****	-0.004	**4.8**	**1.2**	**0.10**

Bold indicates p≤0.1; * p<0.05; ** p<0.01; *** p<0.001; † p<0.2

[1] Exponential covariance structure modelled;

[2] Default covariance structure modelled only for Fuji due to non-convergence of exponential model;

[3] Insufficient mammograms were available during these time segments (re: Table 7-16) hence growth parameters are: not tabulated or not applicable

[4] RE intercept variance estimates for model were not significant (i.e. 2*SE> RE variance estimate)

[5] Four episodes of unknown Version are not modelled

The mixed model weights participants with higher numbers of measurements more heavily than participants contributing fewer episodes. For example, the totals for number of distinct participants for the time segment from baseline to year 1 for film, KE52, and KE54 mammograms were 45, 37 and 39 respectively, Table 7-16. The total episodes during the same time segment for film KE52 and KE54 mammograms were 59, 46 and 31 respectively (Table

7-16). The maximum number of participants with repeated measurements in each time segment is the difference between the number of total episodes and number of distinct participants for that time segment. Therefore the maximum number of participants contributing repeated (more than one) measurements for the baseline to year 1 time segment for film, KE52 and KE54 mammograms are 14, 9 and 3 participants, respectively. This means that very few participants contributed multiple episodes to the baseline to year one PD relationships tabulated in Table 7-17.

Fitting of a paired Version model with both KE52 and Fuji mammograms (controlled for mammogram Version) yielded a combined annual PD change for years 1 to 5 of 0.025/year, p=0.37 (Table 7-17). In comparison, another paired Version model fitted with KE54 and Fuji mammograms (controlled for mammogram Version) yielded a combined annual PD change for years 1 to 5 of 0.1, $p<0.001$ (Table 7-17). This PD change over time coefficient for years 1 to 5 is approximately halfway between the individual Version coefficient values of 0.2 for KE54 and 0.01 for Fuji mammograms. It is not surprising that the coefficient for the combined Version model is halfway between the individual Version estimates because the number of distinct participants and total mammograms are similar for years 1 to 5 for KE54 and Fuji mammograms, Table 7-16. The results from these two paired Version models further supports an assertion that the significant increase in PD for years 1 to 5 noted in the digital-mammogram only parsimonious square root transformed PD model, Table 7-7, is likely due to change in PD on KE54 mammograms only. PD change over time for the KE52 and Fuji mammograms for years 1 to 5 is effectively flat (no change). The growth curve for the KE54 mammograms for the baseline to year 1 period also differs from that of the film and KE52 mammograms (Table 7-17). Whilst these differences in PD growth between the different mammogram Versions might be just to chance (e.g. small sample sizes/repeated measurements re: Table 7-16), errors in measurement or other reasons, potentially other technical factors could be affecting the appearance of the KE54 mammograms compared to other Versions which were not controlled for during the analysis.

To further examine the noted longitudinal PD differences for KE54 mammograms compared to other Versions, participants for whom sequential mammograms were available for the KE54 mammogram Version were reviewed within the Sante DICOM Editor viewer. One participant visibly decreased in PD from baseline to year 1. Although some of the increases in PD during the 1 to 5 year time period were likely due to (random) measurement error (because no other reason was found), a few participants visibly increased in PD. The reasons for this are unknown, and did not appear to be explained by changes in mammogram exposure compared to other participants or trends in the data (e.g. cease of randomised treatment, symptoms such as hot flushes or arthralgia). An increased sample from participants with KE54 mammograms may help to clarify if the observed significant increase over time during year 1 to 5 is genuine.

7.6 Discussion

Overall, the aggregate (treated + control) models for PD and DA appeared to provide a reasonable fit for the observed mammographic data, although the power to detect a difference in MD change over time is low due to the small sample size including repeated measurements within any of the different mammographic Versions. Contrary to early expectations, both PD and DA showed significant ($p<0.05$) average increases over the 12 to 60 month (years 1 to 5) follow up period, which are likely due to the transition from lower PD digital mammograms (KE52) to higher PD digital mammograms (Fuji). Untransformed PD showed an average increase of 0.4% per year (+1.6% total during years 1 to 5) and untransformed DA showed an average increase of 100mm^2/year (about 400mm^2 or 4cm^2 total during years 1 to 5). Very small (–0.2% for PD and –4mm^2 for DA) non-significant decreases in the aggregate growth curve coefficients for PD and DA were noted after the cease of trial treatment at 5 year post-randomisation.

The modelled small effects on longitudinal PD and DA for baseline to year 1 noted for the aggregate (treated + control groups) in Aim 4 (this chapter, Table 7-7) could potentially be

composed of the predicted MD reduction due to anastrozole during the first 12 months of treatment and a smaller longitudinal MD change for the control group.

The models of square root transformed PD and DA appeared to better meet the assumptions of normally distributed residuals than did untransformed models of PD and DA. Square root transformation is a popular technique used during modelling of longitudinal modelling of MD to improve normality assumptions [148, 184, 188, 191, 401, 487, 495]. Although additional calculations are required in order to back-transform square root transformed data (with the attendant possibility that error will be introduced whilst doing so), the improvement in the linear prediction ('xb') plots (Figure 7-12) offset the appeal of using native units of measurement for PD and DA for the Primary Aim. The other diagnostic post estimation plots implied that the square root transformed PD and DA models met the assumptions of residual symmetry for all RE levels better than models with untransformed PD and DA, and that the predicted values provided good estimates of the observed (measured) values for PD and DA.

The magnitude and sign of the covariates in the baseline characteristics multivariable regression models of PD, DA and BA (Chapter 6, Table 6-11— 85 participants, 85 episodes) were generally similar to those in the multivariable models of aggregate longitudinal MD, Table 7-7 (120 participants, 540 episodes). Age at randomisation was significantly ($p<0.05$) and inversely associated with square root transformed PD in baseline (Chapter 6) and aggregate longitudinal models (this chapter), as expected [137]. BMI was also significantly ($p<0.001$) and negatively associated with PD in all baseline and longitudinal models, as expected [178]. The associations between these important BC risk and MD confounders for this sample of CMN IBIS-II participants therefore appear to be consistent with other populations [189, 192].

The inverse association with age was not significant ($p<0.05$) for any DA model (untransformed or transformed, baseline or longitudinal), however the p-value was 0.09 for the parsimonious longitudinal model of square root DA. Models with larger numbers of IBIS-II participants and

episodes would likely show a significant inverse association between DA and age at randomisation as seen in another high risk population [192]. BMI was not significantly associated with DA, which was not surprising as use of DA has been suggested as an alternative to PD, especially when BMI is not available— as often occurs for mammographic data collected from BC screening programs. However, other reports have found associations between BMI and DA (although less strongly than with PD) [189, 192].

Age at first birth (≥30 years vs <30 years) and age at menopause were not significantly associated with PD in baseline models (Aim 3, Chapter 6), however both parameters had a positive and significant ($p<0.01$) association with square root transformed PD in the aggregate longitudinal models (this chapter). Since the coefficients for age at first birth and age at menopause in baseline and longitudinal models are of similar size (~1 and ~0.6, respectively), the lack of significance in the baseline models is likely due to the smaller sample size of participants and mammograms used in the baseline models. Age at first birth ≥30 years relative to age at first birth <30 years was significantly and positively associated with (untransformed and square root transformed) baseline and longitudinal DA and BA; similar trends in other reports have been noted for PD and DA [189] and for BA [188].

Current smoking status was only significantly associated with MD in parsimonious models of BA (untransformed and square root transformed); current smoking status was not significantly associated with PD and DA. This is similar to another report [188] for both PD and DA in the (high risk) IBIS-I population, and for DA but not PD in another IBIS-I study [192]. The association with significantly lower PD for current and previous smokers for IBIS-I participants appears to be due to (non-significantly) lower DA but (non-significantly) higher BA. BA was positively associated with current smoking status in the sampled IBIS-II population: a ~4000mm^2 (40cm^2) increase compared to never smokers; however there is no biological rationale for this. BA was also associated strongly with BMI (as per other reports of BA and/or AA [188, 192, 296]), which is expected. Age at menarche was also positively but weakly (0.05

$<p \leq 0.1$) associated with BA in parsimonious models. As for smoking status, there is no biological rationale for this so the association may be just due to chance.

BMI, age, risk of BC and smoking status were also associated with longitudinal PD for IBIS-I participants [192] as well as for a general screening population (normal-risk population) [189].

One important unknown covariate, number of children, is associated both with cross-sectional PD and DA, and longitudinal change in PD [145, 188, 189]. Inclusion of this covariate if it had been available may have improved the baseline and longitudinal MD models.

The unconditional means model ICCs in Table 7-4 for PD and DA range from 0.89 to 0.90, hence the correlation between follow ups for each participant is high, and only a small proportion of the variability of the density measurements is due to within person differences in MD. This likely implies that the change in MD over time is small per person compared to the between person differences in MD, which is in keeping with the small longitudinal MD changes noted for the full and parsimonious aggregate models (Table 7-7). The ICC for BA is high (0.98). Taken in conjunction with the non-significant outcomes for the Wilcoxon rank sum test (film vs digital mammograms) and Kruskal-Wallis test (digital mammogram Version comparison) for BA, Table 7-3, the high ICC for BA is likely due to relative stability of the measured breast area for each participant over time (despite the longitudinal variations in mammogram Version).

The relatively strong (for an AI) unconditional growth model change of ~ –2 to –3% between baseline and up to year 3 for PD was a pleasant surprise. However, this unconditional result was likely in part due to the transition from film (highest average PD mammogram Version) at baseline to KE 52 mammograms (lowest average PD mammogram Version) at the first year visit for half the participants. This likely created an artificial drop in PD for the film to KE52 transition, because mammogram Version was not accounted for in the unconditional growth

model. This trend is in competition with participants whose mammograms transitioned from KE52 to KE54 and from KE54 to Fuji mammograms, who should show an increase in PD. Similarly, the statistically significant increase in PD and DA continuous linear change over time from baseline for the unconditional growth models reflects the tendency of PD to increase over time for the digital mammograms. Hence the results from the unconditional model needed to be treated with caution. Mammogram Version needed to be taken into account.

The unconditional growth models revealed a significant negative covariance (correlation) for PD and DA; this indicated women with higher initial PD and DA had a greater decrease over time than women with lower initial PD. This is not unexpected, and may be related to age of the participant [189]. However, the negative correlation seen between higher initial PD and slope may also be partly attributable to the film to KE52 transition. Separate parsimonious PD models for participants ≤ 60 years (n=47 participants) vs >60 years of age at randomisation (n=73 participants) were fitted—adjusted for mammogram Version and the other significant parsimonious model parameters age at randomisation, BMI, age at menopause and parous age, Table 7-10; these models yielded square root transformed PD first year growth coefficients of –0.20 vs –0.05 respectively. Further restriction of the models to film-only mammograms yielded PD first year growth coefficients of –0.29 for ≤ 60 years vs 0.12 for >60 years. A similar trend was also noted for most digital-only models (Table 7-10), and for different dichotomisations for age at randomisation, Table 7-11. The coefficient for the interaction between PD change over time and age at randomisation modelled as a continuous parameter was also significant for baseline to year 1 (p=0.032, 0.01 square root PD/year increase in age, Table 7-12). This implies that age of the participant may also be related to the rate of change in PD during the first year of trial treatment, independent of the film to digital transition.

Even after the addition of time to the model, the within person RE estimates for the unconditional growth model MD parameters were significant (Table 7-5), hence a large proportion of the within- person variation was still unexplained. This implied further time-

varying parameters could be added to the model to explain this variability. One potentially important parameter which may vary over time is BMI; longitudinal BMI was not recorded for the IBIS-II participants hence (baseline) BMI is only utilised once in the model. Time (modelled in years, as a three segment linear continuous spline) was the principal parameter which varied over time utilised in the Aim 4 longitudinal aggregate (treated + control) MD models. However, the mammogram Versions also changed over time and its addition as both a FE and RE to the model reduced estimates of the (residual) within person variability.

It was not possible to distinguish differences in MD— nor potentially, differences in MD change over time— for participants with higher BC risk (i.e. women with a stronger family history of BC and ovarian cancer) from the participants at lower risk (i.e. fewer number of relatives with BC and/or ovarian cancer) within this small sample of high risk women; this was due to the non-linearity of the relationship noted during the baseline characteristics analysis in Aim 3 (last chapter). The inability to utilise this parameter, as well as related factors such as Gail or Tyrer-Cuzick BC risk, is another limitation for the longitudinal models in this analysis. Incorporation of family history (or other metric of BC risk beyond the included parameters of BMI, age, parous status and menopausal status) may have improved the model.

The statistics for model fit—log likelihood (LL), AIC and BIC— were smaller (closer to 0) for the unconditional growth model vs the unconditional means model. The differences between statistics for the models ranged from ~40 to 70. This is far greater than the BIC 10 unit difference which denotes very strong evidence for the model with the lower number [493][6] [478]. Similarly, use of the Deviance statistic (-2LL for a multilevel model) shows that –78 difference for the unconditional growth model greatly exceeded the 0.001 critical value of a

[6] The BIC cutoffs for model selection of 0-2, 2-6, 6-10 and >10 [weak, positive, strong and very strong evidence, respectively] are for models with independent observations. This is not the case for mixed longitudinal models [Jones 2011]. Singer and Willett [2003] recommend use of -2(log likelihood), the Deviance statistic for a multilevel model, rather than the AIC or BIC when assessing model fit.

chi–squared test with 3 degrees of freedom[7] (16.27) [490]. The results of these tests underscored the improvement in fit by the addition of time to the model.

The removal or addition of one covariate in the models sometimes yielded differences in AIC greater than 2, but would change the BIC by less than 3. To create more parsimonious models, selection of covariates based on the BIC outcome was undertaken. Selection of covariates based on differences in AIC may have resulted in the retention of more covariates, and provided models which are more generalisable to other high-risk populations and with more predictive capability.

Unlike the film and Fuji mammogram Version random slopes, the RE slopes for KE52 and KE54 mammograms are likely not significant because the mean structure for the model (i.e. the FE part of the model) is similar to the lower PD and DA means of the KE52 and KE54 Version mammograms. The RE for film and Fuji mammograms improve the model because PD and DA tend to be higher for film and Fuji mammograms, and thus deviate more from the mean structure (population average) than do KE52 and KE54 mammograms.

The change in DA from baseline to year 1 post-randomisation also was largely driven by the DA decrease in film mammograms compared to digital mammograms (Table 7-7). Dense area like PD also increased from years 1 to 5 post-randomisation. The slight increase in BA for all time periods between baseline and year 7 could be due to changes in mammogram Version over time, or possibly an increase in the compressibility of the breast tissue with increasing age. It is unlikely to be due to a systematic increase in breast compression pressure applied during mammography, which should vary randomly across episodes.

The apparent decreases in PD (–1%) and DA (–190mm^2) from baseline to year 1 were in opposition to the annual increase in BA of ~190mm^2 during this time, Table 7-7. These may be

[7] Three degrees of freedom for the chi–squared test are needed because three additional parameters (one fixed effect and two random effects— a random slope, and a covariance) were added to the model.

genuine trends, as two of the three mammogram Versions showed decreases in PD and DA during this time period, Table 7-17.

The reasons for the lack of the expected decrease in PD during baseline to year 1 as well as years 1 to 5 for KE54 mammograms compared to film, KE52 and Fuji mammograms are unknown. The difference in response did not appear to be explained by changes in mammogram exposure compared to other participants or trends in the data (e.g. cease of randomised treatment, symptoms such as hot flushes or arthralgia).

Other AI and MD studies have shown changes >3% PD for some participants after AI treatment [28, 31, 311]. It is not known if increase in MD due to AI treatment might potentially be associated with increases, decreases or no change in BC risk. The WHI trial of progestins + estrogen HRT observed both significant increases in BC risk and MD [246, 496]. A nearly significant trend towards lower BC incidence for women (without uteri) randomised to estrogen-only HRT was observed in an WHI trial [497]; average PD also tended to increase for the WHI participants treated with estrogen-only HRT [272]. Prior to the use of estrogen-reducing endocrine therapy for early and metastatic BC, progression of metastatic BC was slowed by high-dose estrogen therapy. Presumably MD increased for the women treated with estrogen therapy, as observed for women treated with HRT. Further research is required to ascertain if MD increases and/or decreases in response to AI treatment are associated with change in BC risk.

The numbers of participants were too small to assess the age x time interaction within each PD stratum (<10%, 10 to <25%, & ≥25%), however the PD stratum with the most participants and episodes, PD 10% to <25%, showed a marginally significant baseline to year 1 interaction between age and time (p=0.05). Because these models are the aggregate of the treated and control participants, if a treatment effect is present this will also affect all aggregate models fitted with the age x time interaction terms. It is not known whether the rate of change due to AI treatment is also associated with age, but given the slower rate of decline in PD with

increasing age it is likely that if a rate of decline relative to controls is associated with anastrozole treatment in the CMN IBIS-II population, the rate of change is likely to inversely associated with age (i.e. younger anastrozole treated participants will have greater rates of decline than older anastrozole treated participants).

The sensitivity analysis in this chapter for the comparison of the MI model with imputed age at menopause and a model with an 'as reported' age at menopause or hysterectomy was performed after the Primary analysis (Chapter 8) was completed. During this sensitivity analysis comparison it was noted that the rate of change in PD for the different segments of time from baseline to year 7 do not differ for the MI and 'as reported' age at hysterectomy models. Although models for the Primary Aim (Chapter 8, unblinded analysis of treated vs control participants) included imputations for age at menopause, it is possible that the outcomes for the Primary Aim noted in Chapter 8 would be unchanged if the 'as reported' age at menopause or hysterectomy was utilised as a covariate instead of the imputed age at menopause covariate. Comparisons of the MI and non-MI models (including the 'as reported' model) for within and between person variability also revealed that use of the MI age at menopause covariate introduced additional within person and between person variability to the model. This further supports the idea that the parsimonious models tabulated in Table 7-7 could have been further simplified by substitution of the 'as reported' age at menopause or hysterectomy for the MI age at menopause covariate.

7.7 Conclusion

Parsimonious mixed models of square root transformed PD and DA provided adequate longitudinal growth curve modelling of the observed (measured) MD for CMN IBIS-II participants. The estimated mean aggregate (treated + control) annual change from baseline to year 1 for PD (%) was –1% and –200 mm^2 (0.2cm^2) for DA. These MD changes are quite small, and unlikely to be detected visually; however more sensitive and (probably) accurate

techniques to measure MD such as Hologic's Qantra and Matakina's Volpara may be able to reliably measure such small changes in MD. Examination of the PD parsimonious models' growth coefficients for each mammogram Version (film, KE52, KE54 and Fuji) implied that the decrease in PD noted between baseline and year 1 were not due (solely) to the film to digital transition.

The addition of terms for treatment group (Chapter 8, Primary Aim) was expected to clarify if longitudinal MD differences existed for treated vs control participants in this sample of IBIS-II participants, despite some of the unusual (i.e. increasing PD) results found during characterisation of longitudinal MD for the aggregate group.

8. Primary Aim treated vs control longitudinal MD

The first part of this chapter describes the aims and methods for the Primary Aim of this thesis. The results from the unblinded (treated vs control) analysis are reported, followed by a discussion of the results, and a brief summary of the key points of the chapter.

8.1 Aim for Chapter 8

The Aim of this chapter (Aim 5, Primary Aim) was to compare longitudinal changes in MD for treated vs control CMN IBIS-II participants. This chapter completes the work commenced in Chapter 7 (Aim 4), the unblinded (aggregate) MD longitudinal analysis. The Primary Aim was accomplished by adding terms for treatment group to the statistical model developed in Aim 4, to ascertain if anastrozole treatment was associated with greater decreases in MD over time for the treated group compared to the control group. A series of sensitivity analyses were undertaken to examine differences in longitudinal MD for film only and digital mammogram only models, as well as different subgroups of participants such as those with known age at menopause, and younger (≤60 years) vs older (>60 years) participants.

8.2 Methods

Treatment allocation for the IBIS-II trial remains blinded until at least 2022 when the final participants will complete 10 years of follow up. Direct access to the treatment allocation for CMN IBIS-II participants was not possible. Hence this analysis was undertaken in collaboration with statisticians at the Queen Mary University of London (QMUL), including the IBIS-II trial statistician who has access to the unblinded trial data. Approval for this unblinded analysis was sought and received from the IBIS-II trial steering committee. The statistical plan was developed in conjunction with the QMUL statisticians.

8.2.1 Study population

The mammographic density measurements utilised for the Primary Aim consisted of all trial mammographic episodes up to year 7 collected for the 120 CMN IBIS-II participants randomised to treatment with anastrozole or placebo treatment. All mammographic episodes were utilised regardless of whether the participant remained on treatment or not (intention to treat analysis). A sensitivity analysis was undertaken using the per-protocol population (participants who remained on the protocol-allocated treatment).

8.2.2 Measures

The outcome measures selected for the Primary Aim were percent density and (absolute) dense area. These mammographic density attributes ("phenotypes" [469]) were selected for a number of reasons. In the aggregate group model of longitudinal MD change (Chapter 7, Aim 4), both PD and DA were associated with significant ($p<0.05$) longitudinal MD changes during baseline to year 1, and years 1 to 5. As described in Chapter 2, percent density is the most commonly utilised and reported mammographic density attribute. PD is therefore reported in this chapter for compatibility with other MD studies. PD was also the principal MD attribute utilised in sensitivity analyses because it is a more widely recognisable MD attribute than DA; the aggregate treatment group models (Chapter 7) also suggested longitudinal PD and DA were similar. Furthermore, large (≥10% or more) reductions in PD due to tamoxifen treatment are associated with lower BC risk [17, 18, 297]. Women undertaking AI treatment in general show slight reductions in PD compared to their counterparts taking no treatment [28, 29]. As for tamoxifen, PD may ultimately prove to be a biomarker for endocrine therapy with AI. Hence PD was selected as an outcome measure for the Primary Aim (this chapter).

Dense area is a less utilised MD attribute, because it cannot be effectively assessed using visual techniques. However given the growing predominance of digital mammography and automated

methods to measure MD, DA may become a more frequently utilised MD attribute. Like PD, studies have suggested that higher DA is positively associated with BC risk e.g. [401, 498-500]. DA alone may be a better predictor of BC risk than PD alone (unadjusted by BMI or weight) [184], and hence may be more useful in longitudinal studies of MD and BC risk where BMI, weight or other measures of body size are not available. Reductions in DA due to tamoxifen treatment have also been associated with reductions in BC risk [297]. Like PD, DA may ultimately be shown to be a useful biomarker for endocrine treatment, including AI therapy. Therefore DA was also selected as an outcome measure for the Primary Aim.

The MD attribute adipose area (AA, where AA = total breast area (BA) - DA) has been shown to be associated positively [501] but primarily negatively with BC risk [469]. Due to the slight expected changes in PD (and therefore DA) due to anastrozole treatment, changes in AA were deemed unlikely to provide additional AI treatment longitudinal information compared to PD or DA within this small study sample. Although use of AA as an outcome measure could potentially provide useful insights into the effects of endocrine treatment for MD and BC risk, this project (due to the small sample size) was not able compare changes in MD and BC risk. Hence AA was not utilised as an outcome for the Primary Aim.

Anastrozole treatment was not expected to affect BA differentially compared to controls. This assumption was supported by the results of the aggregate group longitudinal model (Aim 4, Chapter 7) for BA. The aggregate models showed BA was associated only with non-significant, slight (~200 mm^2/year) average annual increases from baseline to year 7. Therefore BA was not included as an outcome measure.

The set of covariates included in the Primary Aim models are described further below.

8.3 Statistical methods

A set of statistical scripts in Stata v12.1 was prepared to perform the analysis at QMUL on a securely transferred dataset containing the relevant MD and covariate data for the CMN IBIS-II participants. The QMUL collaborators performed testing of the scripts prior to running the analysis on the unblinded data, to ensure the scripts ran smoothly in a later version of Stata, and to ensure the output would not divulge the treatment status of participants to researchers. The statistical output including log files, graphs and model estimates was securely transferred from QMUL when the analysis was completed. Tables of the model coefficients, standard errors and significance were generated after the using log file output from the analysis.

8.3.1 Descriptive analysis

Baseline characteristics by treatment group for all covariates of interest were quantified where possible, and the results reviewed for marked differences between the groups. Summary statistics were presented for continuous covariates— age at randomisation, height, weight, BMI, age at menarche, age at menopause (natural, ovariectomy, hysterectomy), oral contraceptive duration, HRT duration, age at first birth, number of relatives with ovarian cancer and/or BC, weighted number of relatives[8], time between baseline mammogram and randomisation, PD, DA, breast area, adipose area, time between baseline/first mammogram to last mammogram, time between randomisation and last mammogram. Summary statistics were obtained for: the numbers of observations (n), mean, standard deviation, median and first (Q1) and third quartile (Q3). Although not all of the continuous parameters have skewed distributions, only median (Q1 to Q3) for the baseline characteristics are tabulated. This was performed for compatibility with the reported IBIS-II trial results in Chapter 3 (Table 3-1 [26]) and the subsequent CMN baseline characteristics presented in Chapter 6 (Table 6-3).

[8] The sum of : 1 for a first degree relative, 0.5 for a second degree relative

The number of follow up episodes for each treatment group was tabulated. Due to the potential for unblinding of participants, it was not possible to know the numbers of observations in each treatment group for the following categorical parameters: age at first birth as a categorical parameter (<30y vs 30+y vs nulliparous), parity status, smoking status, previous oral contraceptive use, previous HRT use, previous participation in the IBIS-I trial, randomisation year (2007 to 2011), baseline mammogram status, baseline episode (mammogram) Version (n=85), first episode Version (n=120), follow up episode Version. Although inappropriate because allocation to treatment group was performed randomly, chi-squared and Fisher's exact tests were therefore used to assess for differences in number of treated vs control participants allocated to these categorical parameters.

Longitudinal lowess plots (line only, dots suppressed) of percent density (<50%) by mammogram Version for each treatment group were used to qualitatively assess for differences in MD change over time for each mammogram Version.

8.3.2 Primary Analysis

To achieve the Primary Aim of this thesis, a mixed model with a parsimonious set of covariates was used to model change in MD over time for the CMN participants. MD was modelled as square root transformed percent mammographic density (PD) and square root transformed dense area (DA), to improve normality assumptions. Terms for treatment group and treatment group by time interaction were added to the aggregate mixed model of longitudinal MD change of both treated and control CMN IBIS-II participants developed in Chapter 7 (Aim 4), to ascertain if longitudinal MD differences exist for treated vs control groups.

As described in Chapter 7, the parsimonious set of covariates modelled were: age at randomisation, BMI, parous age (3 categories: <30 vs 30+ years vs nulliparous), and age at menopause. Age at menopause was imputed for 27 participants for whom only age at hysterectomy was known. Age at randomisation and age at menopause were centred at 50

years, whilst BMI was centred at 25 kg/m^2. Time (in years) was centred at the date of randomisation for each participant, and modelled as a piecewise continuous spline consisting of 3 segments with change points (knots) at 1 and 5 years. Mammogram Version was modelled as a 4-category fixed effect (film, KE52, KE54, Fuji). The upper level of the model ('level 2') was participant, and the bottom level of the model ('level 1') was the average of PD or DA for each mammographic episode. Two significant RE slopes[9], one for film mammograms and one for Fuji mammograms, were also modelled at the level of participant; these provided additional 'between person' RE to help explain the variability in the model. An exponential covariance structure was utilised as well as robust SE.

8.3.3 Regression diagnostics

Regression diagnostics were undertaken on the two Primary models (PD and DA), and included the following:

1. Linear prediction (xb) vs observed (measured) MD, with a line of equality ("x=y" passing through graph origin) for reference. Comparison of the fixed effects (FE) linear predictions ('xb' in Stata) vs the observed values is performed to assess how well the FE portion of the model predicts the observed data, and to check for heteroskedasticity (e.g. increased variability at larger predicted values: a funnel shaped plot). Non-constant variance over the range of observed values suggests the linear model might not be as efficient as it could be (i.e. the standard errors (SE) might be larger than necessary). The presence of heteroskedasticity might also indicate that linearity assumptions between the outcome parameters (i.e. PD and DA) and the (FE) model components have not been met, or that there are other problems with the model such as non-independence between the errors (residuals) and observed data [502]. As described in Chapter 7, comparison of the observed vs FE predicted values the fixed effect (FE) coefficients (βs) produced by the mixed model

[9] As stated in Chapter 7, random effects (RE) were considered to be significant if the SE were ≤ half of the RE variance estimate. (Stata does not provide a p-value for RE parameters.)

describe the mean structure of the model. These coefficients represent the average MD response for all episodes from the CMN participants which were used to build the model. The 'predicted' MD values are the FE MD linear predictions ('xb') for each participant: $\hat{y}$ (predicted value) = xb [10]. The predicted values for each participant are therefore generated as a function of the observed (measured) MD values for each participant and the FE coefficients generated by the model. If the model has perfect predictive ability, then the (predicted) values output by the model will exactly equal the observed (measured) MD input into the model. All values would then coincide with a line of equality (x=y passing through graph origin), because the values predicted by the model would be identical to the observed values (x=y). Perfect prediction is unlikely to occur; nonetheless, the (residual) scatter of the predicted vs observed values around the line of equality should be symmetric, and not show greater dispersion for some values vs others (i.e. the predicted values should be homoscedastic not heteroskedastic). Due to an initial, early error in this thesis, the convention to plot observed values on the horizontal (x) axis, and the predicted values on the vertical (y) axis has not been followed. Please note the axes on this type of scatter plot are reversed throughout the thesis: the observed values are plotted on the vertical axis whilst the predicted values are plotted on the horizontal axis. Hence the line of equality has been substituted for the more typical reference of a line of best fit for the observed vs predicted values (y=bx); this is because this relationship is not the same as the x=by, which was predicted by the statistical program during creation of the graphs (with an automated script). Hence a line of equality has been presented rather than the line of best fit.

2. Linear prediction vs measured MD comparisons of MI and non-MI models. Post estimation commands in Stata for multiply imputed (MI) models are limited. In order to utilise the full

[10] The Primary models include more than one FE coefficient, hence the predicted value (expectation) for a particular participant for a particular episode more closely resembles $\hat{y} = x_1\beta_1 + x_2\beta_{2+}x_2\beta_2 + \ldots.$, where x_1 is age at randomisation for that participant and β_1 the coefficient for age at randomisation, x_2 is the participant's BMI and β_2 is the coefficient for BMI, etc. $\hat{y}$ is actually a vector of responses (predictions of the model); xb is commonly denoted as $\mathbf{X}\beta$, where β is a vector of fixed effects coefficients and **X** is a matrix of known variables (e.g. the IBIS-II participants' baseline characteristics and MD measurements) input into the model.

range of post-estimation commands available in Stata for mixed models, the Primary PD and DA models were refitted utilising the average of the 25 imputations used to estimate age at menopause for 27 CMN participants. This approach had previously been successfully undertaken in Chapter 7 during post-estimation checking of the aggregate group models. Although use of the average of 25 imputations for the age at menopause will likely result in underestimation of the SE for this parameter, it did not appear to greatly alter the aggregate group model estimates, SE, or predictive ability. To ascertain whether the addition of the terms for treatment group did not invalidate use of the non-MI models during RE residual checking for the Primary Aim, the multiply imputed linear predictions (xb) for each participant from the mixed models were compared graphically with the predicted values which utilised an average of the MI estimates (non-MI models). Estimates from the non-MI models were utilised during checking of model normality assumptions for the Primary model RE residuals (discussed further below). Predictions for the RE values— and therefore plots of the fitted model predictions (described next)— were then possible to generate with the non-MI models.

3. Fitted prediction (xb + random effects (RE)) vs measured MD & line of equality. As for the graphs of the FE linear prediction vs observed MD, graphs of the fitted vs observed values are used to check model fit for predicted values using both the FE and RE generated by the mixed model. The FE linear 'predicted' (xb) values for mixed models in Stata represent the model expectation for the observed values without the addition of values estimated for the random effects. Stata's mixed model 'fitted' values are generated by adding estimates of the RE values (best linear unbiased predictions, 'BLUP's [503]) to the predicted FE (xb) values[11]. As described in Chapter 7, the mean structure for a mixed model (the fixed effect

[11] The mixed model which contains both fixed and random is often depicted as $\hat{y} = \mathbf{X}\beta + \mathbf{Z}u + \varepsilon$. As described previously, $\hat{y}$ is a vector of responses (predictions), **X** is a matrix of known variables input into the model, and β is a vector of estimated FE coefficients. $\mathbf{X}\beta$ is the structural (fixed) portion of the mixed model. $\mathbf{Z}u + \varepsilon$ are the random effects, which comprise the stochastic (random) portion of the mixed model. **Z** is the design matrix for the estimated random effects u. ε is a vector of (level-1, within person) random errors. Predictions of $\mathbf{Z}u$ are the BLUPs; predictions of $\mathbf{X}\beta$ are the FE.

β's) represents the average response (growth curve) for all CMN participants over time. This average longitudinal response is composed of individual responses over time (individual growth curves for each CMN participant). Each individual participant's response deviates (to a greater or lesser extent) from the average response (Grand mean growth curve, e.g. Figure 7-10) on average by a certain amount. In the absence of other random effects, this deviation from the average response is the between-person RE constant (intercept). As described in Chapter 7, the final Primary Aim model fit was greatly improved[12] through the incorporation of between-person random effects slopes for film and Fuji mammograms (but not other combinations of RE for mammogram Version); hence the between person RE each participant (the offset from the FE mean structure for each participant) are composed of BLUPs from three different RE (a constant and two random slopes). Although RE for time were modelled, these lost significance after incorporating an exponential covariance structure for the (level-1, within person) residuals into the model. Presumably, if more variability in response over time existed between the CMN IBIS-II participants (re: Chapter 7), RE for time (which represent different, random slopes (changes over time) for each participant) may have been retained in the model. As such— because the current model does not contain a RE for time— fitted growth curves for CMN participants without any RE for mammogram Version are essentially parallel to the Grand mean growth curve generated by the FE part of the model; MD change over time estimates for participants who do have RE for film and/or Fuji mammograms however, will vary from the Grand mean growth curve. Each participant's individual fitted growth curve is offset from the Grand mean growth curve by the estimates for all RE (level-2, between person residuals) for each participant. Therefore, to generate fitted values for the mixed model, RE values for each participant (for the between-person (BP) constant, BP film mammogram slope, and BP Fuji mammogram slope) are added by Stata to the predicted values (xb) for each time point for each participant. The amount the observed PD or DA value for a

[12] The Bayesian information criterion (BIC), used to compare the fit of different models, decreased by more than 35 units

participant's mammographic episode deviates from that participant's fitted growth curve is the fitted (level-1, within person 'measurement error') residual for that episode. Because the RE for each participant tends to account for most of the residual difference between the linear prediction (xb) and the observed values for each participant, the fitted (within person, level-1) residuals are much smaller than residuals for the FE predicted vs observed values. Hence the (fitted) residuals for fitted linear predictions utilising estimates for both the FE (xb) and RE (BLUPs) vs observed values tend to fall much closer to the line of equality than do the residuals for comparisons made using the observed vs FE predictions (xb) only.

4. Normality examination of the RE constant distributions and fitted model residuals.

Multivariate normality of the RE constants (level-2, between person residuals) and multivariate normality of the model's fitted (level-1, within person) residuals are important assumptions for mixed models [490]. As for the residuals of the linear FE predictions (xb), the magnitude of the residuals for the fitted model (FE + RE) should be independent of the values of the data input into the model ("x", independent variables); this means that the residuals should have constant variance across the range of the independent variables. If the residuals are independent and random, they will generally form a normal distribution. Lack of normality of the residuals likely means that heteroskedasticity is present, and there are probably issues with the model, e.g. non-linearity, or non-independence of observations (which should be accounted for by the groups (levels) in the mixed model). The normality of the residuals was qualitatively reviewed via examination of histogram distributions of the residuals, and by creating residual plots as per Singer & Willet [490] Figure 4.5:

- Normal score (normal probability plots vs the raw residuals for different levels of the model)
- Standardised residuals at the different RE model levels

8.3.4 Subgroup and sensitivity analyses

A series of sensitivity analyses were undertaken, to further investigate the potential impact of certain subgroupings of the data such as film only vs digital only mammograms, and participants of different ages. These sensitivity analyses were mainly undertaken for the PD Primary model, due to the more frequent use of PD in studies of MD— which makes it a more widely recognised and understood MD attribute. The results for DA in the aggregate model (Chapter 7) appeared to be very similar to those for PD; hence repetition of the sensitivity analyses for DA was likely to result in redundant information. The following sensitivity analyses are for PD only, unless otherwise specified.

1. PD and DA: Film mammogram only models vs digital mammogram only models. Most extant studies of MD have utilised film mammograms; hence measurement and use of film mammograms is well validated. Known issues with digital mammograms such as differences in post-processing across mammography machine and software versions, as discussed in previous chapters, makes use of this type of imaging data more problematic. Although a number of studies have successfully used digital mammograms e.g. [182, 504], the variability in distributions between the three digital mammographic Versions utilised in this study was likely to impact the longitudinal results for this project. Ideally, an analysis utilising each mammographic Version separately should have been undertaken, however the sample size was too small to do this. The treatment x time interaction terms were assessed for statistical significance ($p<0.05$). Due to the potentially small sample size relative to effect size, p-values between 0.05 and 0.1 were also reported. For the DA digital-only model, mammographic Version was modelled with film as the reference category instead of the KE52 Version resulting in a different parameterization of the model but no impact on association.

2. PD and DA: Full set of covariates vs parsimonious set of covariates. The full set of covariates (Parsimonious covariates with: age at menarche, smoking status, HRT use, oral

contraceptive use, IBIS-I participation) was compared to the parsimonious model to assess if any marked changes in coefficients occurred. Although the development of a good model for longitudinal change before the addition of important covariates (like treatment) is a suggested strategy [36], this comparison was undertaken in the event addition of terms for anastrozole treatment greatly altered the behaviour of the model.

3. Check change points from 1.25 to 2.0 years in 0.25 year increments. As acknowledged above, inclusion of terms for treatment group were unlikely to alter the overall longitudinal structure of the model. However, a set of models with change points for the linear spline at 1.25, 1.5, 1.75 and 2.0 years instead of 1.0 years was generated, to check that the selected change point at 1.0 years still best suited the data for the Primary Aim model. An additional set of aggregate models was created in response to the results of this sensitivity analysis. The methods for this are described in 8.3.6 Additional methods.

4. Participants with natural & ovariectomy (known age at) menopause only. This analysis was undertaken to check that the imputed ages for age at natural menopause for 27 participants did not change the conclusions from the Primary models.

5. Parous participants only. In order to model both non-parous and parous women parsimoniously in the model, a categorical parameter was created. This resulted in loss of information for the relationship between MD and age at first birth, because the group of parous women was dichotomised. This check was performed to ascertain if this loss of information impacted upon the Primary model.

6. Age at Randomisation: Women <=60 vs women >60 years at baseline. MD declines with age, however this relationship may be attenuated or reversed for older women [189, 471]. The degree to which treatment with an AI can alter MD may therefore also be smaller for older post-menopausal women compared to younger post-menopausal women. The small sample size precluded examination of the effect of age with a three-way interaction between time, treatment group and age.

7. Censored at time of participant cease of treatment "per protocol analysis". Cessation of randomised treatment with anastrozole prior to five years may dilute the subtle effects of the AI on MD, hence this subgroup was modelled to compare to the Primary PD model. Episodes were omitted from this analysis at the time treatment was (prematurely) ceased . Mammograms up to 3 months after the treatment stop date were considered on-treatment mammograms.

8. Mixed model with autoregressive (AR1) covariance structure to be compared against GEE model with AR1 correlation structure. A generalized estimating equation (GEE) approach to the modelling was also undertaken, and results compared to the linear mixed model, as GEEs are robust to covariance structure misspecification. Generalized estimating equation (GEE) models were developed for use on longitudinal data with binary outcomes [505], but are also appropriate for continuous outcomes. GEE is a type of 'marginal' or 'population-average' model. Marginal regression models estimate the "marginal expectation of a discrete or continuous response… as a function of explanatory variables" [492]; in other words, marginal models estimate the population average response for a categorical or continuous dependent variable as a function of set of explanatory covariates. Given that the Primary Aim of this thesis was to characterise the average response of the treated vs control CMN IBIS-II participants, a marginal model was of potential interest. The term 'marginal' emphasizes that the estimated mean response at each longitudinal time point "depends on the only on the covariates of interest, not on any random effects or previous responses" [506]. For instance, it is not necessary to model random (latent) effects with GEE, as the correlation is treated as a nuisance parameter. Hence, marginal models differ from mixed effect models which specifically model both fixed and random effects (FE and RE). The FE estimates of the mixed model in Chapter 7 represent the mean structure of the modelled data, and are analogous to the (population-averaged) estimates that would result if the same data were modelled using a GEE approach. Marginal models also differ from another popular type of model, the "transition or generally conditional models (e.g., Markov

models), where the mean response depends also on previous responses." [506] GEE estimates (the coefficients or 'βs') are robust to misspecifications of the correlations among the repeated measurements data. It is still important to carefully select a working correlation structure to improve efficiency [507], but the GEE method provides valid standard errors for the estimates even if the correlation structure is misspecified if robust ('sandwich estimator') SE are used. GEEs are also able to handle unbalanced data (different number of follow ups for each participant) and arbitrary patterns of missing data [506]. Despite these excellent properties, a GEE model was not selected for use with the Primary Aim because 40% of the data was dropped during GEE modelling in Stata due to unequal spacing of observations and unequal numbers (unbalanced data) for 47 participants and 1 participant with only one episode. In contrast, the implementation of the linear mixed model in Stata was able to estimate model coefficients without loss of data, even though some participants had unequal spacing of episodes (i.e. missing episodes). The baseline episodes and other trial episodes which were 'missing' from the dataset (other than those not collected for participants who ceased trial follow up) did not appear to be associated with the trial treatment effect or other systematic and/or unobserved factors besides— possibly— mammogram Version[13] and/or date of mammography (which are unrelated to the expected effect of anastrozole). Therefore the missing data was considered to be missing at random, which would not affect the assumptions of the mixed linear model. The AR covariance structure was selected for the GEE analysis because this most closely resembles the theoretical covariance/correlation structures within the longitudinal data, of the different correlation/covariance structures common to both GEE and mixed models in Stata. For this sensitivity analysis, the mixed model was restricted to the 310 episodes from 72 participants which were retained by Stata in the GEE model.

[13] Some of issues resulting from use of episodes with different mammogram Versions are discussed later in this chapter. Potentially, a selection model or pattern mixture model strategy (re: Singer and Willet [2003]) could be employed to correct the problems which resulted from the frequent change in Version and/or missing episodes in the dataset.

9. PD and DA: mixed model with default standard errors (vs robust error models). These models were compared to examine if use of the robust errors compared to default SE (original information matrix, OIM) had an impact on the Primary model outcomes.

10. Models of PD and DA – treatment group RE testing

    a.) Treatment group as a RE slope at the level of id:

    b.) As above + by(group) residual structure

Although the FE terms for treatment group may (fully) account for differences in variability between the anastrozole and control group, potentially, additional between-person variability due to the effects of anastrozole could be accounted for with a RE term for treatment group (as was seen for the film and Fuji mammogram Version RE slopes). Within-person correlation (rho) may also differ between treatment groups, hence a further additional term for treatment group was added to the covariance structure specification to explore this possibility.

8.3.5 Changes in RE variability, MD growth, with increasing model complexity

The changes in between- and within-person variability for the models in Aim 4 (Chapter 7) and the Primary Aim (this chapter) were quantified, in order of increasing model complexity. To assess the impact of the parsimonious set of covariates on the variability, an additional set of unconditional growth models for PD and DA were fitted. These new models utilised the same time and RE structure as the parsimonious aggregate model (Chapter 7) but the BC risk factor covariates were omitted— the established MD confounders age at randomisation, BMI, age at menopause and age at first birth. The 'technical' covariate mammogram Version was the sole FE covariate retained in this model. This new model was given the appellation 'unconditional aggregate growth model'. This new model was between the unconditional growth and parsimonious aggregate models in complexity. The new model was fitted specifically to examine if removal of the established MD confounders from the aggregate parsimonious model

had any effect on the RE variances or the coefficients and SE for MD change over time (MD growth curve). FE coefficients for the models were compared for differences >10% relative to the less complex model's coefficients, and whether the estimated confidence intervals overlapped (± 2SE).

Quantification of the change in RE variance with increasing model complexity was performed with a formula similar to equation 7–1 in Chapter 7: the percentage change in RE variability relative to the less complex (earlier) model was calculated. For example, the RE variance for the (more complex, unblinded) Primary model was subtracted from the corresponding (blinded) parsimonious aggregate RE variance; this difference was then divided by the parsimonious aggregate RE variance to provide a value for the relative change in variability with the addition of terms for treatment group in the mixed model. Modelling was undertaken for all mammograms (film + digital), as well as the film-only and digital only subsets of mammograms.

8.3.6 Additional methods

Back transformation of model coefficients from models with square root transformed dependent variables (i.e. sqrtPD, sqrtDA) was performed by using the original intercept value as representative of a participant who entered the trial at age 50 years, underwent menopause at age 50 who had a BMI of 25 kg/m^2. As described in the Methods for Aim 4 (Chapter 7) , squaring the original intercept value provided a back-transformed value for PD (%) or DA (mm^2) for the reference participant. The square of the intercept was then subtracted from the square of the sum of each coefficient and the intercept:

Equation 8-1

$$Back\ transformed\ coefficient = (\text{Coefficient} + \text{intercept})^2 - (\text{intercept})^2$$

This process was repeated to generate back transformed values for the SE, and upper and lower values for the 95% confidence interval (95% CI). As stated in Chapter 7, models with film

mammograms as the reference mammogram Version category have higher back transformed MD values than the median baseline MD for CMN participants. An additional set of back transformed values for treated group MD change were calculated for PD (reference value 20%) and DA (reference value 2700mm^2). Except for the Primary models of all, film-only and digital-only mammograms, the remainder of the tabulated coefficients in this chapter report square root transformed PD and DA model coefficients. This was done for ease of reporting as well as accuracy.

Estimates of the MD difference (time point 2 – time point 1) from baseline to year 1 were calculated by multiplying the estimated MD rate of change for the first continuous time segment (baseline to year 1) by 1 year. Subsequent differences over time were calculated by multiplying the estimated MD change per year by the time elapsed during a particular continuous time segment, and adding this to the estimated MD difference calculated for the previous time segment/s. The value for the anastrozole (FE) coefficient was interpreted as the relative difference in intercept between the treatment groups to create graphs of MD growth.

PD growth curves, selected by first year within-person mammogram Version

To further examine the impact of different mammogram Versions upon estimates of mean longitudinal PD generated by the mixed model, a number of different PD change over time trajectories for different sets of participants were examined qualitatively. The models used the non-multiply imputed square root transformed parsimonious PD model developed in Aim 4, Chapter 7 (aggregate (blinded) model of treated + control participants), but models were fitted with a parameter for categorical time (months since randomisation: 0, 6, 12, 24, 36, 48, 60, 72, 84) as per the unconditional growth models in Aim 4, instead of the three segment continuous spline of time. Use of categorical time allowed the PD change trajectory to vary freely across categories of time. As mentioned previously, the categorical time coefficients model the average PD difference from baseline, which is the value of the intercept in the model.

The relatively large changes in PD due to within-person Version transitions during baseline to year 1 may obfuscate the much smaller changes in PD due to AI treatment. The largest differences in average PD and DA occurred during the film to digital transition (7% difference, Table 7-3, Chapter 7); of the 45 total participants with film mammograms, the majority (n=28) transitioned from film to digital mammograms after the baseline episode (see Figure 7-5, Chapter 7, randomisation year 2008). MD also increased by 2% to 3%, on average, during the intra-digital Version transitions (KE52 to KE54 to Fuji; Table 7-3, Chapter 7); intra-digital Version change is therefore also likely to mask relative change in MD due to AI treatment.

Most of the expected difference in PD change between the treated and control groups may occur from baseline to year (as per the IBIS-I trial [296]). A number of different groups were selected based on the Version of within-person mammograms present for baseline to year 1. This was undertaken to investigate how inclusion and exclusion of different groups of participants with different first year Versions affected the PD growth trajectory of the first year and later years.

The different groups are described in Table 8-1. Line plots were created for PD estimated by the model for each group, and the graph was also fitted with a square root transformed PD decrease of –0.05 per time category as the 'ideal' trajectory for the control group. A change of –0.05 square root PD per year equates to a decline of approximately –0.5% PD/year. This decline in PD is less marked than the typical decline of about –1%/year for women aged 40+. However –0.5% PD per year for the CMN IBIS-II control group is likely appropriate because median age of the CMN participants was 62 years, and the decline in PD eases with age [144].

Although each group has been selected based on the within-person Version/s of mammograms present during baseline to year 1, all episodes after year 1 are modelled, regardless of their Version. Inclusion of all post year 1 episodes provides qualitative insights into the contributions of each group of participants to the years 1 to 7 (aggregate) PD growth curve.

Table 8-1 First year (baseline to year 1) groups, selected by Version of first year mammograms

First Year group	Rationale for group, and first year mammogram Versions
All Film First Year (Group 1) n=14 (participants)	The film only Primary model displayed the expected reduction in PD for the anastrozole treated group relative to the control group. This group of participants has only film mammograms from baseline to year 1, to emulate the PD change over time estimated by the film only Primary model.
Same version, film + digital, in the First Year (Group 2) n=31	This group of participants has repeated (within person) episodes from the same Version during the first year. In other words, only participants with multiple episodes of film (Group 1) OR KE52 mammograms OR with KE54 mammograms during the first year are included in this group. (No Fuji baseline or 6 month mammograms were collected). This group was selected to review the PD trajectory for participants whose Version was stable from baseline to year 1.
Film to digital transition in Year 1 (Group 3) n=36	This group of participants had baseline or 6 month films but subsequent mammograms which were digital. This group was selected to show the effects on PD change for participants transitioning from film to digital in the first year.
Same version + film to digital transition in Year 1 (Group 4) n=61	This group contains all participants from Group 2, but adds in participants who had baseline or 6 month films but whose subsequent mammograms were digital (Group 3). This group was created to examine the PD growth effects of women transitioning from film to digital during the first year relative to those whose Versions were stable.
All film, including all film to digital transition in the First Year (Group 5) n=45	This group has Group 1 participants, as well as the participants who transitioned from film to digital within the first year. This growth curve shows the effects of including episodes transitioning from film to digital in the first year (in comparison to the Group 1 first year trajectories)
Digital to digital transition during the First Year (Group 6) n=32	This group includes only participants who transitioned between digital Versions in the First Year. Hence only participants transitioning from KE52 to KE54 (or Fuji) mammograms OR participants transitioning from KE54 to Fuji mammograms are included in this group. This growth curve provides insights into the PD changes caused by intra-digital transitions
Same digital Version, First Year (Group 7) n=17	This group is a subset of Group 2, but does not include participants with films only in the first year. This group therefore has women with repeated KE52 and KE54 mammograms during the first year.
All digital Versions in the First Year (Group 8) n=45	This group is the combination of Group 5 and Group 6, to show the effects of merging these groups on PD growth. Groups 5 and 6 overlap slightly because some participants changed digital Versions after the 6 month follow up.
All digital Versions in the First Year, including Unknown Versions (Group 9) n=46	This group is the combination of Group 5 and Group 6 (i.e. Group 7), with the addition of three episodes of Unknown digital Version. (One episode of unknown digital Version is not included, because it followed a baseline film episode).

8.4 Results

8.4.1 Descriptive characteristics at baseline

Baseline characteristics for all 120 CMN IBIS-II participants who contributed mammograms to this analysis are shown in Table 8-2; these baseline characteristics appear to be balanced between the treatment groups. Median age at randomisation was 61.1 years for control and 62.3 years for anastrozole treated participants. Median BMI (the other strong MD confounder) was 28.0 for controls and 28.6 for anastrozole treated participants.

Table 8-2 CMN IBIS-II participant baseline characteristics, continuous covariates[1], by treatment group

Parameter	CONTROL GROUP				ANASTROZOLE GROUP			
	N	Median	25^{th}–75^{th} percentile (Q1–Q3)		N	Median	25^{th}–75^{th} percentile (Q1–Q3)	
Age (years)	58	61.1	57.1	64.7	62	62.3	56.8	65.8
Height (cm)	58	162	157	167	62	162	157	166
Weight (kg)	58	75	66	92	62	74.5	66	83
BMI (kg/m^2)	58	28.0	26	34	62	28.6	26.0	31.6
Age at menarche (years)	58	13	12	14	62	13	12	14
Age at menopause (years)								
Natural menopause	27	49	47	51	30	52	48	53
Hysterectomy age	31	46	39	51	33	41	36	46
Ovarian oblation age	20	46	40	52	20	45	37	50.5
Age at first birth (years)	56	23	21	25	56	22	19.5	24.5
Previous HRT (months)	38	43	12	84	36	60	24	114
Oral Contraceptive (months)	54	108	60	130	58	78	36	120
Total # relatives with BC and/or ovarian cancer	58	2	1	3	62	2	1	3

[1] Due to the potential for unblinding of participants, the number of participants in each category of categorical covariates could not be reported by treatment group.

Comparisons between the groups were also made for categorical parameters but the exact numbers (and percentages) allocated to categorical parameters by treatment group could not be provided by the QMUL statisticians, because of the potential for unblinding of researchers to participant group allocation. Although inappropriate because treatment group was randomly allocated, only non-significant ($p>0.25$, Fisher's exact test) differences were found between treatment group for the following categorical parameters: parous vs non-parous women, age at first birth (3 category parameter <30 years, ≥ 30 years, non-parous), previous HRT use (true/false), previous oral contraceptive use (true/false), smoking status (current, past, never), and previous IBIS-I participation (true/false).

Baseline MD related parameters for the two treatment groups for the 85 CMN participants who had baseline mammograms are provided in Table 8-3. Although baseline MD parameters appear to be balanced between the treatment groups, median PD was slightly higher in the control group than the treated group — 18.3% vs 16.7%. This difference in median baseline PD of 1.6% is approximately the same as expected for the change in PD due to anastrozole treatment. Median DA (mm^2) was also marginally higher in the control vs treated group— $2757mm^2$ vs $2684mm^2$.

Table 8-3 Measures of central tendency for MD baseline parameters, Control vs treated group

Parameter	CONTROL GROUP				ANASTROZOLE GROUP			
	N	Median	25^{th}–75^{th} percentile (Q1–Q3)		N	Median	25^{th}–75^{th} percentile (Q1–Q3)	
Months before randomisation, baseline episode	41	-3	-5	-1	44	-2	-6.5	-1
Percent Density (PD, %)	41	18.3	9.4	29.1	44	16.7	8.0	25.2
Dense Area (mm^2)	41	2757	1455	5033	44	2684	1356	4408
Breast Area (mm^2)	41	17061	13347	21804	44	16682	14340	21896
Adipose Area (mm^2)	41	13122	9836	18773	44	14098	10749	19811
Percent Adipose (%)	41	81.7	70.9	90.6	44	83.3	74.8	92.0
PD, square root transf.	41	4.3	3.0	5.4	44	4.1	2.8	5.0
DA, square root transf.	41	52.2	38	70	44	51.4	36	66

transf. transformed (square root transformed)
Only the 85 participants with baseline mammograms are included in this table

The number of mammograms for each treatment group appeared to be balanced across all follow ups from baseline (year 0) to year 7, Table 8-4. No statistical difference in randomisation year or frequency of mammogram Version was detected between treatment groups ($p>0.05$, Fisher's exact test: exact numbers not divulged for this analysis due to potential for unblinding).

Table 8-4 Number of mammographic episode for each follow up, by Treatment group

Treatment Group	Follow up number (years)								
	0	0.5	1	2	3	4	5	6	7
Control	41	11	31	43	43	41	29	14	5
Anastrozole	44	14	29	50	48	41	35	16	5

The lowess plots of (raw, 'unconditional') longitudinal PD for each mammogram Version by treatment group did not appear markedly different, Figure 8-1 and Figure 8-2. Much of the variability in the distributions of the mammogram Versions between the treatment groups is likely because different sets of participants contribute to the lowess plots at different times.

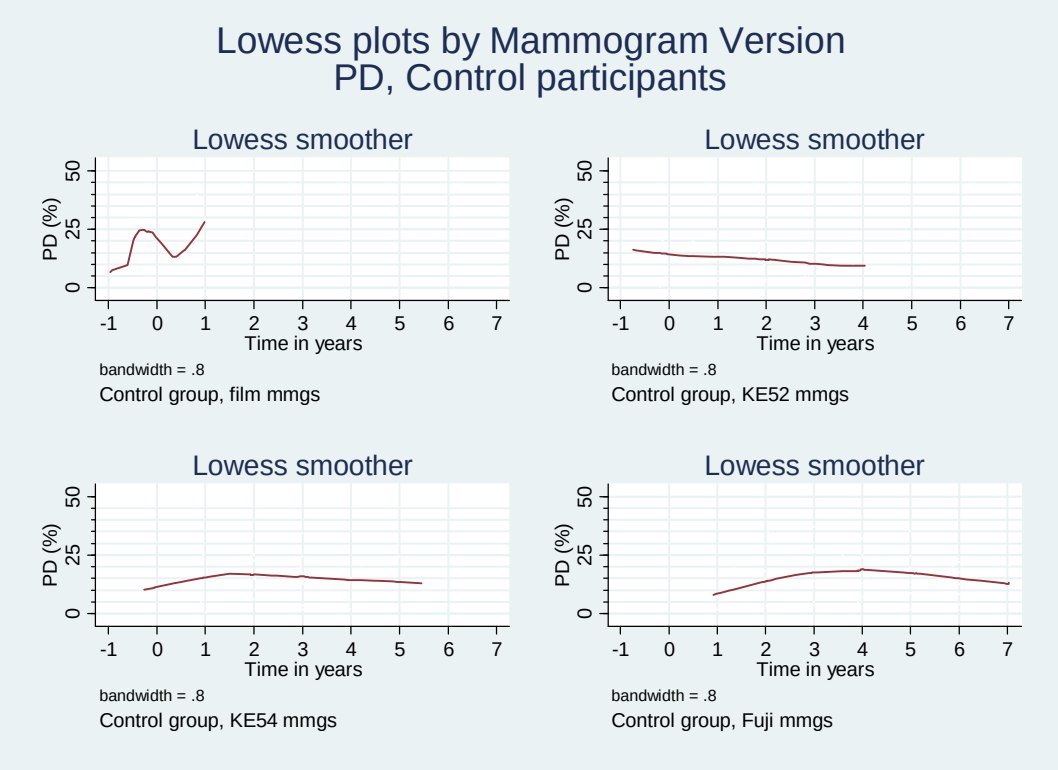


Figure 8-1 Lowess plots of longitudinal PD by mammogram Version, Control group

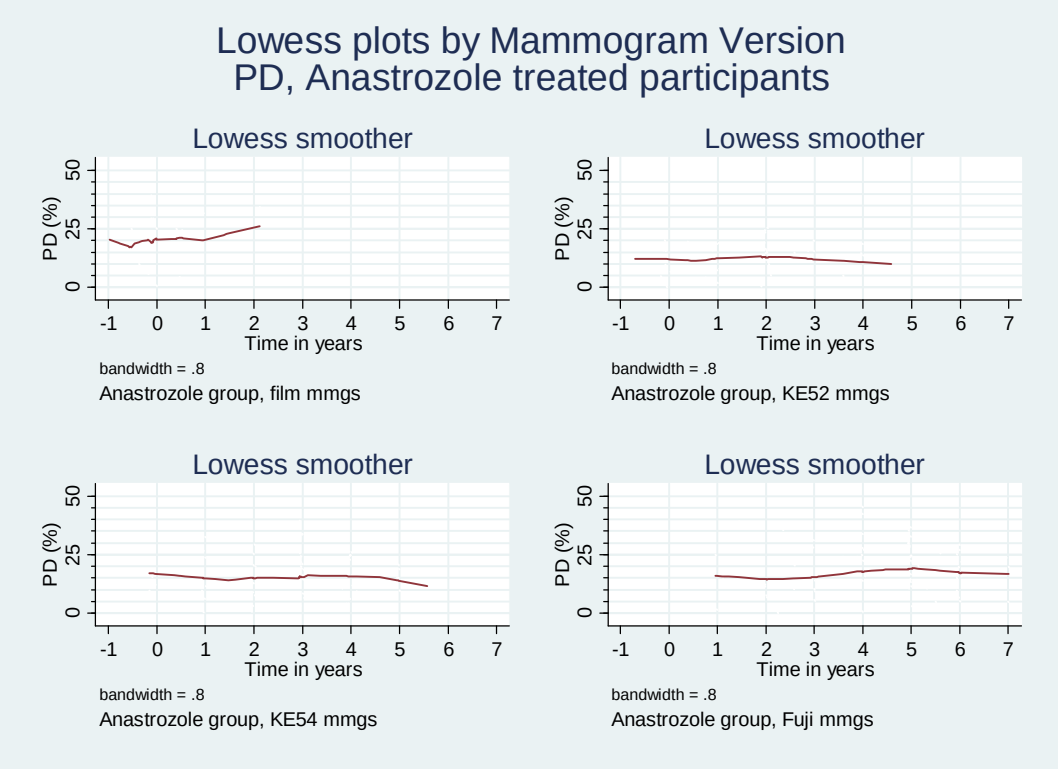


Figure 8-2 Lowess plots of longitudinal PD by mammogram Version, Anastrozole group

8.4.2 Primary analysis

The coefficients, standard errors (SE), p-values (P), and 95% confidence intervals (95% CI) for the Primary Model results for square root transformed PD and DA are tabulated in Table 8-5.

Anastrozole treatment vs no treatment was associated a non-significantly (p=0.13) lower average intercept of –0.33 in square root PD for the Primary model. Whilst a significant (p=0.03) mean annual decrease in square root transformed PD of –0.14 was observed between baseline and year 1 for the control group, the change for the anastrozole treated participants relative to controls over the same time period was not significant (0.02/year, p=0.8). Similarly, whilst a significant annual increase in square root transformed PD was observed for the control participants for years 1 to 5 (0.05/year, p=0.04), the corresponding difference in annual change for anastrozole treated participants relative to controls was not significant (–0.02/year, p=0.6). Change over time in square root transformed PD for years 5 to 7 was non-significant for both the control group and the difference in change for anastrozole treated participants relative to controls.

The results for DA were similar, except the annual decline in square root transformed DA from baseline to year 1 for the control group was marginally non-significant at the 5% level (p=0.07). No significant interaction between treatment group and annual change in DA during the three time periods (baseline to year 1, years 1 to 5, years 5 to 7) was observed. Hence anastrozole treatment was not significantly associated with increases or decreases, relative to the control group, in PD or DA for years 0 to 7 for the all mammogram Primary models.

The coefficient for age at randomisation showed the expected inverse (negative) relationship for PD, and was statistically significant (p=0.007). The relationship between age at randomisation and DA was also negative, but not significant (p=0.12). As noted previously, this may have occurred because the median age of the participants was 61 to 62 years, and the inverse

Table 8-5 Square root transformed Primary Model results for square root transformed PD DA, All mammograms (film+digital), 540 episodes

	Square root transformed Percent Density					Square root transformed Dense Area				
	ALL MAMMOGRAMS 120 participants 540 episodes (follow ups)									
Covariate	**Coefficient (β)**	**SE**	**P**	**95% CI**		**Coefficient (β)**	**SE**	**P**	**95% CI**	
Age at Randomisation (years)	-0.04	0.02	0.007	-0.07	-0.01	-0.38	0.25	0.12	-0.87	0.10
BMI (kg/m^2)	-0.12	0.02	<0.001	-0.15	-0.08	--	--	--	--	--
Age at Menopause (years)	0.06	0.02	0.003	0.02	0.10	0.88	0.28	0.001	0.34	1.4
Age First Birth, ≥30y vs <30y (ref)	1.06	0.33	0.001	0.41	1.70	21.4	6.5	0.001	8.7	34.1
Non-parous vs <30y (ref)	0.03	0.29	0.92	-0.54	0.60	6.3	4.3	0.15	-2.3	14.8
Mammogram Version, KE52 v Film(ref)	-0.76	0.12	<0.001	-1.00	-0.52	-10.3	1.5	<0.001	-13.4	-7.3
KE54 v Film (ref)	-0.48	0.13	<0.001	-0.74	-0.22	-6.5	1.8	<0.001	-10.0	-3.1
Fuji v Film (ref)	-0.29	0.15	0.06	-0.59	0.02	-4.6	2.2	0.033	-8.8	-0.36
Anastrozole (no=ref)	-0.33	0.22	0.13	-0.76	0.10	-2.8	3.3	0.39	-9.3	3.6
Intercept	5.42	0.30	<0.001	4.84	6.00	61.4	4.0	<0.001	53.6	69.2
	Annual change in MD (All mammograms)									
Baseline to Year 1	-0.14	0.07	0.03	-0.27	-0.01	-1.5	0.83	0.073	-3.1	0.14
Years 1 to 5	0.05	0.03	0.04	0.00	0.10	1.0	0.37	0.006	0.28	1.7
Years 5 to 7	-0.02	0.07	0.75	-0.16	0.12	0.30	1.0	0.77	-1.7	2.3
Anastrozole x Baseline to Year 1	0.02	0.08	0.80	-0.14	0.18	-0.17	1.1	0.87	-2.3	1.9
Anastrozole x Year 1 to Year 5	-0.02	0.03	0.56	-0.07	0.04	-0.39	0.3	0.27	-1.1	0.30
Anastrozole x Year 5 to Year 7	0.01	0.12	0.94	-0.22	0.24	-0.48	1.6	0.77	-3.7	2.7
	Random effects – estimates for between- and within-person change									
Between person variance										
Film mammograms	0.29	0.09	*	0.16	0.52	50.2	13.7	*	29.4	85.6
Fuji mammograms	0.15	0.04	*	0.09	0.24	27.9	7.0	*	17.2	45.5
Intercept	1.24	0.26	*	0.82	1.88	274.1	41.9	*	203.2	369.8
Within person correlation (rho)	0.47	0.14	*	0.23	0.72	0.25	0.11	*	0.10	0.52
Within person variance	0.09	0.03	*	0.05	0.16	12.3	2.2	*	8.7	17.4

P p-value; x interaction; y years ; ref reference category; * p<0.05 because the SE are one-half the size of the variance estimate or less

Table 8-6 Back transformed Primary Model results for PD (%) and DA (mm^2), All mammograms (film+digital), 540 episodes

	Back transformed Percent Density (PD %)					Back transformed Dense Area (DA mm^2)				
	ALL MAMMOGRAMS 120 participants 540 episodes (follow ups)									
Covariate	**Coefficient (β)**	**SE**	**P**	**95% CI**		**Coefficient (β)**	**SE**	**P**	**95% CI**	
Age at Rand. (years)	-0.46	0.17	0.007	-0.80	-0.13	-47	30	0.12	-106	12
BMI (kg/m^2)	-1.27	0.19	<0.001	-1.63	-0.91	--	--	--	--	--
Menopause (years)	0.63	0.21	0.003	0.22	1.05	109	34	0.001	42	177
Age First Birth, ≥30y vs <30y (ref)	12.57	3.67	0.001	4.65	21.32	3085	839	0.001	1142	5351
Non-parous vs <30y (ref)	0.31	3.22	0.92	-5.55	6.82	807	553	0.15	-273	2033
Mammogram Version, KE52 v Film(ref)	-7.67	1.33	<0.001	-9.82	-5.40	-1163	192	<0.001	-1464	-845
KE54 v Film (ref)	-4.99	1.46	<0.001	-7.50	-2.35	-762	219	<0.001	-1128	-372
Fuji v Film (ref)	-3.01	1.69	0.06	-6.02	0.18	-542	269	0.033	-1004	-44
Anastrozole (no=ref)	-3.47	2.43	0.13	-7.67	1.10	-340	415	0.39	-1054	457
Intercept	29.4	3.29	<0.001	23.5	36.0	3773	505	<0.001	2875	4793
Annual change in MD (All mammograms), model intercept is the reference value										
Baseline to Year 1	-1.48	0.71	0.03	-2.81	-0.11	-180	102	0.073	-371	17
Years 1 to 5	0.58	0.28	0.04	0.02	1.14	125	46	0.006	35	216
Years 5 to 7	-0.24	0.77	0.75	-1.71	1.26	37	126	0.77	-206	288
Anastrozole x Baseline to Year 1	0.22	0.89	0.80	-1.49	1.98	-21	133	0.87	-273	241
Anastrozole x Year 1 to Year 5	-0.17	0.29	0.56	-0.72	0.39	-47	43	0.27	-130	37
Anastrozole x Year 5 to Year 7	0.09	1.28	0.94	-2.34	2.63	-58	205	0.77	-441	344

P p-value; x interaction; These back transformed values are representative for a participant randomised at age 50y, BMI of 25, menopause at age 50y, and AFB <30y.

Table 8-7 Anastrozole group MD change, back transformed Primary Model results, all mammograms, 20% PD and DA 2700mm^2 reference values

Covariate	**Treatment group annual change in MD (All mammograms)**	
	PD (%) , reference value 20% PD	**DA mm^2, reference value 2700mm^2 (27cm^2)**
Anastrozole x Baseline to Year 1	0.18%/year	-18 mm^2/year
Anastrozole x Year 1 to Year 5	-0.14%/ year	-40 mm^2/year
Anastrozole x Year 5 to Year 7	0.08%/year	-50 mm^2/year

association between age and PD (and to a lesser extent, DA) weakens with age. Body mass index showed the expected inverse and statistically significant relationship with PD. Age at menopause and age at first birth ≥30 years relative to women <30 years showed the expected positive relationship with PD and DA. For age at first birth, the coefficient for non-parous women (relative to first birth at <30 years of age) was positive but not significant for both PD and DA. Each digital mammogram Version was negatively associated with PD and DA relative to film; these associations were all significant except for PD and Fuji mammograms which were marginally non-significant at the 5% level (p=0.06).

The variance for each random effect (RE) in the PD and DA models was significant (the coefficients were at least twice as large as the corresponding SE). This significant variability of the RE intercepts showed that further covariates could be added to the models to help explain both between and within person variability.

FE model coefficients were back transformed and are tabulated in Table 8-6, above. The coefficients in this table are representative of the average response for a CMN IBIS-II participant who was randomised at age 50, experienced menopause at age 50, and who has a BMI of 25. Longitudinal change by treatment group for back transformed PD is shown in Figure 8-3. Very little difference in longitudinal change is noted between the groups.

The difference of the mean FE intercept for back transformed PD (in %) for anastrozole treated participants relative to controls was –3.5% (SE 2.4%), whilst the corresponding (adjusted) for the back transformed anastrozole intercept for DA was –340 mm^2 (–3.4cm^2). The use of the relatively high intercept PD and DA values (e.g. almost 30% PD) during back transformation in Table 8-6 has slightly inflated the estimate of the intercepts; substitution of more representative reference values for PD (20%, Table 6-6 and Table 8-3) and DA (2700mm^2) yields (adjusted) anastrozole intercept values of –3% for PD and –283mm^2 for DA. These latter back transformed estimates are larger than the (raw 'unadjusted') baseline differences of –1.6% for

PD and –73mm^2 for DA, Table 8-3. The back transformed anastrozole intercepts for PD and DA in Table 8-6 also may be larger than the treated vs control difference in median PD and DA at baseline because the model uses film mammograms as the base category (Version) which has higher median and Q1 to Q3 values for PD and DA (e.g. Table 6-6) compared to median and Q1 to Q3 PD and DA for all mammograms, Table 8-3.

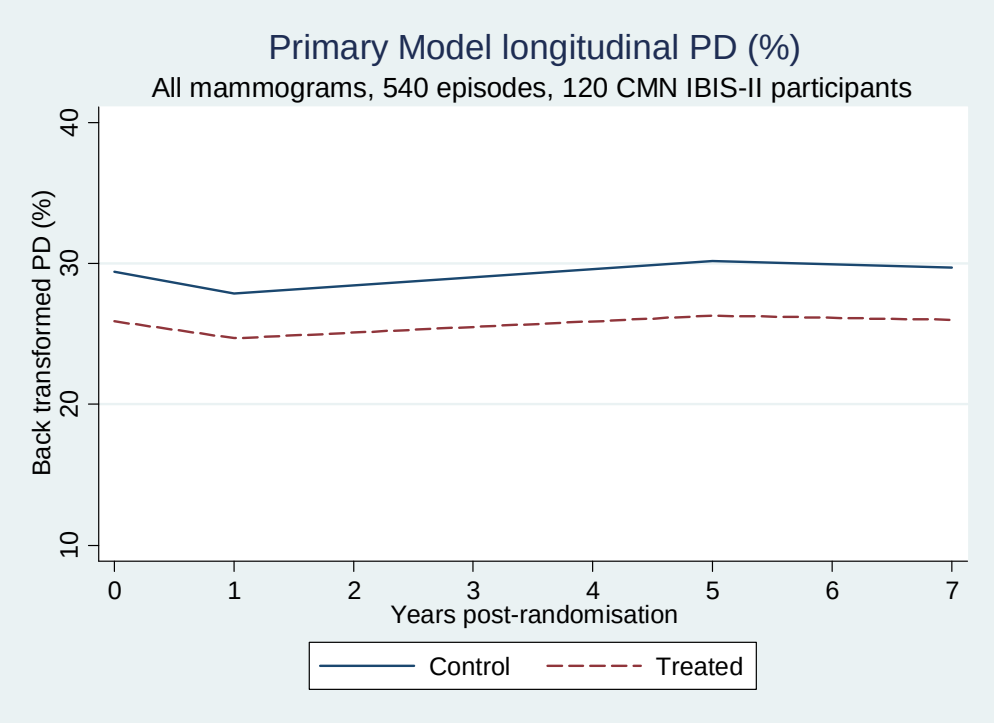


Figure 8-3 Longitudinal back transformed PD by Treatment Group, all mammograms (540 episodes)
The estimated growth curves for the treated and control groups are almost parallel to each other, reflecting the small annual PD change for the treatment group relative to the control group

8.4.3 Regression diagnostics

8.4.3.1 *Linear prediction (xb) vs observed (measured) MD*

The Primary model for PD appeared to fit the observed (measured) values reasonably well, Figure 8-4. The predicted values (xb) from the FE part of the mixed model on the x-axis tend to be similar to their observed counterparts on the y-axis; however the line of equality does not evenly bisect the plotted data. Square root transformed PD predicted values >5 tend to be smaller than the observed values. The model may therefore under predict PD for women with >25% density. The model may also slightly over predict PD at smaller (<4) observed values.

Little heteroskedasticity appears to be present.

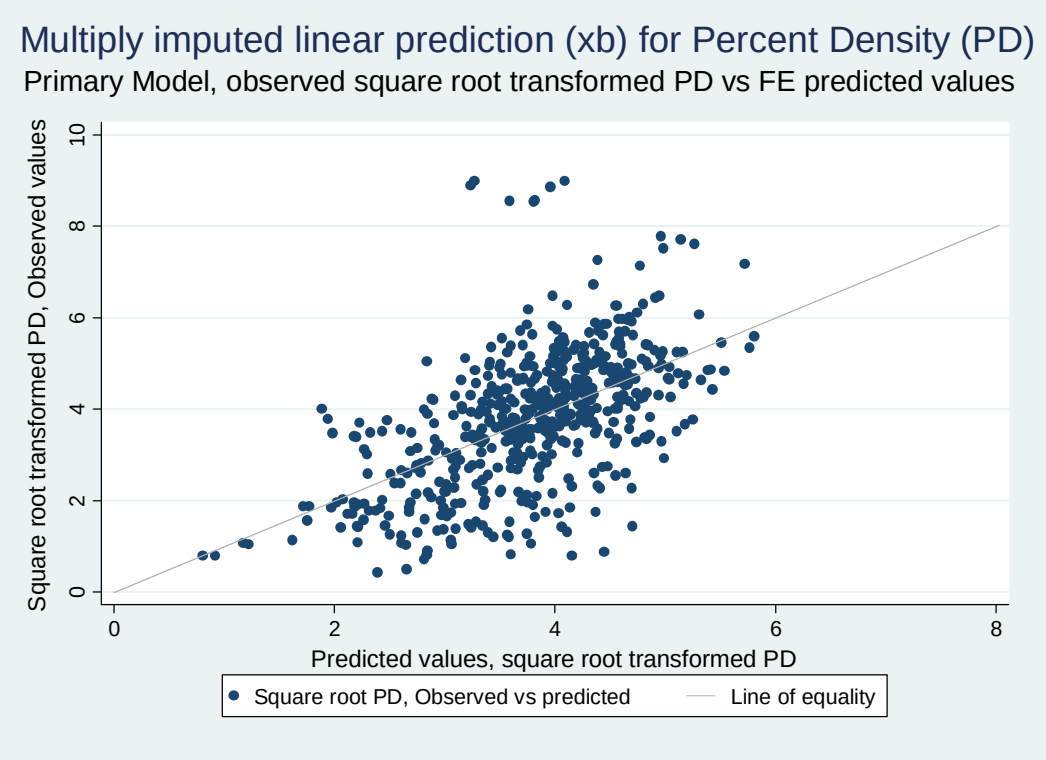


Figure 8-4 PD Primary Model, linear prediction (xb) vs observed (measured) values

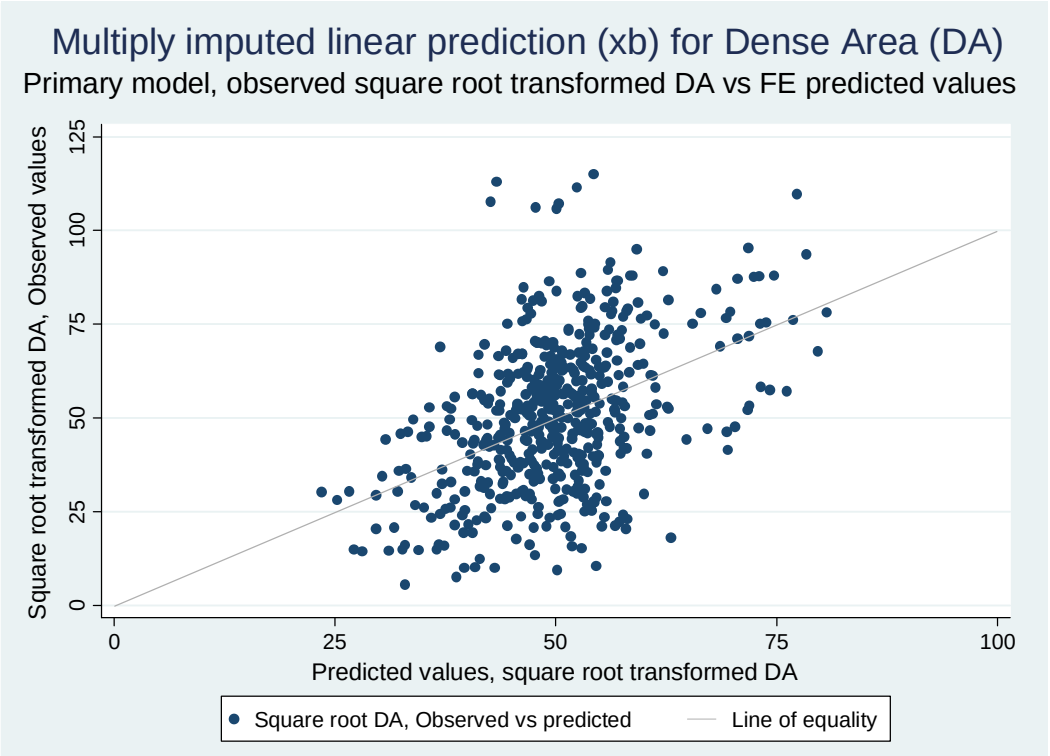


Figure 8-5 DA Primary Model, linear prediction (xb) vs observed (measured) values

There are a few outliers where predicted and observed values for the participant with very high density (PD ~80%, observed square root transformed values of ~9) are particularly mismatched. This indicates the Primary model is likely to under predict PD for women with high PD (≥50%). The model may not be a good fit for participants with highly dense breasts.

The Primary Model for DA likewise appeared to fit the observed values for (square root transformed) DA reasonably well without much evidence of heteroskedasticity, Figure 8-5. However, the scatter plot of Figure 8-5 appears to have curved clustering of data in the right half of the central cluster (at predicted values >40 (on the x-axis)) which tapers out at around 60; there seems to be a compression of DA for values which should be more evenly distributed between values of 40 to 75 on the x-axis. In contrast the PD plot scatter plot (Figure 8-4) is more evenly distributed over most more of its higher predicted values (>4). The reasons for this compression and curvature of the DA central cluster are not clear. This curvature may be due to chance but may indicate a higher order (non-linear) relationship in DA model. The PD and DA models differ, because some of the factors associated with PD (primarily BMI) are not associated as strongly with DA as they are PD; as mentioned previously, women with high PD do not always have high DA, and the reverse can be true as well. The model seems to fit the observed DA reasonably well for most women with observed square root transformed DA values of <60, which equates to <3,600 mm^2 or <36cm^2 of dense area on the mammograms. The sparseness for predicted DA between 50 and 70 may be due to the relatively small sample size of women in the data (n=120), especially those with DA (>5000mm^2, or 71mm in square root transformed units). As for the PD Primary model, the predicted DA for the woman with high DA (whose square root transformed DA is >75mm (5,600mm^2)) is underestimated by the DA Primary model.

8.4.3.2 *MD prediction comparisons of MI and non-MI models*

The MI and non-MI linear predictions are visually similar for both the PD and DA Primary models, Figure 8-6 and Figure 8-7. No systematic or unusual differences appear to exist between the pairs of graphs. These observed vs predicted graphs are quite similar, thus the fitted RE residuals for the MI models would likely resemble the fitted RE residuals for the non-MI Primary models, had they been possible to predict within the statistical program. This is partly because age at menopause is only imputed for 27 of 120 (23%) of total participants.

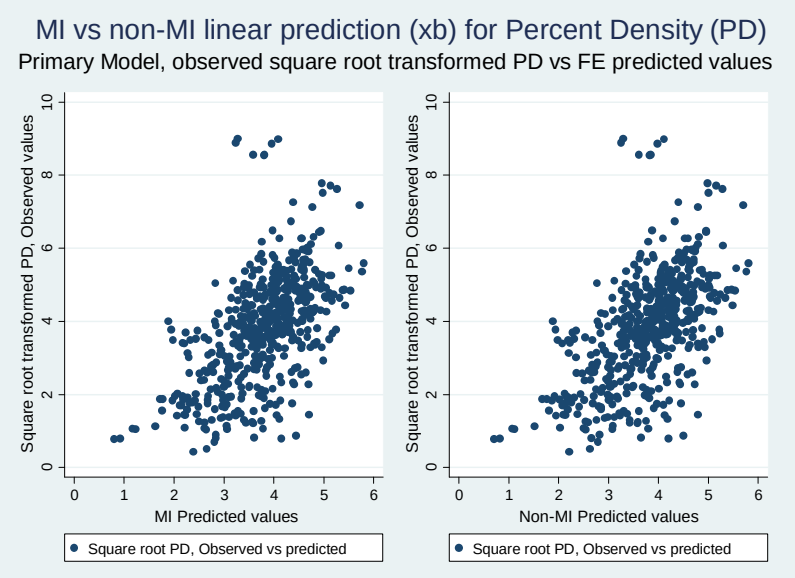


Figure 8-6 PD Primary model, MI (left) vs non-MI (right) predicted vs observed values

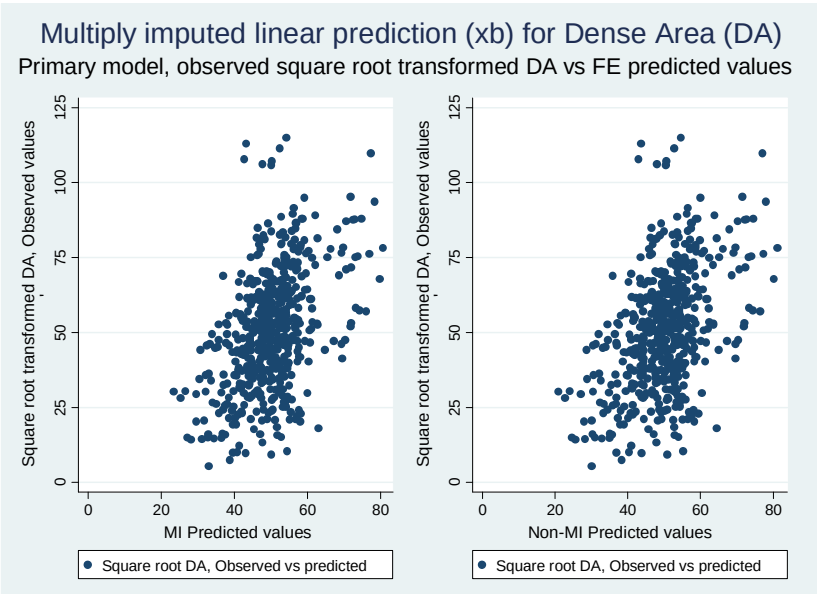


Figure 8-7 DA Primary model, MI (left) vs non-MI (right) predicted vs observed values

8.4.3.3 *Fitted prediction vs measured MD*

The Primary model best predictions for the outcome based on both the RE and FE (the best linear unbiased predictions (BLUPs) of the random effects plus the FE linear predictions) provide a very good fit for the observed values, Figure 8-8 and Figure 8-9. The PD Primary model may still slightly under predict the fitted values at lower observed values of PD and slightly over predict at higher PD. The fitted values for DA are slightly more evenly distributed around the line of equality than are the ones for PD over the entire range of observed values; no trend to under or over predict is visible. The mixed model appears to fit well for the fitted Primary outcomes.

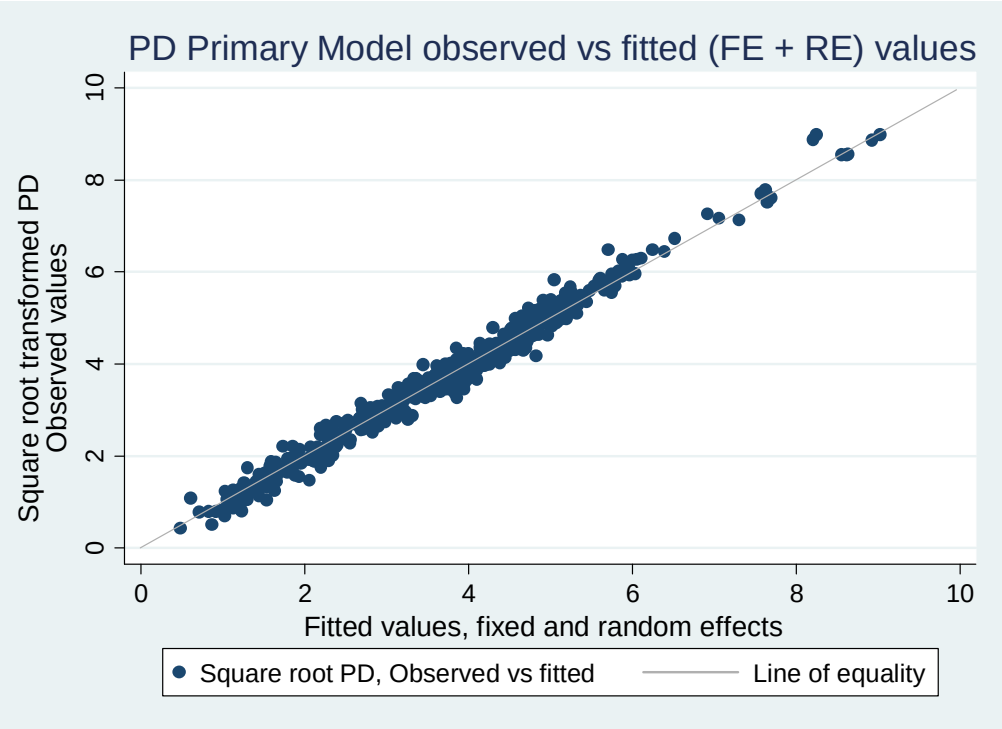


Figure 8-8 PD Primary model fitted (FE + RE) vs observed values

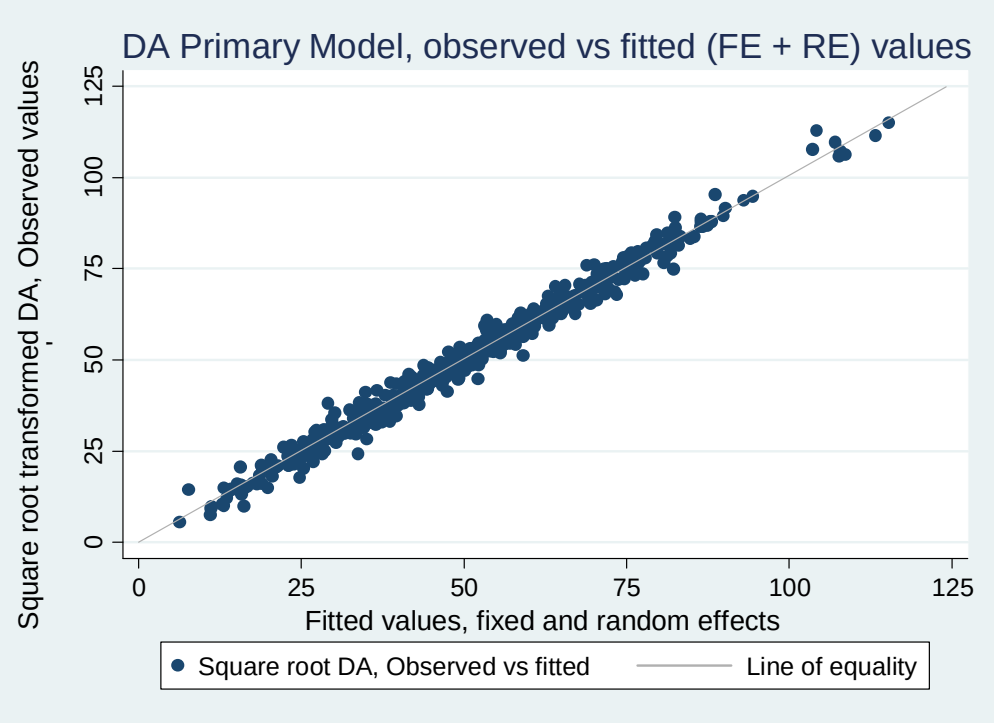


Figure 8-9 DA Primary model fitted (FE + RE) vs observed values

8.4.3.4 *Normality assessment of the fitted model RE distributions*

Histograms of the best linear unbiased predictions (BLUPs) for the between person (level 2) random effects and model residuals (at level 1) for the PD (Figure 8-10) and DA (Figure 8-12) Primary models show that the distribution of the BLUPs for each RE and for the within person residuals are approximately normally and symmetrically distributed. Participants with high PD and/or high DA have caused a small number of right skewed BLUPs; the peak at 0 for the film BLUPs are due to few film mammograms in the dataset. The film mammogram RE residuals for the PD and DA Primary models are shown in the upper row of Figure 8-11 and Figure 8-13. The flattened area in the inverse normal plot (left column, upper row) and the high number of 0 values on the standardised residual plots (right column upper row) also reflect the few participants with film mammograms in the dataset. The remainder of the normal probability and standardised residuals plots for PD and DA (Fuji mammogram Version RE, between person and within person residuals) demonstrate that the residuals are approximately normally distributed.

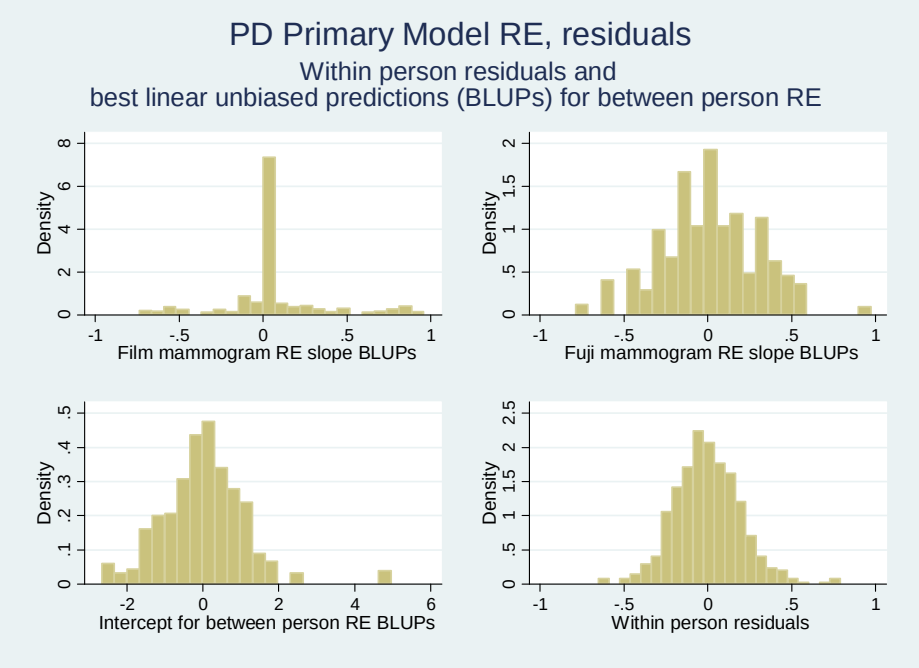


Figure 8-10 PD Primary model histograms of between person RE and within person residuals
The histograms of the predicted RE and residuals show approximately normal distributions
The large peak at 0 for the film Version RE is due to few film episodes in the data set

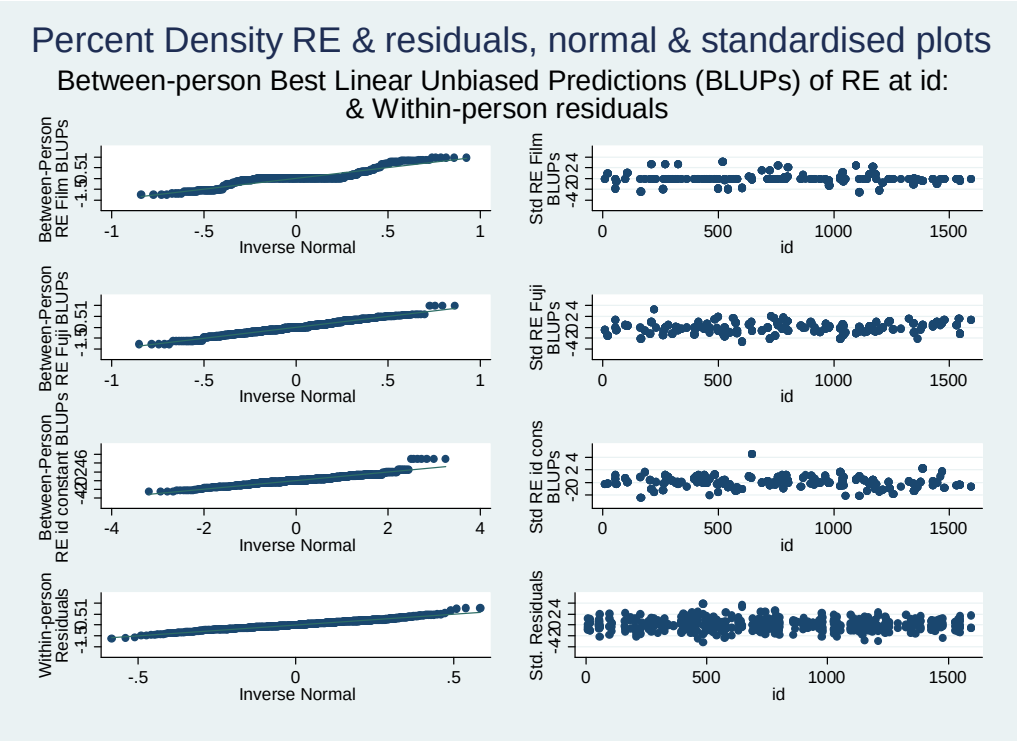


Figure 8-11 PD Primary model RE and residual assumption checking
Left column– normal probability plots for the best linear unbiased predictions (BLUPs) of the between person random effects and within person residuals; Right column– standardised residual plots of the RE BLUPs and residuals. Between person (level 2) RE: Top row– Film Version; Second row– Fuji Version; Third row– between person RE intercept. Bottom row– Within person (level 1) residuals

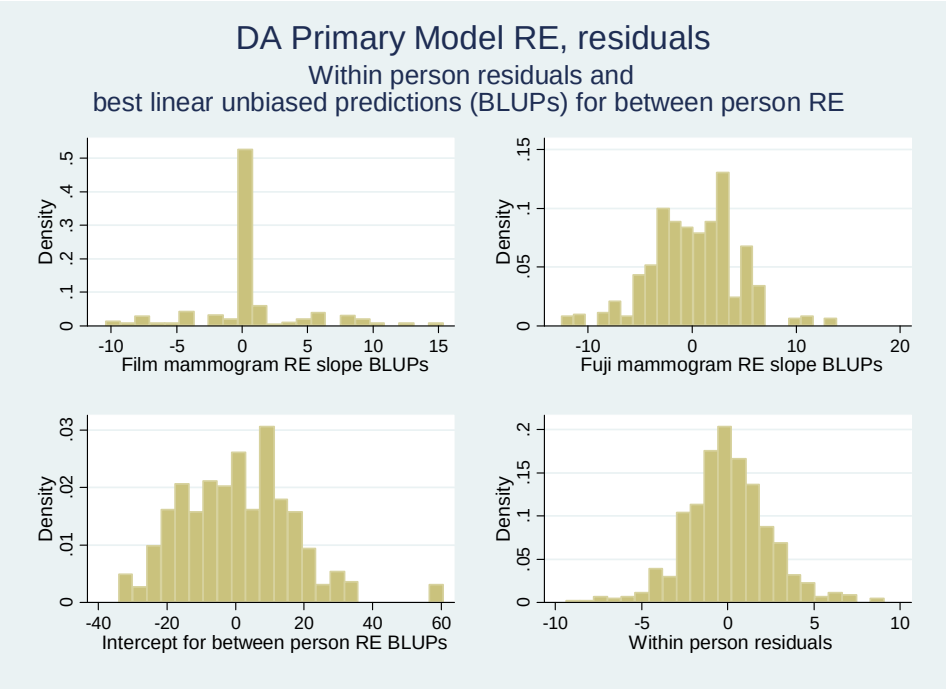


Figure 8-12 DA Primary model histograms of between person RE and within person residuals
The histograms of the predicted RE and residuals show approximately normal distributions
The large peak at 0 for the film Version RE is due to few film episodes in the data set

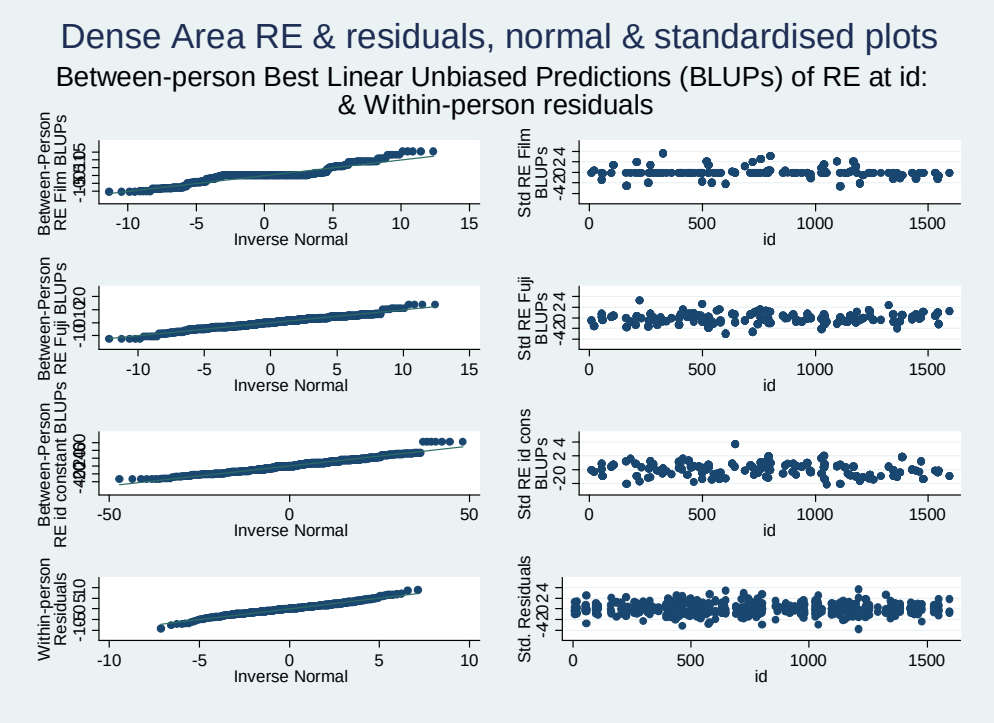


Figure 8-13 DA Primary model RE and residual assumption checking
Left column– normal probability plots for the best linear unbiased predictions (BLUPs) of the between person random effects and within person residuals; Right column– standardised residual plots of the RE BLUPs and residuals. Between person (level 2) RE: Top row– Film Version; Second row– Fuji Version; Third row– between person RE intercept. Bottom row– Within person (level 1) residuals

8.4.4 Subgroup and sensitivity analyses

This section contains the subgroup analyses by mammogram Type (film only and digital only models), as well as other sensitivity analyses.

8.4.4.1 *Primary model, by mammogram Type (film only, digital only)*

8.4.4.1.1 Film mammogram only Primary Model

The results for the PD and DA Primary models with film mammograms only are shown in Table 8-8. Sixty-three mammographic episodes from 45 participants were included in the film mammogram only subset used to generate the models. Back transformed FE model outcomes for PD (in %) and DA (in mm^2) are shown in Table 8-9. The back transformed values correspond to the average response for a representative CMN IBIS-II participant who was randomised at age 50, underwent menopause at age 50 with a BMI of 25 kg/m^2.

The intercept for the back transformed PD values is 33%, and was calculated by squaring the intercept for the square root transformed PD model (5.71). The value for the back transformed intercept implies that average baseline PD for the representative participant (randomisation and menopause at age 50y and BMI of 25) with film mammograms is 33%. This is slightly higher than the back transformed value of 29% for the representative participant in the all mammograms mixed model, Table 8-6. Film mammograms have higher PD on average than digital mammograms in this dataset, which explains the difference in intercept between the film only and all mammograms models.

Interestingly, the PD interaction term of anastrozole treatment with the term for time between baseline to year 1 was significant (p=0.017) with coefficient values of –0.33/year for square root transformed PD and –3.7% /year for back transformed PD (%). The interaction term between anastrozole treatment and change in DA over the first year was also significant (p=0.016); the interaction term coefficient for square root transformed DA was –4.6/year and –547 mm^2/year

Table 8-8 Square root transformed Primary Model of PD and DA, film mammograms only (63 episodes)

	Square root transformed Percent Density					Square root transformed Dense Area				
	FILM MAMMOGRAMS ONLY 45 participants, 63 total episodes (follow ups)†									
Covariate	**Coefficient (β)**	**SE**	**P**	**95% CI**		**Coefficient (β)**	**SE**	**P**	**95% CI**	
Age at Randomisation (years)	-0.03	0.04	0.39	-0.11	0.04	-0.24	0.58	0.68	-1.4	0.90
BMI (kg/m^2)	-0.20	0.05	<0.001	-0.29	-0.10	--	--	--	--	--
Age at Menopause (years)	0.03	0.05	0.63	-0.08	0.13	0.25	0.74	0.74	-1.2	1.7
Age First Birth, ≥30y vs <30y (ref)	0.94	0.69	0.17	-0.41	2.29	26.5	17.7	0.14	-8.3	61.3
Non-parous vs <30y (ref) [1]	--	--	--	--	--	--	--	--	--	--
Anastrozole (no=ref)	-0.30	0.44	0.49	-1.16	0.56	-5.40	6.8	0.43	-18.7	7.9
Intercept	5.7	0.54	<0.001	4.7	6.8	62.1	7.3	<0.001	47.7	76.5
Annual change in MD (All mammograms)										
Baseline to Year 1	0.01	0.09	0.10	-0.17	0.19	0.50	1.11	0.65	-1.7	2.7
Years 1 to 5 [2]	--	--	--	--	--	--	--	--	--	--
Years 5 to 7 [2]	--	--	--	--	--	--	--	--	--	--
Anastrozole x Baseline to Year 1	-0.33	0.14	0.017	-0.60	-0.06	-4.6	1.9	0.016	-8.29	-0.86
Anastrozole x Year 1 to Year 5 [2]	--	--	--	--	--	--	--	--	--	--
Anastrozole x Year 5 to Year 7 [2]	--	--	--	--	--	--	--	--	--	--
Random effects – estimates for between- and within-person change										
Between person variance										
Intercept	2.2	0.63	*	1.25	3.84	0.0064	NC	ns		
Within person correlation (rho)	0.014	0.046	ns	<0.0001	0.90	0.99	1.5	ns	2.1E-98	1
Within person variance	0.039	0.012	*	0.021	0.071	506	58457	ns	2.77E-96	9.2E+100

P p-value; x interaction; ref reference category; ns not significant; NC not calculated (by the statistical program); y years;

† 45 participants contributed 63 mammographic episodes to this analysis (~1.4 episodes per participant)

* p<0.05, SE are ≤0.5 times the variance estimate

[1] results for non-parous women are not applicable because there were no non-parous women with film mammograms

[2] only a single film episode was available after the first year of follow up; hence results for years 1 to 5 were omitted. Results for years 5 to 7 are not applicable

Table 8-9 Back transformed Primary Model of PD (%) and DA (mm^2), film mammograms only (63 episodes)

	Back transformed Percent Density (%)					Back transformed Dense Area (mm^2)				
	FILM MAMMOGRAMS ONLY 45 participants 63 total episodes (follow ups)†									
Covariate	**Coefficient (β)**	**SE**	**P**	**95% CI**		**Coefficient (β)**	**SE**	**P**	**95% CI**	
Age at Randomisation (years)	-0.37	0.43	0.39	-1.2	0.47	-30.3	73.0	0.68	-171	113
BMI (kg/m^2)	-2.2	0.54	<0.001	-3.2	-1.2	--	--	--	--	--
Age at Menopause (years)	0.29	0.60	0.63	-0.89	1.49	30.6	91.9	0.74	-147	213
Age First Birth, ≥30y vs <30y (ref)	11.6	8.3	0.17	-4.5	31.4	3989	2517	0.14	-960	11356
Non-parous vs <30y (ref)[1]	--	--	--	--	--	--	--	--	--	--
Anastrozole (no=ref)	-3.3	5.2	0.49	-11.9	6.7	-641	888	0.43	-1971	1043
Intercept	32.6	6.4	<0.001	21.7	45.8	3854	965	<0.001	2275	5846
	Annual change in MD (All mammograms)									
Baseline to Year 1	0.11	1.05	0.10	-1.9	2.2	62.3	139.5	0.65	-206	340
Years 1 to 5[2]	--	--	--	--	--	--	--	--	--	--
Years 5 to 7[2]	--	--	--	--	--	--	--	--	--	--
Anastrozole x Baseline to Year 1	-3.7	1.6	0.017	-6.5	-0.68	-546.9	238.9	0.016	-960	-106
Anastrozole x Year 1 to Year 5[2]	--	--	--	--	--	--	--	--	--	--
Anastrozole x Year 5 to Year 7[2]	--	--	--	--	--	--	--	--	--	--

P p-value; x interaction; ref reference category; y years;

† 45 participants contributed 63 mammographic episodes to this analysis (~1.4 episodes per participant)

[1] results for non-parous women are not applicable because there were no non-parous women with film mammograms

[2] only a single film episode was available after the first year of follow up; hence results for years 1 to 5 were omitted. Results for years 5 to 7 are not applicable

These back transformed values are representative for a participant randomised at age 50y, BMI of 25, menopause at age 50y, and AFB <30y

Table 8-10 Anastrozole group MD change, back transformed Primary Model results, film mammograms, 20% PD and DA 2700mm^2 reference values

Covariate	**Treatment group annual change in MD (Film mammograms only)**	
	PD (%) , reference value 20% PD	**DA mm^2, reference value 2700mm^2 (27cm^2)**
Anastrozole x Baseline to Year 1	-2.9%/year	-455 mm^2/year

(–5.5cm^2/year) for back transformed DA. Treatment with anastrozole appeared to be associated with a reduction in both PD and DA during the first year of therapy for IBIS-II imaging undertaken with film mammograms.

The baseline to year 1 annual change in PD coefficient for the control group was non-significant (p=0.1) but positive. This was rather unexpected. The back transformed, slight change of +0.11% per year is not likely to be clinically significant however. The corresponding, non-significant (p=0.65) annual control group increase in DA (62mm^2 (= 0.6cm^2)) is similarly unlikely to be clinically significant. Longitudinal change by treatment group for back transformed PD for the film only Primary Model is shown in Figure 8-14.

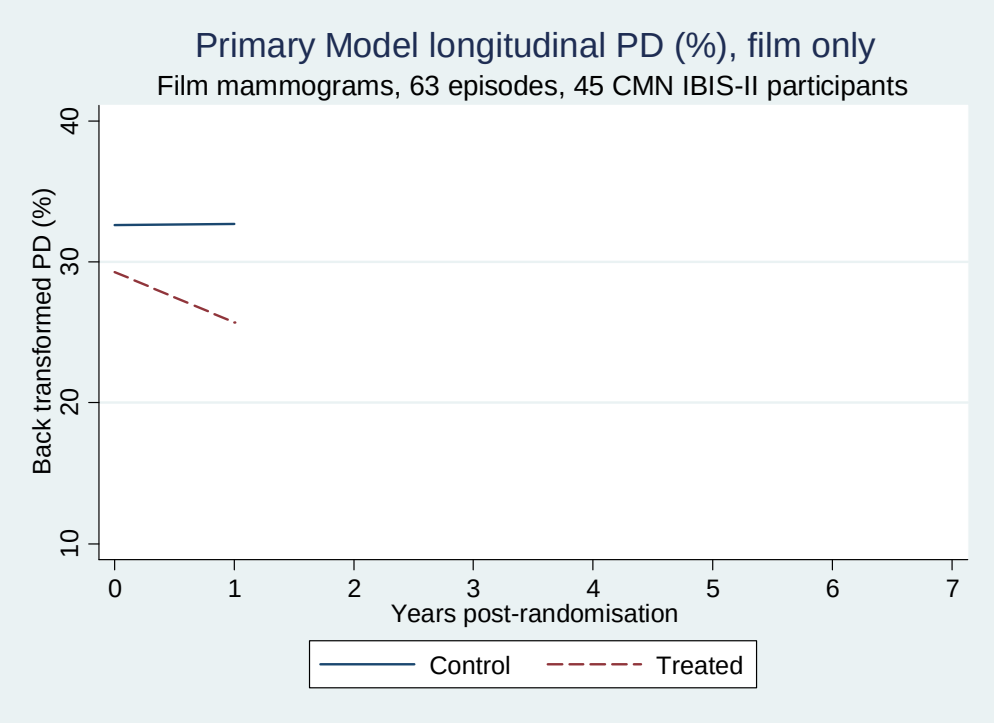


Figure 8-14 Back transformed longitudinal PD by Treatment Group, film only Primary Model
The film only model estimated growth curves diverge from baseline to year 1 for the treatment groups, reflecting the significant decline in PD for the treated group relative to the control group.

The remainder of the coefficient signs and magnitude are generally in keeping with the expected direction of the relationship. A negative relationship between age at randomisation and anastrozole treatment is seen for both PD and DA; BMI is also negatively associated with PD. Age at menopause and age at first birth are positively associated with PD and DA. All these

relationships are not statistically significant however ($p \geq 0.05$), with the exception of BMI for PD ($p < 0.001$). The non-significance for most coefficients is likely due to the small sample size (63 episodes from 45 participants).

The film mammograms PD RE variance estimates are significant for the between person and within person variance intercepts, Table 8-8. This implies that further covariates could be modelled to explain the variability both between and within each participant for PD. The within person correlation (rho) for PD was not significant, which is perhaps due to the small sample size which had few repeated measurements within the film mammogram subset of the data. All the DA RE estimates are not significant, Table 8-8. A standard error (SE) was not calculated for the between person variance for the DA intercept. The DA between person RE variance is very small (0.0064), hence a SE may not have been possible to calculate (i.e. it exceeded the limits of the statistical program). Although the non-significant variance estimate for the DA within person residual could mean that most of the within person variability has been explained by the model, the extremely wide 95% CI for the within person correlation belies this and suggests as per the film only PD model that insufficient data was present to correctly estimate a the variance-covariance structure for the within person residuals. The exponential covariance structure provides an estimate of the correlation for the within person measurements (episodes), but a correlation cannot be calculated if insufficient repeated measurements are in the dataset. The default covariance structure (the independent structure, i.e. no correlation) would have better suited the DA film only Primary model.

8.4.4.1.2 Digital mammogram only Primary Model

The results for the PD and DA Primary models with digital mammograms only are shown in Table 8-11. Four hundred and seventy-seven (477) mammographic episodes from all 120 participants were included in the digital mammogram only data subset used to generate the models. Back transformed FE model outcomes for PD (in %) and DA (in mm^2) are shown in

Table 8-12. The back transformed values correspond to the average response for a representative CMN IBIS-II participant who was randomised at age 50, underwent menopause at age 50 with a BMI of 25 kg/m^2. Longitudinal change by treatment group for back transformed PD is shown in Figure 8-15. Back transformed values using the representative values of 20% PD and 2700mm^2 DA are shown in Table 8-13.

The intercept for the back transformed PD values is 21.4%, which is lower than the back transformed intercept for the film mammogram model (33%) and back transformed value of 29% for the all mammograms Primary model, Table 8-6. Digital mammograms have lower PD on average than film mammograms in this dataset, which explains the difference in intercept between the digital only compared to the film only and all mammograms models.

Whilst the baseline to year 1 and years 1 to 5 coefficients for annual change in PD and years 1 to 5 coefficient for DA annual change in the control group were statistically significant ($p<0.05$), the interaction terms between anastrozole treatment and change over time were not significant. The interaction term for change in DA for anastrozole treated participants relative to control participants for years 1 to 5 (back transformed value of –66mm^2/year) had the lowest p value ($p=0.14$). All other anastrozole and change over time interaction terms had p-values >0.2.

Both the PD and DA interaction terms for baseline to year 1 were (non-significantly) positive, whilst both interaction terms for years 1 to 5 were (non-significantly) negative. If anastrozole treatment truly is associated with a decrease over time relative to control treatment for the first and/or subsequent years of treatment, this effect appears to be masked by the instability of the MD measurements and/or representation of MD on the mammograms. For instance, although the model accounts for the average difference between the different digital Versions, the increase in PD between a baseline KE52 mammogram and a higher MD KE54 mammogram at year 1 due to the change in mammogram Version may on average be higher than is accounted

Table 8-11 Square root transformed Primary Model of PD and DA, digital mammograms only (477 episodes)

	Square root transformed Percent Density					Square root transformed Dense Area				
	DIGITAL MAMMOGRAMS ONLY 120 participants, 477 total episodes (follow ups)†									
Covariate	**Coefficient (β)**	**SE**	**P**	**95% CI**		**Coefficient (β)**	**SE**	**P**	**95% CI**	
Age at Randomisation (years)	-0.04	0.02	0.011	-0.07	-0.01	-0.37	0.25	0.13	-0.85	0.11
BMI (kg/m^2)	-0.12	0.02	<0.001	-0.15	-0.08	--	--	--	--	--
Age at Menopause (years)	0.06	0.02	0.003	0.02	0.09	0.86	0.27	0.001	0.33	1.39
Age First Birth, ≥30y vs <30y (ref)	1.2	0.40	0.003	0.39	1.94	21.9	6.67	0.001	8.86	35.0
Non-parous vs <30y (ref)	0.08	0.27	0.77	-0.44	0.60	6.55	4.13	0.11	-1.54	14.6
Mammogram Version, KE54vKE52(ref)	0.24	0.04	<0.001	0.16	0.32	-6.09	2.65	0.02	-11.3	-0.90
Fuji v KE52 (ref)	0.41	0.08	<0.001	0.26	0.57	-2.67	2.76	0.33	-8.08	2.74
Anastrozole (no=ref)	-0.36	0.22	0.11	-0.79	0.08	-3.07	3.23	0.34	-9.40	3.27
Intercept	4.6	0.29	<0.001	4.1	5.2	56.8	4.17	<0.001	48.7	65.0
Annual change in MD (All mammograms)										
Baseline to Year 1	-0.16	0.08	0.048	-0.32	0.00	-1.50	0.97	0.12	-3.41	0.40
Years 1 to 5	0.08	0.03	0.002	0.03	0.14	1.35	0.38	<0.001	0.60	2.10
Years 5 to 7	-0.02	0.07	0.80	-0.15	0.12	0.35	0.99	0.73	-1.60	2.29
Anastrozole x Baseline to Year 1	0.09	0.10	0.35	-0.10	0.28	0.47	1.22	0.70	-1.93	2.87
Anastrozole x Year 1 to Year 5	-0.04	0.03	0.23	-0.09	0.02	-0.58	0.39	0.14	-1.35	0.18
Anastrozole x Year 5 to Year 7	0.02	0.12	0.85	-0.20	0.25	-0.31	1.61	0.85	-3.48	2.85
Random effects – estimates for between- and within-person change										
Between person variance										
Fuji mammograms	0.15	0.04	*	0.09	0.25	28.3	7.2	*	17.2	46.6
Intercept	1.2	0.27	*	0.81	1.9	270.8	42.1	*	199.7	367.2
Within person correlation (rho)	0.50	0.16	*	0.23	0.78	0.28	0.145	ns	0.08	0.61
Within person variance	0.10	0.03	*	0.05	0.19	12.4	2.71	*	8.1	19.0

P p-value; x interaction; ns not significant; ref reference category; y years; * p<0.05 because the SE are ≤0.5 times the variance estimate

† 120 participants contributed 477 episodes to this analysis (~4 episodes per participant)

Table 8-12 Back transformed Primary Model of PD (%) and DA (mm^2), digital mammograms only (477 episodes)

	Back transformed Percent Density (%)					Back transformed Dense Area (mm^2)				
	DIGITAL MAMMOGRAMS ONLY 120 participants 477 total episodes (follow ups)									
Covariate	**Coefficient (β)**	**SE**	**P**	**95% CI**		**Coefficient (β)**	**SE**	**P**	**95% CI**	
Age at Randomisation (years)	-0.38	0.15	0.011	-0.67	-0.09	-42.3	28.0	0.13	-96.4	12.3
BMI (kg/m^2)	-1.1	0.16	<0.001	-1.4	-0.76	--	--	--	--	--
Age at Menopause (years)	0.51	0.17	0.0030	0.17	0.86	98.6	30.5	0.001	38.1	160
Age First Birth, ≥30y vs <30y (ref)	12.1	3.8	0.0030	3.7	21.7	2973	802	0.001	1085	5203
Non-parous vs <30y (ref)	0.71	2.5	0.77	-3.9	5.9	788	486	0.11	-173	1879
Mammogram Version, KE54vKE52(ref)	2.3	0.39	<0.001	1.5	3.1	-296	321	0.33	-853	319
Fuji v KE52 (ref)	4.0	0.73	<0.001	2.5	5.6	-118	334	0.71	-713	540
Anastrozole (no=ref)	-3.2	2.1	0.11	-6.7	0.73	-339	378	0.34	-980	382
Intercept	21.4	2.8	<0.001	16.5	27.0	3230	492	0.00	2367	4226
	Annual change in MD (All mammograms)									
Baseline to Year 1	-1.5	0.76	0.048	-2.9	-0.01	-169	111	0.12	-376	46
Years 1 to 5	0.8	0.25	0.0020	0.3	1.3	155	43.6	<0.001	68.6	243
Years 5 to 7	-0.2	0.63	0.80	-1.4	1.1	39.4	114	0.73	-179	266
Anastrozole x Baseline to Year 1	0.8	0.90	0.35	-0.9	2.7	53.2	141	0.70	-216	334
Anastrozole x Year 1 to Year 5	-0.3	0.27	0.23	-0.9	0.2	-65.7	44.5	0.14	-151	21.0
Anastrozole x Year 5 to Year 7	0.2	1.08	0.85	-1.8	2.4	-35.4	186	0.85	-383	332

P p-value; x interaction; ns not significant; ref reference category; y years;

These back transformed values are representative for a participant randomised at age 50y, BMI of 25, menopause at age 50y, and AFB <30y

† 120 participants contributed 477 episodes to this analysis (~4 episodes per participant)

Table 8-13 Anastrozole group MD change, back transformed Primary Model results, digital mammograms only, 20% PD and DA 2700mm^2 reference values

Covariate	**Treatment group annual change in MD (Digital mammograms only)**	
	PD (%) , reference value 20% PD	**DA mm^2, reference value 2700mm^2 (27cm^2)**
Anastrozole x Baseline to Year 1	0.8%/year	48 mm^2/year
Anastrozole x Year 1 to Year 5	-0.3%/ year	-60 mm^2/year
Anastrozole x Year 5 to Year 7	0.2 %/year	-33 mm^2/year

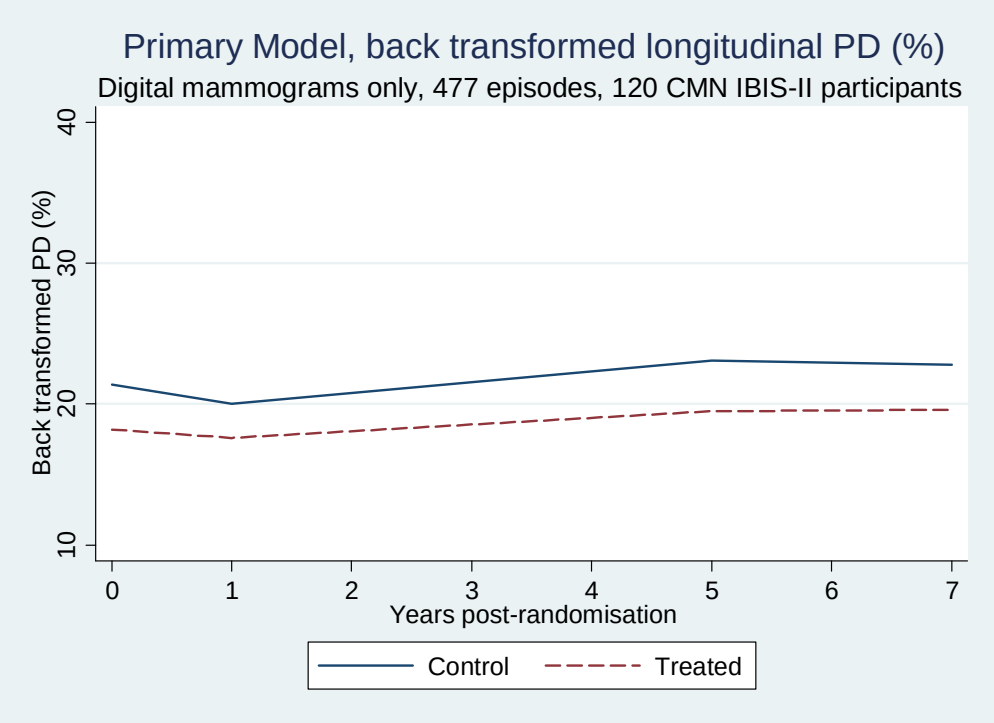


Figure 8-15 Back transformed longitudinal PD by Treatment Group, digital only Primary Model
The digital mammogram only estimated growth curves for the treated and control groups are almost parallel to each other, reflecting the small relative annual PD change for treatment vs control

for by the model. If so, any decreases in PD from anastrozole treatment will be masked due to the change in mammogram Version, and PD will appear to increase instead of decreasing.

8.4.4.2 *Full vs parsimonious Primary Model comparison*

The coefficients and SE for the Primary (square root transformed) PD and DA models are compared with Primary models with a full set of covariates in Table 8-14. This comparison was undertaken to check for unexpected differences between the parsimonious and full set of covariates in the Primary models. The PD parsimonious coefficients were largely similar to their counterparts in the full model (±10%), except for the anastrozole treatment indicator which showed a 15% difference (Primary model coefficient –0.33 vs full model coefficient –0.38). A few of the DA parsimonious coefficients changed more than 10%: age at randomisation (Primary –0.38 vs –0.57 Full), anastrozole treatment indicator (Primary –0.28 vs –0.48 Full),

and FE intercept (Primary 61.4 vs 74.8 Full), but all of the Primary model coefficients retained the same sign as their full model counterparts.

Table 8-14 Primary (parsimonious) models vs Full covariate models, PD and DA

Covariate	PD (square root) Primary model		PD (square root) Full model		DA (square root) Primary model		DA (square root) Full model	
	Coef.(β)	SE	Coef.(β)	SE	Coef.(β)	SE	Coef.(β)	SE
ALL MAMMOGRAMS 120 participants 540 episodes (follow ups)								
Age at Rand. (yrs)	-0.04**	0.02	-0.05*	0.02	-0.38‡	0.25	-0.57‡	0.31
BMI (kg/m^2)	-0.12***	0.02	-0.12***	0.02	--	--	-0.36	0.28
Menarche (yrs)	--	--	0.04	0.02	--	--	1.18‡	0.79
Menopause (yrs)	0.06**	0.02	0.06**	0.05	0.88***	0.28	0.96**	0.28
AgeFirstBirth<30y	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
≥ 30 years	1.06**	0.33	1.02**	0.35	21.4***	6.5	21.8***	5.9
Non-parous	0.03	0.29	0.03	0.30	6.3‡	4.3	5.7	4.8
OC use - Never	--	--	ref.	ref.	--	--	ref.	ref.
Ever users	--	--	-0.28	0.37	--	--	-5.0	6.0
HRT – Never	--	--	ref.	ref.	--	--	ref.	ref.
Ever users	--	--	0.02	0.23	--	--	-0.93	3.3
Smoking- never	--	--	ref.	ref.	--	--	ref.	ref.
Current	--	--	0.34	0.29	--	--	9.0**	3.3
Ex-smoker	--	--	0.10	0.22	--	--	0.24	3.4
IBIS-1 – No	--	--	ref.	ref.	--	--	ref.	ref.
Yes	--	--	0.23	0.26	--	--	3.1	3.7
Mammogram Version- film	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
KE52	-0.76***	0.12	-0.76***	0.12	-10.3***	1.5	-10.3***	1.5
KE54	-0.48***	0.13	-0.48***	0.13	-6.5***	1.8	-6.5***	1.7
Fuji	-0.29‡	0.15	-0.28‡	0.15	-4.6*	2.2	-4.5*	2.1
Anas. (no=ref)	-0.33‡	0.22	-0.38‡	0.22	-2.8	3.3	-4.8‡	3.3
Intercept	5.42***	0.30	5.86***	0.90	61.4***	4.0	74.8***	14.3
Annual change in MD (All mammograms)								
Baseline to Year 1	-0.14*	0.07	-0.14*	0.07	-1.5‡	0.83	-1.5‡	0.8
Years 1 to 5	0.05*	0.03	0.05*	0.03	1.0**	0.37	1.0**	0.37
Years 5 to 7	-0.02	0.07	-0.02	0.07	0.30	1.0	0.3	1.0
Anas. x Baseline-Year 1	0.02	0.08	0.02	0.08	-0.17	1.1	-0.14	1.1
Anas. x Year 1—Year 5	-0.02	0.03	-0.02	0.03	-0.39	0.3	-0.38	0.35
Anas. x Year 5—Year 7	0.01	0.12	0.01	0.12	-0.48	1.6	-0.48	1.6
Random effects – estimates for between- and within-person change								
Between person variance								
Film mmgs	0.29*	0.09	0.29*	0.09	50.2*	13.7	49.6*	13.4
Fuji mmgs	0.15*	0.04	0.15*	0.04	27.9*	7.0	28.0*	7.0
Intercept	1.24*	0.26	1.21*	0.28	274.1*	41.9	253.4*	45.9
Within person correlation (rho)	0.47*	0.14	0.47*	0.14	0.25*	0.11	0.25*	0.11
Within person variance	0.09*	0.03	0.09*	0.03	12.3*	2.2	12.3*	2.2

Coef. coefficient; ref. reference category; Anas. anastrozole treatment; y yrs years; mmg mammograms; * p<0.05; ** p<0.01; *** p<0.001; ‡ p≤0.15;

8.4.4.3 *Change points from 1.25 to 2.0 years*

Increasing the PD Primary cut point (knot) in 0.25 year increments from 1.0 to 2.0 years revealed a trend toward an easing in decrease of annual change in PD for control treated participants for the first segment of the continuous spline, Table 8-15. A similar trend was noted during modelling of the aggregate (treated + control) data (Aim 4, Chapter 7), and was part of the rationale for selecting the cut point of 1.0 year for the first segment of time. This is because the easing of PD rate of change with increasing time to the first cut point may have indicated a substantial change in the aggregate (treated + control) growth curve was missed by increasing the time at which the spline first was allowed to bend. Maximising the annual rate of decline for the aggregate longitudinal model at 1.0 years also appeared to suit the maximum theoretical change in MD expected for anastrozole treated participants. However, in Table 8-15, relative annual change for the treated group for the first time segment is the most negative, albeit non-significantly, for the cut point at 2.0 years.

It is not clear why treated participants in the film only Primary model show a clear decrease in relative PD and DA annual change relative to the control group from baseline to year one, yet the all mammogram and digital only models do not. Potentially, the pattern in relative PD annual change between the treated and control groups with the different cut points in Table 8-15 may hint that the cut points for the model may be incorrectly positioned. In theory, the growth curve for the control group should show a consistent decline of about –1% per year from baseline to year 7. However results from each cut point show a decline in annual PD rate of change for the first time segment, an increase during the second time segment, and a decline for years 5 to 7. With the exception of the decline between years 5 to 7, this is a pattern similar to the unconditional growth model of categorical time, which had a local minimum at 2.0 years (Table 7-5, Chapter 7). The pattern also matches the PD change over time pattern of the aggregate model fitted with categorical time (not tabulated). The pattern is caused by participants transitioning from baseline film to subsequent digital mammograms (which

Table 8-15 PD Primary model (1.0 year cut point) compared with cut points at 1.25, 1.5, 1.75 & 2 years

Covariate	PD Primary Model (ref)		PD 1.25 year cut point		PD 1.5 year cut point		PD 1.75 year cut point		PD 2.0 year cut point	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE
ALL MAMMOGRAMS[1] 120 participants, 540 episodes										
Age at Rand (y)	-0.04	0.02	-0.04	0.02	-0.04	0.02	-0.04	0.02	-0.04	0.02
BMI (kg/m^2)	-0.12	0.02	-0.12	0.02	-0.12	0.02	-0.12	0.02	-0.12	0.02
Menopause (y)	0.06	0.02	0.06	0.02	0.06	0.02	0.06	0.02	0.06	0.02
AFB ≥30y v <30y (ref)	1.06	0.33	1.06	0.33	1.06	0.33	1.06	0.33	1.06	0.34
Non-parous	0.03	0.29	0.03	0.29	0.03	0.29	0.04	0.29	0.04	0.29
Mmg Version KE52 v Film(ref)	-0.76	0.12	-0.76	0.12	-0.76	0.12	-0.77	0.12	-0.78	0.12
KE54	-0.48	0.13	-0.48	0.13	-0.49	0.13	-0.49	0.13	-0.50	0.13
Fuji	-0.29	0.15	-0.28	0.15	-0.29	0.15	-0.29	0.15	-0.30	0.15
Anas. (no=ref)	-0.33	0.22	-0.33	0.22	-0.33	0.22	-0.32	0.22	-0.32	0.22
Intercept	5.4	0.30	5.4	0.30	5.4	0.30	5.4	0.30	5.4	0.30
Annual change in PD (All mammograms)[2]										
Baseline to Cut point	-0.14*	0.07	-0.11‡	0.06	-0.09‡	0.05	-0.06	0.04	-0.04	0.04
Cut point to 5y	0.05*	0.03	0.06*	0.03	0.06*	0.03	0.05‡	0.03	0.05‡	0.03
Years 5 to 7	-0.02	0.07	-0.02	0.07	-0.02	0.07	-0.02	0.07	-0.02	0.07
Anastrozole x Baseline–Cutpt.	0.02	0.08	0.01	0.07	0.01	0.06	-0.002	0.06	-0.01	0.05
Anastrozole x Cutpt. to Year 5	-0.02	0.03	-0.02	0.03	-0.02	0.03	-0.01	0.03	-0.01	0.03
Anastrozole x Year 5 to Year 7	0.01	0.12	0.01	0.12	0.01	0.12	0.01	0.12	0.01	0.12

y yr year; Coef. coefficient; ref. reference category; SE standard error; Pts participants; eps episodes; AFB age at first birth; Cutpt. cut point; x interaction term; mmg mammogram/s;

[1] p-values for Age at randomisation, BMI, Menopause, AFB, mammogram Version, anastrozole and the intercept are similar for all models (i.e. are identical or very similar to the p-values for the PD Primary all mammogram model).

[2] p-values as marked: * $p<0.05$; ‡ $p\leq0.15$;

contributes to the drop from baseline to year 1) as well as the intra-digital Version transition, re: Figure 7-7, Chapter 7. The cut point at 1.0 year does appear to suit the growth curve of the control group for this particular model (with all participants and episodes), because the growth curve of the control group is unaffected by a treatment effect and should resemble that of the aggregate model because the expected change in PD due to anastrozole treatment is small.

However, the aggregate (treated + control) model average PD at each time point is highly dependent upon the Version/s of mammogram included in the model, Figure 8-16, as well which participants have data at each time point, e.g. the PD lowess plots Figure 8-1 and Figure 8-2. The PD trajectories (growth curves) in Figure 8-16 reveal the varying effects

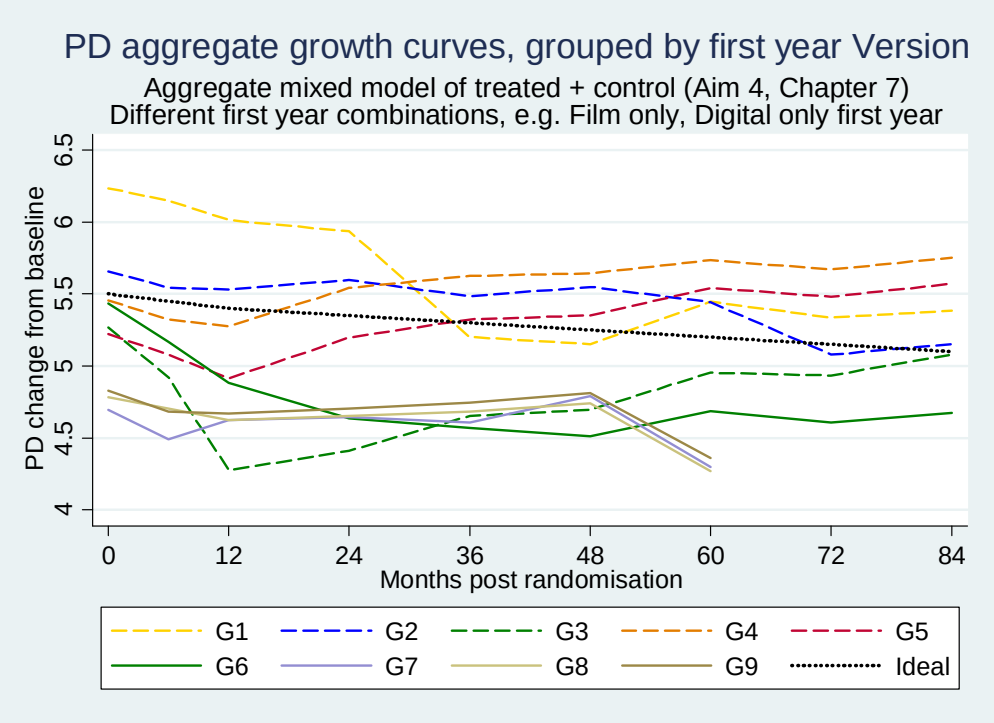


Figure 8-16 Square root transformed PD aggregate growth curves, different first year Version groups
G Group; Models with film mammograms have dashed lines, models with digital mammograms only are graphed as solid lines; an 'ideal' control group trajectory is represented by a line of black dots.
The participants are grouped by Version of mammogram present from baseline to year 1 (Table 8-1): G1 Film only; G2 Film, KE52 or KE54 only; G3 Film to digital only; G4 (G2 + G3); G5 (G1 + G3); G6 Intra-digital transition only; G7 KE52 or KE54 only (G2 – G1); G8 All known digital Versions (G6 + G7); G9 All digital, includes 3 episodes of unknown Version (G8 + unknown Version); Ideal 'ideal' control group trajectory –0.5% PD/year (untransformed PD)

different mammogram Versions have on PD change in the first year (baseline to year 1). The greatest PD decline during the first year is seen for participants who transitioned from film to a digital Version (G2, green dashed line); the next largest PD decline during the first year resulted from participants who transitioned from one digital Version to another digital Version in the first year (solid green line). The group of participants with a PD growth trajectory which most closely mirrors the 'ideal control group' trajectory have the same within-person mammogram Version during the first year (G2, blue dashed line).

8.4.4.4 *Participants with known age at menopause only; parous participants only*

Coefficients for the PD Primary model fitted for women with known age at menopause only, Table 8-16, are similar to the coefficients for the Primary Model fitted for all participants, Table 8-5 and Table 8-16. All subgroup coefficients are within ± 1 SE of the Primary model coefficients, except the intercept which has changed because age at menopause was not centred in the participants with known age at menopause only model. Use of an imputed age at menopause for the 27 participants does not appear to have greatly affected the PD Primary model outcomes.

Parous participants only

Coefficients for the PD Primary model fitted for parous women only, Table 8-16, are similar to the coefficients for the Primary Model fitted for all participants, Table 8-5. With the exception of the parameter which had been redefined for the subgroup model (age at first birth is now a continuous not categorical parameter), the other coefficients are within ± 1 SE of the Primary model coefficients. Parous age was modelled as a continuous parameter, and was not centred so both the age at first birth and intercept differ for the subgroup model with parous women only in Table 8-16.

Table 8-16 PD Primary model sensitivity and subgroup analysis, All mammograms (film + digital)

Covariate	PD Primary Model (ref)		PD, known age at menopause		PD, parous only		PD, participants <=60 years only		PD, participants >60 years only		PD, "per protocol" treatment	
	Coef.(β)	SE	Coef.(β)	SE	Coef.(β)	SE	Coef.(β)	SE	Coef.(β)	SE	Coef.(β)	SE
# Pts, # episodes	**120 pts, 540 eps**		**93 pts, 424 eps**		**112 pts, 509 eps**		**47 pts, 214 eps**		**73 pts, 326 eps**		**117 pts, 447 eps**	
Age at Rand. (years)	-0.04**	0.02	-0.03‡	0.02	-0.05*	0.02	-0.06	0.04	-0.05	0.04	-0.05**	0.02
BMI (kg/m^2)	-0.12***	0.02	-0.12***	0.02	-0.11***	0.02	-0.10***	0.02	-0.14***	0.03	-0.12***	0.02
Menopause (years)	0.06**	0.02	0.08[1]***	0.02	0.05*	0.02	0.08**	0.02	0.05*	0.02	0.06**	0.02
Age First Birth, ≥30y vs <30y (ref)	1.06**	0.33	0.75*	0.31	0.02[2]	0.03	1.45**	0.47	0.48‡	0.31	1.06**	0.33
Non-parous vs <30y (ref)	0.03	0.29	0.05	0.31	--	--	0.68*	0.29	-0.41	0.48	0.07	0.32
Mammogram Version, KE52 v Film(ref)	-0.76***	0.12	-0.83***	0.14	-0.83***	0.11	-0.93***	0.15	-0.73***	0.15	-0.73***	0.12
KE54 v Film (ref)	-0.48***	0.13	-0.54***	0.15	-0.59***	0.12	-0.60***	0.15	-0.49**	0.17	-0.42**	0.14
Fuji v Film (ref)	-0.29‡	0.15	-0.32‡	0.17	-0.42*	0.14	-0.49**	0.18	-0.27	0.20	-0.17	0.16
Anas. (no=ref)	-0.33‡	0.22	-0.28	0.25	-0.28	0.23	-0.23	0.29	-0.48‡	0.31	-0.35‡	0.23
Intercept	5.42***	0.30	1.6‡	0.93	5.1***	0.71	5.4***	0.34	5.7***	0.68	5.4***	0.31
Annual change in PD (All mammograms)												
Baseline to Year 1	-0.14*	0.07	-0.10‡	0.06	-0.13‡	0.07	-0.11	0.08	-0.15‡	0.10	-0.16*	0.07
Years 1 to 5	0.05*	0.03	0.04‡	0.03	0.07**	0.03	0.04	0.03	0.07*	0.04	0.06‡	0.03
Years 5 to 7	-0.02	0.07	-0.08	0.07	-0.01	0.07	-0.03	0.09	0.04	0.11	--	--
Anas. x Baseline to Year 1	0.02	0.08	0.00	0.09	0.00	0.08	-0.15[3]	0.10	0.17‡	0.11	0.03	0.08
Anas. x Year 1 to Year 5	-0.02	0.03	-0.01	0.03	-0.02	0.03	0.01	0.04	-0.04	0.03	-0.03	0.03
Anas. x Year 5 to Year 7	0.01	0.12	0.07	0.12	-0.01	0.12	0.06	0.16	-0.07	0.17	--	--

Coef. coefficient; ref. reference category; SE standard error; Pts participants; eps episodes; y years; AFB age at first birth; Anas. anastrozole treatment; x interaction term;

* p<0.05; ** p<0.01; *** p<0.001; ‡ p≤0.15; [1] Menopause age is not centred

[2] Parous age is a continuous variable (not centred)

[3] p-value 0.154

[4] p-value 0.172

8.4.4.5 *Participants ≤60 vs >60 years at randomisation; 'per protocol analysis'*

Analysis of the Primary model by age at randomisation (≤60 y vs >60y) produced models with fairly similar results. Most coefficients were within ± 1 SE of the Primary model coefficients; all were within ± 2 SE. The most affected coefficients were for a variable which had a category with small cell sizes— age at first birth. The coefficients for mammogram Version were also more affected than most other parameters, which may be a reflection of the smaller sample size for women ≤ 60 year at randomisation.

PD change over time during the first year of treatment for the anastrozole group relative to the control group for the subgroup of younger participants at randomisation was non-significant but decreasing (–0.15, p=0.154) whilst annual PD change for their older counterparts showed an non-significant annual increase (0.17, p<0.15). The remaining anastrozole interaction terms (years 1 to 5, and years 5 to 7) did not show as marked differences between the different age sub-groups. Age may impact on first year PD response to anastrozole treatment; however this study did not have adequate sample size or statistical power to investigate this further.

Censored at cease of treatment: "per protocol analysis"

Restricting the PD Primary model to only 'on treatment' episodes (prior to year 6) did not greatly affect the model coefficients compared to the PD model with all episodes, Table 8-16. In particular, the growth coefficients (including the anastrozole treatment interaction terms) were nearly identical.

8.4.4.6 *Mixed model vs GEE model; Default SE vs robust SE*

The FE coefficients are fairly similar for the PD Primary model, PD Primary model with autoregressive (AR) covariance structure, and PD fitted as a GEE model with an AR covariance structure, Table 8-17. As expected, the FE coefficients are nearly identical for the PD Primary AR and PD GEE AR model, because GEE models for continuous outcomes are effectively

identical to their (mixed) linear regression counterparts. Some FE coefficients differ between the two PD AR models and the (exponential covariance structure) PD Primary model due to differences in the number of participants included in the models; this likely accounts for the >1 SE (but < 2 SE) difference in coefficients for the age at first birth categorical covariate. It may also explain the decrease in size for the (ns) anastrozole treatment covariate (–0.33 for all 120 participants to about –0.05 in the two AR models).

The PD Primary and two AR PD models produced similar estimates for the RE between person variance intercept and within person variability (± 1SE). The within person correlation (rho) estimates were also similar, as was the Fuji mammogram RE slope estimate. The film mammogram slope estimate differed the most between the PD Primary covariance structures. This was likely due to the fewer number of film episodes in the AR data set (n=14) compared to the Primary model dataset (n=63 film episodes). Given that none of the PD Primary model coefficients differed by more than 2 SE compared to the PD AR and GEE AR model coefficients, and the more stable (larger cell size) covariate coefficients were <1 SE apart, the selected modelling approach (mixed model with an exponential covariance structure) appears appropriate.

Default SE vs robust SE

There was no consistent pattern for the SE of the PD and DA Primary models with the default (original information matrix, OIM) vs the robust SE utilised in the PD and DA Primary models, Table 8-17. The default SE are sometimes smaller, sometimes identical, and also sometimes larger than the corresponding robust SE. Growth parameters which were significant with the default SE are generally still significant with robust SE, and use of either SE does not alter interpretation of the model.

Table 8-17 PD Primary model , PD Primary model and GEE model with AR1 covariance structure, DA Primary model, and PD and DA Primary models with default SE

Covariate	PD Primary Model (ref)		PD Primary, AR1 covariance		PD, GEE model AR1 covariance		PD Primary, default SE		DA Primary Model (ref)		DA Primary, default SE	
	Coef.(β)	SE	Coef.(β)	SE	Coef.(β)	SE	Coef.(β)	Default SE	Coef.(β)	SE	Coef.(β)	Default SE
# Pts, # episodes	**120 pts, 540 eps**		**72 pts, 310 eps**		**72 pts, 310 eps**		**120 pts, 540 eps**		**120 pts, 540 eps**		**120 pts, 540 eps**	
Age at Randomisation (years)	-0.04**	0.02	-0.05*	0.02	-0.05*	0.02	-0.04*	0.02	-0.38‡	0.25	-0.38	0.28
BMI (kg/m^2)	-0.12***	0.02	-0.10***	0.02	-0.11***	0.02	-0.12***	0.02	--	--	--	--
Menopause (years)	0.06**	0.02	0.07**	0.02	0.07**	0.02	0.06**	0.02	0.88**	0.28	0.88**	0.28
Age First Birth, ≥30y v <30y(ref)	1.06**	0.33	0.59‡	0.33	0.68*	0.31	1.06**	0.46	21.4**	6.5	21.4**	6.8
Non-parous vs <30y (ref)	0.03	0.29	0.11	0.31	0.12	0.31	0.03	0.43	6.3‡	4.3	6.3	6.3
Mammogram Version,												
KE52 v Film (ref)	-0.76***	0.12	-0.73***	0.14	-0.69***	0.15	-0.76***	0.10	-10.3***	1.5	-10.3***	1.3
KE54 v Film (ref)	-0.48***	0.13	-0.45**	0.14	-0.41**	0.15	-0.48***	0.11	-6.5***	1.8	-6.5***	1.5
Fuji v Film (ref)	-0.29‡	0.15	-0.22	0.16	-0.17	0.18	-0.29*	0.14	-4.6*	2.2	-4.6**	1.9
Anastrozole Tx (no=ref)	-0.33‡	0.22	-0.04	0.29	-0.08	0.30	-0.33‡	0.22	-2.8	3.3	-2.8	3.2
Intercept	5.42***	0.30	5.4***	0.44	5.4***	0.44	5.4***	0.32	61.4***	4.0	61.4***	4.1
Annual change in PD (All mammograms)												
Baseline to Year 1	-0.14*	0.07	-0.11‡[1]	0.07	-0.15‡[1]	0.08	-0.14*	0.06	-1.5‡	0.83	-1.5‡	0.77
Years 1 to 5	0.05*	0.03	0.08*	0.03	0.07*	0.03	0.05‡	0.03	1.0**	0.37	1.0*	0.40
Years 5 to 7	-0.02	0.07	-0.10*	0.05	-0.08‡	0.04	-0.02	0.07	0.30	1.02	0.30	0.99
Anastrozole x Baseline to Year 1	0.02	0.08	0.01	0.09	0.06	0.09	0.02	0.07	-0.17	1.07	-0.17	0.96
Anastrozole x Year 1 to Year 5	-0.02	0.03	-0.03	0.03	-0.02	0.04	-0.02	0.03	-0.39	0.35	-0.39	0.43
Anastrozole x Year 5 to Year 7	0.01	0.12	0.13	0.14	0.13	0.12	0.01	0.10	-0.48	1.64	-0.48	1.37
Random effects – estimates for between- and within-person change												
Between person variance												
Film mammograms	0.29*	0.09	0.06	0.08	--	--	0.29*	0.09	50.2*	13.7	50.2*	14.6
Fuji mammograms	0.15*	0.04	0.11*	0.04	--	--	0.15*	0.03	27.9*	7.0	27.9*	5.9
Intercept	1.2*	0.26	1.2*	0.36	--	--	1.2*	0.17	274*	42	274*	37
Within person correlation (rho)	0.47*	0.14	0.57*	0.13	--	--	0.47*	0.11	0.25*	0.11	0.25*	0.11
Within person variance	0.09*	0.03	0.12*	0.04	--	--	0.09*	0.02	12.3*	2.2	12.3*	1.9

Coef. coefficient; ref. reference category; SE standard error; Pts participants; eps episodes; y years; Tx treatment; x interaction term;
* p<0.05; ** p<0.01; *** p<0.001; ‡ p≤0.15;
[1] Baseline to Year 1 p-value for PD AR1 model is 0.14, p-value for corresponding GEE coefficient is 0.055;

8.4.4.7 *Treatment group RE assessment*

The RE term for anastrozole (a RE slope for anastrozole treatment) is non-significant when added to the Primary models of PD and DA (i.e. the estimate for the anastrozole RE slope variance is not greater than two times its SE), Table 8-18.

Modelling of a separate covariance structure for each treatment group was enabled by adding a term for treatment group to the covariance structure specification, Table 8-18. Treatment group may be an effect modifier for within-person correlation (rho) for both PD and DA— the control group had a higher within person correlation than the aggregate correlation (rho) of the Primary models, and the within person aggregate correlation in the Primary models was higher than the correlation of the anastrozole treated group. Because of the overlap of the within person correlation covariance estimates (<1 SE apart), the within person correlations by treatment group are unlikely to be significantly different.

The lower within-person correlation (rho) for the treated group corresponded to a higher within-person variance estimate compared to the aggregate within person variance estimate of the Primary PD and DA models. This seemed appropriate, as higher within person variability is likely to be associated with a lower within person correlation. The differences in estimated within person variance between the treatment groups approached > 2SE, so the variance estimates for the two treatment groups may significantly differ.

In a future analysis, to reduce the potential impact of the non-significant RE anastrozole intercept on the reported changes in within person correlation and variance by treatment group, the PD and DA Primary models could be re-fitted without the anastrozole treatment RE term. The resulting by treatment group within person correlation and variance estimates could then be compared to those tabulated in Table 8-18.

Table 8-18 PD and DA Primary models ± RE for treatment group ± by treatment group RE covariance residual structure (All mammograms, all 540 episodes)

Covariate	PD Primary Model (ref)		PD Primary + RE anastrozole Tx		PD + RE anastrozole + by group cov.		DA Primary Model (ref)		DA Primary + RE anastrozole Tx		DA + RE anastrozole + by Tx group cov.	
	Coef.(β)	SE	Coef.(β)	SE	Coef.(β)	SE	Coef.(β)	SE	Coef.(β)	SE	Coef.(β)	SE
Age at Randomisation (years)	-0.04**	0.02	-0.04*	0.02	-0.04*	0.02	-0.38‡	0.25	-0.37‡	0.25	-0.39‡	0.24
BMI (kg/m^2)	-0.12***	0.02	-0.11***	0.02	-0.11***	0.02	--	--	--	--	--	--
Menopause (years)	0.06**	0.02	0.06**	0.02	0.06**	0.02	0.88**	0.28	0.89**	0.28	1.01***	0.24
Age First Birth, ≥30y vs <30y (ref)	1.06**	0.33	1.12**	0.37	1.11**	0.37	21.4**	6.5	21.7**	6.6	21.6**	6.5
Non-parous vs <30y (ref)	0.03	0.29	0.08	0.28	0.08	0.28	6.3‡	4.3	6.2‡	4.3	6.5‡	3.7
MammogramVersion, KE52v Film(ref)	-0.76***	0.12	-0.76***	0.12	-0.82***	0.12	-10.3***	1.5	-10.3***	1.5	-11.0***	1.5
KE54 v Film (ref)	-0.48***	0.13	-0.48***	0.13	-0.55***	0.13	-6.5***	1.8	-6.6***	1.8	-7.3***	1.7
Fuji v Film (ref)	-0.29‡	0.15	-0.28‡	0.15	-0.38**	0.15	-4.6*	2.2	-4.6*	2.2	-5.7**	2.1
Anastrozole Tx (no=ref)	-0.33‡	0.22	-0.33‡	0.22	-0.32‡	0.22	-2.8	3.3	-2.8	3.3	-2.9	3.2
Intercept	5.4***	0.30	5.3***	0.29	5.4***	0.29	61.4***	4.0	61***	4.0	62.1***	4.0
Annual change in PD (All mammograms)												
Baseline to Year 1	-0.14*	0.07	-0.14*	0.06	-0.12‡	0.06	-1.5‡	0.83	-1.5‡	0.82	-1.2‡	0.81
Years 1 to 5	0.05*	0.03	0.05*	0.03	0.06*	0.03	1.0**	0.37	1.0**	0.37	1.2*	0.38
Years 5 to 7	-0.02	0.07	-0.02	0.07	-0.01	0.07	0.30	1.02	0.30	1.0	0.42	0.93
Anastrozole x Baseline to Year 1	0.02	0.08	0.02	0.08	0.01	0.08	-0.17	1.07	-0.17	1.1	-0.30	1.1
Anastrozole x Year 1 to Year 5	-0.02	0.03	-0.02	0.03	-0.01	0.03	-0.39	0.35	-0.38	0.35	-0.37	0.35
Anastrozole x Year 5 to Year 7	0.01	0.12	0.01	0.12	0.00	0.11	-0.48	1.64	-0.48	1.6	-0.57	1.6
Random effects – estimates for between- and within-person change												
Between person variance												
Film mammograms	0.29*	0.09	0.29*	0.09	0.31*	0.08	50.2*	13.7	50.2*	13.7	52.6*	13.2
Fuji mammograms	0.15*	0.04	0.15*	0.04	0.15*	0.04	27.9*	7.0	27.9*	7.0	28.1*	6.8
Anastrozole treatment	--	--	0.54	0.54	0.53	0.54	--	--	48.9	101.3	51.2	79.2
Intercept	1.2*	0.26	0.96*	0.21	0.97*	0.20	274*	42	247*	42	243*	38
Within person correlation (rho)[1]	0.47*	0.14	0.47*	0.13	0.49*	0.23	0.25*	0.11	0.25*	0.11	0.27	0.22
Within person variance[1]	0.09*	0.03	0.09*	0.03	0.07*	0.03	12.3*	2.2	12.3*	2.2	9.7*	2.7
Within person correlation (rho) Tx[2]	--	--	--	--	0.42*	0.16	--	--	--	--	0.21	0.12
Within person variance, Tx[2]	--	--	--	--	0.11*	0.03	--	--	--	--	14.2*	3.0

Coef. coefficient; ref. reference category; SE standard error; Pts participants; eps episodes; y years; Tx treatment group; cov covariance residual structure; x interaction term; * p<0.05; ** p<0.01; *** p<0.001; ‡ p≤0.15; [1]without the by Tx group covariance term in model = aggregate group estimate, else is by treatment group covariance estimate for the control group; [2] By treatment group covariance estimate for the anastrozole group

8.4.5 Change in RE Variability, MD growth as model complexity increases

This section reviews the changes in the random effects variability as the mixed model increases in complexity. The effect of increasing model complexity on the MD growth parameters is also assessed.

Addition of terms for time to the model (i.e. to the unconditional growth model (model 2) and unconditional aggregate growth (model 3)) decreased within person variability relative to the unconditional means model (model 1), Table 8-19. Incorporation of FE terms for BC risk covariates— age at randomisation, age at menopause, age at first birth (and BMI for PD)— decreased between person variability in the parsimonious aggregate model (model 4) compared to model 3. The FE covariates in model 4 did not markedly change mean annual MD change over time nor the associated SE compared to MD change over time for (the unconditional but otherwise similar) model 3. Inclusion of terms for anastrozole treatment and the anastrozole interaction (the latter terms are not shown in the table) to the Primary model (model 5) induced little change in RE variability ($<\pm 3\%$) or MD change over time compared to model 4 with the exception of an approximate doubling of the SE for baseline to year 1 annual MD change for model 5 (Primary model) compared to model 4. The remainder of this section reviews the changes in RE variability and MD change over time between the models listed in Table 8-19 in more detail.

Within person variability

As described in Singer and Willet [490], covariates which change over time reduce intra-person (within person) variability. Thus due to the addition of time as a covariate, the reductions in within person variability for PD and DA were expected between the unconditional means model (model 1, least complex model) and the unconditional growth model (model 2, low complexity), as well as the unconditional growth model (model 2) and unconditional aggregate growth model (model 3, intermediate complexity), Table 8-19. PD within person variability declines by 33%

Table 8-19 Annual change in MD and RE variance comparisons, All mammograms models (540 episodes), in order of increasing complexity

Covariate	1. Unconditional Means (Ch 7)			Unconditional Growth (Ch 7)			3. Unconditional Aggregate Growth (new)			4. Parsimonious Aggregate (Ch 7)			5. Primary Model Treated vs Control		
	Estimate	SE	Change[1]	Estimate	SE	Change[1]	Estimate	SE	Change[1]	Estimate	SE	Change[1]	Estimate	SE	Change[1]
Square root PD annual change over time (interaction terms not shown for Primary model (model 5)															
Annual change	--	--		0.05*	0.02		--	--		--	--		--	--	
Baseline to Year 1	--	--		--	--		-0.12**	0.04		-0.13**	0.04		-0.14*	0.07	
Years 1 to 5	--	--		--	--		0.05*	0.02		0.04‡	0.02		0.05*	0.03	
Years 5 to 7	--	--		--	--		-0.02	0.06		-0.02	0.06		-0.02	0.07	
Random effects – estimates for between- and within-person change															
Between person variance[2]	2.0			2.32		+16%	2.35		+1%	1.71		-27%	1.68		-2%
Time (years)	--	--		0.02*	0.005		--	--		--	--		--	--	
Film mammograms	--	--		--	--		0.30*	0.08		0.29*	0.08		0.29*	0.09	
Fuji mammograms	--	--		--	--		0.15*	0.04		0.15*	0.04		0.15*	0.04	
Intercept	2.0*	0.27		2.3*	0.31	+15%	1.90*	0.29	-17%	1.27*	0.25	-33%	1.24*	0.26	-2%
Within person corr (rho)	--	--		--	--		0.46*	0.13		0.47*	0.13		0.47*	0.14	
Within person variance	0.24*	0.02		0.16*	0.01	-33%	0.09*	0.03	-44%	0.09*	0.03	0%	0.09*	0.03	0%
Square root DA annual change over time (interaction terms not shown for Primary model (model 5)															
Annual change	--	--		0.64**	0.23		--	--		--	--		--	--	
Baseline to Year 1	--	--		--	--		-1.56**	0.54		-1.57**	0.54		-1.48‡	0.83	
Years 1 to 5	--	--		--	--		0.85*	0.34		0.81*	0.34		1.01**	0.37	
Years 5 to 7	--	--		--	--		0.07	0.84		0.05	0.84		0.30	1.02	
Random effects – estimates for between- and within-person change															
Between person variance[2]	357			420.2		+18%	418.6		-0.4%	353.7		-16%	352.1		-0.5%
Time (years)	--	--		3.2*	0.77		--	--		--	--		--	--	
Film mammograms	--	--		--	--		49.6*	12.8		49.7*	12.9		50.2*	13.7	
Fuji mammograms	--	--		--	--		27.0*	6.7		27.0*	6.8		27.9*	7.0	
Intercept	357*	47		417*	57	+17%	342*	44.8	-18%	277*	39.9	-19%	274*	42	-1%
Within person corr (rho)	--	--		--	--		0.24*	0.11		0.25*	0.11		0.25*	0.11	
Within person variance	39*	2.7		27*	2.1	-31%	12.3*	2.2	-54%	12.4*	2.2	+1%	12.3*	2.2	-1%

Ch 7 Chapter 7; SE standard error; corr correlation; * p<0.05; ** p<0.01; *** p<0.001; ‡ p≤0.15;

[1] % change in RE estimated variance relative to less complex (earlier) model

[2] Total between person variance (e.g. RE variances for the Intercept + film + Fuji) shown in row

between models 1 to 2 (relative to model 1), and by 44% between models 2 and 3 (relative to model 2). The decline in within person variability for DA is similar between models 1 and 2 (–31%) but higher between models 2 and 3 (–54%). Because no further time varying parameters are added to the more complex models— parsimonious aggregate (model 4) and Primary models (model 5, most complex model)— negligible changes of <±1% are seen for within person variability during the transition from model 3 to model 4, and from model 4 to 5.

Between person variability

Addition of a (single) FE term for continuous time and a RE parameter for continuous time to model 2 compared to model 1 corresponded to increases of 15 to 18% in between person variability between the PD and DA unconditional means and unconditional growth models, Table 8-19. As described in Chapter 7, the RE parameter for time allows each person to have their own growth curve (slope) which can vary from the mean growth curve fitted by the FE term for time. This likely accounts for the increase in between person variability noted during the transition from models 1 to 2. Most of the total between person variability in model 2 (e.g. the estimate of 2.32 for PD) is comprised of the variability in the between person intercept (e.g. 2.3 for PD) and only a very small proportion is due to the RE parameter for time (0.02 for PD). The freedom to have a unique growth curve likely affected the RE intercept for some participants, increasing the variability of the between person RE intercept.

The increase in between person intercept variability between models 1 and 2 may particularly have been affected by the single continuous FE parameter for time (e.g. FE annual change in PD of 0.05 for all participants, model 2 Table 8-19), because it does not suit the growth curves for some participants very well (e.g. the participants transitioning from film (highest average PD) to KE52 (lowest average PD) mammograms and then to KE54 mammograms will have a U- or V-shaped (nonlinear) growth curve which cannot be modelled using a single straight line). Terms for mammogram Version are also not incorporated into the unconditional growth model (model

2) as either fixed or random effects, which could potentially increase the variability in participant response. Most participants transitioning from film to KE52 mammograms will decline in PD and DA, and then increase in PD and DA for KE54 and then Fuji mammograms (a U-shaped growth curve). The between person RE parameters have likely changed to compensate for the lack of fit to the single FE parameter for time.

Model 3 (unconditional aggregate growth) included FE terms for continuous time modelled as a spline with three segments and cut points at years 1 and 5. Model 3 also incorporated RE slopes for film and Fuji mammograms at the person level, and an exponential covariance structure. Although these changes from model 2 to model 3 increased the degrees of freedom from 5 to 12, model fit improved significantly; the corresponding changes in log likelihood (LL), AIC and BIC for PD were –565 to –400 (LL), 1140 to 824 (AIC) and 1161 to 875 (BIC). The unconditional aggregate model for DA showed similar improvements in model fit, e.g. decrease in BIC from 3939 (model 2) to 3727 (model 3).

Improvement of model fit to the mean growth curve for the CMN IBIS-II participants in model 3 decreased between person intercept variability in model 3 relative to both models 2 and 1. Both PD and DA show a ~18% decline in between person intercept variability for model 3 compared to model 2. Both models have an approximate 5% lower between person intercept variance compared to model 1.

Although between person intercept variability declined for model 3 relative to both models 1 and 2, overall (total) between person variability increased for model 3 compared to model 1. This is due to the addition of the RE slopes for film and Fuji mammogram Versions, which allow the growth curves of participants to randomly vary from the (FE) population average growth curve. The change in total between person variability between models 2 and 3 is

negligible (<±1%), hence the increase in total between person variability for models 2 or 3 compared to model 1 is about 17%.

The addition of the FE BC risk covariates to the parsimonious aggregate model (model 4) markedly reduced variability in the between person random intercept as well as total between person variability, but almost all of this decrease was due to drop in variability of the between person intercept. The decrease in variability of the between person intercept between models 3 and 4 was –33% for PD, and –19% for DA. Variability of the mammogram Version RE intercepts was essentially unchanged. [As stated earlier in this section, within person variance was similarly unaffected, because the FE BC risk covariates are all constant (they do not change over time within each person).]

Incorporation of FE terms for anastrozole treatment (a constant and an interaction term with the spline for continuous time) lowered the total between person variability negligibly between model 4 and the Primary model. The very small decrease of –2% for the PD Primary model relative to model 4 was solely due to a –2% decrease in the between person RE intercept. An even smaller decrease of –1% in between person RE intercept in the DA model was noted. Very small increases in the film and Fuji mammogram Version RE slope variability offset the small –1% decrease in DA RE intercept variability to yield a –0.5% change in total between person variability for the DA model.

MD change over time

As described in Chapter 7, alteration of the FE and RE modelling of continuous time from a single linear parameter to a continuous spline with three segments and cut points at year 1 and 5 improved model fit for both PD and DA. For example, changing continuous time from the single linear growth curve to the three segment spline significantly decreased the BIC for the PD model from 1161 to 1100.

Whilst the years 1 to 5 growth curve segment in model 3 retained the sign and magnitude of the single linear parameter fitted in model 2 (0.05, SE 0.02, $p<0.05$, Table 8-19), the baseline to year 1 and years 5 to 7 continuous time segments now replicated the expected mean decrease in MD over time.

The coefficients for (aggregate group) MD annual change for models 3 and 4 are not large (e.g. baseline to year 1 for PD is –0.12, Table 8-19). Although the SE increase (presumably due to halving of the sample size), the MD annual change coefficients for the control group in the Primary models (model 5) hardly differ from those of models 3 and 4. In other words, little MD change over time is occurring. The small changes in MD growth for PD and DA seem to be reflected by the small estimated within person variance relative to total RE variance. Within person variability as a proportion of total variability is low— in model 5 (Primary model) about 5% for PD (0.09/(0.09+1.68)) and 3.5% for DA (12.3/(12.3+352)). The intraclass correlations— the proportion of the "total outcome variation" which exists "between people" [490]— for PD and DA are correspondingly high. The ICC for PD is 0.95 (1.68/(1.68+0.09)) and the ICC for DA is 0.97 (352/(352+12.3)). Very little change in MD over time appeared to be present in the PD and DA models. However, the high ICCs also imply that the (subjective) MD measurements (the average of the four mammographic Views per episode) are also not too variable.

Addition of the BC risk covariates to model 4 did not appear to greatly affect the coefficients for PD and DA change over time compared to model 3. The small change in MD growth between models 3 and 4 for both PD and DA implies that the BC risk covariates are unlikely to impact modelling of MD change over time. Although very important for models of MD change and BC risk (and therefore clinically important), the BC risk covariates could potentially be ignored during measurement of MD if characterisation of change in MD was all that was required.

Addition of the anastrozole treatment parameters to model 5 had a larger effect on the DA model than for PD. This may have occurred because the relative size of the anastrozole interaction terms (not shown in Table 8-19) for DA were much larger (e.g. ~20 to 60mm^2, Table 8-6) compared to the size of the interaction terms for PD (<±0.25%, Table 8-6).

8.5 Discussion

Treatment with anastrozole was expected to lower PD in the treated participants by 1 to 2 % in the first 12 months relative to CMN IBIS-II participants randomised to placebo treatment [29]. However, a greater reduction in MD for the anastrozole treated group compared to controls was not apparent in the all mammogram (540 episode) Primary models of PD and DA. The PD Primary model interaction terms showed very small and non-significant PD changes for the anastrozole group relative to controls from baseline to year 1 for the all mammogram model (e.g. less than ± 0.25%/year, back transformed PD, Table 8-6). Whilst an expected, negative sign for the coefficients of DA change for anastrozole treated participants relative to controls was noted, the differences in DA change over time for the first year of treatment were not significant ($p>0.25$, Table 8-6, back transformed values of –21 to –58 mm^2/year).

The all mammogram Primary models are comprised of both film and digital mammograms. Whilst none of the digital only models showed significant change in MD over time for anastrozole treated participants relative to controls, fitting the Primary model with 63 film-only episodes yielded a significant –3.7% back transformed mean PD difference in annual change at 12 months between the treated and control participants, Table 8-9. The film mammogram only results for DA were also significant; back transformation of the coefficients yielded a mean difference in first year annual change of –547 mm^2 (–5.5cm^2) per year in DA between the treated and control groups for the film-only Primary model.

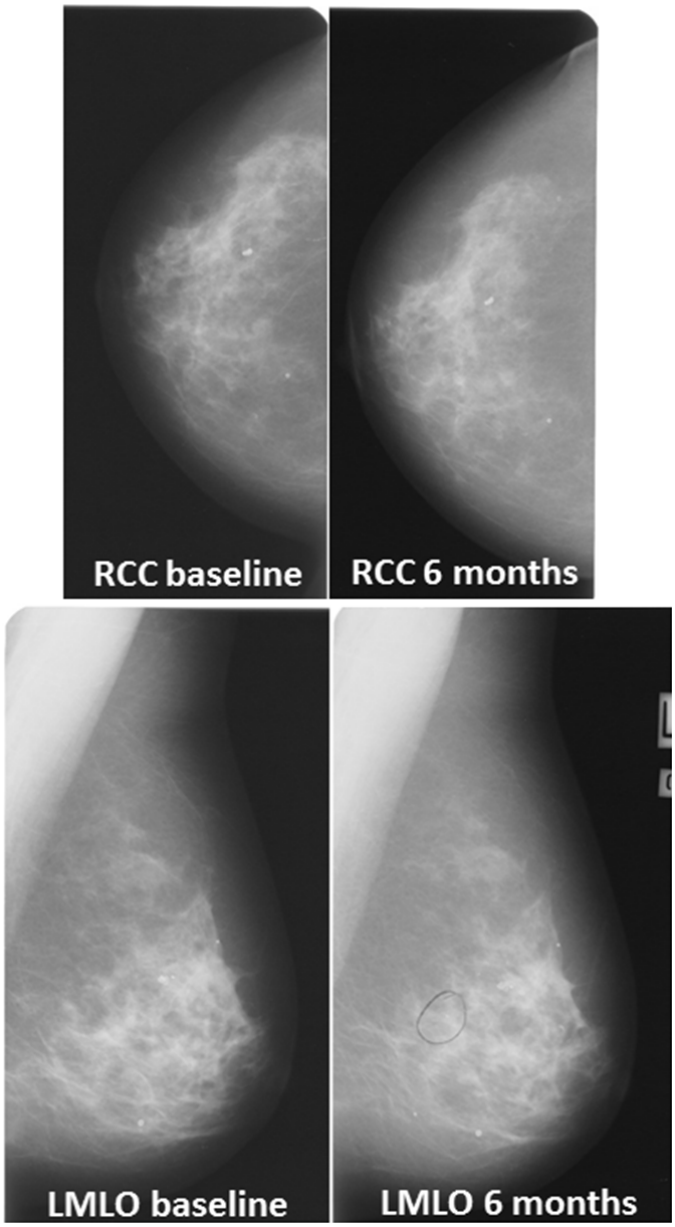


Figure 8-17 Example baseline and 6 month film mammograms from one CMN participant
Upper row – RCC View baseline and six month mammograms; Lower row – LMLO View mammograms. This participant's treatment status is blinded, however the average Cumulus assessed difference in PD between the participant's baseline and 6 month mammograms is –4.3%, similar to the average difference in one year change scores between treatment and control for PD on film mammograms between baseline to year 1 for the mixed model (–3.7%/year, Table 8-9). Although the Cumulus assessed average episode difference in density for this participant is nearly 5%, the visual difference between the film baseline mammograms (left column, taken 10 months prior to randomisation) and mammograms taken 6 months after randomisation (right column) is difficult to discern.
The participant's baseline characteristics are similar to that of the representative film mammogram participant (randomised at age 50y, menopause at age 50y, BMI of 25, baseline film episode average PD of 33%).

The back transformed difference in PD average change of –3.7% for the first year of anastrozole treatment relative to that of controls approaches a one category difference of 5% on the CRUK 21–category visual scale. A change of <5% is likely difficult to detect clinically using only visual assessment. A Cumulus measured decrease of ~4% is shown in Figure 8-17. The subtle differences in PD seen between these baseline and 6 month film mammograms are similar to the mean response to anastrozole (–3.7% PD) relative to controls for a representative anastrozole treated participant with film mammograms from the sample IBIS-II population, Table 8-9. The visual difference in PD in Figure 8-17 is not pronounced. Stating definitively that a visual decrease in PD has occurred is challenging.

Although very modest (<5%/year), this is the first time, to our knowledge, a significant difference in annual MD change per year has been reported for women treated with an AI relative to controls in a randomised controlled trial. One other study has reported a rate of change in PD during the first year of AI treatment [28]. An average (adjusted) annual rate of –0.12%/year (95%CI –0.84 to 0.59) for 48 women randomised to placebo vs –0.68 (95%CI –1.34 to –0.02) for 56 women randomised to letrozole treatment in the MA.17 trial was reported. Adjusted mean first year annual change for the letrozole group was therefore –0.56%/year lower compared to that for controls (p=0.23, two sample t-test). The rate reported for MA.17 may have been smaller than the rate reported for this project because the population had previously completed 5 years of TAM treatment for early BC within 3 months of randomisation to extended treatment with placebo or letrozole. Tamoxifen is known to reduce MD [248, 296] as well as BC risk, so the women likely began the extended (AI) treatment phase with reduced density due to previous endocrine therapy. Given that the group sizes for each arm were larger in MA.17 (a baseline and 1 year mammogram were collected for all women in the study), it is also possible that the annual rate of change from this larger sample of women in MA.17 may be more representative of the true response to AI treatment.

The film-only, estimated first year back transformed –4% in relative annual PD change for anastrozole treated participants is at least double what was expected, although use of a more typical baseline PD reference value of 20% alters this back transformed estimate to –3%/year (Table 8-10). Declines of greater than 2% at 12 months have been reported during AI treatment in single arm studies. Prowell et al reported baseline median PD of 13.4% which declined to 10.3% in a cohort of 54 women with early BC treated with 1 year of anastrozole [310]. A median PD difference of –3.4% at 12 months was reported for 35 high risk women treated with exemestane [30]. A decline of –3.1% at 12 months was also seen in an AI treated subgroup of women with early BC [18]. However, most AI studies report <2% mean decline at 12 months of treatment [28, 29, 311-313].

Further research is needed to determine if subtle (<5%) changes in PD due to AI treatment as noted in the film-only Primary model are related to reductions in BC risk. This project was not powered to address this important clinical issue. However the Primary aim of this project was to help ascertain if a detectable, possibly significant difference in MD exists between anastrozole treated participants and controls, to support the larger, long term goal of utilising MD as a biomarker for AI treatment efficacy.

A very recent study utilising percent *volumetric* density measured using two commercially available programs compared volumetric density for women diagnosed with breast cancer treated with an AI (cases, N=403) to volumetric density in post-menopausal matched controls without breast cancer (N=1618) [33]. This case-control volumetric density project found statistically significant ($p<0.05$) differences in adjusted annualized change in volumetric percent density (VPD) for cases vs controls of approximately –0.25% (Volpara) and –0.5% (Quantra). Although statistically significant differences between cases and controls were found for annual change in volumetric percent density, interestingly, statistically significant changes in dense volume (cm^3) were not found using either measurement program. Even more interestingly,

perhaps, stratification of the cases and controls by baseline VPD of <10% vs ≥10% found no significant differences between cases and controls for annualised changes in VPD for women with baseline VPD <10%; however both Volpara and Quantra measured statistically significant *increases* in annualised change in dense volume (DV) for women with baseline VPD <10% (N>200 cases; N>1000 for controls).

The VPD, AI and TAM project is the first to report a statistically significant change in MD for AI treated women compared to controls. The significant findings of the case-control VPD, AI and TAM project underscores the necessity of measuring AI induced change in MD with precise, preferably fully automated techniques. The adjusted annual reductions in VPD for AI treated women relative to controls were –0.30% to –0.58%, which are quite small. If these were (two dimensional) PD changes they would be almost impossible to detect using semi-automated thresholding techniques like Cumulus without extremely large number of cases and controls, due to the inherent variability of any subjective measurement method.

Although volumetric density change was found to be approximately linear with time, another reason why the VPD, AI and TAM project annualised average VPD change estimates may be quite small is because change in volumetric density was measured using the last available mammogram on AI treatment (median 31.5 months). If most of the change in MD due to AI treatment occurs during the first year of treatment, then annualised MD change measured later will be smaller than if measured at one year on treatment, (Figure 8-18, slope of green dashed line is smaller than slope of segment 1 (first red dashed line)). However, longitudinal growth for AI treated vs control participants has never been characterised adequately, so it is not known if the growth curve for AI treated participants (relative to controls) is similar to that modelled for the tamoxifen treated participants in IBIS-I [296]. The comparatively marked relative reduction in first year annual PD change estimated for treated vs control participants for the film only Primary model in this thesis (CMN MD and AI substudy) is likely to be unusual. Although

the strong relative reduction in PD modelled for film mammograms hints at a strong reduction during the first year of anastrozole treatment, the (non-significant) relative increase in annual PD change for the treated group in the larger participant sample of the Primary (all mammograms) and digital only Primary models opposes this finding.

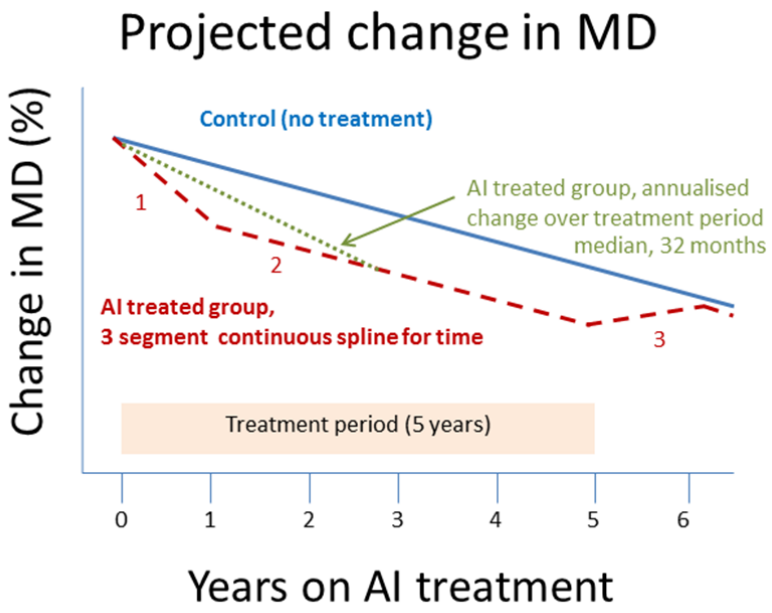


Figure 8-18 Different time modelling strategies for AI treated participants

Because the VPD, AI and TAM project utilised raw mammographic data, the MD measurements made in that study were not affected by differences in digital image post processing. The measurements made in the VPD, AI and TAM project were also fully automated, hence their measurements are much more reliable than subjective measurements made in this project (CMN MD and AI substudy). Additionally, the VPD, AI and TAM also used paired pre-treatment and on treatment mammograms for each case with similar timing of mammograms for the controls, further adding to the validity of their model. The results for the Primary Aim of this thesis in contrast, were affected by a small sample size, variable numbers of episodes collected per participant, and frequent changes in mammographic Version over time, in addition to the variability introduced by the subjective measurement method.

The results of the VPD, AI and TAM project suggest it might be worthwhile stratifying the Primary Model into two groups of participants with higher and lower PD, as participants with low VPD in that project showed a significant increase in DV for the AI treated women compared to controls. Whilst stratification by initial PD in the aggregate longitudinal model did show an easing in baseline to year 1 PD annual decline for women with lower initial PD (Chapter 7, Table 7-13), this comparison was not examined for the Primary Aim due to the small cell sizes for stratification by PD and treatment group.

The modelled change in DA at 12 months for the film only Primary model was also higher than the one study which reported change in DA during AI treatment [31]. Mean decline in (absolute) DA for 259 women treated for early BC with exemestane or letrozole at 24 months was $-180mm^2$ for this earlier study, compared to the back transformed annual DA decrease (film only Primary model) of about $-550mm^2$ at 12 months of treatment for the anastrozole treated group relative to controls for this project (Table 8-9, DA intercept^2 value of $3850mm^2$). Back transformed DA using a more characteristic value of $2700mm^2$ yields a slightly smaller annual first year rate of DA change of $-455mm^2$/year. The smallest average DA difference which can reliably be measured is not well characterised, however an automated method of assessing both PD and DA may be able to reliably detect subtle longitudinal differences on mammographic images, as shown recently for DV using volumetric techniques.

Most of the MD and AI literature cited above were undertaken using film mammograms only, so this is unlikely to be the source of the differences between other studies and the estimated rate of change on MD due to AI treatment in this project (CMN MD and AI substudy).

Although the annual change in MD from baseline to 12 months for the anastrozole treated group relative to the control group for film only mammograms (about –3%) is much higher than expected, the model appears to be internally valid. The 'raw' median PD difference (subsequent

PD–baseline PD) for the 14 participants with unique pairs of film mammograms at baseline and 6 months (n=7) or baseline and 12 months (n=7) was –1.9% (aggregate treated+control data, not tabulated). This is not very different from the modelled aggregate change in (untransformed) PD for film mammograms of –1.8% between baseline and year 1 (Table 7-6, Chapter 7). The aggregate rates of annual first year PD change for film mammograms for raw PD difference (–1.9%/year), mixed model of untransformed PD (–1.8%/year) and back transformed PD (–1.5%/year to –2.3%/year) are all quite similar.

Participants who contribute more data are more heavily weighted in the mixed model. Restriction of the aggregate film-only square root transformed PD model to just the 32 film episodes from 14 participants with 2 or more film episodes yielded a significant change in PD of –0.21 (p=0.022, SE 0.092, 95%CI –0.39 to –0.03, data not tabulated). These results are almost identical to the results of the aggregate square root transformed PD film only model with all 63 film episodes: baseline to year 1 annual PD change of –0.20 (Table 7-6, p-value 0.023, SE 0.089, 95%CI –0.37 to –0.03 (latter statistics not tabulated)). The similarities in (film only) coefficients and SE between the 63 episode, 45 participant aggregate model and the restricted 32 episode, 14 participant model implies that the significant film-only baseline to year 1 result in the Primary model is likely based on the data from these 14 participants. This equates to an effective sample size of 14. The estimated power to detect a difference in means of 4, standard deviation of 8, significance level of 0.05 with 7 participants in each group (assuming that the 14 participants are equally distributed between the two treatment groups) is 0.155. This further underscores the utility of the mixed model to detect alterations in MD growth (change over time) between treatment groups, and indicates that the study is quite underpowered to detect even a relatively large (for an AI) difference in mean MD between treated and control groups.

Weighting within the mixed model may also have skewed the actual difference in (raw) mean MD between treatment groups in the film only Primary models. Of the fourteen participants

with two or more film mammograms, three participants contributed baseline, 6 months and 12 month film episodes. Longitudinal data from these three participants are likely to be more heavily weighted in the model than data from participants with only 2 film episodes. Median difference in PD at 6 months for the 3 participants was –2.7%, and –3.6% at 12 months. Due to blinding, it is not known if these participants were randomised to anastrozole treatment, but if so this may have also inflated the modelled film-only back transformed Primary model value of –3.7% (–3%, 20% PD reference value) in first year annual rate of PD decline for participants randomised to anastrozole relative to controls.

The results from the digital mammogram only Primary model reflected the results from the all mammogram Primary models: only non-significant interactions between anastrozole treatment and MD change over time were found for the digital only PD and DA Primary models, Table 8-11. Although some of the p-values were relatively low (p=0.23 for PD and p=0.14 for DA for the years 1 to 5 interaction term), a larger sample size may however show that a significant difference between the treatment groups also exists for models utilising only digital mammograms.

A sensitivity analysis undertaken for this project, the CMN MD and AI substudy, for different continuous time cut points (section 8.4.4.3 Change points from 1.25 to 2.0 years) showed that the rate of annual PD change for the first continuous time segment was maximised for the control group at a cut point of 1.0 years, Table 8-15. However, the rate of annual PD change for the IBIS-II control group should, in theory, be consistent from baseline to year 7 (as discussed in section 8.4.4.3 (Change points from 1.25 to 2.0 years)). The lack of a consistent annual rate of decrease for the control group implies PD changes from the film to digital transition (Figure 7-5, Chapter 7) and the intra-digital mammogram transition (from KE52 to KE54 to Fuji; Table 7-2, Chapter 7) have been strongly influenced PD mean growth in the model, as illustrated by the widely varying growth curves of the different first year Version groups in Figure 8-16.

The sensitivity analysis for the Primary Aim (this chapter) would have benefitted from a model with time fitted as a categorical parameter for each treatment group. This may have provided additional insights into the growth curves for the anastrozole treated and control participants in this project. Further stratification by mammogram Version would be desirable, however, the number of CMN participants with repeated measurements within any Version is small (re: Table 7-16, Chapter 7), making interpretation of such models problematic.

One of the main difficulties encountered in the longitudinal analysis for this project (the CMN MD and AI substudy) was use of processed digital mammographic images. The instability of the amount of MD retained on the processed digital mammographic images is likely the largest source of potential error and confounding for assessment of longitudinal MD in this project. Although no difference in BIRADS MD categorisation between film and digital mammograms in a very large dataset was reported [508], comparison of the film and digital distributions (Figure 7-7 and Table 7-3 in Chapter 7) made with Cumulus assessed PD in this project (the CMN MD and AI substudy) implies that much of the longitudinal density information was removed by post-processing of the Kodak Elite mammograms.

Why the aggregate results for the KE54 mammograms do not show a significant decrease from baseline to year 1 is likely due to few repeated measurements (n=3, Table 7-15, Chapter 7) for participants from baseline to year 1. However, the reasons why the KE54 mammogram Version showed a significant increase during years 1 to 5 compared to the other mammogram Version (Table 7-14, Chapter 7) is not known. Other MD and AI studies have shown increases in MD in response to AI treatment in approximately 10 to 20% of subjects [28, 31, 311]; the prognostic value of these increases in MD is not known. Increases in MD from hormonal treatments are typically associated with increases in BC risk [246, 496, 509, 510], however they may be associated with lower BC risk in certain circumstances [272, 497].

Adjustment with typical BC risk factors appears to have little impact on the rate of annual change in PD due to AI treatment, similar to that seen in other studies (MA.17 trial [28] and [29], [313]). Hence adjustment for BC risk factors seems to have negligible effect upon the rate of annual change in PD due to AI treatment. This implies that—for the purposes of ascertaining if differences in longitudinal PD and DA exist between the AI treated and control groups in the high risk population of the IBIS-II trial—measuring PD and DA only might be sufficient for this purpose. Modelling of/adjustment for BC risk covariates is not required. However, not all known MD and BC risk confounders were modelled in this thesis; confounders such as duration of breast feeding and number of children were not available to model. Therefore further investigation of the apparent lack of effect of BC risk factors on longitudinal MD change in post-menopausal, high risk women, as well as other populations of women, is warranted.

Baseline BC risk and MD characteristics were similar between treatment groups, Table 8-2 and Table 8-3; hence any differences in MD due to the transition between film and digital mammograms and between digital Versions are likely balanced between the anastrozole and control groups.

Because the estimated average treatment effect of anastrozole noted for film mammograms is small (<5% PD/year) relative to the clinically (visually) detectable minimum change of 10%, this implies that methods other than visual assessment which are able to reliably detect small differences are required to assess mammograms for potentially clinically important, but quite small and difficult to discern differences in MD during AI treatment. The related, and also relatively small, back transformed mean DA difference of ~5.5cm^2 (film only model, Table 8-9) for the difference in change over 12 months for anastrozole treated participants relative to controls at 12 months is also likely to require an automated (non-subjective) method to reliably detect this small change.

A non-proprietary automated program to remeasure MD on the CMN IBIS-II mammographic images could be used to generate new MD measurements, for use in a mixed model and/or mean difference in MD statistical tests. A number of automated programs exist, but in general the creators of the image processing techniques have only described (sometimes in great detail) the steps and algorithms used during image processing, e.g. [410, 511, 512]. Few have made a full, downloadable program freely available for public use. However, as of 2016 the new LIBRA program is freely available. The program has been validated on both ("analogue like") raw and processed digital images, and its measurements are comparable with Cumulus measurements made on the same set of raw and processed digital images [513]. Another program, STRATUS, has not yet been made publicly available but also works on both raw and processed mammographic images [415, 514]. It would be interesting, in the future, to remeasure the CMN IBIS-II mammograms using these programs and assess for differences in the treated and control groups. If the programs are easy to use, and also provides reliable measurements of MD, they could prove a clinically useful method to assess for AI induced changes of MD. The results of this project imply, however, mammogram Version would need to be stable between mammographic episodes for any measurement technique utilising post-processing digital images to provide reasonable estimates of longitudinal MD change.

More coefficients changed between the aggregate model and treated vs control model for DA than for PD, however the differences were also very small; most differed by <10% except for the coefficient for non-parous women. Differences in annual MD change for the control group compared to those for the aggregate model is expected due to the interaction term. Given the small impact on MD annual change noted for the addition of BC risk covariates to the mixed model (i.e. Model 3 vs Model 4, Table 8-19) the change in coefficient value for non-parous women is unlikely to be clinically significant.

The model appeared to be a reasonable fit to the data, without any egregious departures of normality for the RE residuals. The mixed model appears to fit the PD data slightly better than the DA data, Figure 8-4 and Figure 8-5, because the PD predicted vs observed values more closely follow the line of equality. The linear prediction (generated using the estimated FE coefficient values) appears to somewhat under predict both PD and DA at higher observed values and over predict them at lower observed values. This implies that the model could be improved, perhaps through use of additional covariates although (as mentioned) incorporation of covariates does not seem to change the rate of MD change substantially. Use of larger dataset may also improve the predictive abilities of the model by including more women with very high MD.

In the general population, the annual decrease in average PD over time noted at younger ages tapers off for women over age 65 [145, 189]. In this CMN MD and AI project, younger treated participants may have experienced a greater between treatment group difference in decline in PD during baseline to year 1 than their older (>60 years) counterparts, Table 8-16. Greater declines in PD due to AI treatment have previously been associated with lower age and/or higher baseline PD [18, 31, 310, 312], however an association between higher age and/or PD was not found in all studies [30, 31].

Differences in longitudinal DA for older vs younger women, and higher vs lower initial DA were not modelled. It would be interesting to compare the age-related and initial PD and DA results from this project, since the age-related decrease in mean DA appears to be relatively stable until age 70 [189], however this was considered to be beyond the scope of the current project, and would potentially require a large sample.

Many of the earlier studies which examined the effects of AI treatment on MD also examined other potential biomarkers, e.g. Ki-67—a tissue proliferative index [303]; expression of trefoil

protein 1 (TFFI)—an estrogen response gene present in breast tissue [30]; serum insulin-like growth factor (IGF) levels [29]; reductions in bone mineral density [29, 313]; and serum estrogen levels [29, 30, 310]. Although small changes in MD due to AI treatment on their own may be difficult to utilise as biomarker for treatment efficacy, potentially the combination of changes in MD with other biomarkers of treatment effectiveness could be used to predict which women will respond to treatment with AI.

The models fitted in Chapters 7 (Aim 4, aggregate models) and this chapter (Primary Aim, treated vs control groups) displayed the expected decreases in RE variability with the addition of fixed and time varying covariates, Table 8-19. Information on some important confounders of BC risk and MD were not collected for the IBIS-II participants— e.g. number of full term pregnancies, duration of breast feeding, change in BMI over time. Use of these covariates may have improved model fit. The addition of the (BC risk) covariates was found to reduce estimates of the between person variability in the mixed model (i.e. total between person RE variability decreased –16% between model 3 and model 4, Table 8-19), although the growth curve coefficients were not greatly affected.

Differences in the continuous time coefficients, Table 8-19, as well as higher within person variability estimates, between the unconditional growth model (model 2, fitted with an uncut continuous spline) and the unconditional aggregate growth model (model 3, fitted with a three segment continuous spline), show that modelling of time parameters which do not suit the (mean) growth curves in the data can causes errors in model outcomes.

This analysis has a number of strengths. The studied population were participants in a large international, investigator-led clinical trial run by well-established organisations with expertise in designing and coordinating BC randomised controlled trials. The majority of CMN IBIS-II trial participants were willing to contribute their mammograms for use in approved MD

projects. The population was well characterised; for example, a complete set of baseline data was available for every CMN IBIS-II participant included in the sample population. BC risk factors were well balanced between the treatment groups. Many randomisation and all post-randomisation mammograms were undertaken on a single mammography machine at the CMN hospital. The selected statistical model, a hierarchical linear mixed model, is a popular technique for assessing longitudinal growth [490]. The use of processed digital mammograms reflects the circumstances facing most MD research as well as clinical conditions in the post-film era of mammography.

The study was limited by typical difficulties in collecting mammographic data, such as expense, time, and geographical distance. The lack of a freely available (at the time the measurements were undertaken), consistent and fully automated MD measurement technique introduced variability into the MD measurements. The film to digital transition and ongoing changes to digital mammogram acquisition and post-processing at CMN impaired the likelihood of accurately assessing MD change over time for IBIS-II participants with digital mammograms, which comprised most of the available mammographic data. Although statistically significant differences in PD and DA longitudinal growth for the treated group compared to controls were observed during the first year of anastrozole treatment for participants with film mammograms, the modelled outcomes resulted from a small sampled population (as many as 45, but likely as few as 14 women due to weighting within the model) with unequal numbers of episodes contributed per participant. The power for this CMN MD and AI study was low due to the small sample size, and adequate power was only possible for moderate to large effect sizes.

Some limitations on the generalisability of this study exist. The CMN participants in this study are likely to originate from the lower Hunter (greater Newcastle) and other regional areas of NSW, which have a primarily Caucasian population (i.e. approximately 95% are likely to be of European ancestry [515]); therefore these results may only be applicable to primarily Caucasian

populations. All IBIS-II participants are at high risk (1.5 to 2-fold or greater) of BC; hence these results are likely only applicable to other populations of post-menopausal women at high risk of BC (1.5 to 2-fold or greater). The estimated back-transformed intercept values for the Primary and film-only Primary models for PD (~30%) and DA (~3800mm^2) are higher than median PD and DA estimated at baseline for the CMN participants (e.g. 18.9% PD) and median PD for other postmenopausal populations (e.g. 18.7% PD [144]). The back-transformed model coefficient estimates in Table 8-6 and Table 8-9 therefore have larger absolute values than have been tabulated using more representative values for PD and DA during back-transformation of the coefficients (i.e. 20% PD and 2700mm^2 DA, in Table 8-7 and Table 8-10).

8.6 Conclusion

Treatment with anastrozole may be associated with a small, but significant reduction in change in PD and DA over time compared to PD and DA change in similar women not undertaking treatment with anastrozole. Further research is required to ascertain if the AI associated modelled reductions on film mammograms for PD and DA are seen in the broader IBIS-II population, and if they can be consistently measured in other groups of women treated with AI for BC prevention and treatment.

Validation of the results noted for film mammograms is also required for PD and DA measured on digital mammograms, as film mammography is now rarely utilised in developed regions. An additional analysis of treated vs control participants baseline to year 1 mammograms separately by digital mammogram type (KE52, KE54) may provide interesting (hypothesis generating), if non-significant, results for MD measured on digital mammograms.

The clinically important issue of whether these subtle changes in MD associated with AI treatment are also associated with lower of rates of BC incidence in the prevention and treatment settings needs to be addressed. Changes in MD could potentially be combined with

changes in other potential biomarkers for AI treatment, to provide better insight into which women will respond to AI treatment.

Use of post-processed digital mammograms introduced undesirable and probably insurmountable variability in the amount of MD retained on these images, disrupting accurate representation and therefore longitudinal measurement of MD on the CMN IBIS-II participant digital mammograms. Use of a semi-automated measurement technique (i.e. assessor inconsistency) further added to measurement variability. If a treatment effect from anastrozole on MD was present, both of these issues likely contributed to the inability to detect it. Re-measurement of the IBIS-II mammograms using another technique, preferably one that is readily accessible, fully automated, and reliable, may improve the reported estimations of MD change for the anastrozole treated participants relative to controls (for both film and digital mammograms).

The likelihood of detecting changes in MD due to treatment with AI could be improved if both the baseline and one year post-treatment mammograms are taken on mammography machines with identical software and hardware configurations. This would eliminate the differences in image post-processing over time which affected the digital, but not the film MD measurement results in this project. Adoption of another strategy such as collection of the raw mammographic images may also be beneficial, but this latter approach is likely to require additional manipulation of the images prior to measurement. However, use of commercially available programs which used raw mammographic data to measure changes in volumetric breast density for AI (and tamoxifen) treated women diagnosed with BC have provided evidence that this approach may be suitable to detect change due to AI treatment.

Mammographic density may prove to be a valid biomarker for AI treatment efficacy; reductions in MD due to AI treatment in combination with other biomarkers may yield a useful clinical

tool to predict women who will respond to treatment. The major impediment to substantiation of MD as an endocrine therapy biomarker, especially in the age of digital mammography, appears to be lack of a widespread (unified), well validated, reliable and automated method with which to measure it, as well as stability of the (processed) digital images over time. Wider implementation of automated measurement techniques may assist with future studies of MD as a biomarker for AI and other treatments, especially if measurements from the different techniques are interchangeable. Standardisation of the retention (removal) of the dense tissues from digital mammograms, and/or the routine retention and use of 'raw' digital data, would also improve the validity of longitudinal MD/BD studies.

9. Discussion

The Primary Aim of this thesis was to compare longitudinal changes in mammographic density for treated vs control participants in the IBIS-II trial. This was undertaken to discover if measurable MD differences existed between IBIS-II treatment groups, as lower MD in response to anastrozole treatment may be an early indication that treatment is effective. Previous research has shown women who experienced lower levels of MD in response to estrogen-lowering endocrine therapies such as tamoxifen and aromatase inhibitors are at lower risk of developing hormone-sensitive BC [17, 18]. Results from the Primary Aim Models, comprised of MD measured from both film and digital mammograms, did not show a significant difference in response over time for anastrozole treated participants compared to control participants in the CMN MD and AI substudy. However, a possible treatment effect was noted in a sensitivity analysis of the Primary Model utilising film mammograms only— anastrozole treated participants had a small (<5%) but significantly greater rate of annual decline in percent density and dense area from baseline to year one compared to control participants. This is the first time a significant difference in MD change over time has been detected in a randomised controlled clinical trial for women treated with an AI relative to a control group. This result implies MD may be a useful biomarker of treatment efficacy in AI treated women.

Prior to completing the Primary Aim, a number of related preliminary activities were undertaken. The factors affecting BC and MD were reviewed (Chapter 2) because they are useful to know: probable confounders need to be taken into account when preparing models of longitudinal change in MD. It is also necessary to know if important confounders are missing (e.g. age or BMI). The techniques available to measure MD were appraised (Aim 1). Three techniques were selected to measure MD on the Calvary Mater Newcastle IBIS-II mammograms; repeatability results from two of the methods, the visual and Cumulus techniques, are reported in Chapter 5 (Aim 2). The repeatability analysis revealed the Cumulus

MD assessment technique was more reliable than the visual assessment technique. Correlation of the repeated measurements with Cumulus was high; repeated measurements ICCs ranged from 0.94 for digital mammograms to 0.97 for film mammograms. The variability in PD measurement using the Cumulus technique, however, was about PD ±10%, which is approximately equal to the smallest meaningful PD change associated with a difference in clinical outcomes during the IBIS-I (tamoxifen) BC prevention trial [17].

Baseline characteristics of the CMN IBIS-II participants, and associations between the participants' baseline MD and BC risk factors, are described in Chapter 6 (Aim 3). Baseline characteristics of the 120 CMN IBIS-II participants who contributed MD measurements to this study did not differ greatly from the cohort of all international IBIS-II participants. Typical baseline associations between MD and BC risk factors were found for the CMN participants, e.g. higher MD was associated with lower BMI, and lower MD was correlated with increasing age. Film mammograms were found to have significantly higher PD and DA compared to MD measured on digital mammograms.

A longitudinal model of the aggregate (treated and control) MD response to anastrozole treatment was developed in Chapter 7 (Aim 4) using a linear mixed regression model to account for the repeated (longitudinal) measurements in the data. Square root transformations of the outcome parameters PD, DA and BA were undertaken to improve residual normality assumptions for the regression models. The BC risk factors age at randomisation, body mass index, age at menopause and age at first birth were found to be significantly associated with square root transformed PD. The BC risk factors age at menopause and age at first birth were found to be significantly associated with square root transformed DA. Significant differences in PD and DA were found for all digital mammographic Versions relative to film mammograms in all PD and DA models. Adjustment of the longitudinal models by the BC risk factors reduced between person variability in the PD model by 27%, and the DA model by 16%, however these

adjustments had negligible effects upon the PD and DA annual rate of change estimated by the models. Although adjustment for BC risk factors is very important during modelling of MD and BC risk [178], this result implies if the outcome of interest is solely the response of longitudinal MD to AI treatment (and not BC risk/BC outcomes) it may not be necessary to adjust for BC risk factors during modelling of the treatment effect. This feature could be useful during community studies of AI prevention, if the BC characteristics for the women are not reported or known.

The Primary Aim (Aim 5, Chapter 8) was accomplished by adding terms for anastrozole treatment to the aggregate (blinded) model from Chapter 7 in an unblinded analysis of the CMN IBIS-II MD response to trial treatment. The analysis was performed with the assistance of collaborators at QMUL, who have access to the treated vs control (unblinded) IBIS-II trial data. Although a non-significant rate of annual change in PD between baseline and year 1 for treated participants vs control of +0.2% was estimated by the Primary Model (both digital and film mammograms, 120 participants) and +0.8% (non-significant) for the digital mammogram only Primary model (120 participants), the film mammogram only Primary model (45 participants) showed an anticipated and significant average annual rate of decline of approximately –3% from baseline to year 1 for anastrozole treated participants relative to placebo treated participants. The baseline to year 1 changes for DA for treated vs control participants were similar to those for PD: DA increased non-significantly for the Primary model (all mammograms) and digital mammogram-only Primary model, but annual rate of change DA decreased significantly for the film-only Primary model (–5 cm^2/year on average).

No other significant decreases (or increases) in the rate of annual change for the treated participants relative to controls were noted for PD or DA for the Primary model, digital mammogram-only Primary Model for the other time periods modelled (years 1 to 5, years 5 to 7). (All film episodes bar one occurred at or prior to year 1, hence no reliable data after the first

year post-randomisation were available for the film-only Primary model.) Given the small sample size (120 participants), a suggestion of DA reduction for anastrozole treated participants relative to controls was observed: treated vs control annual DA change for years 1 to 5 for the digital mammogram-only Primary models was about $-0.5cm^2$/year ($-66mm^2$/year), p=0.14.

These results are based on a very small sample size of women, especially the results for the film-only Primary model which (due to weighing of repeated measurements in the mixed linear regression model) may be based on a sample of as few as 14 participants. This is one of the principal limitations of this project. Hence this project has insufficient power to detect differences if they exist; because of the small sample size and unequal number of episodes per participant, the significant baseline to year 1 result for the film-only Primary model should be viewed cautiously. Further investigation of the Primary model results within a larger population is warranted. Fortunately, a mammographic density study is planned for the international cohort of IBIS-II participants; this other, international project will hopefully confirm and expand upon the results presented in this thesis. This international study may also be able to address the issue of monitoring individual women, e.g. what is the smallest AI-induced MD change possible to reliably detect for the MD assessment methods chosen for the analysis.

As described in Chapter 8, other limitations besides the small sample size/limited power include the variability of the subjective assessment technique, and the changes in dense appearance of the mammograms due changes in mammographic Versions (e.g. image post-processing software). The differences in MD distribution and appearance of images from the three different digital mammographic Versions utilised in this project infers that adjustment of the post processing algorithms to maximise BC detection in digital mammography is ongoing, The approach taken during post-processing by the various machine vendors also appears to differ.

It is understandable that mammography machine manufacturers would incorporate software algorithms to remove as much of the dense appearing tissues of the breast on digital mammographic images as possible, without (hopefully) affecting the ability to detect cancerous lesions. Mammographic density has long been one of the prime causes of missed BC detection on mammograms. The ability to increase BC detection through manipulation of the digital mammographic image is one of the reasons why digital mammography has overall superior to rates of BC detection compared to film [516, 517]. Sensitivity in women with dense breasts, and younger women (who tend to have denser breasts) is enhanced for digital mammograms compared to film. However, film may be superior for BC detection for women with older women (who tend to have lower MD) and women with fatty breasts [459]. Given that the main purpose of mammography is to detect BC, manipulation of the resulting images to enhance BC detection makes sense. This manipulation however has caused loss of information for secondary use of the mammograms, such as detection of longitudinal differences in MD due to AI treatment in this project.

Given the promising results of the VPD, AI and TAM study, and those for a another AI study which successfully utilised longitudinal raw mammographic images in conjunction with film images [31], use of the raw mammographic data for the CMN IBIS-II participants would likely have provided a better insight into the MD changes for the treated group relative to the control group.

Processed digital images have been successfully utilised in projects comparing MD and BC risk [504, 512, 513]. The correlation between PD measured on processed vs raw digital image data is typically high, and the associations for MD and BC risk produced with either type of image similar. However, a general recommendation from these studies is to use the same processed digital image type (i.e. Version) throughout population or longitudinal studies. The Primary model was not repeated for subsets of mammograms with a single digital Version only (i.e.

separate models were not fitted for KE52, KE54 and Fuji only mammograms), because of the small sample sizes.

Because all IBIS-II participants are at high risk of BC, the results of this study are likely only generalisable to other populations of post-menopausal women who are also at high risk (1.5 to 2–fold or more) of BC. Therefore the results of this thesis may not be applicable to populations of average-risk women (who may or may not consider AI treatment for BC prevention as their risk of BC is lower than that for IBIS-II participants). Strengths of the study include that the comparisons of MD change over time were made for two similar groups of women randomly assigned to anastrozole or placebo treatment, and that the BC risk characteristics for these women were well documented, because they are participants of randomised trial coordinated by experienced, investigator-led cooperative BC trials groups.

The results from the Primary models are similar to those for other projects which compared change in aromatase inhibitors in treated compared to control populations. Non-significant, but lower PD at around one year of treatment has generally been observed for AI treated women compared to control women in other randomised studies [28, 29, 313]. Very recently published results, however, demonstrated small but generally significant annual reductions in volumetric density (–0.25 to –0.75%/year) for cases (women treated with AI for BC) compared to healthy controls from the general screening population [33].

In this project, the CMN MD and AI substudy, the first year rate of change in PD for anastrozole treated participants relative to controls noted for participants with film mammograms in this project is relatively small, about –3%. This small annual change in PD is unlikely to be reliably detected using visual assessment of mammograms (i.e. during mammography clinics); other methods of detecting the subtle changes which are likely to be imparted on the breasts during AI treatment should be employed, such as the commercial

volumetric techniques utilised in the recent volumetric density project. Potentially, other biomarkers besides changes in MD could be used either in lieu of, or in addition to this potential AI biomarker.

The reviews of BC and MD in this thesis confirmed BC is very heterogenous disease, with many associated risk factors. Few strong (>3-fold) risk factors for BC exist. It is difficult to predict which women will develop BC and which will not, even among populations at very high risk such as women with BRCA1 mutations. Hence screening for populations of women at higher risk of BC is essential, to increase the likelihood of detection of BC at an early stage for the best treatment and survival outcomes. BC risk increases with age. MD (a strong BC risk factor) is typically too extensive in younger women to make mammography worthwhile (BC is concealed by the dense breast tissues); however MD tends to decline with age, particularly over the menopausal transition. The combination of increasing BC risk with age and declining MD (in conjunction with other factors) is why many population-based, mammographic screening programs, including Breast Screen Australia, commence invitations to screen around the average age of menopause (50 to 51 years). High MD is an important issue clinically [201], as it often masks cancer detection on mammograms. The best screening paradigm for women with high MD, however, has not yet been established.

The methods review in this thesis (Aim 1, Chapter 4) re-confirmed that MD is difficult to measure repeatably and reliably; this fact was also demonstrated during Aim 2 (Chapter 2, reliability analysis) and Aims 4 and 5 (Chapters 7 and 8, longitudinal MD analyses). This difficulty in reliably measuring MD is part of the reason a woman's MD is not generally disclosed to her during BC screening, and is likely the primary reason why MD is not studied more broadly as a potential biomarker for treatment efficacy. Many techniques exist to measure breast density, and more are developed every decade. However no single method has been

shown to be superior to others, hence no true standards exist to measure MD and many different methods are in current use.

The lack of a precise, quick, reliable and easy to use method to measure mammographic density impacted upon the ability of this project to assess MD differences in the CMN treated vs control IBIS-II participants. Additionally, the differences in post-processing of digital mammograms caused variable amounts of density to be removed from the post-processed digital mammograms. Besides the measurement errors introduced by the subjective MD assessment method, changes to the appearance of MD on the (processed) digital mammograms were likely the other major source of error in assessment of the true MD difference between IBIS-II anastrozole treated participants and controls. Two extant commercial measurement programs integrated with digital mammography machines (which utilise pre-processing ('raw') mammographic data) may have provided longitudinal data suitable for measurement of the small average percent density change (–1 to –2%) expected from anastrozole treatment in IBIS-II participants, if it had been possible to use these methods during this project.

Visual methods are adequate for epidemiological purposes to categorise women into levels of BC risk based on the amount of MD on their mammograms, and for categorising the likelihood that BC has been missed on a mammogram. This thesis confirms, however, previous research which shows consistent measurements of MD are difficult to reproduce, and semi-automated computer programs provide consistency which is superior to visual methods. The fluctuations in MD assessment noted even with utilisation of the more consistent semi-automated technique imply a fully automated method would enhance the likelihood of a definitive outcome. Given that BC prevention therapies are highly desirable, especially those that may be particularly efficacious in women with high MD (e.g. [20]), use of imaging techniques to improve use of MD as a biomarker for treatment efficacy are urgently required.

9.1 Future directions

Ongoing research is required for techniques to identify women who will develop breast cancer and those who will not. The evidence provided in this thesis indicates breast density may be modestly reduced in women treated with anastrozole for BC prevention compared to similar women who do not receive treatment. The reviews of BC and MD in earlier chapters of this thesis indicate that MD may be an underutilised BC biomarker in many different areas, such as 1.) characterisation of MD from puberty to first full-term pregnancy as a biomarker for BC risk (this is an area of active research); 2.) characterisation of intra-, post- and inter-pregnancy changes in MD as a biomarker for BC risk (can changes in MD be utilised as a pregnancy-associated BC predictor?); 3.) high vs low MD and risk of BC recurrence with and without radiotherapy. A recent review of MD as a potential biomarker for adjuvant and preventive BC treatments outlines the steps which might be undertaken to enable broader use of MD as an endocrine therapy biomarker [518].

Specific recommendations for future research, based on the work included in this thesis are:

1. *Confirmation that reduction in MD during (the first year of?) anastrozole treatment for BC prevention is associated with a reduction in BC incidence,* such as what has been observed for preventive treatment with tamoxifen [17] as well as AI treatment for BC [18]. The very recent results of the VPD, AI and TAM project [33] provide good evidence that MD may be a useful biomarker for AI treatment efficacy. Results from the international IBIS-II MD and AI project coordinated by QMUL will hopefully provide insight into whether reductions in MD due to AI treatment are associated with reductions in BC incidence within the IBIS-II randomised control trial population.

2. *Use of a mixed model to model MD change on unblinded IBIS-II data, to ascertain if certain characteristics are associated with greater or lesser changes in MD, such as those*

due to AI treatment. If certain characteristics (such as higher baseline MD, recent menopause) are associated with MD change, they may add to the predictive capacity to ascertain women who will most benefit from AI preventive therapy. One of the strengths of a mixed model is its ability to provide insights into associations with longitudinal change in the model. Individual change trajectories could potentially be examined for common characteristics to see if certain traits are associated with rates of change in response to AI treatment, or perhaps why some women in the control group experienced a more rapid decline in MD than others.

3. *Methods to reliably and repeatably measure MD over time are required.* This has been an ongoing issue for MD research, as well as clinical practice, since the inception of Wolfe's MD categories. Although this is not a new issue raised first by this thesis, this issue impacted upon the ability of this project to differentiate longitudinal change between the IBIS-II treatment groups. The clinically approved, fully automated volumetric density measurement tools (Volpara, Quantra) are the most likely candidates for immediate implementation to resolve this issue, as demonstrated for example by the VPD, AI and TAM project [33] . However, these commercial methods require access to the raw mammographic digital data, which may not be possible to collect in many circumstances. Other techniques, e.g. [513], may also be useful for measurement of post-processed digital mammograms.

4. *Standardisation of the mammographic output of conventional (two-dimensional) mammography and three-dimensional breast tomography* to assist with longitudinal assessment of breast density as a biomarker during clinical therapy for BC prevention and treatment. Standardisation of the image output has other potential purposes as well, such as during MD stratification into higher risk and lower risk categories for risk assessment, and characterisation of longitudinal MD in specific groups of women (i.e. different risk factors for BC). A naïve suggestion would be to retain a set of pre-processed images along with the

set of processed images used for clinical purposes. Given that this would double the amount of storage space needed to save the electronic images, perhaps pre-processed images of (greatly) reduced resolution could be preserved as it might still be feasible to obtain adequate cross sectional and longitudinal MD information from lower quality images. It is likely that removal of variable amounts of MD from mammographic images (as noted for the longitudinal images utilised in this thesis) adds to the inherent variability [519] of clinical (visual) assessments of MD. Access to a consistent set of images (unaffected by the vagaries of digital post processing) may assist with accuracy of MD in clinical visual assessment, especially given the increasing requirements to report breast density (re: the legislation in the USA [520]).

5. *Elucidation of the earliest time/s of change in breast density in response to treatment, as well as the time at which the maximum rate of breast density change occurs* to assist with characterisation of the early treatment response as a biomarker for treatment efficacy. It is not known how quickly the dense tissues of the breast change in response to different stimuli such as hormonal treatments, pregnancy, cessation of breast feeding, and menopause. It is plausible women who have a visible or —via some other assessment technique— detectable change in breast density three months after the start of endocrine treatment for BC prevention (or treatment) are at a lower risk of BC in comparison to women who do not have detectable changes in breast density at 6 or 12 months. The rapidity and magnitude of the BD response might not only depend upon a woman's age or menopausal status, but might also differ by the treatment given (e.g. AI, tamoxifen, chemotherapy) and other characteristics of the woman (such as number of children, duration of breast feeding, genetic composition). Given that mammography is potentially harmful, use of non-ionising imaging techniques such as ultrasound, non-contrast MRI or other more novel approaches [521, 522] could be used to precisely characterise the short and long term breast tissue response to treatment and other stimuli.

9.2 Conclusion

Whilst the information presented in this thesis provides insight into the longitudinal changes relative to controls experienced by high risk women who take anastrozole for BC prevention, these results were confounded by the technical changes in image acquisition which occurred during the IBIS-II trial. The mammographic image acquisition and MD assessment issues posed for this analysis however, are likely representative of the issues facing most modern, retrospective projects to assess longitudinal MD, for example in response to AI treatment. Prospective collection of pre-processed digital images, had they been available for CMN IBIS-II trial participants from baseline, likely would have produced more reliable quantification of the effects of anastrozole on longitudinal MD. Since the advent of xeromammography for breast imaging, ongoing improvements to and expansion of the techniques used to image breasts has led to changes to in routine clinical practice, therefore regular changes to the techniques and standards of practice are likely to continue. This provides hope that consistent and reliable techniques may become widely available to adequately assess MD as a biomarker, such as during AI treatment for BC prevention.

The small, but significant reduction in MD annual rate of change during the first year of therapy for anastrozole treated CMN IBIS-II participants relative to controls for participants undertaking film mammography provides an indication that MD may be a biomarker for AI clinical efficacy. Further research is required to ascertain the status MD as a biomarker for treatment with AI and other hormonal therapies.

References

1. Australian Institute of Health and Welfare 2014 *Cancer in Australia: an overview, 2014.* 2014. **Cancer series no. 90. Cat. no. CAN 88.**
2. World Health Organization. *Global Health Observatory data repository*. [cited 2016 6 September]; Available from: http://apps.who.int/gho/data/node.main.CODWORLD?lang=en.
3. DeVita, V.T. and E. Chu, *A History of Cancer Chemotherapy.* Cancer Research, 2008. **68**(21): p. 8643-8653.
4. Chabner, B.A. and T.G. Roberts, *Chemotherapy and the war on cancer.* Nat Rev Cancer, 2005. **5**(1): p. 65-72.
5. National Breast and Ovarian Cancer Centre (NBOCC), *Breast cancer risk factors: a review of the evidence.*, Australian Government Department of Health and Ageing, Editor. 2009, NBOCC: Surry Hills, NSW.
6. Sherratt, M.J., J.C. McConnell, and C.H. Streuli, *Raised mammographic density: causative mechanisms and biological consequences.* Breast Cancer Res, 2016. **18**(1): p. 45.
7. Price, E.R., et al., *The California breast density information group: a collaborative response to the issues of breast density, breast cancer risk, and breast density notification legislation.* Radiology, 2013. **269**(3): p. 887-92.
8. Wolfe, J.N., *Breast patterns as an index of risk for developing breast cancer.* AJR Am J Roentgenol, 1976. **126**(6): p. 1130-7.
9. Egan, R.L. and R.C. Mosteller, *Breast cancer mammography patterns.* Cancer, 1977. **40**(5): p. 2087-90.
10. Berg, W.A., *Current Status of Supplemental Screening in Dense Breasts.* Journal of Clinical Oncology, 2016. **34**(16): p. 1840-1843.
11. Melnikow, J., et al., *U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews*, in *Supplemental Screening for Breast Cancer in Women With Dense Breasts: A Systematic Review for the U.S. Preventive Service Task Force*. 2016, Agency for Healthcare Research and Quality (US): Rockville (MD).
12. Cancer Prevention and Research Institute (Italy) *Tailored Screening for Breast Cancer in Premenopausal Women (TBST), NCT02619123*. 2018 [cited 2018 22 April 2018]; Available from: https://clinicaltrials.gov/ct2/show/NCT02619123.
13. UMC Utrecht *Breast Cancer Screening With MRI in Women Aged 50-75 Years With Extremely Dense Breast Tissue: the DENSE Trial, NCT01315015*. 2018 [cited 2018 22 April 2018]; Available from: clinicaltrials.gov/ct2/show/NCT01315015.
14. University of California (Davis) *Assessing Breast Density's Value in Imaging - A Comparative Effectiveness Study (BCSC-ADVANCE), NCT02980848*. 2018 [cited 2018 22 April 2018]; Available from: clinicaltrials.gov/ct2/show/NCT02980848.
15. University of California (San Francisco) *Women Informed to Screen Depending on Measures of Risk (Wisdom Study) (WISDOM),NCT02620852*. 2018 [cited 2018 22 April 2018]; Available from: clinicaltrials.gov/ct2/show/NCT02620852.
16. Cecchini, R.S., et al., *Baseline mammographic breast density and the risk of invasive breast cancer in postmenopausal women participating in the NSABP study of tamoxifen and raloxifene (STAR).* Cancer Prev Res (Phila), 2012. **5**(11): p. 1321-9.
17. Cuzick, J., et al., *Tamoxifen-induced reduction in mammographic density and breast cancer risk reduction: a nested case-control study.* J Natl Cancer Inst, 2011. **103**(9): p. 744-52.

18. Kim, J., et al., *Breast density change as a predictive surrogate for response to adjuvant endocrine therapy in hormone receptor positive breast cancer.* Breast Cancer Res, 2012. **14**(4): p. R102.
19. DeFilippis, R.A., et al., *CD36 repression activates a multicellular stromal program shared by high mammographic density and tumor tissues.* Cancer Discov, 2012. **2**(9): p. 826-39.
20. Sun, X., et al., *CCL2-driven inflammation increases mammary gland stromal density and cancer susceptibility in a transgenic mouse model.* Breast Cancer Research, 2017. **19**(1): p. 4.
21. Luo, T., et al., *Aspirin use and breast cancer risk: a meta-analysis.* Breast Cancer Res Treat, 2012. **131**(2): p. 581-587.
22. Early Breast Cancer Trialists' Collaborative Group (EBCTG), *Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials.* The Lancet, 2011. **378**(9793): p. 771-784.
23. Fabian, C.J., *The what, why and how of aromatase inhibitors: hormonal agents for treatment and prevention of breast cancer.* International Journal of Clinical Practice, 2007. **61**(12): p. 2051-2063.
24. Buzdar, A.U., et al., *Summary of aromatase inhibitor clinical trials in postmenopausal women with early breast cancer.* Cancer, 2008. **112**(3 Suppl): p. 700-9.
25. Barnes, B.B., et al., *Population attributable risk of invasive postmenopausal breast cancer and breast cancer subtypes for modifiable and non-modifiable risk factors.* Cancer Epidemiol, 2011. **35**(4): p. 345-52.
26. Cuzick, J., et al., *Anastrozole for prevention of breast cancer in high-risk postmenopausal women (IBIS-II): an international, double-blind, randomised placebo-controlled trial.* Lancet, 2014. **383**(9922): p. 1041-8.
27. Goss, P.E., et al., *Exemestane for Breast-Cancer Prevention in Postmenopausal Women.* New England Journal of Medicine, 2011. **364**(25): p. 2381-2391.
28. Vachon, C.M., et al., *Pilot study of the impact of letrozole vs. placebo on breast density in women completing 5 years of tamoxifen.* Breast, 2007. **16**(2): p. 204-10.
29. Cigler, T., et al., *A randomized, placebo-controlled trial (NCIC CTG MAP1) examining the effects of letrozole on mammographic breast density and other end organs in postmenopausal women.* Breast Cancer Res Treat, 2010. **120**(2): p. 427-35.
30. Gatti-Mays, M.E., et al., *Exemestane Use in Postmenopausal Women at High Risk for Invasive Breast Cancer: Evaluating Biomarkers of Efficacy and Safety.* Cancer Prevention Research, 2016. **9**(3): p. 225-233.
31. Henry, N.L., et al., *Aromatase inhibitor-induced modulation of breast density: clinical and genetic effects.* Br J Cancer, 2013. **109**(9): p. 2331-2339.
32. van Nes, J.G.H., et al., *Minimal impact of adjuvant exemestane or tamoxifen treatment on mammographic breast density in postmenopausal breast cancer patients: A Dutch TEAM trial analysis.* Acta Oncologica, 2014. **0**(0): p. 1-12.
33. Engmann, N.J., et al., *Longitudinal changes in volumetric breast density with tamoxifen and aromatase inhibitors.* Cancer Epidemiology Biomarkers & Prevention, 2017.
34. Are You Dense? *Dr. Nancy Cappello's Story*. 2008 [cited 2016 3 Nov 2016]; Available from: https://www.areyoudense.org/stories/nancy/.
35. Cappello, N.M., *Decade of 'normal' mammography reports--the happygram.* J Am Coll Radiol, 2013. **10**(12): p. 903-8.
36. Weiss, R.E., *Modelling Longitudinal Data*. Springer Texts in Statistics. 2005, New York, NY USA: Springer.
37. Macias, H. and L. Hinck, *Mammary gland development.* Wiley Interdiscip Rev Dev Biol, 2012. **1**(4): p. 533-57.

38. Pandya, S. and R.G. Moore, *Breast development and anatomy.* Clin Obstet Gynecol, 2011. **54**(1): p. 91-5.
39. Ginsburg, O.M., L.J. Martin, and N.F. Boyd, *Mammographic density, lobular involution, and risk of breast cancer.* Br J Cancer, 2008. **99**(9): p. 1369-74.
40. Australian Institute of Health and Welfare & National Breast Cancer Centre 2006 *Breast cancer in Australia: an overview, 2006.* 2006. **Cancer series no. 34. cat. no. CAN 29**.
41. National Breast Cancer Centre and National Cancer Control Initiative, *Clinical Practice Guidelines for Psychosocial Care of Adults with Cancer*, National Breast Cancer Centre, Editor. 2003, National Breast Cancer Centre,: Camperdown, NSW.
42. National Breast Cancer Centre, *Cancer - How are you travelling? Understanding the emotional and social impact of cancer*, N.B.C. Centre, Editor. 2007, National Breast Cancer Centre: Camperdown, NSW.
43. Gorey, K.M., et al., *Breast cancer survival in ontario and california, 1998-2006: socioeconomic inequity remains much greater in the United States.* Ann Epidemiol, 2009. **19**(2): p. 121-4.
44. Australian Institute of Health and Welfare & Cancer Australia 2012 *Breast cancer in Australia: an overview.* 2012. **Cancer series no. 71. Cat. no. CAN 67.**, 1-202.
45. Albrektsen, G., I. Heuch, and S.O. Thoresen, *Histological type and grade of breast cancer tumors by parity, age at birth, and time since birth: a register-based study in Norway.* BMC Cancer, 2010. **10**: p. 226.
46. Ijaduola, T.G. and E.B. Smith, *Pattern of breast cancer among white-American, African-American, and nonimmigrant west-African women.* J Natl Med Assoc, 1998. **90**(9): p. 547-51.
47. Li, C.I., K.E. Malone, and J.R. Daling, *Differences in breast cancer hormone receptor status and histology by race and ethnicity among women 50 years of age and older.* Cancer Epidemiol Biomarkers Prev, 2002. **11**(7): p. 601-7.
48. Sakamoto, G. and H. Sugano, *Pathology of breast cancer: present and prospect in Japan.* Breast Cancer Res Treat, 1991. **18 Suppl 1**: p. S81-3.
49. Bhikoo, R., et al., *Systematic review of breast cancer biology in developing countries (part 1): Africa, the middle East, eastern europe, Mexico, the Caribbean and South america.* Cancers (Basel), 2011. **3**(2): p. 2358-81.
50. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* International Journal of Cancer, 2014: p. n/a-n/a.
51. World Health Organization (WHO) *World Health Statistics 2014*. 2014.
52. Peters, D.H., et al., *Poverty and Access to Health Care in Developing Countries.* Annals of the New York Academy of Sciences, 2008. **1136**(1): p. 161-171.
53. Yip, C.H. and N.A. Taib, *Breast health in developing countries.* Climacteric, 2014: p. 1-6.
54. Remennick, L., *The Challenge of Early Breast Cancer Detection among Immigrant and Minority Women in Multicultural Societies.* The Breast Journal, 2006. **12**: p. S103-S110.
55. Bhikoo, R., et al., *Systematic review of breast cancer biology in developing countries (part 2): asian subcontinent and South East Asia.* Cancers (Basel), 2011. **3**(2): p. 2382-401.
56. Smith, R.A., et al., *Breast Cancer in Limited-Resource Countries: Early Detection and Access to Care.* The Breast Journal, 2006. **12**: p. S16-S26.
57. Maher, D., N. Ford, and N. Unwin, *Priorities for developing countries in the global response to non-communicable diseases.* Global Health, 2012. **8**: p. 14.
58. Lancaster, G.A., M. Green, and S. Lane, *Reducing bias in ecological studies: an evaluation of different methodologies.* Journal of the Royal Statistical Society: Series A (Statistics in Society), 2006. **169**(4): p. 681-700.

59. Che, Y., et al., *Comparison of survival rates between Chinese and Thai patients with breast cancer.* Asian Pac J Cancer Prev, 2014. **15**(15): p. 6029-33.
60. Bigaard, J., et al., *Breast cancer incidence by estrogen receptor status in Denmark from 1996 to 2007.* Breast Cancer Res Treat, 2012. **136**(2): p. 559-64.
61. McCormack, V.A., et al., *International Consortium on Mammographic Density: Methodology and population diversity captured across 22 countries.* Cancer Epidemiology, 2016. **40**: p. 141-151.
62. Australian Institute of Health and Welfare (AIHW), *Australian Cancer Incidence and Mortality (ACIM) book: Breast Cancer*, Australian Institute of Health and Welfare, Editor. 2017, AIHW: Canberra.
63. National Cancer Institute (NCI). *NCI Dictionary of Cancer Terms*. 2016; Available from: http://www.cancer.gov/publications/dictionaries/cancer-terms.
64. Stoll, B.A., *Hormonal management of advanced breast cancer.* Br Med J, 1969. **2**(5652): p. 293-7.
65. National Breast and Ovarian Cancer Centre and Australian Cancer Network, *The pathology reporting of breast cancer. A guide for pathologists, surgeons, radiologists and oncologists*, National Breast and Ovarian Cancer Centre, Editor. 2008, NBOCC: Surry Hills, NSW.
66. Bertos, N.R. and M. Park, *Breast cancer - one term, many entities?* J Clin Invest, 2011. **121**(10): p. 3789-96.
67. Perou, C., et al., *Molecular portraits of human breast tumours.* Nature, 2000. **406**: p. 747 - 752.
68. Lord, S.J., et al., *Breast cancer risk and hormone receptor status in older women by parity, age of first birth, and breastfeeding: a case-control study.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(7): p. 1723-30.
69. Barnes, B.B., et al., *Population attributable risk of invasive postmenopausal breast cancer and breast cancer subtypes for modifiable and non-modifiable risk factors.* Cancer Epidemiol, 2010.
70. Althuis, M.D., et al., *Etiology of hormone receptor-defined breast cancer: a systematic review of the literature.* Cancer Epidemiol Biomarkers Prev, 2004. **13**(10): p. 1558-68.
71. Collins, L., et al., *Pathologic features and molecular phenotype by patient age in a large cohort of young women with breast cancer.* Breast Cancer Res Treat, 2012. **131**(3): p. 1061-1066.
72. Higgins, M.J. and V. Stearns, *Pharmacogenetics of Endocrine Therapy for Breast Cancer.* Annual Review of Medicine, 2011. **62**(1): p. 281-293.
73. Bombonati, A. and D.C. Sgroi, *The molecular pathology of breast cancer progression.* J Pathol, 2011. **223**(2): p. 308-18.
74. Graubard, B.I., A.N. Freedman, and M.H. Gail, *Five-year and lifetime risk of breast cancer among U.S. subpopulations: implications for magnetic resonance imaging screening.* Cancer Epidemiol Biomarkers Prev, 2010. **19**(10): p. 2430-6.
75. Australian Institute of Health and Welfare & National Breast Cancer Centre 2006 *Breast cancer in Australia: an overview, 2009.* 2009. **Cancer series no. 50. Cat. no. CAN 46**.
76. Wooster, R. and B.L. Weber, *Breast and Ovarian Cancer.* New England Journal of Medicine, 2003. **348**(23): p. 2339-2347.
77. Nolan, E., et al., *RANK ligand as a potential target for breast cancer prevention in BRCA1-mutation carriers.* Nat Med, 2016. **22**(8): p. 933-9.
78. Islami, F., et al., *Breastfeeding and breast cancer risk by receptor status--a systematic review and meta-analysis.* Ann Oncol, 2015. **26**(12): p. 2398-407.
79. Manrique Tejedor, J., M.I. Figuerol Caldero, and A. Cuellar De Frutos, *[BREASTFEEDING AS A METHOD OF BREAST CANCER PREVENTION].* Rev Enferm, 2015. **38**(12): p. 32-8.

80. Collaborative Group on Hormonal Factors in Breast Cancer, *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): p. 187-95.
81. Russo, I.H. and J. Russo, *Pregnancy-induced changes in breast cancer risk.* J Mammary Gland Biol Neoplasia, 2011. **16**(3): p. 221-33.
82. Kobayashi, S., et al., *Reproductive history and breast cancer risk.* Breast Cancer, 2012. **19**(4): p. 302-8.
83. Lyons, T., P. Schedin, and V. Borges, *Pregnancy and Breast Cancer: when They Collide.* J Mammary Gland Biol Neoplasia, 2009. **14**(2): p. 87-98.
84. Russo, J., et al., *Breast differentiation and its implication in cancer prevention.* Clin Cancer Res, 2005. **11**(2 Pt 2): p. 931s-6s.
85. Pike, M., et al., *'Hormonal' risk factors, 'breast tissue age' and the age-incidence of breast cancer.* Nature, 1983. **303**(5920): p. 767-770.
86. Boyd, N.F., et al., *Mammographic density.* Breast Cancer Res, 2009. **11 Suppl 3**: p. S4.
87. Russo, J., et al., *The protective role of pregnancy in breast cancer.* Breast Cancer Res, 2005. **7**(3): p. 131-42.
88. Russo, J. and I.H. Russo, *The role of estrogen in the initiation of breast cancer.* J Steroid Biochem Mol Biol, 2006. **102**(1-5): p. 89-96.
89. Howell, A., et al., *Risk determination and prevention of breast cancer.* Breast Cancer Res, 2014. **16**(5): p. 446.
90. Wiechmann, L., et al., *Presenting features of breast cancer differ by molecular subtype.* Ann Surg Oncol, 2009. **16**(10): p. 2705-10.
91. ESHRE, C.W.G., *Hormones and breast cancer.* Hum Reprod Update, 2004. **10**(4): p. 281-93.
92. Heldring, N., et al., *Estrogen Receptors: How Do They Signal and What Are Their Targets.* Physiol. Rev., 2007. **87**(3): p. 905-931.
93. Forbes, J.F., et al., *Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 100-month analysis of the ATAC trial.* Lancet Oncol, 2008. **9**(1): p. 45-53.
94. Sestak, I. and J. Cuzick, *Preventive therapy for breast cancer.* Curr Oncol Rep, 2012. **14**(6): p. 568-73.
95. Strassmer-Weippl, K. and P.E. Goss, *Prevention of breast cancer using SERMs and aromatase inhibitors.* J Mammary Gland Biol Neoplasia, 2003. **8**(1): p. 5-18.
96. Davies, C., et al., *Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial.* The Lancet, 2013(0).
97. Vogel, V.G., *The NSABP Study of Tamoxifen and Raloxifene (STAR) trial.* Expert Rev Anticancer Ther, 2009. **9**(1): p. 51-60.
98. Early Breast Cancer Trialist's Collaborative Group (EBCTG), *Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials.* The Lancet, 2015. **386**(10001): p. 1341-1352.
99. Martino, S., et al., *The role of selective estrogen receptor modulators in the prevention of breast cancer: comparison of the clinical trials.* Oncologist, 2004. **9**(2): p. 116-25.
100. Ellis, A.J., et al., *Selective estrogen receptor modulators in clinical practice: a safety overview.* Expert Opin Drug Saf, 2015. **14**(6): p. 921-34.
101. Keogh, L., et al., *Australian clinicians and chemoprevention for women at high familial risk for breast cancer.* Hereditary Cancer in Clinical Practice, 2009. **7**(1): p. 9.
102. Kaplan, C., et al., *Willingness to use tamoxifen to prevent breast cancer among diverse women.* Breast Cancer Res Treat, 2012: p. 1-10.

103. Li, F., et al., *The selective estrogen receptor modulators in breast cancer prevention.* Cancer Chemother Pharmacol, 2016. **77**(5): p. 895-903.
104. Bambhroliya, A., M. Chavez-MacGregor, and A.M. Brewster, *Barriers to the Use of Breast Cancer Risk Reduction Therapies.* J Natl Compr Canc Netw, 2015. **13**(7): p. 927-35.
105. Fontein, D.B.Y., et al., *Specific Adverse Events Predict Survival Benefit in Patients Treated With Tamoxifen or Aromatase Inhibitors: An International Tamoxifen Exemestane Adjuvant Multinational Trial Analysis.* Journal of Clinical Oncology, 2013.
106. Cuzick, J., et al., *Treatment-emergent endocrine symptoms and the risk of breast cancer recurrence: a retrospective analysis of the ATAC trial.* Lancet Oncol, 2008. **9**(12): p. 1143-8.
107. Liedke, P.E.R. and P.E. Goss, *Aromatase inhibitors and musculoskeletal adverse events.* Lancet Oncol, 2011(0).
108. Mieog, J.S., et al., *Carpal tunnel syndrome and musculoskeletal symptoms in postmenopausal women with early breast cancer treated with exemestane or tamoxifen after 2-3 years of tamoxifen: a retrospective analysis of the Intergroup Exemestane Study.* Lancet Oncol, 2012. **13**(4): p. 420-32.
109. Waltman, N.L., et al., *Vitamin D insufficiency and musculoskeletal symptoms in breast cancer survivors on aromatase inhibitor therapy.* Cancer Nurs, 2009. **32**(2): p. 143-50.
110. Napoli, N., et al., *High prevalence of low vitamin D and musculoskeletal complaints in women with breast cancer.* Breast J, 2010. **16**(6): p. 609-16.
111. Russo, J., G.A. Balogh, and I.H. Russo, *Full-term pregnancy induces a specific genomic signature in the human breast.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(1): p. 51-66.
112. Cuzick, J., *Chemoprevention of breast cancer.* Breast Cancer, 2008. **15**(1): p. 10-6.
113. Kumar, P. and R. Aggarwal, *An overview of triple-negative breast cancer.* Arch Gynecol Obstet, 2016. **293**(2): p. 247-69.
114. Newman, L.A., et al., *The 2014 Society of Surgical Oncology Susan G. Komen for the Cure Symposium: triple-negative breast cancer.* Ann Surg Oncol, 2015. **22**(3): p. 874-82.
115. Murphy, R.A., et al., *Beyond breast cancer: mammographic features and mortality risk in a population of healthy women.* PLoS One, 2013. **8**(10): p. e78722.
116. American College of Radiology, *Breast imaging reporting and data system (BI-RADS)*. Edition 1 ed. 1992, Reston, VA: American College of Radiology.
117. American College of Radiology *ACR BI-RADS Atlas Fifth Edition Atlas - Mammography. II Reporting System*. 2013.
118. American College of Radiology and Breast Cancer Surveillance Consortium *(American College of Radiology and the Breast Cancer Surveillance Consortium) THE ACR BREAST IMAGING REPORTING AND DATA SYSTEM (BI-RADS®)*. 2003.
119. American College of Radiology *ACR BI-RADS Atlas Fifth Edition Quick Reference*. 2013.
120. Dehkordy, S.F. and R.C. Carlos, *Dense breast legislation in the United States: state of the states.* J Am Coll Radiol, 2013. **10**(12): p. 899-902.
121. Monticciolo, D.L., et al., *Breast Cancer Screening in Women at Higher-Than-Average Risk: Recommendations From the ACR.* J Am Coll Radiol, 2018. **15**(3 Pt A): p. 408-414.
122. Boyd, N.F., et al., *Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study.* J Natl Cancer Inst, 1995. **87**(9): p. 670-5.
123. Lipson, J., et al. *Frequently Asked Questions About Breast Density, Breast Cancer Risk, and the Breast Density Notification Law in California: A Consensus Document*. 2016 2016].

124. Li, H., et al., *Comparative analysis of image-based phenotypes of mammographic density and parenchymal patterns in distinguishing between BRCA1/2 cases, unilateral cancer cases, and controls.* J Med Imaging (Bellingham), 2014. **1**(3): p. 031009.
125. Torres-Mejia, G., et al., *Mammographic features and subsequent risk of breast cancer: a comparison of qualitative and quantitative evaluations in the Guernsey prospective studies.* Cancer Epidemiol Biomarkers Prev, 2005. **14**(5): p. 1052-9.
126. Daye, D., et al., *Mammographic parenchymal patterns as an imaging marker of endogenous hormonal exposure: a preliminary study in a high-risk population.* Acad Radiol, 2013. **20**(5): p. 635-46.
127. Haberle, L., et al., *Characterizing mammographic images by using generic texture features.* Breast Cancer Res, 2012. **14**(2): p. R59.
128. Byng, J.W., et al., *The quantitative analysis of mammographic densities.* Phys Med Biol, 1994. **39**(10): p. 1629-38.
129. Highnam, R., et al. *Robust Breast Composition Measurement - VolparaTM*. in *IWDM*. 2010.
130. John, B., *QantraTM Volumetric Assessment for Digital Mammography*, in *QUANTRA DS USA Rev3:Layout 1*, Hologic, Editor. 2009, Hologic.
131. Hartman, K., et al., *Volumetric Assessment of Breast Tissue Composition from FFDM Images.* IWDM 2008: Proceedings of the International Workshop on Digital Mammography, 2008. **LNCS 5116**: p. 33-39.
132. Saftlas, A.F. and M. Szklo, *Mammographic parenchymal patterns and breast cancer risk.* Epidemiol Rev, 1987. **9**: p. 146-74.
133. Bland, K.I., et al., *A clinicopathologic correlation of mammographic parenchymal patterns and associated risk factors for human mammary carcinoma.* Ann Surg, 1982. **195**(5): p. 582-94.
134. Goodwin, P.J. and N.F. Boyd, *Mammographic parenchymal pattern and breast cancer risk: a critical appraisal of the evidence.* Am J Epidemiol, 1988. **127**(6): p. 1097-108.
135. Boyd, N.F., et al., *Mammographic patterns and breast cancer risk: methodologic standards and contradictory results.* J Natl Cancer Inst, 1984. **72**(6): p. 1253-9.
136. McCormack, V.A. and I. dos Santos Silva, *Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(6): p. 1159-69.
137. Boyd, N., et al., *A longitudinal study of the effects of menopause on mammographic features.* Cancer Epidemiol Biomarkers Prev, 2002. **11**(10 Pt 1): p. 1048-53.
138. Stomper, P.C., et al., *Analysis of parenchymal density on mammograms in 1353 women 25-79 years old.* AJR Am J Roentgenol, 1996. **167**(5): p. 1261-5.
139. Woolcott, C.G., et al., *Mammographic density change with 1 year of aerobic exercise among postmenopausal women: a randomized controlled trial.* Cancer Epidemiol Biomarkers Prev, 2010. **19**(4): p. 1112-21.
140. Buchanan, J.B., et al., *Selected prognostic variables for mammographic parenchymal patterns.* Cancer, 1981. **47**(9): p. 2135-7.
141. Hart, B.L., et al., *Age and race related changes in mammographic parenchymal patterns.* Cancer, 1989. **63**(12): p. 2537-9.
142. Flook, D., et al., *Changes in Wolfe mammographic patterns with aging.* Br J Radiol, 1987. **60**(713): p. 455-6.
143. Brisson, J., et al., *Mammographic features of the breast and breast cancer risk.* Am J Epidemiol, 1982. **115**(3): p. 428-37.
144. Kelemen, L.E., et al., *Age-specific trends in mammographic density: the Minnesota Breast Cancer Family Study.* Am J Epidemiol, 2008. **167**(9): p. 1027-36.
145. Maskarinec, G., et al., *A longitudinal investigation of mammographic density: the multiethnic cohort.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(4): p. 732-9.

146. McCormack, V.A., et al., *Comparison of a new and existing method of mammographic density measurement: intramethod reliability and associations with known risk factors.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(6): p. 1148-54.
147. McCormack, V.A., et al., *Sex steroids, growth factors and mammographic density: a cross-sectional study of UK postmenopausal Caucasian and Afro-Caribbean women.* Breast Cancer Res, 2009. **11**(3): p. R38.
148. Ghosh, K., et al., *Association between mammographic density and age-related lobular involution of the breast.* J Clin Oncol, 2010. **28**(13): p. 2207-12.
149. Byrne, C., et al., *Mammographic features and breast cancer risk: effects with time, age, and menopause status.* J Natl Cancer Inst, 1995. **87**(21): p. 1622-9.
150. Graham, S.J., et al., *Quantitative correlation of breast tissue parameters using magnetic resonance and X-ray mammography.* Br J Cancer, 1996. **73**(2): p. 162-8.
151. Lee, N.A., et al., *Fatty and fibroglandular tissue volumes in the breasts of women 20-83 years old: comparison of X-ray mammography and computer-assisted MR imaging.* AJR Am J Roentgenol, 1997. **168**(2): p. 501-6.
152. Wei, J., et al., *Correlation between mammographic density and volumetric fibroglandular tissue estimated on breast MR images.* Med Phys, 2004. **31**(4): p. 933-42.
153. Eng-Wong, J., et al., *Effect of raloxifene on mammographic density and breast magnetic resonance imaging in premenopausal women at increased risk for breast cancer.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(7): p. 1696-701.
154. Nie, K., et al., *Development of a quantitative method for analysis of breast density based on three-dimensional breast MRI.* Med Phys, 2008. **35**(12): p. 5253-62.
155. Boyd, N., et al., *Breast-tissue composition and other risk factors for breast cancer in young women: a cross-sectional study.* Lancet Oncol, 2009. **10**(6): p. 569-80.
156. Thompson, D.J., et al., *Assessing the usefulness of a novel MRI-based breast density estimation algorithm in a cohort of women at high genetic risk of breast cancer: the UK MARIBS study.* Breast Cancer Res, 2009. **11**(6): p. R80.
157. Ko, E.S., et al., *Background enhancement in breast MR: correlation with breast density in mammography and background echotexture in ultrasound.* Eur J Radiol, 2011. **80**(3): p. 719-23.
158. Nie, K., et al., *Quantitative analysis of breast parenchymal patterns using 3D fibroglandular tissues segmented based on MRI.* Med Phys, 2010. **37**(1): p. 217-26.
159. Moon, W.K., et al., *Comparative study of density analysis using automated whole breast ultrasound and MRI.* Med Phys, 2011. **38**(1): p. 382-389.
160. Maskarinec, G., et al., *Comparison of breast density measured by dual energy X-ray absorptiometry with mammographic density among adult women in Hawaii.* Cancer Epidemiology, 2011. **35**(2): p. 188-193.
161. Maskarinec, G., et al., *A comparison of breast density measures between mothers and adolescent daughters.* BMC Cancer, 2011. **11**(330).
162. Novotny, R., et al., *Puberty, body fat, and breast density in girls of several ethnic groups.* American Journal of Human Biology, 2011. **23**(3): p. 359-365.
163. Shepherd, J.A., et al., *Breast density assessment in adolescent girls using dual-energy X-ray absorptiometry: a feasibility study.* Cancer Epidemiology, Biomarkers & Prevention, 2008. **17**(7): p. 1709-13.
164. Dorgan, J.F., et al., *Height, adiposity and body fat distribution and breast density in young women.* Breast Cancer Res, 2012. **14**(4): p. R107.
165. Preston, D.L., et al., *Solid Cancer Incidence in Atomic Bomb Survivors: 1958-1998.* Radiat Res, 2007. **168**(1): p. 1-64.
166. Boice, J.D., Jr., et al., *Frequent Chest X-Ray Fluoroscopy and Breast Cancer Incidence among Tuberculosis Patients in Massachusetts.* Radiat Res, 1991. **125**(2): p. 214-222.

167. Ronckers, C.c.M., et al., *Multiple Diagnostic X-rays for Spine Deformities and Risk of Breast Cancer.* Cancer Epidemiology Biomarkers & Prevention, 2008. **17**(3): p. 605-613.
168. Preston, D.L., et al., *Radiation Effects on Breast Cancer Risk: A Pooled Analysis of Eight Cohorts.* Radiat Res, 2002. **158**(2): p. 220-235.
169. Boyd, N.F., et al., *Breast tissue composition and susceptibility to breast cancer.* J Natl Cancer Inst, 2010. **102**(16): p. 1224-37.
170. Denholm, R., et al., *Growth Trajectories, Breast Size, and Breast-Tissue Composition in a British Pre-Birth Cohort of Young Women.* Am J Epidemiol, 2017.
171. Gaskins, A.J., et al., *Dairy intake in relation to breast and pubertal development in Chilean girls.* Am J Clin Nutr, 2017. **105**(5): p. 1166-1175.
172. Jung, S., et al., *Dietary Fat Intake During Adolescence and Breast Density Among Young Women.* Cancer Epidemiol Biomarkers Prev, 2016. **25**(6): p. 918-26.
173. Yaffe, M.J., *Mammographic density. Measurement of mammographic density.* Breast Cancer Res, 2008. **10**(3): p. 209.
174. Boyd, N.F., et al., *The relationship of anthropometric measures to radiological features of the breast in premenopausal women.* Br J Cancer, 1998. **78**(9): p. 1233-8.
175. Sala, E., et al., *High-risk mammographic parenchymal patterns and anthropometric measures: a case-control study.* Br J Cancer, 1999. **81**(7): p. 1257-61.
176. McCormack, V.A., et al., *Life-course body size and perimenopausal mammographic parenchymal patterns in the MRC 1946 British birth cohort.* Br J Cancer, 2003. **89**(5): p. 852-9.
177. Haars, G., et al., *Measurements of breast density: no ratio for a ratio.* Cancer Epidemiol Biomarkers Prev, 2005. **14**(11 Pt 1): p. 2634-40.
178. Boyd, N.F., et al., *Body size, mammographic density, and breast cancer risk.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(11): p. 2086-92.
179. Masala, G., et al., *Dietary and lifestyle determinants of mammographic breast density. A longitudinal study in a Mediterranean population.* Int J Cancer, 2006. **118**(7): p. 1782-9.
180. Guthrie, J.R., et al., *Mammographic densities during the menopausal transition: a longitudinal study of Australian-born women.* Menopause, 2007. **14**(2): p. 208-15.
181. Masala, G., et al., *Physical activity and mammographic breast density in a Mediterranean population: the EPIC Florence longitudinal study.* Int J Cancer, 2009. **124**(7): p. 1654-61.
182. Sung, J., et al., *High-density lipoprotein cholesterol, obesity, and mammographic density in korean women: the healthy twin study.* J Epidemiol, 2011. **21**(1): p. 52-60.
183. Tseng, M. and C. Byrne, *Adiposity, adult weight gain and mammographic breast density in US Chinese women.* Int J Cancer, 2011. **128**(2): p. 418-25.
184. Stone, J., et al., *Using mammographic density to predict breast cancer risk: dense area or percent dense area.* Breast Cancer Research, 2010. **12**(6): p. R97.
185. Reeves, K.W., et al., *Longitudinal association of anthropometry with mammographic breast density in the Study of Women's Health Across the Nation.* Int J Cancer, 2009. **124**(5): p. 1169-77.
186. Lokate, M., et al., *Mammographic density and breast cancer risk: the role of the fat surrounding the fibroglandular tissue.* Breast Cancer Research, 2011. **13**(5): p. R103.
187. Li, T., et al., *The association of measured breast tissue characteristics with mammographic density and other risk factors for breast cancer.* Cancer Epidemiol Biomarkers Prev, 2005. **14**(2): p. 343-9.
188. McCormack, V.A., et al., *Ethnic variations in mammographic density: a British multiethnic longitudinal study.* Am J Epidemiol, 2008. **168**(4): p. 412-21.
189. Lokate, M., et al., *Age-related changes in mammographic density and breast cancer risk.* Am J Epidemiol, 2013. **178**(1): p. 101-9.

190. Heng, D., et al., *Risk factors for breast cancer associated with mammographic features in Singaporean chinese women.* Cancer Epidemiol Biomarkers Prev, 2004. **13**(11 Pt 1): p. 1751-8.
191. Woolcott, C., et al., *Mammographic density, parity and age at first birth, and risk of breast cancer: an analysis of four case–control studies.* Breast Cancer Res Treat, 2012: p. 1-9.
192. Stone, J., et al., *Determinants of percentage and area measures of mammographic density.* Am J Epidemiol, 2009. **170**(12): p. 1571-8.
193. Ramon y Cajal, T., et al., *Mammographic density and breast cancer in women from high risk families.* Breast Cancer Res, 2015. **17**: p. 93.
194. Warwick, J., et al., *Breast density and breast cancer risk factors in a high-risk population.* Breast, 2003. **12**(1): p. 10-6.
195. van Gils, C.H., et al., *Parity and mammographic breast density in relation to breast cancer risk: indication of interaction.* Eur J Cancer Prev, 2000. **9**(2): p. 105-11.
196. Grove, J.S., et al., *Factors associated with mammographic pattern.* Br J Radiol, 1985. **58**(685): p. 21-5.
197. Dite, G.S., et al., *Predictors of mammographic density: insights gained from a novel regression analysis of a twin study.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(12): p. 3474-81.
198. Douglas, J.A., et al., *Mammographic breast density--evidence for genetic correlations with established breast cancer risk factors.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(12): p. 3509-16.
199. Loehberg, C.R., et al., *Assessment of mammographic density before and after first full-term pregnancy.* Eur J Cancer Prev, 2010. **19**(6): p. 405-12.
200. Brisson, J., A.S. Morrison, and N. Khalid, *Mammographic parenchymal features and breast cancer in the breast cancer detection demonstration project.* J Natl Cancer Inst, 1988. **80**(19): p. 1534-40.
201. Boyd, N.F., et al., *Mammographic density and the risk and detection of breast cancer.* N Engl J Med, 2007. **356**(3): p. 227-36.
202. Titus-Ernstoff, L., et al., *Breast cancer risk factors in relation to breast density (United States).* Cancer Causes Control, 2006. **17**(10): p. 1281-90.
203. Vachon, C.M., et al., *Association of mammographically defined percent breast density with epidemiologic risk factors for breast cancer (United States).* Cancer Causes Control, 2000. **11**(7): p. 653-62.
204. Gram, I.T., E. Funkhouser, and L. Tabar, *The Tabar classification of mammographic parenchymal patterns.* Eur J Radiol, 1997. **24**(2): p. 131-6.
205. Lowery, J.T., et al., *Complementary approaches to assessing risk factors for interval breast cancer.* Cancer Causes Control, 2011. **22**(1): p. 23-31.
206. de Stavola, B.L., et al., *Relationship of mammographic parenchymal patterns with breast cancer risk factors and risk of breast cancer in a prospective study.* Int J Epidemiol, 1990. **19**(2): p. 247-54.
207. Frydenberg, H., et al., *Insulin-like growth factor-1, growth hormone, and daily cycling estrogen are associated with mammographic density in premenopausal women.* Cancer Causes Control, 2014. **25**(7): p. 891-903.
208. Louhiala, P., *How tall is too tall? On the ethics of oestrogen treatment for tall girls.* Journal of Medical Ethics, 2007. **33**(1): p. 48-50.
209. Brinks, S., *Estrogen's history as a growth limiter*, in *LA Times*. 2007: Los Angeles.
210. Varghese, J.S., et al., *The heritability of mammographic breast density and circulating sex-hormone levels: two independent breast cancer risk factors.* Cancer Epidemiol Biomarkers Prev, 2012. **21**(12): p. 2167-75.

211. Brown, S.B. and S.E. Hankinson, *Endogenous estrogens and the risk of breast, endometrial, and ovarian cancers.* Steroids, 2015. **99**(Pt A): p. 8-10.
212. Hankinson, S.E. and A.H. Eliassen, *Endogenous estrogen, testosterone and progesterone levels in relation to breast cancer risk.* J Steroid Biochem Mol Biol, 2007. **106**(1-5): p. 24-30.
213. Lee, E., et al., *Hormone metabolism genes and mammographic density in Singapore Chinese women.* Cancer Epidemiol Biomarkers Prev, 2013. **22**(5): p. 984-6.
214. Bremnes, Y., et al., *Endogenous sex hormones, prolactin and mammographic density in postmenopausal Norwegian women.* Int J Cancer, 2007. **121**(11): p. 2506-11.
215. Greendale, G.A., et al., *The association of endogenous sex steroids and sex steroid binding proteins with mammographic density: results from the Postmenopausal Estrogen/Progestin Interventions Mammographic Density Study.* Am J Epidemiol, 2005. **162**(9): p. 826-34.
216. Yong, M., et al., *Associations between endogenous sex hormone levels and mammographic and bone densities in premenopausal women.* Cancer Causes Control, 2009. **20**(7): p. 1039-53.
217. Iversen, A., et al., *Cyclic endogenous estrogen and progesterone vary by mammographic density phenotypes in premenopausal women.* Eur J Cancer Prev, 2016. **25**(1): p. 9-18.
218. Gierach, G.L., et al., *Relationship of serum estrogens and metabolites with area and volume mammographic densities.* Horm Cancer, 2015. **6**(2-3): p. 107-19.
219. Tamimi, R.M., et al., *Endogenous hormone levels, mammographic density, and subsequent risk of breast cancer in postmenopausal women.* J Natl Cancer Inst, 2007. **99**(15): p. 1178-87.
220. Tamimi, R.M., et al., *Endogenous sex hormone levels and mammographic density among postmenopausal women.* Cancer Epidemiol Biomarkers Prev, 2005. **14**(11 Pt 1): p. 2641-7.
221. Warren, R., et al., *Associations among mammographic density, circulating sex hormones, and polymorphisms in sex hormone metabolism genes in postmenopausal women.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(8): p. 1502-8.
222. Simpson, E.R., *Aromatase: biologic relevance of tissue-specific expression.* Semin Reprod Med, 2004. **22**(1): p. 11-23.
223. Labrie, F., et al., *Endocrine and Intracrine Sources of Androgens in Women: Inhibition of Breast Cancer and Other Roles of Androgens and Their Precursor Dehydroepiandrosterone.* Endocr Rev, 2003. **24**(2): p. 152-182.
224. Simpson, E.R., et al., *Estrogen--the Good, the Bad, and the Unexpected.* Endocr Rev, 2005. **26**(3): p. 322-330.
225. Hong, C.C., et al., *Cytochrome P450 1A2 (CYP1A2) activity, mammographic density, and oxidative stress: a cross-sectional study.* Breast Cancer Res, 2004. **6**(4): p. R338-51.
226. Riza, E., et al., *Urinary estrogen metabolites and mammographic parenchymal patterns in postmenopausal women.* Cancer Epidemiol Biomarkers Prev, 2001. **10**(6): p. 627-34.
227. Noh, J.J., et al., *Mammographic densities and circulating hormones: a cross-sectional study in premenopausal women.* Breast, 2006. **15**(1): p. 20-8.
228. Wong, C.C.Y. and E. al., *A longitudinal study of epigenetic variation in twins.* Epigenetics, 2010. **5**(6).
229. Ursin, G., et al., *The relative importance of genetics and environment on mammographic density.* Cancer Epidemiol Biomarkers Prev, 2009. **18**(1): p. 102-12.
230. Stone, J., et al., *Mammographic density and candidate gene variants: a twins and sisters study.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(7): p. 1479-84.
231. Boyd, N.F., et al., *Heritability of mammographic density, a risk factor for breast cancer.* N Engl J Med, 2002. **347**(12): p. 886-94.

232. Kataoka, M., et al., *Genetic models for the familial aggregation of mammographic breast density.* Cancer Epidemiol Biomarkers Prev, 2009. **18**(4): p. 1277-84.
233. Martin, L.J., et al., *Family history, mammographic density, and risk of breast cancer.* Cancer Epidemiol Biomarkers Prev, 2010. **19**(2): p. 456-63.
234. Wolfe, J.N., A.F. Saftlas, and M. Salane, *Mammographic parenchymal patterns and quantitative evaluation of mammographic densities: a case-control study.* AJR Am J Roentgenol, 1987. **148**(6): p. 1087-92.
235. Maskarinec, G. and L. Meng, *A case-control study of mammographic densities in Hawaii.* Breast Cancer Res Treat, 2000. **63**(2): p. 153-61.
236. Maskarinec, G., et al., *Ethnic and geographic differences in mammographic density and their association with breast cancer incidence.* Breast Cancer Res Treat, 2007. **104**(1): p. 47-56.
237. Maskarinec, G., et al., *Mammographic density and breast cancer risk by family history in women of white and Asian ancestry.* Cancer Causes Control, 2015. **26**(4): p. 621-6.
238. del Carmen, M.G., et al., *Mammographic breast density and race.* AJR Am J Roentgenol, 2007. **188**(4): p. 1147-50.
239. Ursin, G., et al., *Mammographic density and breast cancer in three ethnic groups.* Cancer Epidemiol Biomarkers Prev, 2003. **12**(4): p. 332-8.
240. Turnbull, C., *Genome-wide association study identifies five new breast cancer susceptibility loci.* Nature Genetics, 2010.
241. Hunter, D.J. and S.J. Chanock, *Genome-Wide Association Studies and "The Art of the Soluble".* J Natl Cancer Inst, 2010. **102**(12): p. 836-837.
242. Varghese, J.S., et al., *Mammographic Breast Density and Breast Cancer: Evidence of a Shared Genetic Basis.* Cancer Res, 2012. **72**(6): p. 1478-1484.
243. Stone, J., et al., *Novel Associations between Common Breast Cancer Susceptibility Variants and Risk-Predicting Mammographic Density Measures.* Cancer Res, 2015. **75**(12): p. 2457-67.
244. Mitchell, G., et al., *Mammographic density and breast cancer risk in BRCA1 and BRCA2 mutation carriers.* Cancer Res, 2006. **66**(3): p. 1866-72.
245. Passaperuma, K., et al., *Is mammographic breast density a breast cancer risk factor in women with BRCA mutations?* J Clin Oncol, 2010. **28**(23): p. 3779-83.
246. McTiernan, A., et al., *Estrogen-plus-progestin use and mammographic density in postmenopausal women: Women's Health Initiative randomized trial.* J Natl Cancer Inst, 2005. **97**(18): p. 1366-76.
247. Atkinson, C., et al., *Mammographic patterns as a predictive biomarker of breast cancer risk: effect of tamoxifen.* Cancer Epidemiol Biomarkers Prev, 1999. **8**(10): p. 863-6.
248. Brisson, J., et al., *Tamoxifen and mammographic breast densities.* Cancer Epidemiol Biomarkers Prev, 2000. **9**(9): p. 911-5.
249. Chow, C.K., et al., *Effect of tamoxifen on mammographic density.* Cancer Epidemiol Biomarkers Prev, 2000. **9**(9): p. 917-21.
250. Gao, J., *The Effect of Tamoxifen on Mammographic Density in Women with High Risk of Breast Cancer*, in *Surgical Oncology*. 2006, University of Newcastle: Newcastle. p. 224.
251. Spicer, D.V., et al., *Changes in mammographic densities induced by a hormonal contraceptive designed to reduce breast cancer risk.* J Natl Cancer Inst, 1994. **86**(6): p. 431-6.
252. Tehranifar, P., et al., *Reproductive and menstrual factors and mammographic density in African American, Caribbean, and white women.* Cancer Causes & Control, 2011. **22**(4): p. 599-610.
253. Boyd, N.F., et al., *Mammographic density, response to hormones, and breast cancer risk.* Journal of Clinical Oncology, 2011. **29**(22): p. 2985-2992.

254. Banks, E. and K. Canfell, *Invited Commentary: Hormone Therapy Risks and Benefits-"The Women's Health Initiative Findings and the Postmenopausal Estrogen Timing Hypothesis.* Am J Epidemiol, 2009. **170**(1): p. 24-28.
255. The Women's Health Initiative Study Group, *Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group.* Control Clin Trials, 1998. **19**(1): p. 61-109.
256. Bland, K.I., et al., *The effects of exogenous estrogen replacement therapy of the breast: breast cancer risk and mammographic parenchymal patterns.* Cancer, 1980. **45**(12): p. 3027-33.
257. Stomper, P.C., et al., *Mammographic changes associated with postmenopausal hormone replacement therapy: a longitudinal study.* Radiology, 1990. **174**(2): p. 487-90.
258. McNicholas, M.M., et al., *Pain and increased mammographic density in women receiving hormone replacement therapy: a prospective study.* AJR Am J Roentgenol, 1994. **163**(2): p. 311-5.
259. Persson, I., E. Thurfjell, and L. Holmberg, *Effect of estrogen and estrogen-progestin replacement regimens on mammographic breast parenchymal density.* J Clin Oncol, 1997. **15**(10): p. 3201-7.
260. Cyrlak, D. and C.H. Wong, *Mammographic changes in postmenopausal women undergoing hormonal replacement therapy.* AJR Am J Roentgenol, 1993. **161**(6): p. 1177-83.
261. Rutter, C.M., et al., *Changes in breast density associated with initiation, discontinuation, and continuing use of hormone replacement therapy.* Jama, 2001. **285**(2): p. 171-6.
262. Harvey, J.A., et al., *Evaluating hormone therapy-associated increases in breast density comparison between reported and simultaneous assignment of BI-RADS categories, visual assessment, and quantitative analysis.* Acad Radiol, 2005. **12**(7): p. 853-62.
263. Harvey, J., et al., *Hormone replacement therapy and breast density changes.* Climacteric, 2005. **8**(2): p. 185-92.
264. Sterns, E.E. and B. Zee, *Mammographic density changes in perimenopausal and postmenopausal women: is effect of hormone replacement therapy predictable?* Breast Cancer Res Treat, 2000. **59**(2): p. 125-32.
265. Kaufman, Z., et al., *The mammographic parenchymal patterns of women on hormonal replacement therapy.* Clin Radiol, 1991. **43**(6): p. 389-92.
266. Ursin, G., et al., *Post-treatment change in serum estrone predicts mammographic percent density changes in women who received combination estrogen and progestin in the Postmenopausal Estrogen/Progestin Interventions (PEPI) Trial.* J Clin Oncol, 2004. **22**(14): p. 2842-8.
267. Gambrell, R.D. and P.K. Natrajan, *Moderate dosage estrogen-androgen therapy improves continuation rates in postmenopausal women: impact of the WHI reports.* Climacteric, 2006. **9**(3): p. 224-233.
268. Watson, J., L. Wise, and J. Green, *Prescribing of hormone therapy for menopause, tibolone, and bisphosphonates in women in the UK between 1991 and 2005.* European Journal of Clinical Pharmacology, 2007. **63**(9): p. 843-849.
269. Canfell, K., et al., *Decrease in breast cancer incidence following a rapid fall in use of hormone replacement therapy in Australia.* Med J Aust, 2008. **188**(11): p. 641-4.
270. Vankrunkelsven, P., et al., *Reduction in hormone replacement therapy use and declining breast cancer incidence in the Belgian province of Limburg.* Breast Cancer Res Treat, 2009. **118**(2): p. 425-32.

271. Verkooijen, H.M., et al., *The incidence of breast cancer and changes in the use of hormone replacement therapy: a review of the evidence.* Maturitas, 2009. **64**(2): p. 80-5.
272. McTiernan, A., et al., *Conjugated equine estrogen influence on mammographic density in postmenopausal women in a substudy of the women's health initiative randomized trial.* J Clin Oncol, 2009. **27**(36): p. 6135-43.
273. Greendale, G.A., et al., *Effects of estrogen and estrogen-progestin on mammographic parenchymal density. Postmenopausal Estrogen/Progestin Interventions (PEPI) Investigators.* Ann Intern Med, 1999. **130**(4 Pt 1): p. 262-9.
274. Greendale, G.A., et al., *Postmenopausal hormone therapy and change in mammographic density.* J Natl Cancer Inst, 2003. **95**(1): p. 30-7.
275. Vachon, C.M., et al., *Case-control study of increased mammographic breast density response to hormone replacement therapy.* Cancer Epidemiol Biomarkers Prev, 2002. **11**(11): p. 1382-8.
276. Crandall, C.J., et al., *Increases in serum estrone sulfate level are associated with increased mammographic density during menopausal hormone therapy.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(7): p. 1674-81.
277. Nielsen, M., et al., *Breast density changes associated with postmenopausal hormone therapy: post hoc radiologist- and computer-based analyses.* Menopause, 2010. **17**(4): p. 772-8.
278. Lee, E., et al., *Progestogen levels, progesterone receptor gene polymorphisms, and mammographic density changes: results from the Postmenopausal Estrogen/Progestin Interventions Mammographic Density Study.* Menopause, 2012. **19**(3): p. 302-10.
279. Freeman, M.E., et al., *Prolactin: Structure, Function, and Regulation of Secretion.* Physiological Reviews, 2000. **80**(4): p. 1523-1631.
280. Lowry, S.J., et al., *Predictors of breast density change after hormone therapy cessation: results from a randomized trial.* Cancer Epidemiol Biomarkers Prev, 2011. **20**(10): p. 2309-12.
281. Ellingjord-Dale, M., et al., *Polymorphisms in hormone metabolism and growth factor genes and mammographic density in Norwegian postmenopausal hormone therapy users and non-users.* Breast Cancer Res, 2012. **14**(5): p. R135.
282. Hayden, C., *GnRH analogues: applications in assisted reproductive techniques.* Eur J Endocrinol, 2008. **159**(suppl_1): p. S17-25.
283. Jonat, W., *Luteinizing hormone-releasing hormone analogues--the rationale for adjuvant use in premenopausal women with early breast cancer.* Br J Cancer, 1998. **78 Suppl 4**: p. 5-8.
284. Gnant, M., et al., *Endocrine therapy plus zoledronic acid in premenopausal breast cancer.* N Engl J Med, 2009. **360**(7): p. 679-91.
285. Hackshaw, A., et al., *Long-term effectiveness of adjuvant goserelin in premenopausal women with early breast cancer.* J Natl Cancer Inst, 2009. **101**(5): p. 341-9.
286. Goldhirsch, A., et al., *Progress and promise: highlights of the international expert consensus on the primary therapy of early breast cancer 2007.* Annals of Oncology, 2007. **18**(7): p. 1133-1144.
287. Rabaglio, M., S. Aebi, and M. Castiglione-Gertsch, *Controversies of adjuvant endocrine treatment for breast cancer and recommendations of the 2007 St Gallen conference.* Lancet Oncol, 2007. **8**(10): p. 940-9.
288. Regan, M.M., et al., *Premenopausal endocrine-responsive early breast cancer: who receives chemotherapy?* Ann Oncol, 2008. **19**(7): p. 1231-41.
289. Ursin, G., et al., *The detection of changes in mammographic densities.* Cancer Epidemiol Biomarkers Prev, 1998. **7**(1): p. 43-7.

290. Gram, I.T., et al., *Reversal of gonadotropin-releasing hormone agonist induced reductions in mammographic densities on stopping treatment.* Cancer Epidemiol Biomarkers Prev, 2001. **10**(11): p. 1117-20.
291. Weitzel, J.N., et al., *Reduced mammographic density with use of a gonadotropin-releasing hormone agonist-based chemoprevention regimen in BRCA1 carriers.* Clin Cancer Res, 2007. **13**(2 Pt 1): p. 654-8.
292. Cuzick, J., et al., *First results from the International Breast Cancer Intervention Study (IBIS-I): a randomised prevention trial.* Lancet, 2002. **360**(9336): p. 817-24.
293. Fisher, B., et al., *Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study.* J Natl Cancer Inst, 1998. **90**(18): p. 1371-88.
294. Nelson, H.D., et al., *Systematic Review: Comparative Effectiveness of Medications to Reduce Risk for Primary Breast Cancer.* Ann Intern Med, 2009. **151**(10): p. 703-W.235.
295. Lienart, V., et al., *Effect of Preventive Hormonal Therapy on Breast Density: A Systematic Qualitative Review.* ScientificWorldJournal, 2014. **2014**: p. 942386.
296. Cuzick, J., et al., *Tamoxifen and breast density in women at increased risk of breast cancer.* J Natl Cancer Inst, 2004. **96**(8): p. 621-8.
297. Li, J., et al., *Mammographic Density Reduction Is a Prognostic Marker of Response to Adjuvant Tamoxifen Therapy in Postmenopausal Patients With Breast Cancer.* Journal of Clinical Oncology, 2013.
298. Nyante, S.J., et al., *Prognostic Significance of Mammographic Density Change After Initiation of Tamoxifen for ER-Positive Breast Cancer.* Journal of the National Cancer Institute, 2015. **107**(3).
299. Pearman, L., et al., *The effects of raloxifene on mammographic breast density: a review of clinical trials.* Menopause, 2010. **17**(3): p. 654-9.
300. Fabian, C.J., et al., *Clinical Trial of Acolbifene in Premenopausal Women at High Risk for Breast Cancer.* Cancer Prevention Research, 2015. **8**(12): p. 1146-1155.
301. Hind, D., et al., *Hormonal therapies for early breast cancer: systematic review and economic evaluation.* Health Technol Assess, 2007. **11**(26): p. iii-iv, ix-xi, 1-134.
302. Goss, P.E., et al., *Exemestane versus anastrozole in postmenopausal women with early breast cancer: NCIC CTG MA.27--a randomized controlled phase III trial.* J Clin Oncol, 2013. **31**(11): p. 1398-404.
303. Fabian, C.J., et al., *Reduction in proliferation with six months of letrozole in women on hormone replacement therapy.* Breast Cancer Res Treat, 2007. **106**(1): p. 75-84.
304. Mousa, N.A., et al., *Aromatase inhibitors and mammographic breast density in postmenopausal women receiving hormone therapy.* Menopause, 2008. **15**(5): p. 875-84.
305. Liberati, A., et al., *The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration.* PLoS Med, 2009. **6**(7): p. e1000100.
306. Vernon, D., J. Jobling, and J. Toouli, *Surgical versus endoscopic treatment of bile duct stones.* The Cochrane Database of Systematic Reviews, 2001(4).
307. Ekpo, E.U., et al., *Relationship Between Breast Density and Selective Estrogen-Receptor Modulators, Aromatase Inhibitors, Physical Activity, and Diet: A Systematic Review.* Integr Cancer Ther, 2016. **15**(2): p. 127-44.
308. Higgins J.P.T. and Green S. (Editors) *Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 (online)*. 5.1.0 ed. Vol. Version 5.1.0 [updated March 2011]. 2011, http://handbook.cochrane.org.: The Cochrane Collaboration.
309. Cuzick, J., et al., *The effect of anastrozole, tamoxifen or a combination of both drugs on mammographic breast density in post menopausal women with breast cancer, 29th SABACS (2006), Abstract #5033.* Breast Cancer Res Treat, 2006. **100 Suppl 1**.

310. Prowell, T.M., et al., *Changes in breast density and circulating estrogens in postmenopausal women receiving adjuvant anastrozole.* Cancer Prev Res (Phila), 2011. **4**(12): p. 1993-2001.
311. Smith, J., et al., *A pilot study of letrozole for one year in women at enhanced risk of developing breast cancer: effects on mammographic density.* Anticancer Res, 2012. **32**(4): p. 1327-31.
312. Vachon, C.M., et al., *Mammographic Breast Density Response to Aromatase Inhibition.* Clinical Cancer Research, 2013.
313. Cigler, T., et al., *A randomized, placebo-controlled trial (NCIC CTG MAP.2) examining the effects of exemestane on mammographic breast density, bone density, markers of bone metabolism and serum lipid levels in postmenopausal women.* Breast Cancer Research and Treatment, 2011.
314. Alliance for Clinical Trials in Oncology *Changes in Breast Density and Blood Hormone Levels in Postmenopausal Women Receiving Anastrozole or Exemestane for Breast Cancer, NCT00516698*. 2018 [cited 2018 22 April 2019]; Available from: clinicaltrials.gov/ct2/show/study/NCT00516698.
315. Seoul National University Hospital *Breast Density Change Predicting Response to Adjuvant Aromatase Inhibitor (DEAR), NCT01765049*. 2018 [cited 2018 22 April 2018]; ClinicalTrials.gov listing]. Available from: clinicaltrials.gov/ct2/show/NCT01765049.
316. Case Comprehensive Cancer Centre *Changes in Breast Density and Breast Cancer Risk in Women With Breast Cancer and in Healthy Women, NCT00445445*. 2018 [cited 2018 22 April 2018]; Available from: clinicaltrials.gov/ct2/show/NCT00445445.
317. Destounis, S., et al., *Qualitative Versus Quantitative Mammographic Breast Density Assessment: Applications for the US and Abroad.* Diagnostics, 2017. **7**(2): p. 30.
318. Hinton, C.P., et al., *Mammographic parenchymal patterns: value as a predictor of hormone dependency and survival in breast cancer.* AJR Am J Roentgenol, 1985. **144**(6): p. 1103-7.
319. Hinton, C.P., et al., *The relationship of background mammographic pattern to hormone dependency in breast cancer.* Br J Surg, 1984. **71**(5): p. 357-9.
320. Conroy, S.M., et al., *Mammographic density and hormone receptor expression in breast cancer: The Multiethnic Cohort Study.* Cancer Epidemiol, 2011.
321. Antoni, S., et al., *Is mammographic density differentially associated with breast cancer according to receptor status? A meta-analysis.* Breast Cancer Res Treat, 2013. **137**(2): p. 337-347.
322. Arora, N., et al., *Impact of Breast Density on the Presenting Features of Malignancy.* Ann Surg Oncol, 2010. **17**(0): p. 211-218.
323. Ma, H., et al., *Is there a difference in the association between percent mammographic density and subtypes of breast cancer? Luminal A and triple-negative breast cancer.* Cancer Epidemiol Biomarkers Prev, 2009. **18**(2): p. 479-85.
324. Kurian, A., et al., *Lifetime risks of specific breast cancer subtypes among women in four racial/ethnic groups.* Breast Cancer Research, 2010. **12**(6): p. R99.
325. Gierach, G.L., et al., *Relationship Between Mammographic Density and Breast Cancer Death in the Breast Cancer Surveillance Consortium.* J Natl Cancer Inst, 2012. **104**(16): p. 1218-1227.
326. Masarwah, A., et al., *Very low mammographic breast density predicts poorer outcome in patients with invasive breast cancer.* Eur Radiol, 2015. **25**(7): p. 1875-82.
327. Cil, T., et al., *Mammographic density and the risk of breast cancer recurrence after breast-conserving surgery.* Cancer, 2009. **115**(24): p. 5780-7.
328. Maskarinec, G., et al., *Mammographic density as a predictor of breast cancer survival: the Multiethnic Cohort.* Breast Cancer Res, 2013. **15**(1): p. R7.

329. Swann, C.A., et al., *Mammographic density and physical assessment of the breast.* AJR Am J Roentgenol, 1987. **148**(3): p. 525-6.
330. Woolcott, C.G., et al., *Associations of overall and abdominal adiposity with area and volumetric mammographic measures among postmenopausal women.* Int J Cancer, 2010: p. n/a-n/a.
331. Peters, T.M., et al., *Physical activity and mammographic breast density in the EPIC-Norfolk cohort study.* Am J Epidemiol, 2008. **167**(5): p. 579-85.
332. Brisson, J., et al., *Diet, mammographic features of breast tissue, and breast cancer risk.* Am J Epidemiol, 1989. **130**(1): p. 14-24.
333. Boyd, N.F., et al., *Effects at two years of a low-fat, high-carbohydrate diet on radiologic features of the breast: results from a randomized trial. Canadian Diet and Breast Cancer Prevention Study Group.* J Natl Cancer Inst, 1997. **89**(7): p. 488-96.
334. Tamburrini, A.L., et al., *Associations between mammographic density and serum and dietary cholesterol.* Breast Cancer Res Treat, 2011. **125**(1): p. 181-9.
335. Sprague, B.L., et al., *Prevalence of mammographically dense breasts in the United States.* J Natl Cancer Inst, 2014. **106**(10).
336. Ellingjord-Dale, M., et al., *Breast cancer susceptibility variants and mammographic density phenotypes in norwegian postmenopausal women.* Cancer Epidemiol Biomarkers Prev, 2014. **23**(9): p. 1752-63.
337. Trinh, T., et al., *Background risk of breast cancer influences the association between alcohol consumption and mammographic density.* Br J Cancer, 2015. **113**(1): p. 159-65.
338. Trinh, T., et al., *Background risk of breast cancer and the association between physical activity and mammographic density.* Breast Cancer Res, 2015. **17**: p. 50.
339. Assi, V., et al., *A case-control study to assess the impact of mammographic density on breast cancer risk in women aged 40-49 at intermediate familial risk.* Int J Cancer, 2015. **136**(10): p. 2378-87.
340. Rudolph, A., et al., *A comprehensive evaluation of interaction between genetic variants and use of menopausal hormone therapy on mammographic density.* Breast Cancer Res, 2015. **17**: p. 110.
341. Li, J., et al., *Breast cancer genetic risk profile is differentially associated with interval and screen-detected breast cancers.* Ann Oncol, 2015. **26**(3): p. 517-22.
342. Kerlikowske, K., et al., *Breast Cancer Risk by Breast Density, Menopause, and Postmenopausal Hormone Therapy Use.* J Clin Oncol, 2010. **28**(24): p. 3830-3837.
343. Tice, J.A., et al., *Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model.* Ann Intern Med, 2008. **148**(5): p. 337-47.
344. Barlow, W.E., et al., *Prospective breast cancer risk prediction model for women undergoing screening mammography.* J Natl Cancer Inst, 2006. **98**(17): p. 1204-14.
345. Brentnall, A.R., 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 Res, 2015. **17**(1): p. 147.
346. Gail, M.H., et al., *Projecting individualized probabilities of developing breast cancer for white females who are being examined annually.* J Natl Cancer Inst, 1989. **81**(24): p. 1879-86.
347. Claus, E.B., N. Risch, and W.D. Thompson, *Autosomal dominant inheritance of early-onset breast cancer. Implications for risk prediction.* Cancer, 1994. **73**(3): p. 643-51.
348. Tyrer, J., S.W. Duffy, and J. Cuzick, *A breast cancer prediction model incorporating familial and personal risk factors.* Statistics in Medicine, 2004. **23**(7): p. 1111-1130.
349. National Cancer Institute (NCI). *Breast Cancer Risk Assessment Tool*. 2016 [cited 2016; Gail Model 2]. Available from: http://www.cancer.gov/bcrisktool/Default.aspx.

350. Jacobi, C.E., et al., *Differences and similarities in breast cancer risk assessment models in clinical practice: which model to choose?* Breast Cancer Res Treat, 2009. **115**(2): p. 381-90.
351. Chen, J., et al., *Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density.* J Natl Cancer Inst, 2006. **98**(17): p. 1215-26.
352. Mavaddat, N., et al., *Incorporating tumour pathology information into breast cancer risk prediction algorithms.* Breast Cancer Research, 2010. **12**(3): p. 1-12.
353. National Cancer Institute (NCI). *About the (Breast Cancer Risk Assessment, Gail Model 2) Tool*. 2016 [cited 2016; Available from: http://www.cancer.gov/bcrisktool/about-tool.aspx.
354. Tice, J.A., et al., *Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population.* Breast Cancer Res Treat, 2005. **94**(2): p. 115-22.
355. Darabi, H., et al., *Breast cancer risk prediction and individualised screening based on common genetic variation and breast density measurement.* Breast Cancer Res, 2012. **14**(1): p. R25.
356. Warwick, J., et al., *Mammographic breast density refines Tyrer-Cuzick estimates of breast cancer risk in high-risk women: findings from the placebo arm of the International Breast Cancer Intervention Study I.* Breast Cancer Res, 2014. **16**(5): p. 451.
357. Wald, N.J., A.K. Hackshaw, and C.D. Frost, *When can a risk factor be used as a worthwhile screening test?* BMJ, 1999. **319**(7224): p. 1562-1565.
358. Cheddad, A., et al., *Enhancement of mammographic density measures in breast cancer risk prediction.* Cancer Epidemiol Biomarkers Prev, 2014. **23**(7): p. 1314-23.
359. Gierach, G.L., et al., *Relationships between computer-extracted mammographic texture pattern features and BRCA1/2 mutation status: a cross-sectional study.* Breast Cancer Res, 2014. **16**(4): p. 424.
360. Caldwell, C.B., et al., *Characterisation of mammographic parenchymal pattern by fractal dimension.* Phys Med Biol, 1990. **35**(2): p. 235-47.
361. Gierach, G.L., et al., *Emerging Concepts in Breast Cancer Risk Prediction.* Curr Obstet Gynecol Rep, 2013. **2**(1): p. 43-52.
362. Lindstrom, S., et al., *Common variants in ZNF365 are associated with both mammographic density and breast cancer risk.* Nature Genetics, 2011. **43**(3): p. 185-7.
363. Vachon, C.M., et al., *Common Breast Cancer Susceptibility Variants in LSP1 and RAD51L1 Are Associated with Mammographic Density Measures that Predict Breast Cancer Risk.* Cancer Epidemiology Biomarkers & Prevention, 2012.
364. Hopper, J.L., *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): p. 863-867.
365. Fasching, P.A., et al., *Breast Cancer Risk - From Genetics to Molecular Understanding of Pathogenesis.* Geburtshilfe Frauenheilkd, 2013. **73**(12): p. 1228-1235.
366. Byrne, C. and S. Spernak, *What is breast density?* Breast Cancer Online, 2005. **8**(10): p. null-null.
367. Cuzick, J., et al., *Long-term results of tamoxifen prophylaxis for breast cancer--96-month follow-up of the randomized IBIS-I trial.* J Natl Cancer Inst, 2007. **99**(4): p. 272-82.
368. Forbes, J.F., *Tokyo conference presentation*. 2010: Tokyo.
369. Howell, A., et al., *Results of the ATAC (Arimidex, Tamoxifen, Alone or in Combination) trial after completion of 5 years' adjuvant treatment for breast cancer.* Lancet, 2005. **365**(9453): p. 60-2.

370. Mitchell, M.N., *Interpreting and Visualizing Regression Models Using Stata*. 2012, College Station, Texas USA: Stata Press.
371. Snijders, T.A.B. and R.J. Bosker, *Multilevel Analysis. An introduction to basic and advanced multilevel modelling*. 2000, London UK: SAGE Publications Ltd.
372. Chen, J.H., G. Gulsen, and M.Y. Su, *Imaging Breast Density: Established and Emerging Modalities.* Transl Oncol, 2015. **8**(6): p. 435-45.
373. Falcon, S., et al., *Imaging Management of Breast Density, a Controversial Risk Factor for Breast Cancer.* Cancer Control, 2017. **24**(2): p. 125-136.
374. Wolfe, J., *A study of breast parenchyma by mammography in the normal woman and those with benign and malignant disease.* Radiology, 1967. **89**: p. 201-5.
375. Wolfe, J.N., *The prominent duct pattern as an indicator of cancer risk.* Oncology, 1969. **23**(2): p. 149-58.
376. Garrido Estepa, M., et al., *Evaluation of mammographic density patterns: reproducibility and concordance among scales.* BMC Cancer, 2010. **10**(1): p. 485.
377. Gram, I.T., et al., *Percentage density, Wolfe's and Tabar's mammographic patterns: agreement and association with risk factors for breast cancer.* Breast Cancer Res, 2005. **7**(5): p. R854-61.
378. Boehm, H.F., et al., *Automated classification of breast parenchymal density: topologic analysis of x-ray attenuation patterns depicted with digital mammography.* AJR Am J Roentgenol, 2008. **191**(6): p. W275-82.
379. Mudigonda, N.R., R.M. Rangayyan, and J.E. Desautels, *Detection of breast masses in mammograms by density slicing and texture flow-field analysis.* IEEE Trans Med Imaging, 2001. **20**(12): p. 1215-27.
380. Li, H., et al., *Fractal analysis of mammographic parenchymal patterns in breast cancer risk assessment.* Acad Radiol, 2007. **14**(5): p. 513-21.
381. Nielsen, M., et al., *Mammographic texture resemblance generalizes as an independent risk factor for breast cancer.* Breast Cancer Res, 2014. **16**(2): p. R37.
382. Kontos, D., et al., *Analysis of Parenchymal Texture with Digital Breast Tomosynthesis: Comparison with Digital Mammography and Implications for Cancer Risk Assessment.* Radiology, 2011. **261**(1): p. 80-91.
383. Gastounioti, A., E.F. Conant, and D. Kontos, *Beyond breast density: a review on the advancing role of parenchymal texture analysis in breast cancer risk assessment.* Breast Cancer Res, 2016. **18**(1): p. 91.
384. Boyd, N.F., et al., *Relationship between mammographic and histological risk factors for breast cancer.* J Natl Cancer Inst, 1992. **84**(15): p. 1170-9.
385. American College of Radiology, *Breast imaging reporting and data system (BI-RADS), 4th Edition*. 2003, Reston, VA: American College of Radiology.
386. Burnside, E.S., et al., *The ACR BI-RADS(®) Experience: Learning From History.* Journal of the American College of Radiology : JACR, 2009. **6**(12): p. 851-860.
387. Kerlikowske, K., et al., *Longitudinal measurement of clinical mammographic breast density to improve estimation of breast cancer risk.* J Natl Cancer Inst, 2007. **99**(5): p. 386-95.
388. Kerlikowske, K., et al., *One versus Two Breast Density Measures to Predict 5- and 10-Year Breast Cancer Risk.* Cancer Epidemiol Biomarkers Prev, 2015. **24**(6): p. 889-97.
389. Warren, R., *Hormones and mammographic breast density.* Maturitas, 2004. **49**(1): p. 67-78.
390. Byng, J.W. and et al., *The quantitative analysis of mammographic densities.* Phys Med Biol, 1994. **39**(10): p. 1629.
391. Jeffreys, M., et al., *Breast density: agreement of measures from film and digital image.* Br J Radiol, 2003. **76**(908): p. 561-3.

392. Highnam, R., et al., *Comparing measurements of breast density.* Phys Med Biol, 2007. **52**(19): p. 5881-95.
393. Byng, J.W., et al., *Automated analysis of mammographic densities and breast carcinoma risk.* Cancer, 1997. **80**(1): p. 66-74.
394. Gao, J., et al., *Reproducibility of visual assessment on mammographic density.* Breast Cancer Res Treat, 2008. **108**(1): p. 121-7.
395. Heine, J.J., et al., *An automated approach for estimation of breast density.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(11): p. 3090-7.
396. Verheus, M., et al., *IGF1, IGFBP1, and IGFBP3 genes and mammographic density: the Multiethnic Cohort.* Int J Cancer, 2010. **127**(5): p. 1115-23.
397. Maskarinec, G., et al., *Comparison of breast density measured by dual energy X-ray absorptiometry with mammographic density among adult women in Hawaii.* Cancer Epidemiol, 2010.
398. Stone, J., *Personal communication during Cumulus training at the University of Melbourne*. 2011.
399. Stone, J., et al., *The detection of change in mammographic density.* Cancer Epidemiol Biomarkers Prev, 2003. **12**(7): p. 625-30.
400. Sohn, G., et al., *Reliability of the percent density in digital mammography with a semi-automated thresholding method.* J Breast Cancer, 2014. **17**(2): p. 174-9.
401. Habel, L.A., et al., *Case-control study of mammographic density and breast cancer risk using processed digital mammograms.* Breast Cancer Res, 2016. **18**(1): p. 53.
402. Lillie, E.O., et al., *Polymorphism in the androgen receptor and mammographic density in women taking and not taking estrogen and progestin therapy.* Cancer Res, 2004. **64**(4): p. 1237-41.
403. Lord, S.J., et al., *Polymorphisms in genes involved in estrogen and progesterone metabolism and mammographic density changes in women randomized to postmenopausal hormone therapy: results from a pilot study.* Breast Cancer Res, 2005. **7**(3): p. R336-44.
404. Bremnes, Y., et al., *Different types of postmenopausal hormone therapy and mammographic density in Norwegian women.* Int J Cancer, 2007. **120**(4): p. 880-4.
405. Stuedal, A., et al., *Does breast size modify the association between mammographic density and breast cancer risk?* Cancer Epidemiol Biomarkers Prev, 2008. **17**(3): p. 621-7.
406. Byng, J.W., et al., *Automated analysis of mammographic densities.* Phys Med Biol, 1996. **41**(5): p. 909-23.
407. Martin, K.E., et al., *Mammographic density measured with quantitative computer-aided method: comparison with radiologists' estimates and BI-RADS categories.* Radiology, 2006. **240**(3): p. 656-65.
408. Tagliafico, A., et al., *Mammographic density estimation: comparison among BI-RADS categories, a semi-automated software and a fully automated one.* Breast, 2009. **18**(1): p. 35-40.
409. Zhou, C., et al., *Computerized image analysis: estimation of breast density on mammograms.* Med Phys, 2001. **28**(6): p. 1056-69.
410. Li, J., et al., *High-throughput mammographic-density measurement: a tool for risk prediction of breast cancer.* Breast Cancer Res, 2012. **14**(4): p. R114.
411. Nickson, C., et al., *AutoDensity: an automated method to measure mammographic breast density that predicts breast cancer risk and screening outcomes.* Breast Cancer Res, 2013. **15**(5): p. R80.
412. Keller, B.M., et al., *Breast density and parenchymal texture measures as potential risk factors for Estrogen-Receptor positive breast cancer.* Proc SPIE, 2014. **9035**: p. 90351D.

413. Keller, B.M., et al., *Preliminary evaluation of the publicly available Laboratory for Breast Radiodensity Assessment (LIBRA) software tool: comparison of fully automated area and volumetric density measures in a case–control study with digital mammography.* Breast Cancer Research, 2015. **17**(1): p. 117.
414. Eriksson, M., et al., *A clinical model for identifying the short-term risk of breast cancer.* Breast Cancer Res, 2017. **19**(1): p. 29.
415. Eriksson, M., et al., *A comprehensive tool for measuring mammographic density changes over time.* Breast Cancer Res Treat, 2018.
416. Kopans, D.B., *Basic physics and doubts about relationship between mammographically determined tissue density and breast cancer risk.* Radiology, 2008. **246**(2): p. 348-53.
417. Matakina Solutions. *Volpara Density*. 2017 [cited 2017 10 February 2017]; Available from: http://volparasolutions.com/our-products/volparadensity/.
418. Hologic. *Image Analytics*. 2017 [cited 2017 10 February 2017]; Available from: http://www.hologic.com/products/imaging/mammography/image-analytics.
419. Highnam, R. and M. Brady, *Mammographic Image Analysis*. 1999: Kluwer Academic Publishers, Dordrecht, The Netherlands.
420. Jeffreys, M., et al., *Initial experiences of using an automated volumetric measure of breast density: the standard mammogram form.* Br J Radiol, 2006. **79**(941): p. 378-82.
421. Ding, J., et al., *Evaluating the effectiveness of using standard mammogram form to predict breast cancer risk: case-control study.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(5): p. 1074-81.
422. Aitken, Z., 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): p. 418-28.
423. Boyd, N., et al., *Mammographic density and breast cancer risk: evaluation of a novel method of measuring breast tissue volumes.* Cancer Epidemiol Biomarkers Prev, 2009. **18**(6): p. 1754-62.
424. Chen, J.H., et al., *Reduction of breast density following tamoxifen treatment evaluated by 3-D MRI: preliminary study.* Magn Reson Imaging, 2011. **29**(1): p. 91-8.
425. Ko, E.S., et al., *Background enhancement in breast MR: Correlation with breast density in mammography and background echotexture in ultrasound.* Eur J Radiol, 2010.
426. Moore, S., et al., *Cost-effectiveness of MRI compared to mammography for breast cancer screening in a high risk population.* BMC Health Serv Res, 2009. **9**(1): p. 9.
427. Brooks, J.D., et al., *MRI background parenchymal enhancement, breast density and serum hormones in postmenopausal women.* Int J Cancer, 2018.
428. Hennessey, S., et al., *Bilateral symmetry of breast tissue composition by magnetic resonance in young women and adults.* Cancer Causes Control, 2014.
429. O'Flynn, E.A.M., et al., *Ultrasound Tomography Evaluation of Breast Density: A Comparison With Noncontrast Magnetic Resonance Imaging.* Invest Radiol, 2017. **52**(6): p. 343-348.
430. Klifa, C., et al., *Magnetic resonance imaging for secondary assessment of breast density in a high-risk cohort.* Magn Reson Imaging, 2010. **28**(1): p. 8-15.
431. Kelly, K.M., et al., *Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts.* Eur Radiol, 2010. **20**(3): p. 734-42.
432. Chen, J.H., et al., *Breast density analysis for whole breast ultrasound images.* Med Phys, 2009. **36**(11): p. 4933-43.
433. Zanotel, M., et al., *Automated breast ultrasound: basic principles and emerging clinical applications.* Radiol Med, 2018. **123**(1): p. 1-12.

434. Glide, C., N. Duric, and P. Littrup, *Novel approach to evaluating breast density utilizing ultrasound tomography.* Med Phys, 2007. **34**(2): p. 744-53.
435. Duric, N., et al., *Breast density measurements with ultrasound tomography: a comparison with film and digital mammography.* Med Phys, 2013. **40**(1): p. 013501.
436. Bakic, P.R., et al., *Breast percent density: estimation on digital mammograms and central tomosynthesis projections.* Radiology, 2009. **252**(1): p. 40-9.
437. Skaane, P., et al., *Comparison of Digital Mammography Alone and Digital Mammography Plus Tomosynthesis in a Population-based Screening Program.* Radiology, 2013.
438. Rakowski, J.T. and M.J. Dennis, *A comparison of reconstruction algorithms for C-arm mammography tomosynthesis.* Med Phys, 2006. **33**(8): p. 3018-3032.
439. Tagliafico, A., et al., *Comparative estimation of percentage breast tissue density for digital mammography, digital breast tomosynthesis, and magnetic resonance imaging.* Breast Cancer Res Treat, 2013. **138**(1): p. 311-7.
440. Tagliafico, A., et al., *Mammographic density estimation: one-to-one comparison of digital mammography and digital breast tomosynthesis using fully automated software.* Eur Radiol, 2012: p. 1-6.
441. Vranjesevic, D., et al., *Relationship between 18F-FDG uptake and breast density in women with normal breast tissue.* J Nucl Med, 2003. **44**(8): p. 1238-42.
442. Kuroki-Suzuki, S., et al., *Diagnosis of breast cancer with multidetector computed tomography: analysis of optimal delay time after contrast media injection.* Clin Imaging, 2010. **34**(1): p. 14-9.
443. Shepherd, J.A., et al., *Clinical comparison of a novel breast DXA technique to mammographic density.* Med Phys, 2006. **33**(5): p. 1490-1498.
444. Wang, J., et al., *Agreement of mammographic measures of volumetric breast density to MRI.* PLoS One, 2013. **8**(12): p. e81653.
445. Shepherd, J.A., et al., *Novel use of single X-Ray absorptiometry for measuring breast density.* Technology in Cancer Research and Treatment, 2005. **4**(2): p. 173-182.
446. O'Connor, M.K., et al., *Comparison of radiation exposure and associated radiation-induced cancer risks from mammography and molecular imaging of the breast.* Med Phys, 2010. **37**(12): p. 6187-6198.
447. Shermis, R.B., et al., *Molecular Breast Imaging in Breast Cancer Screening and Problem Solving.* Radiographics, 2017. **37**(5): p. 1309-1606.
448. O'Sullivan, T., et al., *Optical imaging correlates with magnetic resonance imaging breast density and reveals composition changes during neoadjuvant chemotherapy.* Breast Cancer Research, 2013. **15**(1): p. R14.
449. Ruiz, J., et al., *Breast density quantification using structured-light-based diffuse optical tomography simulations.* Appl Opt, 2017. **56**(25): p. 7146-7157.
450. Taroni, P., et al., *Seven-wavelength time-resolved optical mammography extending beyond 1000 nm for breast collagen quantification.* Opt Express, 2009. **17**(18): p. 15932-46.
451. Jud, S.M., et al., *Correlates of mammographic density in B-mode ultrasound and real time elastography.* Eur J Cancer Prev, 2011.
452. Hawley, J.R., et al., *Quantification of breast stiffness using MR elastography at 3 Tesla with a soft sternal driver: A reproducibility study.* J Magn Reson Imaging, 2017. **45**(5): p. 1379-1384.
453. Maskarinec, G., et al., *Bioimpedence to Assess Breast Density as a Risk Factor for Breast Cancer in Adult Women and Adolescent Girls.* Asian Pac J Cancer Prev, 2016. **17**(1): p. 65-71.

454. Greene, T. and E. al., ***Sub-epithelial impedance: a new non-invasive method to measure mammographic density.*** Cancer Research, 2009. **69**(2 Supplement): p. Abstract 4075.
455. Gold, R.H., L.W. Bassett, and B.E. Widoff, *Highlights from the history of mammography.* RadioGraphics, 1990. **10**(6): p. 1111-1131.
456. Van Steen, A. and R. Van Tiggelen, *Short history of mammography: a Belgian perspective.* JBR BTR, 2007. **90**(3): p. 151.
457. Hendrick, R.E., et al., *Comparison of Acquisition Parameters and Breast Dose in Digital Mammography and Screen-Film Mammography in the American College of Radiology Imaging Network Digital Mammographic Imaging Screening Trial.* AJR. American journal of roentgenology, 2010. **194**(2): p. 362-369.
458. Dershaw, D.D., *Status of mammography after the Digital Mammography Imaging Screening Trial: digital versus film.* Breast J, 2006. **12**(2): p. 99-102.
459. Pisano, E.D., et al., *Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST.* Radiology, 2008. **246**(2): p. 376-83.
460. Ooms, E.A., et al., *Mammography: interobserver variability in breast density assessment.* Breast, 2007. **16**(6): p. 568-76.
461. Chang, R.F., et al., *Three comparative approaches for breast density estimation in digital and screen film mammograms.* Conf Proc IEEE Eng Med Biol Soc, 2006. **1**: p. 4853-6.
462. Bland, J.M. and D.G. Altman, *Statistical methods for assessing agreement between two methods of clinical measurement.* Lancet, 1986. **1**(8476): p. 307-10.
463. Van Noorden, R., B. Maher, and R. Nuzzo, *The top 100 papers.* Nature, 2014. **514**: p. 550-553.
464. Bland, J.M. and D.G. Altman, *A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement.* Comput Biol Med, 1990. **20**(5): p. 337-40.
465. Lodder, M.C., et al., *Reproducibility of bone mineral density measurement in daily practice.* Ann Rheum Dis, 2004. **63**(3): p. 285-9.
466. Mello-Thoms, C., et al., *Understanding the Role of Correct Lesion Assessment in Radiologists' Reporting of Breast Cancer*, in *Breast Imaging*, H. Fujita, T. Hara, and C. Muramatsu, Editors. 2014, Springer International Publishing. p. 341-347.
467. Ciatto, S., et al., *Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories.* Breast, 2005. **14**(4): p. 269-75.
468. Stone, J., et al., *Predicting breast cancer risk using mammographic density measurements from both mammogram sides and views.* Breast Cancer Res Treat, 2010. **124**(2): p. 551-4.
469. Pettersson, A., et al., *Mammographic Density Phenotypes and Risk of Breast Cancer: A Meta-analysis.* Journal of the National Cancer Institute, 2014. **106**(5).
470. Cuzick, J., *A Wilcoxon-type test for trend. .* Statistics in Medicine, 1985. **4**: p. 87-90.
471. Modugno, F., et al., *Breast cancer risk factors and mammographic breast density in women over age 70.* Breast Cancer Res Treat, 2006. **97**(2): p. 157-66.
472. Gold, E.B., et al., *Factors Associated with Age at Natural Menopause in a Multiethnic Sample of Midlife Women.* American Journal of Epidemiology, 2001. **153**(9): p. 865-874.
473. Schoenaker, D.A., et al., *Socioeconomic position, lifestyle factors and age at natural menopause: a systematic review and meta-analyses of studies across six continents.* International Journal of Epidemiology, 2014. **43**(5): p. 1542-1562.
474. Do, K.A., et al., *Predictive factors of age at menopause in a large Australian twin study.* Hum Biol, 1998. **70**(6): p. 1073-91.

475. Moorman, P.G., et al., *Effect of Hysterectomy With Ovarian Preservation on Ovarian Function.* Obstetrics & Gynecology, 2011. **118**(6): p. 1271-1279.
476. Lewis-Beck, M.S., *Multiple Regression*, in *Applied Regression*. 1980, SAGE researchmethods. p. 48-74.
477. Zhang, D., et al., *Assessing model fit in joint models of longitudinal and survival data with applications to cancer clinical trials.* Statistics in Medicine, 2014. **33**(27): p. 4715-4733.
478. Jones, R.H., *Bayesian information criterion for longitudinal and clustered data.* Statistics in Medicine, 2011. **30**(25): p. 3050-3056.
479. Maskarinec, G., L. Meng, and G. Ursin, *Ethnic differences in mammographic densities.* Int J Epidemiol, 2001. **30**(5): p. 959-65.
480. Boone, J.M., et al., *Scatter/primary in mammography: comprehensive results.* Med Phys, 2000. **27**(10): p. 2408-16.
481. Hedegard, W. *Mammoguide - Mammography Techniques*. 2017 [cited 2018 23 April 2018]; Available from: https://www.mammoguide.com/mammo-techniques.html.
482. Prentice, R.L., et al., *Estrogen plus progestin therapy and breast cancer in recently postmenopausal women.* Am J Epidemiol, 2008. **167**(10): p. 1207-16.
483. Butler, L.M., et al., *Active, but not passive cigarette smoking was inversely associated with mammographic density.* Cancer Causes Control, 2010. **21**(2): p. 301-11.
484. Gaudet, M.M., et al., *Pooled analysis of active cigarette smoking and invasive breast cancer risk in 14 cohort studies.* Int J Epidemiol, 2016.
485. Macacu, A., et al., *Active and passive smoking and risk of breast cancer: a meta-analysis.* Breast Cancer Research and Treatment, 2015. **154**(2): p. 213-224.
486. Astley, S.M., et al., *A comparison of five methods of measuring mammographic density: a case-control study.* Breast Cancer Res, 2018. **20**(1): p. 10.
487. McCormack, V.A., et al., *Changes and tracking of mammographic density in relation to Pike's model of breast tissue aging: a UK longitudinal study.* Int J Cancer, 2010. **127**(2): p. 452-61.
488. Verheus, M., et al., *Premenopausal insulin-like growth factor-I serum levels and changes in breast density over menopause.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(3): p. 451-7.
489. Gelman, A. and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*. 2006, New York: Cambridge University Press.
490. Singer, J.D. and J.B. Willett, *Applied Longitudinal Data Analysis : Modelling change and event occurrence*. 2003, New York: Oxford University Press.
491. Rabe-Hesketh, S. and A. Skrondal, *Multilevel and Longitudinal Modeling Using Stata*. Vol. Volume I: Continuous Responses. 2012, College Station, Texas USA: Stata Press.
492. Diggle, P.J., K.-Y. Liang, and S.L. Zeger, *Analysis of Longitudinal Data*. 1994, New York, NY USA: Oxford University Press.
493. Raftery, A.E., *Bayesian Model Selection in Social Research.* Sociological Methodology, 1995. **25**: p. 111-163.
494. Chavance, M. and S. Escolano, *Misspecification of the covariance structure in generalized linear mixed models.* Statistical Methods in Medical Research, 2016. **25**(2): p. 630-643.
495. Boyd, N.F., et al., *Mammographic density as a surrogate marker for the effects of hormone therapy on risk of breast cancer.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(5): p. 961-6.
496. Chlebowski, R.T., et al., *Influence of estrogen plus progestin on breast cancer and mammography in healthy postmenopausal women: the Women's Health Initiative Randomized Trial.* JAMA, 2003. **289**(24): p. 3243-53.

497. Anderson, G.L., et al., *Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial.* JAMA, 2004. **291**(14): p. 1701-12.
498. Bertrand, K.A., et al., *Dense and nondense mammographic area and risk of breast cancer by age and tumor characteristics.* Cancer Epidemiol Biomarkers Prev, 2015. **24**(5): p. 798-809.
499. Maskarinec, G., et al., *Mammographic density and breast cancer risk: the multiethnic cohort study.* Am J Epidemiol, 2005. **162**(8): p. 743-52.
500. Yaghjyan, L., et al., *Mammographic Breast Density and Subsequent Risk of Breast Cancer in Postmenopausal Women according to the Time Since the Mammogram.* Cancer Epidemiology Biomarkers & Prevention, 2013.
501. Shepherd, J. and K. Kerlikowske, *Do fatty breasts increase or decrease breast cancer risk?* Breast Cancer Research, 2012. **14**(1): p. 102.
502. Alexopoulos, E.C., *Introduction to Multivariate Regression Analysis.* Hippokratia, 2010. **14**(Suppl 1): p. 23-28.
503. Robinson, G.K., *That BLUP is a Good Thing: The Estimation of Random Effects.* Statistical Science, 1991. **6**(1): p. 15-32.
504. Vachon, C.M., et al., *Comparison of percent density from raw and processed full-field digital mammography data.* Breast Cancer Res, 2013. **15**(1): p. R1.
505. Liang, K.-Y. and S.L. Zeger, *Longitudinal data analysis using generalized linear models.* Biometrika, 1986. **73**(1): p. 13-22.
506. Fitzmaurice, G. and G. Molenberghs, *Advances in longitudinal data analysis: An historical perspective*, in *Longitudinal Data Analysis*, G. Fitzmaurice, et al., Editors. 2008, CRC Press, Taylor & Francis Group: London.
507. Penn State University Online Courses. *Lesson 12: Advanced Topics I - Generalized Estimating Equations (GEE)*. 2017 [cited 2017 23Jan2017]; Available from: https://onlinecourses.science.psu.edu/stat504/book/export/html/179.
508. Harvey, J.A., et al., *Reported Mammographic Density: Film-Screen versus Digital Acquisition.* Radiology, 2012.
509. Chlebowski, R.T., et al., *Estrogen plus progestin and breast cancer detection by means of mammography and breast biopsy.* Arch Intern Med, 2008. **168**(4): p. 370-7; quiz 345.
510. Chlebowski, R.T., et al., *Breast cancer after use of estrogen plus progestin in postmenopausal women.* N Engl J Med, 2009. **360**(6): p. 573-87.
511. Pollan, M., et al., *Validation of DM-Scan, a computer-assisted tool to assess mammographic density in full-field digital mammograms.* Springerplus, 2013. **2**(1): p. 242.
512. Cheddad, A., et al., *Area and volumetric density estimation in processed full-field digital mammograms for risk assessment of breast cancer.* PLoS One, 2014. **9**(10): p. e110690.
513. Busana, M.C., et al., *Impact of type of full-field digital image on mammographic density assessment and breast cancer risk estimation: a case-control study.* Breast Cancer Research, 2016. **18**(1): p. 1-12.
514. Gabrielson, M., et al., *Cohort Profile: The Karolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA).* International Journal of Epidemiology, 2017. **46**(6): p. 1740-1741g.
515. NSW Government, *Multicultural NSW*. 2016, NSW Government: http://multicultural.nsw.gov.au/.
516. Kerlikowske, K., et al., *Comparative effectiveness of digital versus film-screen mammography in community practice in the United States: a cohort study.* Ann Intern Med, 2011. **155**(8): p. 493-502.

517. Perry, N., et al., *The impact of digital mammography on screening a young cohort of women for breast cancer in an urban specialist breast unit.* Eur Radiol, 2011. **21**(4): p. 676-682.
518. Shawky, M.S., et al., *Mammographic density: a potential monitoring biomarker for adjuvant and preventative breast cancer endocrine therapies*. 2016. 2016.
519. Sprague, B.L., et al., *Variation in Mammographic Breast Density Assessments Among Radiologists in Clinical Practice: A Multicenter Observational Study.* Ann Intern Med, 2016. **165**(7): p. 457-464.
520. Richman, I., et al., *Breast Density Notification Legislation and Breast Cancer Stage at Diagnosis: Early Evidence from the SEER Registry.* J Gen Intern Med, 2016.
521. Dean-Ben, X.L., et al., *Volumetric hand-held optoacoustic angiography as a tool for real-time screening of dense breast.* J Biophotonics, 2016. **9**(3): p. 253-9.
522. Taroni, P., et al., *Breast tissue composition and its dependence on demographic risk factors for breast cancer: non-invasive assessment by time domain diffuse optical spectroscopy.* PLoS One, 2015. **10**(6): p. e0128941.
523. Bidgood, W.D., et al., *Understanding and Using DICOM, the Data Interchange Standard for Biomedical Imaging.* Journal of the American Medical Informatics Association, 1997. **4**(3): p. 199-212.
524. Escott, E.J. and D. Rubinstein, *Free DICOM Image Viewing and Processing Software for Your Desktop Computer: What's Available and What It Can Do for You1.* Radiographics, 2003. **23**(5): p. 1341-1357.
525. Gentili, A., C.B. Chung, and T. Hughes, *Use of the MIRC DICOM Service for Clinical Trials to Automatically Create Teaching File Cases from PACS1.* Radiographics, 2007. **27**(1): p. 269-275.
526. Varma, D.R., *Free DICOM browsers.* Indian Journal of Radiology and Imaging, 2008. **18**(1): p. 12-16.
527. Rodríguez González, D., et al., *An open source toolkit for medical imaging de-identification.* Eur Radiol, 2010. **20**(8): p. 1896-1904.
528. McDowall, R.D., *Software out of the box? When an application is referred to as commercial off the shelf (COTS) software, what does this really mean? Do you appreciate what you are buying and can you manage the risks involved? Here, we look at how the term COTS can be used and abused.* Spectroscopy, 2012(11): p. 28.
529. Julian, C., T. Lucy, and J. Farr, *Commercial-Off-The-Shelf Selection Process.* Engineering Management Journal, 2011. **23**(2): p. 63-71.
530. Avery, P., *Calculating Life-Cycle Cost.* Engineered Systems, 2011. **28**(9): p. 44-48.
531. Puech, P.A., et al., *DicomWorks: software for reviewing DICOM studies and promoting low-cost teleradiology.* J Digit Imaging, 2007. **20**(2): p. 122-30.
532. Branstetter, B.F.t., et al., *SimpleDICOM suite: personal productivity tools for managing DICOM objects.* Radiographics, 2007. **27**(5): p. 1523-30.
533. Ruby, T.S., et al., *Is Breast Size a Predictor of Breast Cancer Risk or the Laterality of the Tumor?* Cancer Causes & Control, 1993. **4**(3): p. 203-208.

10. Appendices

A. Deidentification Software and Mammogram Deidentification Method

A.1 Selection of mammogram digitisation and de-identification process

A.1.1 Technical considerations

Visually assessed PD is suitable for any type of mammograms (e.g. film-screen, digital), however use of computer software requires the mammograms to be in a digital (electronic) format. In particular, x-ray films (film-screen mammograms) needed to be converted into a digital format before they could be used electronically. This required access to an x-ray scanner.

All mammograms for the project also required removal of directly identifying information (e.g. participant name, address) prior to density assessment. Unlike digital mammograms, film-screen mammograms contain areas of burned-in identifying information. Whilst parts of images can be masked from within some MD measurement programs, the mammograms are not otherwise de-identified. A quick, easy to use method for de-identifying digitised film-screen mammograms was required, especially for use with visual methods of density assessment.

A.1.2 Scanner selection

During the review of MD measurement techniques it was noted that some programs required film-screen mammograms to be scanned on a digitiser with an optical density of at least 4.0. Higher scanning optical densities provide better contrast within dark areas of the x-ray film — i.e. the scanner is able to capture darker tones (deeper blacks). This is important for accurately capturing the breast edge (typically skin adjacent to fat) and hence important for measuring total breast area on a mammogram.

The lowest cost digitisation option— photographing mammograms with a digital camera [249]— was unlikely to produce the high quality images needed for use with the MD

measurement programs. The existing CMN Department of Surgical Oncology x-ray scanner, a Vidar Diagnostic ProPlus digitiser, scanned at a maximum of 3.85 optical density. Access to another scanner with at least 4.0 optical density capability was required.

Two types of x-ray scanners with ≥4.0 optical density were available in Australia: Array laser scanners (maximum 4.7 optical density); and the Vidar CADPro Advantage scanner, a CCD digitiser with maximum 4.2 optical density. Although previously owned x-ray scanners are often an option in other markets, in Australia this posed potential difficulties for power compatibility (if of US/Canadian origin); safe shipment to Australia; as well as ongoing servicing, support and maintenance. No previously owned x-ray scanners were found for purchase within Australia during 2011.

Outsourcing of digitisation was considered. The Australian Mammographic Density Research Facility at the University of Melbourne offers a paid service to digitise mammograms on an Array scanner. BreastScreen NSW in Sydney owned two recently purchased Array scanners. However, because the ANZ BCTG also required digitisation capability for IBIS-II mammograms, in-house digitisation was preferred.

The Array and Vidar systems differed in capabilities and cost. Two vendors, Hologic and Medilink, offered scanning systems which incorporated an Array laser scanner. The Hologic system (~$75,000) utilised a slightly older version of the Array laser scanner (maximum 4.4 optical density), whilst the Medilink system (a more expensive system) included the newest 4.7 optical density laser scanner. Hologic also offered a Vidar CCD scanning system, with up to 4.2 optical density capabilities (~$50,000).

Batch film scanning capacity differed between the Array and Vidar scanners. Array laser digitisers can scan batches of up to 100 films with an optional autofeeder accessory. The Vidar

CCD scanners had a 25 film capacity. Although CCD scanners theoretically required less ongoing maintenance than laser scanners, laser scanners had a reputation for being more robust.

Due to its higher optical density capability, sturdiness, auto-feed capacity, and similarity to the Array scanner utilised by the IBIS-II international coordinating centre in the UK, the decision was made to purchase an Array laser scanning system with an autofeeder accessory from Medilink. Sufficient funding was obtained through various sources to purchase the Array scanning system in October 2011. Medilink installed the scanner system in the Department of Surgical Oncology Research Office in the Newcastle BreastScreen facility in November 2011.

A.1.3 De-identification method selection

The mammograms used in this pilot project required removal of all participant identifying information (e.g. participant name, address) prior to density assessment. To maintain compatibility between IBIS-II mammograms collected for this substudy and those for the international CRUK MD IBIS-II project, a standard scanning protocol developed by CRUK was used. This generated images in DICOM (Digital Image and Communication In Medicine) format. DICOM is an internationally recognised standard for the creation, storage and transfer of images [523]. Standardisation of the DICOM format is the keystone for compatibility of DICOM images amongst different software programs.

Fully electronic mammograms in DICOM format are easily de-identified because the identifying information is contained in an electronic 'header' (text file) attached to the electronic image. Many free and/or inexpensive programs were readily available to de-identify DICOM images via manipulation of the standard fields contained in the DICOM header [524-527].

However, many popular image manipulation programs do not have native support for DICOM files. DICOM files can be readily transformed into other imaging formats such as .jpg and .tiff

for use with these program, however this may introduce changes to the original image. The information in the DICOM header is also lost during conversion to other image formats.

Images containing burned-in identifying information, such as a typical digitised film-screen mammogram, were more difficult to de-identify. The Array laser scanner used an extremely bright laser to digitise x-rays. Because the light was very strong, identifiable information was still discernible on the digitised image even when covered with a sticky label prior to scanning. Post-scanning de-identification of digitised mammograms was necessary.

The Picture Archiving and Communication System (PACS) clinical software included with the Array scanner appeared capable of masking areas during the scanning process. This software was able to output files to a non-PACS location (e.g. a hard drive), but it did not generate a human-friendly file name. The masking capability did not appear to function when saving to a non-PACS location.

The process used by CRUK to de-identify digitised film-screen mammograms was trialled. Adobe Photoshop CS5 has native compatibility with DICOM files. Files are opened one at a time, the areas to be removed are selected and masked, and the de-identified image is saved. This is a time consuming process. It is also very repetitive, and therefore suitable for automation by computer. However, user input was required by the Photoshop program during the de-identification process, and therefore it could not be automated.

An automated de-identification process was preferred for a number of reasons, including interest in undertaking a larger project (a linkage study between BreastScreen NSW and the NSW 45 and Up Study) which would generate thousands of images to de-identify. During the approvals process for this large project it was discovered that all research data, including mammographic images, needed to be de-identified prior to access by the research team. Hence

a quick and easy to use de-identification technique suitable for use by a competent assistant was required.

Commonly used imaging software was tested for its image de-identification capacity. Windows Paint is not compatible with DICOM files. The free program IrfanView did not produce usable de-identified images. ImageJ (the Java-based image manipulation platform from the National Institutes of Health (NIH) in the USA) is able to automate repetitive transformation of images via an inbuilt programming language. However, it does not have native support for DICOM files, and a suitable and reliable DICOM plugin for ImageJ was not found.

Typically it is better to acquire a commercial off-the-shelf (COTS) product than develop bespoke software [528, 529]. COTS products generally provide lower total cost of ownership, even if an higher initial outlay is incurred [530]. Hence available programs which could de-identify DICOM images with burnt-in areas of patient information with as little user input as possible were scrutinized.

A.1.4 Review of software to de-identify DICOM images

Software capable of quickly and easily batch de-identifying digital images in DICOM format was sought via general internet searches using terms including 'DICOM anonymizer burned in'. Searches in PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) using similar terms were also undertaken. Software was evaluated for its:

- Ability to easily remove DICOM header information from digital mammograms, namely:
    - Batch removal of standard DICOM header attributes
    - Batch removal of custom DICOM attributes (e.g. details added by the manufacturer of the mammography machine which are not part of the DICOM standard)
- Capacity to easily de-identify DICOM images with burned-in identifying areas for:

- A single mammogram
- Multiple mammograms (i.e. batch anonymisation capable)

More than 60 DICOM viewers and 20 de-identifiers/anonymisers were discovered. A particularly thorough listing of DICOM software was available at dclunie.com. Searches in PubMed found two useful, recent (<5 years old) articles discussing free DICOM viewers and anonymisers [526, 527]. Specifications for 34 programs were reviewed; 23 were not further investigated for a variety of reasons including cost and complexity (e.g. MatLab, Offis DICOM toolkit, Ruby DICOM). Table 10-1, below, lists the results of this search. Eleven programs were selected for evaluation of their ease of use and anonymisation capabilities; three programs satisfied at least two of the three evaluation criteria [Table 10-2, below].

Evaluation Results:

- Non-medical imaging software had difficulty with DICOM files.
- Many of the open-source, free programs were Java based; these tended to require modification of the MS Windows PATH environment variable in order to run the Java program, and could be a bit difficult to use (e.g. multiple steps were required to clean the data, the software needed detailed configuration prior to use).
- Three free but complex software packages offered a multitude of capabilities, such as DICOM network and file support and integration with PACS systems, and DICOM header anonymisation:
  - DICOM Confidential (University of Edinborough)
  - DICOM Cleaner (PixelMed by Dr. David Clunie)
  - MIRC DICOM software (Radiological Society of North America)
- The DICOM Anonymizer & Masker, was relatively easy to use. The resulting anonymised files were compatible with MD measurement software. However the software had minor

user interface issues—it failed to run if files were already present in output directory, and the program had to be closed before it could be used again to anonymise files.

- The SanteSoft DICOM Editor software was both simple to install and use, and the resulting de-identified (pseudo-anonymised) DICOM files were compatible with MD measurement software. It was capable of batch anonymising burned-in image data. This program was selected to perform the core anonymisation activities for the project.

- The Tudor DICOM Viewer easily removed data from custom fields in DICOM headers. This program was selected to remove any identifying information remaining in the DICOM header after de-identification with the SanteSoft DICOM Editor.

Table 10-1 Twenty-three DICOM compatible programs which were not further investigated

Name	*Comments*
Ruby DICOM	Command line interface only; both header and burned-in anonymisation capable
MatLab	Computation, analysis and programming software with DICOM anonymisation capabilities
Offis DICOM Toolkit: 'DCMTK'	DICOM server and viewer software, written in the C++ language
DVTk DICOM Toolkit	Large set of Java based tools including DICOM header anonymisation
Escape Medical Viewer: 'EMV'	Commercial software with DICOM viewer (~$300) for both Windows and Apple computers
ezDICOM	DICOM viewer, capable of reading a wide variety of image formats
DicomWorks [531]	Free DICOM viewer with header anonymisation, but was no longer maintained by its developers
GearView QC from PACSGear	Part of commercial suite of PACS software; can remove burned-in data
K-PACS: 'DICOMAnonymize'	Server connection capable; commercial CE/FDA approved 'iQ' version and free version with DICOM anonymiser
OsiriX	Clinical commercial version is FDA approved; Apple computers only. Anonymisation capable.
SimpleDICOM (UPMC) [532]	Complementary program for 1st generation PACS; DICOM header anonymisation
Sha He's DicomAnonymizer	DICOM header anonymisation for files including subfolders
DICOM Anonymizer Pro	Header anonymisation including private tags and subfolders; 49€
Gdcm ('Grassroots DICOM')	C++ library for parsing/writing DICOM files; has DICOM header de-identification capabilities
dicom3tools	Debian command line program which can manipulate (anonymise) DICOM headers
OSIRIS [524]	DICOM viewer—simple header anonymisation and image cropping
Dicom2 converter	Command line image conversion tool (e.g. DICOM file to JPG format); capable of cropping images
PowerDicom	Part of a commercial suite of software; DICOM header batch anonymisation capable
DICOM Parser	Free program from commercial vendor; header anonymisation
ClearCanvas	FDA approved PACS workstation software with free personal edition; header anonymisation capable
iRad Mac OSX Dcm Viewer	Apple computer PACS workstation image viewing software
ANALYZE (Mayo Foundation)	3D image analysis and visualisation software with many complex capabilities
LONI Pipeline (UCLA)	Free workflow software for linking analyses; has graphical interface

Table 10-2 Eleven programs tested for ease use regarding DICOM file anonymisation

Name	Specifications and evaluation criteria										Comments
	Native DICOM support	Medical imaging program	Many capabilities	Easy to use	Commercial program	Inexpensive (≤$500 USD)	1a. Standard header removal	1b.Custom header removal	2a. Single images	2b. Batches of images	
Photoshop CS5	✓				✓		✓		✓		Complex, expensive, no batch save
IrfanView				✓		✓			✓		Image detail lost
ImageJ			✓			✓			✓	✓	Inconsistent DICOM plugins
ShowCase5	✓	✓	✓		✓	✓	✓		✓		Difficult user interface
SanteSoft DICOM Editor	✓	✓	✓	✓	✓	✓	✓		✓	✓	Technological Ed. Institution, Athens, Greece Batch anonymisation capabilities
DICOM Anonymizer & Masker	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Batch anonymises both header and burned-in data $38USD; performance was inconsistent
DICOM Confidential [527]	✓	✓	✓			✓	✓	✓	✓		University of Edinborough, UK Complex, extensible toolkit for teaching and research at multiple sites. Anonymisation capabilities
DICOM Cleaner	✓	✓	✓			✓	✓	✓	✓		PixelMed (dclunie.com) Added a substantial amount of cleaning information to the DICOM header
3TMRI (aka dcm2nn-gui)	✓	✓	✓	✓	✓	✓	✓	✓			Did not removal all private attributes. Software cannot remove burned in data
TUDOR DICOM viewer	✓	✓	✓	✓		✓	✓	✓			Easy to use header anonymiser, but no burned-in data removal capabilities
RSNA MIRC DICOM	✓	✓	✓		✓	✓	✓	✓	✓	✓	Radiological Society of N.America. Complex suite of clinical trials and teaching files software for multiple sites. Easy to use header attribute anonymiser

A.2 Mammogram collection and de-identification

Trial mammograms were collected for all available mammography episodes (visits). The four standard mammographic Views (RCC, LCC, RMLO and LMLO) were collected for each mammography episode.

A.2.1 Digital mammograms

Electronic copies of CMN IBIS-II digital mammograms were collected either directly from the CMN trial coordinators or via the ANZ BCTG. Electronic digital mammograms from the Calvary Mater were provided on DVDs for each participant. The DVDs had patient-friendly software which automatically displayed the mammographic images stored on the disc when the DVD was loaded into a computer.

The DICOM files were stored on the DVD as extensionless files (no *.dcm) with meaningless names: e.g. M000000, M100000. The free software program, Bulk Rename Utility, was used to extract the files for each episode from the nested directory structure on the DVD, and save the files with a meaningful name and DICOM extension: participant ID + date + .dcm. The mammography View for each file (i.e. RCC, LCC, RMLO, LMLO) was recorded later in a tracking file. Fully electronic mammograms from the ANZ BCTG were supplied with a filename comprised of the IBIS-II participant ID, episode date and mammography View. The Bulk Rename Utility was used to replace the IBIS-II ID with a randomly assigned project ID for this project.

The DICOM headers of digital mammograms were de-identified using the Sante DICOM Editor's batch de-identification utility. Residual identifying information in custom DICOM attributes was removed using the Tudor DICOM Viewer's batch de-identifier.

A.2.2 Film-screen and digitised copies of digital mammograms printed to film

De-identification of film mammograms was more labour intensive. The films were first digitised into electronic format on the Array scanner. The first digitised film-screen mammograms collected for the project were de-identified using CRUK's Photoshop procedure, however the majority of digitised film-screen mammograms were de-identified using the Sante DICOM Editor program. Windowing (grey-level) differences in the Photoshop and Sante-Editor de-identified mammograms later appeared to cause the selected MD measurement software program (Cumulus) to crash during batch reads because of 'bad windowing'.

A.2.3 Digitisation procedure for the Array laser scanner

The Array laser scanner was provided with stand-alone (non-PACS) research software called Array View Lite. The software was used to scan batches of mammograms using the Array scanner to output DICOM files with a meaningful name.

An inbuilt feature of the Array View Lite software was utilised to automate systematic naming of the DICOM files output by the program. Array View Lite automatically appended the filename of the first scanned film with a '1'. Subsequent filenames were appended with increments of 1. Hence mammograms sorted into RCC–LCC–RMLO–LMLO order had 1–2–3–4 appended to their respective filenames for the first episode scanned, 5–6–7–8 appended for the second episode scanned, and so on. The Bulk Rename Utility was later used to rename all digitised mammogram filenames appended with a 1, 5, 9... as 'RCC', 2, 6, 10... as 'LCC' and so forth. Episodes were scanned in date order, starting with the oldest episode to assist with subsequent renaming of the files.

The Array View Lite software has an 'autosize' feature to sense and set the size of the mammogram during scanning, however results using this feature were inconsistent. Hence the

size of the mammogram was always entered manually at the start of each batch scan. Film-screen mammograms were typically 18x24 cm in size, but could also be 20x25cm. CR digital mammograms printed to film were usually 20x25cm in size. All mammograms of the same size for a particular woman were sorted into RCC, LCC, RMLO, LMLO order by episode date, and digitised as a batch on the Array scanner. Film-screen mammograms were scanned twice—once at 4.3 optical density for consistency with CRUK mammography scanning guidelines, and once at 4.7 optical density for the CMN IBIS-II MD and AI substudy.

At the time of scanning, the Array View Lite software was utilised to add information such as episode date and study ID into the DICOM header of the digitised mammograms. For episodes scanned one at a time, the DICOM date field was completed with the mammogram date, else the range of episode dates in a batch was entered into a free text field. The optical density configured at the time of scanning was added to the DICOM header and the filename. These DICOM fields provided secondary identifiers to the informative filename for each mammogram, to lessen the potential for later miss-assignment of mammograms to participants post de-identification.

Scanning batches were limited in size not by the autofeeder capacity (100) of the Array scanner but by the amount of memory accessible by the scanning program (~1.6GB of RAM) on the Windows XP computer supplied with the system. Scan settings such as the bit depth and pixel spacing determined the size of each mammogram in memory. The batch size was approximately 50 mammograms when the Array scanner was set to the CRUK IBIS-II configuration of 12 bits and 50 micrometer (μ) pixel spacing.

A.2.4 Deidentification of scanned mammograms using Sante DICOM Editor

The burned-in patient identification details on film mammograms was located in approximately the same location for right and left mammograms of the same size and same modality (film-

screen or digitised digital-printed-to-film). The burned-in identification details for both RCC and RMLO View mammograms was located at the upper left side of the films; identifying information for LCC and LMLO View mammograms was located at the lower right of the films.

Films of different sizes created different sized electronic files; typically files were 34MB (for 18 x 24cm film-screen mammograms) or 56 MB (for 20 x 25cm film-screen mammograms) in size. Each file size had different Sante DICOM Editor X and Y coordinates for the burnt-in information. Prior to batch de-identification, the electronic images were therefore organised into different file sizes for each type of mammogram (digitised film-screen or digitised digital printed-to-film), which were then divided into right or left Views. These divisions by electronic file size and laterality permitted each group of files to be batch de-identified with the same set of Sante DICOM Editor X & Y coordinates.

The X and Y coordinates to mask the burned-in information on each group of files was determined by using the scale displayed on a ruler within the Sante DICOM Editor. Twenty-five units on the ruler corresponded to 500 units in the program's batch anonymisation 'Set Rectangles' entry screen. The sorted files were batch de-identified in Sante Editor using the appropriate coordinates for the mammograms' file size and laterality.

B. Differences in MD measurements, by View, Type and Version

Background and Methods

The expected substantial variation for repeated (longitudinal) Cumulus measurements of both film and digital mammograms in this project (±8%, ±10% respectively, as reported in Aim 2, Chapter 5 (Intra-observer and inter-method reliability)) is large compared to the expected longitudinal decrease in PD of 1% to 2% for treated participants relative to control participants. As described in the thesis Methods (Chapter 1), MD assessment of all four Views was undertaken to reduce the impact of this variability in a small dataset. Typically, however, only a single view is utilised to assess PD. A previous study noted slight differences in BC risk estimation when utilising one, two or four mammographic Views [468]. BA is generally greater on MLO compared to CC Views due to the different perspective used during imaging. BA on left mammograms also tends to be 20 to 30 cm^2 larger than BA for right side mammograms [533]. AA is likely to mirror the differences noted for BA. PD and DA may also vary between right and left mammograms, as well as for CC vs MLO Views. Hence different Views may not be interchangeable for longitudinal analysis of MD, and these differences may need to be taken into account during modelling.

Within episode differences in percent density PD, DA, BA and AA were examined for each mammographic View (RCC, LCC, RMLO, LMLO), by both mammogram Type (film, digital) and mammogram Version (film, KE5.2, KE5.4, Fuji). Comparisons were made between the right and left side for individual Views (RCC vs LCC, and RMLO vs LMLO), for CC vs MLO Views on the same side (RCC vs RMLO, LCC vs LMLO) as well as pairs of Views from each episode divided into right vs left mammograms, and CC vs MLO mammograms. Mammograms from all 540 episodes from Collection 1 and Collection 2 were utilised in preliminary descriptive (graphical) comparisons and quantitative comparisons utilising Wilcoxon (matched-pairs) signed-rank tests for MD differences due to laterality (right vs left), CC vs MLO Views, or different individual Views. The MD parameters were also tested for normality by Type,

View and Version using tests for skewness and kurtosis as well as the Shapiro-Wilk test; normality was also assessed qualitatively (through graphical inspection).

These preliminary paired comparisons treated each pair of observations as if they were independent, when in fact multiple pairs were typically contributed by each participant. The paired differences were therefore likely to be more consistent within participants than between participants. The pairs, however, were treated as independent samples because the variability of the subjective repeated measurements (>8% PD) made for each episode (i.e. on the four mammographic Views for each episode) was expected to be generally greater than between person variability (the difference between pairs of Views within episodes for different women). The number of episodes collected per participant also varied, hence unequal numbers of mammograms from each participant were used in the preliminary comparisons. This effectively allowed data from some participants to influence the results more than others.

A set of comparisons utilising the Wilcoxon signed rank test was also tabulated for composite measurements comprised of the average of both Right, both Left, both CC and both MLO Views from the 'first' episode (baseline or earliest episode, n=120) for each participant. In other words, MD measurements from both right mammograms from an episode were averaged to create a composite (average) measurement for the right side for that episode ((RCC+RMLO)/2), both left side mammograms were averaged to make a composite measurement for the left side for each episode ((LCC+LMLO)/2); the process was repeated for the CC and MLO Views ((RCC+LCC)/2), (RMLO+LMLO)/2). This was done to avoid the issues of repeated measurements, and unequal numbers of episodes and mammograms, encountered during the preliminary comparisons.

Results

The 2,130 Calvary Mater Newcastle IBIS-II mammograms assessed for density comprised 534 RCC, 532 LCC, 534 RMLO and 530 LMLO mammograms (Table 10-3, below). As described previously, the majority of mammograms were in electronic CR digital format.

Table 10-3 Number of mammograms by View (RCC, LCC, RMLO, LMLO) and Type (Film, Digital, DoF)

View	Mammogram Type		Total, Film and Digital Mammograms	Digital Subtotals by Version*		
	Film-screen	Digital		KE52	KE54	Fuji
RCC	61	473	**534**	121	150	198
LCC	63	469	**532**	121	148	196
RMLO	61	473	**534**	120	150	199
LMLO	63	467	**530**	120	148	195
Total	**248**	**1882**	**2130**	**482**	**596**	**788**

* 16 digital mammograms of 'Other' Version are omitted

Tests for normality for all mammograms using the skewness and kurtosis test, and the Shapiro-Wilk test, for PD, DA, BA and AA yielded p-values <0.001. Use of the Shapiro-Wilk test by mammogram View, and Type or Version for each MD measure yielded p-values <0.001. The distribution is skewed for the distribution overall and for each sub group of each MD parameter, as is visible in Figure 10-1.

Figure 10-2 reveals the trend observed in Figure 10-1for higher CC vs MLO for PD and DA, and lower BA and AA for CC vs MLO, continues within each mammogram Version in Figure 10-2; the trend is not as pronounced however, in the less consistent film mammogram distributions for DA and AA (right column, Figure 10-2). Across Versions, the trend for higher overall PD and DA for film compared to digital mammograms is present, whilst the reverse is true for BA and AA. A trend for increasing PD and DA as well as BA and AA is seen for digital mammograms (KE52 to KE54 to Fuji); with the exception of the film to digital transition, this implies that all of the MD parameters tend to increase over time. These trends mirror that seen for PD in Figure 6-1, as well as the trends noted for PD, DA, BA and AA in Figure 7-7.

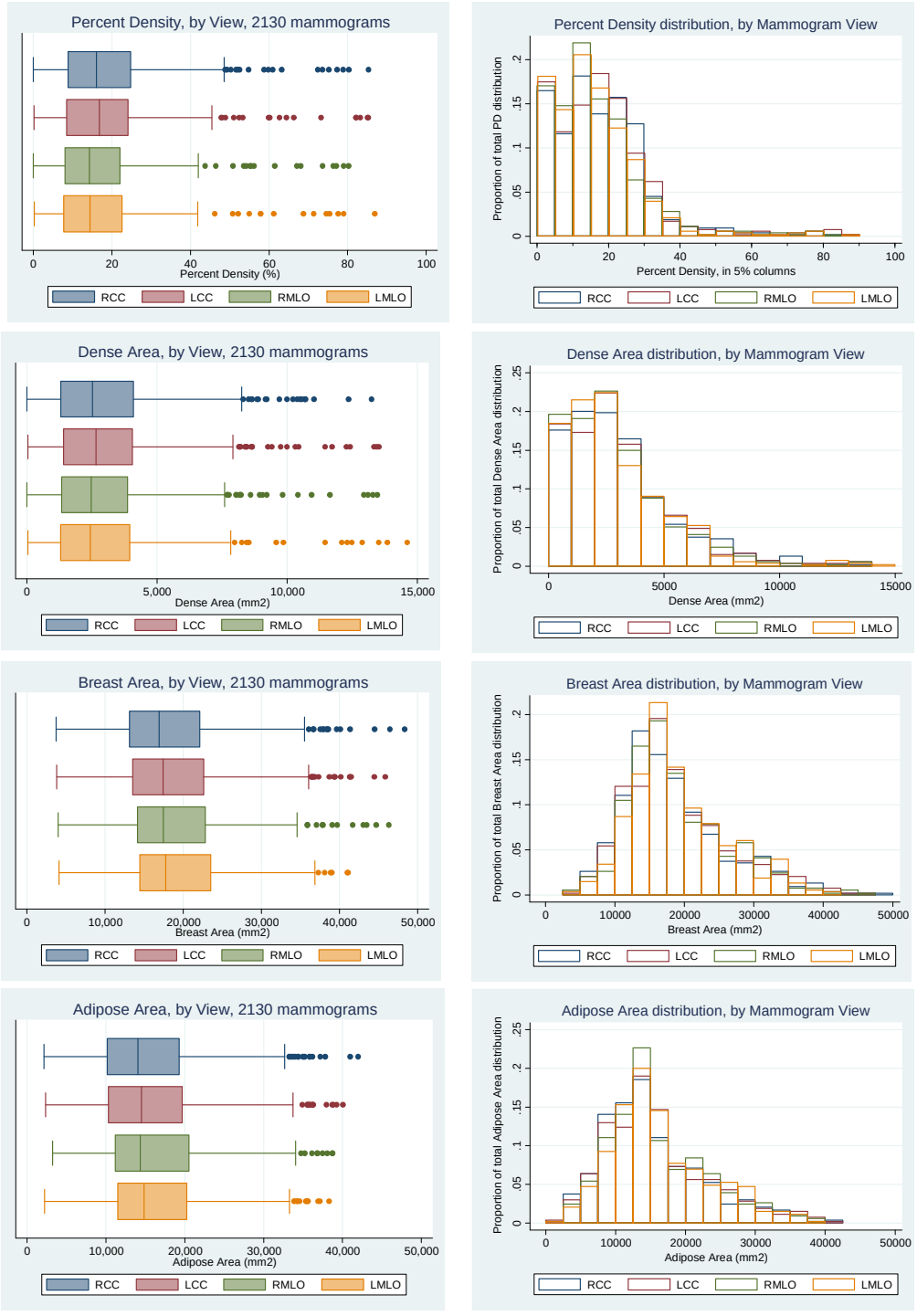


Figure 10-1 Distributions of PD, DA, BA and AA by View.
Left column, box plots by View. Right column, histograms by View. As noted previously, PD and DA have a strong right skew, BA and AA are more normal. Slight differences between CC and MLO views are present for all MD parameters; the difference is most pronounced for CC vs MLO views. PD and DA tend to be higher for CC compared to MLO Views, whilst the reverse is true for BA and AA.

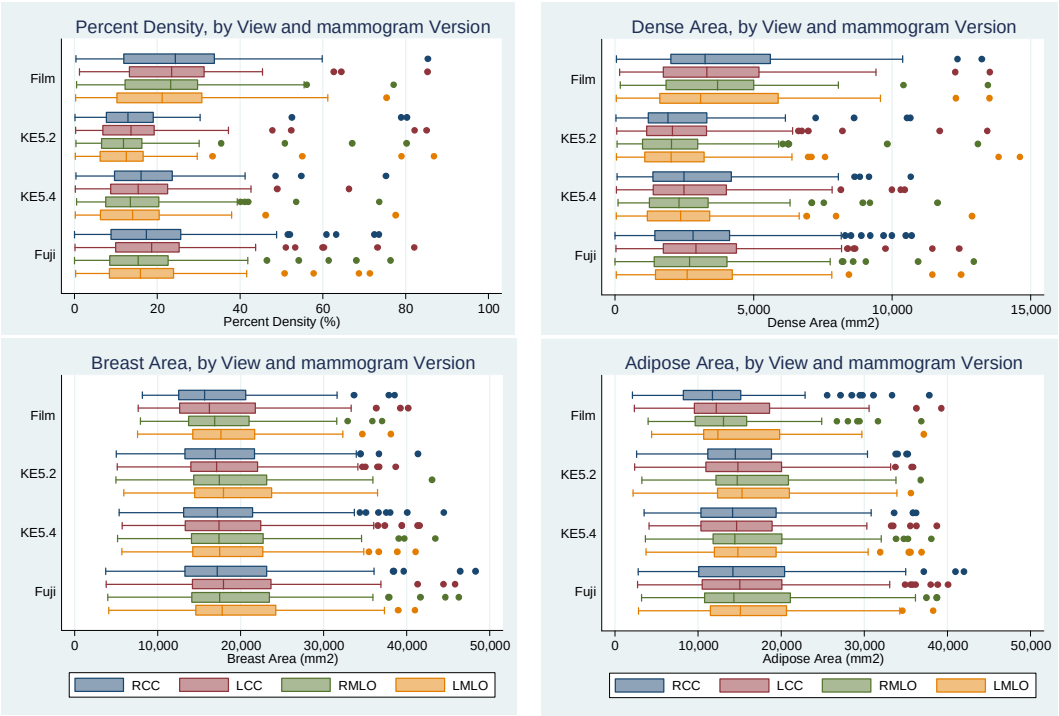


Figure 10-2 Distributions of PD, DA, BA and AA by View and mammogram Version
Versions include film, KE5.2, KE5.4, and Fuji. 16 mammograms of Other Version are not shown. The trends noted in Figure 10-1 between different Views are generally mirrored above within each Version.

Table 10-4 Difference between pairs of Views for PD, DA, BA and AA (overall, by Type & Version)

Comparison[1]	N pairs	Median difference[2]	Diff/ median[3]	Wilcox. p-value	Wilcoxon p-values				
					Film	Digital	KE52	KE54	Fuji
Percent Density (%)		Overall difference all film and digital mammograms			Within Type difference		Within Version difference		
RCC vs LCC	525	0.11	0.07%	ns	ns	ns	ns	ns	ns
RMLO vs LMLO	524	0.06	0.04%	ns	ns	ns	ns	ns	ns
RCC vs RMLO	532	1.10	7%	***	*	***	***	***	***
LCC vs LMLO	530	1.40	9%	***	ns	***	***	***	***
Dense Area (mm^2)		Overall difference			Within Type		Within Version		
RCC vs LCC	525	-36.5	1.5%	ns	ns	ns	ns	ns	ns
RMLO vs LMLO	524	-13.5	0.5%	ns	ns	ns	ns	ns	ns
RCC vs RMLO	532	76.0	3%	***	ns	***	ns	**	ns
LCC vs LMLO	530	128.4	5%	***	ns	***	ns	***	***
Breast Area (mm^2)		Overall difference			Within Type		Within Version		
RCC vs LCC	525	-502	3%	***	ns	***	*	***	***
RMLO vs LMLO	524	-305	2%	***	ns	***	*	ns	*
RCC vs RMLO	532	-596	3%	***	***	***	***	***	***
LCC vs LMLO	530	-510	3%	***	**	***	***	***	***
Adipose Area (mm^2)		Overall difference			Within Type		Within Version		
RCC vs LCC	525	-402	3%	ns	*	***	ns	***	***
RMLO vs LMLO	524	-264	2%	ns	ns	***	ns	ns	**
RCC vs RMLO	532	-789	5%	***	***	***	***	***	***
LCC vs LMLO	530	-689	5%	***	**	***	***	***	**

ns-not significant; *<0.05; ** <0.01, ***<0.001; Wilcoxon's signed-rank test (paired data)
[1]Number of mammogram pairs are 525, 524, 532 & 530 for View comparisons (i.e. RCC v LCC (525), etc.)
[2]Median difference of second in pair subtracted from first, e.g. RCC – LCC
[3]Median difference for each pair divided by the median of its MD parameter, Table 7-3 (all mmgs)

Table 10-5 Difference between pairs of Right vs Left, and CC vs MLO mammograms: PD, DA, BA and AA, overall (all mmgs), and by Type and mammogram Version

Comparison[1]	N pairs	Median diff[2]	Diff/med (%)[3]	Wilcoxon p-value	Wilcoxon p-values				
					Film	Digital	KE5.2	KE5.4	Fuji
Percent Density (%)		Overall difference, all mammograms			Within Type		Within Version		
Right vs Left	1049	0.07	0.45%	ns	ns	ns	ns	ns	ns
CC vs MLO	1062	1.22	8%	***	**	***	***	***	***
Dense Area (mm^2)			Overall difference		Within Type		Within Version		
Right vs Left	1049	-24.6	1%	ns	ns	ns	*	ns	ns
CC vs MLO	1062	107	4%	***	ns	***	**	***	***
Breast Area (mm^2)			Overall difference						
Right vs Left	1049	-384	2%	***	ns	***	**	***	***
CC vs MLO	1062	-542	3%	***	***	***	***	***	***
Adipose Area (mm^2)			Overall difference		Within Type		Within Version		
Right vs Left	1049	-343	2%	ns	*	***	ns	***	***
CC vs MLO	1062	-738	5%	***	***	***	***	***	***

ns-not significant; *<0.05; ** <0.01, ***<0.001; Wilcoxon's paired signed-rank test
[1]Number of mammogram pairs are 1049 for Right v Left, 1062 pairs for CC v MLO
[2]Median difference of second in pair subtracted from first, i.e. Right – Left or CC–MLO
[3]Median difference for each pair divided by the median of its MD parameter Table 7-3 (all mmgs)

Table 10-6 Difference between 120 first episodes, Right vs Left and CC vs MLO, for PD, DA, BA and AA

Comparison[1]	N pairs	Median[2]		Median difference[3]	Wilcoxon p-value[a]
		Right or CC	Left or MLO		
Percent Density (%)					
Right vs Left	119	14.9	16.4	-0.08	0.88
CC vs MLO	120	16.7	14.6	0.72	0.0009
Dense Area (mm^2)					
Right vs Left	119	2451	2538	-68	0.33
CC vs MLO	120	2535	2437	48	0.41
Breast Area (mm^2)					
Right vs Left	119	16590	17152	-399	0.002
CC vs MLO	120	16527	17669	-788	<0.0001
Adipose Area (mm^2)					
Right vs Left	119	13672	14198	-337	0.004
CC vs MLO	120	13562	14288	-792	<0.0001

[a] Wilcoxon's paired signed-rank test for comparison of average Right vs Left or CC vs MLO per episode
[1] Average of the Right and Left, or CC and MLO Views for each participant's 'first' episode (baseline or earliest episode); left side mammograms not available for 1 participant for Right vs Left comparisons
[2] Median value for the Right, CC, Left or MLO average measurements for first episodes
[3] Median difference of second in pair subtracted from first, i.e. Right – Left or CC–MLO

For PD and DA the differences in View for right vs left side were not significant for any comparison— Table 10-4, Table 10-5, and Table 10-6. The right vs left side comparisons were not significant for AA when all mammograms were used (Table 10-4, Table 10-5, n=540 episodes) however when Views from a single episode from each participant were compared (n=120 episodes), the AA right vs left (as well as CC vs MLO) comparisons differed

significantly— by about $-3cm^2$ (right vs left, $p=0.004$) and $-8cm^2$ (CC vs MLO, $p<0.0001$), Table 10-6. The within Type and within Version comparisons for right vs left side mammograms for PD and DA (Table 10-4, Table 10-5) reveal the within Type and Version comparisons for film, KE52, KE54 or Fuji were not significant, like their all mammograms counterparts. However the within Type and within Version comparisons for AA were less consistent; with the exception of KE52 mammograms however, all other AA right vs left comparisons were significantly different. The AA KE52 mammogram comparisons may be different from the other comparisons because of the more extensive post processing removal of dense tissue from KE52 mammograms (i.e. lower PD and DA and concomitantly slightly higher median AA, re: Figure 10-2) compared to other mammogram Versions.

PD tended to be about 1% greater on CC compared to MLO mammograms (Table 10-4, Table 10-5, Table 10-6), a difference which was significant for all comparisons except for the LCC vs LMLO comparison for film, Table 10-4. DA tended to be higher (by 0.5 to 1 cm^2) on CC compared to MLO mammograms (Table 10-4, Table 10-5, Table 10-6), a difference which was not significant when the comparison was restricted to first mammograms only (Table 10-6).

The comparisons for BA showed that right breasts tend to be smaller than left breasts, and that BA on CC Views tends to be smaller than that for MLO Views. These differences were significant for comparisons of all mammograms (Table 10-4, Table 10-5) and first mammograms only (Table 10-6), and significantly different— except for film— for the comparisons by Type and Version (Table 10-4, Table 10-5). The non-significance of the film only comparison for BA may have been due to the smaller sample size (63 episodes) for Film mammograms vs the CR Versions (123, 151 and 200 respectively for KE5.2, KE5.4 and Fuji).

Discussion and Conclusion

During modelling of BA using MD measurements from individual Views, both laterality (right vs left) and CC vs MLO Views may need to be taken into account due to the statistically

significant BA differences in right vs left, as well as CC vs MLO mammograms (Table 10-4, Table 10-5, and Table 10-6). These results reflect the larger cross sectional breast area present on MLO compared to CC View mammograms. These results are also consistent with other reports in the literature, e.g. [533], which show that the left breast tends to be slightly greater in area on mammograms compared to the right breast.

Although BA tends to be greater for left compared to right breasts, this did not translate to significantly more DA on left compared to right breasts, although right side DA was non-significantly smaller than left side DA in all comparisons. A larger sample size may therefore show a significant difference in DA between right and left mammograms. Potentially, because the MLO View provides a greater cross sectional area of the breast (as reflected in the higher BA for MLO compared to CC Views), more DA might also be visible on the MLO compared to CC Views. The dense tissues could potentially be more widely distributed across the greater BA of the MLO View, causing less superimposition of the dense tissues which is measured as increased DA. Given, however, that the total dense volume of the breast is identical whether viewed from the CC or MLO perspective, DA might be the same for both CC and MLO Views. The results from this small sample of high risk women, conversely, imply slightly more DA (~0.5cm^2 to 1cm^2) is visible on CC compared to MLO Views. Although the DA difference for CC vs MLO Views was significant when compared for pairs of measurements of all 2130 mammograms, the slight difference in DA for CC vs MLO Views was not significant when 120 pairs of measurements were compared (1 pair for each of the 120 CMN IBIS-II participants) so the results may be spurious. Further examination of DA measured CC vs MLO View mammograms from a larger population is needed to confirm this result. CC vs MLO Views, and right vs left side may need to be taken into account when modelling longitudinal DA using measurements from individual Views. Because of the small expected treatment effect predicted for anastrozole, even a small difference in DA due to differences in View could potentially affect longitudinal DA trends, hence measurements made on the CC vs MLO, as well as right vs

left sides may not be exchangeable during longitudinal modelling. However, differences in DA were not analysed relative to the total DA present for each participant; the relative within participant DA difference for right vs left, and CC vs MLO Views are likely to be smaller than the variability of the measurements (for subjective measurement techniques). Therefore substitution of one View for another, even if MD on the Views is slightly different, may not result in a different measurement.

PD did not differ greatly (nor significantly) between the right and left sides; the difference between right and left mammogram was around 0.1%, e.g. Table 10-6. PD differed significantly by about 1%, however, between measurements made on the CC compared to MLO Views. As for DA, this difference in PD between CC and MLO Views, although small, may affect the results of longitudinal models. Hence it is likely better to not substitute CC for MLO Views (or MLO for CC Views) during longitudinal modelling. These results imply, however, as for DA very little average difference exists for PD between the right and left breasts, thus– generally– longitudinal modelling is not likely to be greatly affected if CC and MLO Views are substituted for each other. Given that some women are more greatly asymmetric than others in breast size [533] and the appearance of the dense tissues can differ between CC and MLO Views (noted during assessment of mammograms for this project), care may need to be taken during substitution of one View with another.

AA may differ slightly more than PD between measurements made on the right and left side, e.g. 3 cm^2, Table 10-5 and Table 10-6; the difference in right vs left AA was significant for measurements made on first mammograms only, Table 10-6. The CC vs MLO comparisons for AA were significantly different; MLO measurements were approximately 7 to 8 cm^2 larger on average than the CC measurements. These results imply that both laterality (right vs left) and CC vs MLO Views may need to be taken into account when modelling AA longitudinally.

For the purposes of assessing differences in BC risk posed by differences in right, left, CC or MLO View mammograms, all of the MD differences noted in this analysis were small relative to total PD, DA, AA or BA. For instance, PD differed on average by about 1% at most between Views (CC vs MLO) compared to the median baseline PD of 18% (Table 6-6), and DA by generally no more than 1cm^2 relative to the baseline median DA of 28cm^2 (2755 mm^2, Table 6-6). Although MD for some women does differ markedly in appearance on different mammographic Views, these results suggest that on average, MD assessments made on any View are roughly equivalent to those made on other Views. Therefore, the BC risk assessed on any View, or various combinations of individual Views are likely to be equivalent for most women.

This analysis could be improved by examining the range of MD differences between Views for participants, to quantify the measures of central tendency (i.e. not just the median, but also Q1, Q3, the minimum and maximum) for the differences measured for PD, DA as well as AA and BA. The results in Table 10-4 and Table 10-5 are problematic due to the use of repeated and unequal numbers of measurements made from the participants. This study was limited by the small number of women sampled (120). The population sampled were all high risk, post-menopausal women hence these results may not be generalisable to normal risk populations, and/or premenopausal populations. The PD, DA and AA measurements were also affected by differences mammographic Version utilised in this project, which may have made the differences in DA smaller (and those for AA larger) than would have resulted had only mammograms unaffected or less affected by digital post-processing been utilised (e.g. film and/or Fuji mammograms). Strengths of this project include use of a semi-automated technique to assess MD which is a well-established method to measure breast density on mammograms. The ICC results of the reliability analysis (Aim 2, Chapter 5) imply that the variability of the PD measured using the measurement technique for this project is acceptable. Use of all four standard mammographic Views enabled a wide range of comparisons for many different MD attributes (PD, DA, BA and AA).

C. Number of episodes (participants) at each follow up by Version

Table 10-7 Number of episodes (participants) at each follow up, by mammogram Version

Mammogram Version	Follow up number (years post-randomisation)									
	0	.5	1	2	3	4	5	6	7	Total
Film	42	12	8	1	0	0	0	0	0	63
KE52	28	6	20	41	16	11	1	0	0	123
KE54	12	7	20	26	45	28	12	1	0	151
Fuji	0	0	11	25	30	43	51	29	10	199
Other	3	0	1	0	0	0	0	0	0	4
Total	85	25	60	93	91	82	64	30	10	540

Baseline to year 1 mammography was primarily confined to film, KE52 and KE54 mammograms. Fuji mammograms were taken for some participants commencing from follow up 1 onwards, however no Fuji mammograms were taken prior to year 1. Only one film episode occurred after the first year of follow up; film mammograms effectively were taken only from baseline to year 1. A single KE52 episode occurred after year 4, and only one KE54 mammographic episode was taken after the 5th year follow up. The majority of episodes which occurred between years 5 to 7 were the Fuji mammograms. A few mammograms of 'Other' type were taken at baseline and year 1.

D. Simple regression of baseline PD with different combinations of mammographic Views

Table 10-8 Single parameter (simple) regression coefficients for single Views or Averages of 2 or 4 Views, baseline PD in mm^2 and log & square root transformed PD

Covariate	N*	PD in mm^2 (untransformed)									Log transformed PD			Square root PD		
		RCC	LCC	R MLO	L MLO	R+L CC	R+L MLO	Right side	Left side	All Views	RCC	R+L MLO	All Views	RCC	R+L MLO	All Views
		85	85	85	84	85	85	85	84	85	85	85	85	85	85	85
Age at Rand.	85	-0.25[1]	-0.21[1]	-0.31[1]	-0.19[1]	-0.23[1]	-0.27[1]	-0.28	-0.18	-0.32[1]	-0.05[1]	-0.04[1]	**-0.04**[1]	-0.05[1]	-0.05[1]	-0.05[1]
Film-screen	42	0.05[1]	0.34[1]	0.04[1]	0.12[1]	0.19[1]	0.08[1]	0.05[1]	0.23[1]	0.14[1]	-0.03[1]	-0.02[1]	-0.01[1]	-0.02[1]	-0.01[1]	-<0.001
Digital (CR)	43	**-0.6**	**-0.67**	**-0.65**	-0.51	**-0.6**	**-0.6**	**-0.6**	**-0.6**	**-0.6**	**-0.07**	**-0.06**	**-0.07**	**-0.09**	**-0.09**	**-0.09**
KE52	28	**-0.7**	**-0.77**	**-0.57**	-0.54	**-0.7**	**-0.6**	**-0.6**	**-0.7**	**-0.7**	-0.08	**-0.07**	**-0.08**	**-0.09**	**-0.09**	**-0.10**
KE54[2]	12	-0.2	-0.28	-0.44	-0.11	-0.2	-0.3	-0.3	-0.2	-0.3	-0.03	-0.03	-0.03	-0.03	-0.04	-0.04
Fuji	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
BMI (kg/m^2)	85	**-0.9**	**-0.9**	**-1.0**	**-0.8**	**-0.9**	**-0.9**	**-0.9**	**-0.9**	**-0.9**	**-0.10**	**-0.09**	**-0.09**	**-0.13**	**-0.13**	**-0.13**
Film-screen	42	**-1.3**	**-1.5**	**-1.4**	**-1.4**	**-1.4**	**-1.4**	**-1.4**	**-1.4**	**-1.4**	**-0.10**	**-0.12**	**-0.12**	**-0.17**	**-0.19**	**-0.19**
Digital	43	**-0.5**	**-0.6**	**-0.6**	-0.4	**-0.7**	**-0.5**	**-0.6**	**-0.5**	**-0.6**	**-0.09**	**-0.07**	**-0.07**	**-0.09**	**-0.09**	**-0.09**
KE52	28	**-0.6**	**-0.8**	**-0.6**	-0.5	**-0.6**	**-0.6**	**-0.6**	**-0.6**	**-0.6**	**-0.14**	**-0.12**	**-0.12**	**-0.13**	**-0.12**	**-0.13**
KE54[2]	12	-0.5	**-0.4**	-0.7	-0.3	-0.4	-0.5	-0.6	-0.3	-0.5	-0.03	-0.03	-0.03	-0.06	-0.06	-0.06
Height (cm)	85	-0.3	-0.2	-0.2	-0.1	-0.2	-0.2	-0.2	-0.2	-0.2	-0.02	0.02	0.01	-0.01	-0.001	-0.01
Film-screen	42	-0.5	-0.6	-0.4	-0.3	-0.6	-0.4	-0.5	-0.5	-0.5	-0.01	-0.02	-0.02	-0.05	-0.04	-0.04
Digital	43	0.2	0.2	0.1	0.2	0.2	0.1	0.1	0.2	0.2	0.03	**0.06**	0.05	0.03	0.05	0.04
Weight (kg)	85	**-0.3**	**-0.3**	**-0.3**	**-0.3**	**-0.3**	**-0.3**	**-0.3**	**-0.3**	**-0.3**	**-0.03**	**-0.02**	**-0.02**	**-0.04**	**-0.04**	**-0.04**
Film-screen	42	**-0.4**	**-0.5**	**-0.4**	**-0.4**	**-0.5**	**-0.4**	**-0.4**	**-0.4**	**-0.4**	**-0.03**	**-0.03**	**-0.03**	**-0.05**	**-0.05**	**-0.06**
Digital	43	-0.2	-0.2	**-0.2**	-0.1	-0.2	-0.2	-0.2	-0.1	-0.2	-0.02	-0.01	-0.02	-0.03	-0.02	-0.02
Menarche (age in years)	85	0.8	0.7	0.7	0.6	0.8	0.6	0.7	0.7	0.7	0.05	0.07	0.06	0.09	0.10	0.10
Film-screen	42	1.0	1.0	0.8	0.8	1.0	0.8	0.9	1.0	0.9	0.11	0.12	0.11	0.16	0.15	0.15
Digital	43	-1.0	1.2	0.8	-0.9	-1.0	-0.9	-0.9	-0.9	1.2	-0.07	-0.06	-0.07	-0.13	-0.10	-0.10
Age at First Birth (years)	80	0.7	0.5	0.5	0.5	0.6	0.6	0.6	0.5	0.6	0.05	0.04	0.04	0.08	0.07	0.07
Film-screen	42	-0.1	0.03	-0.3	0.1	-0.03	-0.1	-0.2	0.05	-0.1	-0.02	-0.02	-0.02	-0.02	-0.02	-0.01

Table 10-8, continued, baseline simple regression coefficients from models with different combinations of Views

Covariate	N*	PD in mm^2 (untransformed)									Log transformed PD			Square root PD		
		RCC	LCC	R MLO	L MLO	R+L CC	R+L MLO	Right side	Left side	All Views	RCC	R+L MLO	All Views	RCC	R+L MLO	All Views
		85	85	85	84	85	85	85	84	85	85	85	85	85	85	85
Digital	38	**1.0**	0.6	0.9	0.8	**0.8**	**0.9**	**1.0**	0.07	0.8	**0.10**	0.07	0.07	0.14	0.11	**0.11**
Menopause (age in years)	85	**0.6**	**0.6**	**0.5**	**0.5**	**0.6**	**0.5**	**0.6**	**0.6**	**0.6**	**0.05**	**0.05**	**0.05**	**0.08**	**0.07**	**0.08**
Film-screen	42	0.5	0.5	0.3	0.3	0.5	0.3	0.4	0.4	0.4	0.01	0.01	0.01	0.04	0.03	0.04
Digital	43	**0.4**	**0.5**	**0.4**	0.4	**0.4**	**0.4**	**0.4**	**0.4**	**0.4**	**0.06**	**0.07**	**0.07**	**0.07**	**0.08**	**0.08**
HRT (months)	85	-0.02	-0.03	-0.03	-0.02	-0.03	-0.03	-0.03	-0.02	-0.03	**-0.004**	-0.003	-0.003	-0.005	-0.004	-0.004
Film-screen	42	0.03	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.03	-0.003	0.003	-0.003	0.004	-0.005	0.005
Digital	43	**-0.05**	**-0.05**	**-0.05**	-0.04	**-0.05**	**-0.05**	**-0.05**	-0.04	**-0.05**	**-0.006**	**-0.005**	**-0.005**	**-0.008**	**-0.007**	**-0.007**
Oral contra-ceptives (months)	85	0.006	0.01	0.003	0.01	0.006	0.006	0.004	0.006	0.01	0.001	0.001	0.001	0.001	0.001	0.001
Film-screen	42	0.01	0.01	0.001	0.01	0.02	0.005	0.003	0.01	0.01	0.001	<0.001	<0.001	0.001	0.001	0.001
Digital	43	0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.01	0.02	0.002	0.003	0.003	0.003	0.003	0.003
Weighted # rels w BC OC	85	0.6	0.4	0.7	0.4	0.5	0.6	0.6	0.4	0.5	0.02	0.01	0.01	0.06	0.05	0.05
Film-screen	42	1.3[3]	1.1[3]	**1.6[3]**	1.3[3]	1.2[3]	**1.5[3]**	**1.5[3]**	1.2[3]	1.3[3]	0.05	0.06	**0.05**	0.13	0.14	0.13
Digital	43	0.3	0.2	0.5	0.2	0.3	0.3	0.5	0.	0.3	0.02	-<0.001	-<0.001	0.06	0.03	0.04
Nulliparous (yes /no(ref))	85	-5	-4	-5	-6	-4	-5	-5	-5	-5	-0.03	-0.6	-0.4	-0.4	-0.6	-0.5
Smoking status - never	44	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Current	9	2	2.7	0.6	3	2.2	1.8	1.2	3	2	0.41	0.33	0.36	0.34	0.41	0.3
Ex-smoker	32	-2	-1.4	-3	-1.3	-1.5	-2.4	-2.4	-1	-2	-0.10	-0.07	-0.07	-0.19	-0.26	-0.21
Weighted # of relatives[4] One 1st	30	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
One 1st + 1 2nd	21	-3	-2	-3	-2	-3	-3	-3	-2	-3	-0.50	-0.56	-0.53	-0.57	-0.64	-0.59
Two 1st	17	1	-0.7	0.2	0.4	0.2	0.3	0.6	-0.2	0.2	0.07	-0.13	-0.08	0.15	-0.02	-0.002
One 1st+1 2nd+	14	5	3	6	4	4	5	5	3	4	0.05	0.004	-0.005	0.37	0.31	0.28
Mammogram Type -film	42	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.

Table 10-8, continued, baseline simple regression coefficients from models with different combinations of Views

Covariate	N*	PD in mm^2 (untransformed)									Log transformed PD			Square root PD		
		RCC	LCC	R MLO	L MLO	R+L CC	R+L MLO	Right side	Left side	All Views	RCC	R+L MLO	All Views	RCC	R+L MLO	All Views
		85	85	85	84	85	85	85	84	85	85	85	85	85	85	85
Digital	43	**-10.7**	**-9.0**	**-9.5**	**-8.5**	**-9.9**	**-9.2**	**-10.1**	**-8.6**	**-9.5**	**-0.58**	**-0.58**	**-0.57**	**-1.13**	**-1.05**	**-1.06**
Mammogram Version[5]- film	42	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
KE52	28	**-11.7**	**-9.5**	**-11.1**	**-9.3**	**-10.6**	**-11.7**	**-11.5**	**-9.0**	**-10.5**	**-0.69**	**-0.62**	**-0.62**	**-1.26**	**-1.16**	**-1.15**
KE54	12	**-9.7**	**-9.0**	**-8.2**	**-9.6**	**-9.3**	**-9.7**	**-9.0**	**-9.3**	**-9.1**	**-0.38**	-0.36	-0.38	**-0.96**	**-0.92**	**-0.93**
Fuji	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
MLO mmgs vs CC (ref)[6]		N/A	N/A	N/A	N/A	N/A	N/A	-1.0	-1.0	-1.06	N/A	N/A	0.04	N/A	N/A	-1.0
Left mmgs vs Right (ref)[6]		N/A	N/A	N/A	N/A	0.006	0.03	N/A	N/A	0.02	N/A	0.03	0.07	N/A	0.03	0.05
Mammogram View[6]- RCC		N/A	N/A	N/A	N/A	ref.	N/A	ref.	N/A	ref.	N/A	N/A	ref.	N/A	N/A	ref.
LCC		N/A	N/A	N/A	N/A	0.006	N/A	N/A	ref.	0.006	N/A	N/A	0.08	N/A	N/A	0.05
RMLO		N/A	N/A	N/A	N/A	N/A	ref.	-1.0	N/A	-1.1	N/A	ref.	-0.03	N/A	ref.	-0.10
LMLO		N/A	N/A	N/A	N/A	N/A	0.03	N/A	-1.0	-1.0	N/A	0.03	0.03	N/A	-0.57	-0.05

Bold indicates the coefficient (β) is significant (p<0.05); ref. denotes the reference category; N/A = not applicable

*N (down, 2nd column) indicates the number of participants; N (across, 3rd row) indicates the total # of mammograms per group

[1] If one influential participant with PD≥75% excluded (outlier on residual vs fitted plot), the coefficient increases in magnitude and R^2 increases

[2] KE54 baseline mammograms are few in number (n=12), hence coefficients by Type only (film vs digital CR) are subsequently reported

[3] Relationship decreases (smaller coefficient and R^2) when participant with influential PD≥75% is omitted

[4] Three women with a single 2nd degree relative omitted, due to influential diagnostic outliers; p=0.04 for trend for women <60 years at Randomisation (n=45)

[5] Three mammograms of Version 'other' omitted from the comparison

[6] Individual baseline mammograms (Views) utilised for this comparison, not PD averages for each baseline episode as for the rest of the table

E. Three-Level Unconditional Means and Unconditional Growth Models

Mixed modelling coefficients, Unconditional Means models (PD, DA, BA)

Covariate	PD %	PD sqrt	DA mm^2	DA sqrt	BA mm^2	BA sqrt
ALL MMGS, 2126 episodes, 541 episodes, 120 participants						
FE Intercept	**16.5*** **	**3.8*** **	**2906*** **	**50*** **	**19016*** **	**135*** **
Random Effect (RE) variance estimates						
Between–person Random Effects						
Intercept	**126**	**2.0**	**$4.03x10^6$**	**348**	**$51.7x10^7$**	**661**
Mammogram Version	**16.6**	**0.22**	**501073**	**36.4**	**933861**	**9.8**
Within Version	**13.4**	**0.19**	**497031**	**37.2**	**$1.9x10^6$**	**24.8**
ICC[1]	0.81	0.83	0.80	0.83	0.95	0.95
Model fit statistics						
Log Likelihood	-6332	-1826	-17487	-7394	-18908	-6938
AIC	12673	3660	34982	14796	37825	13885
BIC	12695	3683	35005	14818	37847	13907

Bold indicates p≤0.1; * p<0.05; ** p<0.01; *** p<0.001; † p<0.2

FE fixed effects

[1] ICC calculated as Between–person intercept variance/total Variance where total Variance = (Between–person intercept Variance + mammogram Version Variance + Within–person Variance)

Mixed modelling coefficients, Unconditional Growth models (PD, DA, BA)

Covariate	PD %	PD sqrt	DA mm^2	DA sqrt	BA mm^2	BA sqrt
ALL MMGS, 2126 episodes, 541 episodes, 120 participants						
Categorical time FE estimates						
Months	0 (ref)	0 (ref)	0 (ref)	0 (ref)	0 (ref)	0 (ref)
6	**-1.21***	**-0.12**	**-260***	-1.6‡	**341**	1.1‡
12	**-1.87*** **	**-0.22*** **	**-335*** **	**-2.8*** **	234‡	0.9‡
24	**-2.33*** **	**-0.25*** **	**-424*** **	**-3.0*** **	244‡	**1.0**
36	**-1.68** **	**-0.16***	**-296** **	**-1.8**	**489** **	**1.8** **
48	-0.59	-0.04	-153†	-0.41	**387**	**1.4**
60	0.05	0.06	55	1.5†	**622** **	**2.1** **
72	0.62	0.12	127	**2.4**	**789** **	**2.6***
84	**2.46**	**0.32***	359‡	**4.7***	**899***	**3.1***
Intercept	**17.7*** **	**3.9*** **	**3109*** **	**51*** **	**18652*** **	**134*** **
Continuous time FE estimates						
Change/year	0.021	0.013	8.2	0.3	**113** **	**0.4** **
Intercept	16.6***	3.7***	2890***	49.2***	18726***	134***
Continuous time RE variance estimates						
Between–person Random Effects						
Time (years)	**0.55**	**0.008**	13019	**1.3**	**91167**	**1.2**
Intercept	**137**	**2.1**	$4.5x10^6$	**390**	**$51x10^6$**	**652**
Covariance	**-2.7**	**-0.05**	**-118950**	**-11.1**	141987	0.9
Mmg Version	**13.6**	**0.18**	**429519**	**30.2**	**473419**	**4.4**
Within Version	**13.3**	**0.19**	**496658**	**37.1**	**1871391**	**24.5**
Correlation[1]	**-0.31**	**-0.39**	**-0.49**	**-0.49**	0.07	0.03
Continuous time model fit statistics						
Log Likelihood	-6327	-1821	-17483	-7388	-18883	-6911
AIC	12668	3656	34980	14790	37779	13837
BIC	12708	3695	35019	14830	37819	13876

Bold indicates p≤0.1; * p<0.05; ** p<0.01; *** p<0.001; † p<0.2

[1] Correlation calculated as the Between–person (BP) covariance divided by the square root of (BP variance for time x BP intercept variance)

F. Three-Level Full and Parsimonious Models of PD, DA and BA

Covariate	PD %		PD square root[a]		DA mm^2		DA square root[a]		BA mm^2		BA square root[a]	
	Full	Parsimon	Full	Parsimon	Full	Parsimon	Full	Parsimon	Full	Parsimon	Full	Parsimon
ALL MAMMOGRAMS 2126 mammograms, 540 episodes (follow ups) 120 participants												
Age at Rand. (yrs)	-0.28‡	-0.24‡	**-0.05***	**-0.04**	-33	-18	-0.43†	-0.21	123†	--	0.5‡	--
BMI (kg/m^2)	**-0.88*** **	**-0.88*** **	**-0.12*** **	**-0.12*** **	-36	--	**-0.46**	--	**973*** **	**948*** **	**3.5*** **	**3.4*** **
Menarche (yrs)	0.08	--	0.01	--	47	--	0.46	--	381‡	417‡	**1.6**	**1.8**
Menopause (yrs)	**0.37***	**0.35***	**0.06** **	**0.05** **	**81** **	**80** **	**0.89** **	**0.85** **	37	--	0.2	--
Age First Birth <30y	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref
≥ 30 years	**7.3**	**7.6**	**1.1***	**1.1***	**2085** **	**2039** **	**19** **	**19** **	**5115** **	**4826***	**17***	**16***
Non-parous	-0.36	0.11	-0.04	0.05	688	929†	5.1†	7.6	3129‡	2929‡	**11.9**	**11.2**
OC use - Never	ref	--	ref	--	ref	--	ref	--	ref	--	ref	--
Ever users	-1.9	--	-0.38	--	-661	--	-6.2	--	-189	--	-1.3	--
HRT – Never	ref	--	ref	--	ref	--	ref	--	ref	--	ref	--
Ever users	-0.1	--	0.06	--	-236	--	-1.6	--	-996	--	-4.0	--
Smoking- never	ref	--	ref	--	ref	--	ref	--	ref	ref	ref	ref
Current	0.84	--	0.24	--	-44	--	3.1	--	**3700***	**3293***	**13***	**11***
Ex-smoker	0.8	--	0.12	--	-14	--	0.4	--	-1010	-1183	-3.0	-3.8
IBIS-1 No(ref) v Yes	0.95	--	0.23	--	218	--	3.2	--	-444	--	-1.9	--
CC (ref) vs MLO	**-1.6*** **	**-1.6*** **	**-0.18*** **	**-0.18*** **	**-177*** **	**-177*** **	**-1.5*** **	**-1.5*** **	**474*** **	**474*** **	**2.1**	**2.1*** **
Mammogram Version[5]- film	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref
KE52	**-8.9*** **	**-8.9*** **	**-0.90*** **	**-0.91*** **	**-1523*** **	**-1519*** **	**-13*** **	**-13*** **	-44.00	-45.80	0.10	0.1
KE54	**-7.1*** **	**-7.1*** **	**-0.62*** **	**-0.63*** **	**-1181*** **	**-1181*** **	**-9*** **	**-9*** **	-46.64	-50.06	0.10	0.08
Fuji	**-5.3*** **	**-5.3*** **	**-0.45*** **	**-0.45*** **	**-978*** **	**-985*** **	**-8*** **	**-8*** **	-396	-401	-1.1	-1.2
Intercept	**33.4*** **	**29.9*** **	**5.96*** **	**5.39*** **	**6026*** **	**4133*** **	**77*** **	**61*** **	**14691*** **	**13988*** **	**121*** **	**117*** **
Annual change in MD (All mammograms)												
Baseline to Year 1	-0.61†	-0.61†	**-0.11***	**-0.11***	**-139**	**-138**	**-1.19**	**-1.16**	188	193	0.60	0.62
Years 1 to 5	**0.54** **	**0.54** **	**0.06***	**0.06***	**123** **	**126** **	**1.07** **	**1.09** **	**173***	**174***	**0.6**	**0.6**
Years 5 to 7	-0.17	-0.17	0.01	0.01	-9.29	-9.20	0.05	0.05	**323***	**323***	**1.2***	**1.2***

Covariate	PD %		PD square root[a]		DA mm^2		DA square root[a]		BA mm^2		BA square root[a]	
	Full	Parsimon	Full	Parsimon	Full	Parsimon	Full	Parsimon	Full	Parsimon	Full	Parsimon
Random effects estimates for between–person, between–mammogram Version and within–person change												
Between–person random slopes												
Baseline to Year1	**2.4**	**2.4**	**0.08**	**0.08**	**446**	**446**	**2.8**	**2.8**	**696**	**692**	**2.6**	**2.6**
Year 1 to Year 5	--	--	--	--	--	--	--	--	**352**	**353**	**1.2**	**1.2**
Correlation	**-0.35**	**-0.36**	**-0.34**	**-0.34**	**-0.68**	**-0.66**	**-0.65**	**-0.65**	--	--	--	--
Between person Intercept	**10.1**	**10.1**	**1.2**	**1.2**	**2013**	**2028**	**17.7**	**18.1**	**4566**	**4637**	**16.4**	**16.8**
Between Version intercept	**2.3**	**2.4**	**0.28**	**0.27**	**399**	**399**	**3.7**	**3.7**	**617**	**617**	**1.8**	**1.8**
Within Version variance	**3.5**	**3.5**	**0.43**	**0.43**	**694**	**694**	**6.0**	**6.0**	**1340**	**1340**	**4.8**	**4.8**
Statistics of model fit[b]												
Log–likelihood	-6165	-6166	-1666	-1667	-17359	-17361	-7273	-7276	-18794	-18796	-6807	-6809
AIC	12378	12367	3379	3369	34766	34756	14594	14587	37636	37629	13662	13657
BIC	12514	12469	3515	3471	34902	34853	14730	14683	37772	37737	13798	13765
FILM mammograms only: Annual change in MD 248 mammograms, 63 episodes , 45 participants												
Baseline to Year 1	**-1.8****	**-1.8****	**-0.2****	**-0.2****	**-303***	**-306***	**-2.3***	**-2.4***	---	---	---	---
DIGITAL mammograms only: Annual change in MD 1878 mammograms, 434 episodes, 120 participants												
Baseline to Year 1	**-0.88***	**-0.88***	**-0.12***	**-0.12***	**-172***	**-169***	**-1.5***	**-1.5***	216†	219†	0.74‡	0.75†
Years 1 to 5	**0.66****	**0.65****	**0.08****	**0.08****	**138****	**140****	**1.3****	**1.3**	**167***	**168***	**0.54**	**0.55**
Years 5 to 7	-0.10	-0.11	-0.004	-0.005	4.1	5.2	0.25	0.26	**324***	**323***	**1.11***	**1.11***

Parsimon parsimonious model; ref reference category; y yrs years; -- not applicable; mmg mammogram;
Parsimonious model columns list only parameters with p≤0.1 (age-adjusted, except BA)
Bold p≤0.1; * p<0.05; ** p<0.01; *** p<0.001; ‡ p<0.15; † p<0.2;
[a] Coefficients for square root transformed parameters are shown in transformed units; they have not been back-transformed.
[b] The LL, AIC, BIC statistics of model fit were generated by models which utilise an average of the imputed value for menopause age